MONASH University

Neural Mechanisms of Stages of Reward and Avoidance Decision Processes and Clinical Implication in Obsessive-Compulsive Disorder and Gambling Disorder

XIAOLIU ZHANG

Supervisor:

Prof. Murat Yücel – Monash University

Prof. Carsten Murawski – The University of Melbourne

Dr. Chao Suo – Monash University

Dr. Amir Dezfouli – Data61, CSIRO

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2021
(*BrainPark, Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging Facility*)

Copyright notice

Abstract

Reward and avoidance learning form two important components of decision making. Both are two complex processes that involve assigning values to available options, choosing between alternatives-based preferences, assessing consequences for the selected choices, and learning from the outcomes to update future choices. However, the shared or differential neural mechanisms underlying the distinct stages of reward and avoidance decision processes still remains controversial. In the present study, using a novel probabilistic reward and avoidance learning task (involving a probability switch), together with neuroimaging techniques and computational modelling, we showed that both shared and distinct brain regions are involved at distinct stages including: (i) outcome, (ii) expected value and (iii) prediction error (PE) in reward and avoidance-based decision processes.

At the outcome stage, the frontal-subcortical brain areas including the inferior orbitofrontal cortex (OFC), striatum, thalamus, insula, and cingulum were significantly activated by the outcome of reward receipt. Receiving punishment was found to activate the cortical and subcortical brain regions of insula (also active during reward) but also distinct areas including supplementary motor area (SMA) and dorsal striatum were activated by the outcome of getting punished.

At the decision stage of expected value, the activity at the fronto-cortical brain regions including cingulum and superior medial frontal were found associated with the reward expectation. Whereas avoidance expectation recruited broader cortical and subcortical brain areas including not only the cingulum (also active during reward expectation), but also the inferior OFC, insula and dorsal striatum.

At the stage of error processing, a robust PE signal was found associated with activity in the cortical-basal ganglia brain regions under the reward condition; Meanwhile the aversive PE signal covaried with the activity of the frontal-subcortical brain regions (shared with reward processing), and distinct regions of the dorsal striatum. The results demonstrate the dorsal striatum specific role for differential phases of avoidance processing, and existence of the dissociated computational processes underlying reward and avoidance decision processes.

Impulsivity and compulsivity are behavioural traits underpinned by reward and avoidance processes that underlie many aspects of decision-making and are found to be aberrant is many mental health and addictive disorders. For instance, they form the characteristic symptoms of Obsessive-Compulsive Disorder (OCD) and Gambling Disorder (GD). The neural underpinnings of aspects of reward and avoidance learning and their relationship to expression of these clinical symptoms are only partially understood. The present study combined behavioural modelling and neuroimaging techniques to examine brain activity associated with key steps of reward and loss processing in OCD and GD, and its correlations with impulsivity and compulsivity. The findings revealed several regions of altered brain activity underlying the distinct stages of reward- and avoidance-related decision making processes in OCD and GD compared to healthy controls. The OCD group showed the decreased activity in the left operculum part of the inferior frontal, right Opercula part of the inferior frontal and right thalamus at the outcome of getting reward. OCD

participants also showed the increased activity in the left anterior cingulum at the phase of value expectation under avoidance condition. Further, the decreased activity in the left middle cingulum for reward expected value was found negatively correlated with scales of impulsivity measured by the BIS scores in participants with OCD. Meanwhile, participants with GD showed the decreased activity at right cuneus at the outcome of getting reward compared to healthy controls; Further, GD participants showed the increased activity at the brain region including the right triangular part of the inferior frontal for the error processing under avoidance condition.

The present series of studies have demonstrated the shared and distinct neural architecture underpinning the different stages of reward and avoidance decision making processes. Application of this knowledge to the clinical scenario revealed the existence of aberrant reward and avoidance-based decision processes in OCD and GD, and the contribution of these reward and avoidance based neural processes to the impulsivity and compulsivity behavioural traits seen in these clinical conditions.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Print Name: ……XIAOLIU ZHANG…….

Date: ……24/12/2020…….

Main Supervisor name: Murat Yücel

Publications during enrolment

Xiaoliu Zhang, Yann Chye, Leah Braganza, Leonardo F. Fontenelle1, Ben J. Harrison, Linden Parkes, Kristina Sabaroedin, Suzan Maleki, Murat Yücel, Chao Suo, "Severity related neuroanatomical alterations across symptom dimensions in obsessive-compulsive disorder", Journal of Affective Disorder Reports, in press.

Xiaoliu Zhang, Chao Suo, Ben J. Harrison, Leah Braganza, Ben Fulcher, Leonardo Fontenelle, Carsten Murawski*, Murat Yucel*, "Application of Reinforcement Learning Task with Probability Switch on OCD and Problem Gambling participants", abstract, The 2018 Brain conference – Computational Neuroscience of Prediction.

Xiaoliu ZHANG, Chao Suo, Ben J. Harrison, Leah Braganza, Ben Fulcher, Leonardo Fontenelle, Carsten Murawski*, Murat Yucel*, "Delineating reward/avoidance decision processes in the impulsive-compulsive spectrum disorders through a probabilistic reversal learning task", featured oral, abstract index by BMC neuroscience Supplementary, CNS* 2020.

Suzan Maleki, Yann Chye, Xiaoliu Zhang, Linden Parkes, Samuel R. Chamberlain, Leonardo F Fontenelle, Leah Braganza, George Youssef, Valentina Lorenzetti, Ben J Harrison, Murat Yücel, Chao Suo. "Neural correlates of symptoms severity of obsessive-compulsive disorder using magnetization transfer and diffusion tensor imaging". Psychiatry Research: Neuroimaging. 2020 Apr 30; 298:111046. doi: 10.1016/j.pscychresns.2020.111046. Epub 2020 Feb 11.

Acknowledgements

Firstly, I would like to acknowledge the supervision from Prof. Murat Yücel, Dr. Chao Suo, Dr. Amir Dezfouli and Prof. Carsten Murawski. Their suggestions have helped a lot to improve the thesis. Prof. Murat Yucel has helped me guide through the big picture of the whole study, and also provided some specific suggestions of the thesis structures. Dr. Chao Suo has devoted his time to discuss the thesis structures and edit chapters. Dr. Amir Dezfouli has contributed his time to the modelling part. Also, I'd like to appreciate the help from the collaborators. Working with those professional experts was definitely a precious opportunity and valuable experience for me.

As well as an international student, I'd like to show gratitude to the help from the BrainPark team including supervision, meeting arrangement and mentorship. Those meetings helped me survive from some struggles.

Also, I'd like to show my acknowledgement and gratitude for the support from Monash Graduate Scholarship (MGS) and Monash International Postgraduate Research Scholarship (MIPRS).

Finally, I'd like to appreciate the support and love from my family and friends. Their accompaniment was always precious and valuable.

Table of Contents

1 Introduction

1.1 Reinforcement-based decision making

1.1.1 Value-based decision making

Every day human participants are faced with a multitude of decisions, some simple decisions like what to eat or drink or complex decisions like whether or not to spend three years doing a doctoral degree. Decision making is an essential skill to live and manage our life, and it is a complex process that involves assigning value to available options, choosing between alternatives based on preferences, assessing consequences for the selected choices, and learning from the outcomes of decision making to update the future choices (Engel & Cáceda, 2015). Facing the fundamental challenge of the need to survive, the decision behaviour of human participants is directed toward gaining rewards, such as food, money or praise (approach behaviours), and also toward avoiding punishments, such as loss, pain, or humiliation (avoidance behaviours) (Doherty & Pauli, 2017). Thus, the reward and avoidance-related learning are two important components of decision making.

While there are always multiple choices, how do human participants make the choice from the alternatives? For example, consider the choice of career. Do we choose a career doing something that we are passionate about, for example, academia? Or, do we pursue a career such as economist that would be more lucrative? Obviously, everyone has his or her own opinion about the relative value of these particular choices, but ultimately each individual has an inherent desire to seek or maximize reward outcomes according to his or her anticipation. Also, consider the choice of committing a bad behaviour such as crime. Do we choose to give a try with the possibility to get punished? Evidently, we have a desire to avoid the unpleasant outcomes. Thus, when facing multiple decisions, humans always act in a manner that maximizes the prospects of obtaining the resources needed to survive and minimize the probability of encountering situations leading to harm (Doherty et al., 2017).

Specifically, the ability to decide what we want to do with our lives or to make any other decisions for that matter is predicated upon our knowledge of the result or value of the actions available (Krigolson et al., 2014).

1.1.2 Reinforcement learning theory

The problem facing decision making is the learning from the previous experience, through trial and error to improve the future selection. This problem is called reinforcement learning (RL) (Daw & Tobler, 2013; Fearing et al., 1929). Specifically, RL is an adaptive process in which a subject utilizes its previous experience to learn to predict reward, thus improving the outcomes of future choices to reach the goal of maximization of the rewards or minimization of the loss. The famous experiment conducted by Pavlov gave support to how organisms use experience to learn to predict reward. In the experiment, the dogs were exposed to repeated pairings whereby an initially neutral and unconditioned stimulus accompanied with reward such as food. Then, the dogs were found to salivate to the sound of the bell even if it was presented without the good, by virtue of the bell's predictive relationship with the good (Ii, 1927). Based on this experiment, variations of this experiment have been conducted with monkeys (Glimcher, 2011; Niv, 1997; Wolfram Schultz et al., 1997), and human participants (Niv et al., 2012).

RL theory is widely adopted to address the fundamental questions in decision making – 1) how do participants ***acquire their preference*** for different actions and outcomes, and also 2) how do they ***learn*** from the previous experience to update the future choice? Firstly, it is suggested that participants acquire their preference according to the value of the selected action. What does the value of an action reflect? RL theory proposes that the value of an action is a prediction of the subsequent reward or punishment gained by selecting that action (Sutton & Barto, 1998). Secondly, how do we learn from the previous experience? The

learning is realized by trial and error, which is a process of choice selection update following a previous decision that leads to a reward or punishment. For example, imagine we are faced up with two choices – A and B. Initially, we need to make a choice 'A' or 'B', and the prediction of reward (or punishment) associated with each choice is zero. However, if we choose A and are then rewarded, we compute a ***prediction error*** (PE) – the discrepancy between the actual outcome and the predicted value from the selected choice. Importantly, PE is then used to modify the expectation value of choice A, such that over time, the value of choice A comes to accurately reflect the reward gained by making this choice. In contrast, if choosing A leads to being punished, we will generate an ***aversive PE*** to update the expectation value of the choice.

To summarise the learning process after the choice is made, initially the subject will compute PE signals based on reward or punishments are encountered, meaning the discrepancies between the actual value of the reward (or punishment) and the value of the action. By repeating the learning and decision making, it is assumed the system is trying to diminish the magnitude of the PE computed at the time of reward delivery based on the previous experience. In another word, the predicted value of reward is gradually coming to approximate the actual reward value (See ***Figure*** *1-1* for schematic explanation).

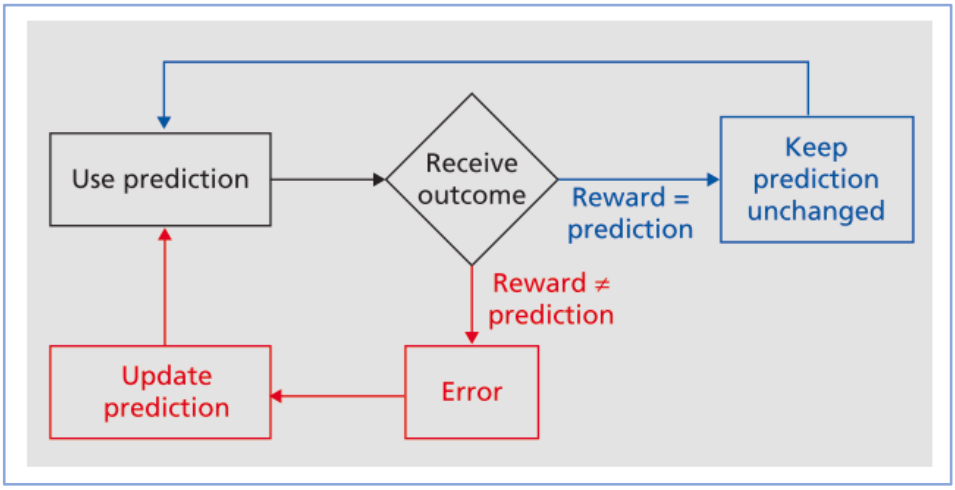

Figure 1-1 *Scheme of learning by PE. Red: a PE exists when the reward differs from its prediction, value of the selected action updated. Blue: no error exists when the outcome matches the prediction, behaviour remains unchanged (Wolfram Schultz, 2016).*

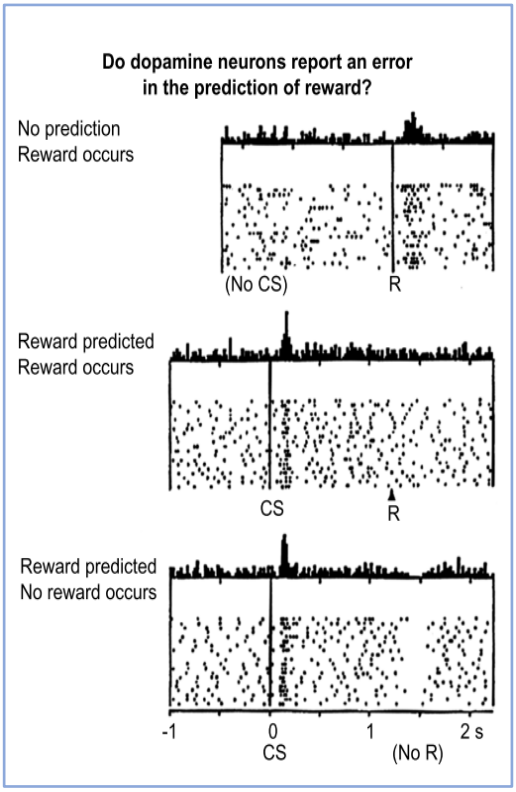

Figure 1-2 *Changes in dopamine neurons' output code for an error in the prediction of appetitive events. (Top) a PE signal before learning as a drop of fruit juice reward occurs. (Middle) No PE signal after learning as a drop of fruit juice reward occurs. (Bottom) a negative PE signal as no reward as expected* (Schultz et al., 1997).

The previous study provided empirical support for the prediction of RL theory through measuring changes in the phasic firing rate of dopaminergic neurons in the monkey's substantia nigra (SN) in classical conditioning experiments (Wolfram Schultz et al., 1997). They demonstrated that, when monkeys are initially given a reward, there is an associated phasic increase in the firing rate of dopaminergic neurons in the substantia nigra pars compacta (SNpc). Further, they also observed that when a reward was consistently paired with a predictive stimulus the phasic increase in dopamine firing rate observed at the time of reward delivery diminished over time, and then a phasic increase in dopamine firing rate was observed again shortly after the onset of the predictive stimulus (See ***Figure 1-2***). This pattern of results could be explained by RL theory specifically: Firstly, a PE was computed early in learning for unexpected rewards as the value of the cue state did not predict the value of the reward. Secondly, the PE at the time of reward was diminished with learning as the value of the cue state approached the value of the reward state – the difference between these states was minimising towards zero, meaning there was no error in prediction. Thirdly, the reward-like neural firing was observed at cue

onset after learning, as the monkey has moved from a state with no value – the state before the cue – to a state with value – the state value, this is a similar concept of the expected value we are going to explored later in this thesis. In summary, the pattern of changes in the dopaminergic response to the predictive cue and the reward is in line with the RL theory, and the *in vivo* human neuroimaging study would be an important extending research area to further support the theory frame. Actually, using some well-designed diagrams and sophisticated functional magnetic resonance imaging (fMRI), similar evidence was observed as we'll discuss later.

1.1.3 Reinforcement learning theory support for reward and avoidance learning

As discussed in ***Chapter*** 1.1.1, reward and avoidance-related learning are critically important parts of decision making as seeking rewards and avoiding punishments is a common propensity of human beings. Central to such behaviour is the ability to reflect the value of rewarding and punishing actions, establishing predictions of when and where such rewards and punishments will occur and use those predictions to form the basis of decisions that guide actions. The RL theory has also been adopted as the main theoretical framework in designing experiments (a typical reinforcement learning task design see ***Figure 1-3***) as well as interpreting the results (Kim et al., 2006). The example task consists of reward and avoidance conditions, which drive the reward and avoidance learning, respectively.

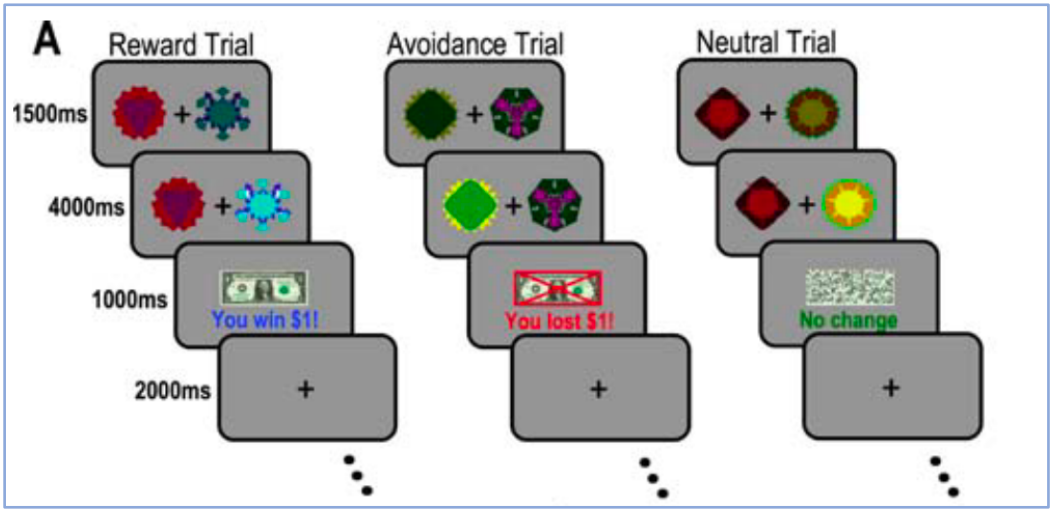

Figure 1-3 *The typical schematic of a RL learning task design included three conditions: Reward, Avoidance and Neutral; each condition has a pair of fractals with different probability of monetary reward, loss or no change (Kim et al., 2006).*

According to RL theory, the actions leading to greater predicted reward will produce a positive PE signal, and as the receipt of a rewarding outcome in a given context serves to strength associations between that context and the response performed, thus the afferent reward PE signal will ensure that such a response is more likely to be selected in the future.

This process well explains the reward-related learning (J. P. O'Doherty et al., 2003; Reynolds et al., 2001; Wolfram Schultz, 2018).

While avoidance is a key characteristic of adaptive and maladaptive fear. Avoidance learning (AL) is one of the instrumental conditioning that an individual learns to increase the frequency of a response with avoidance of an aversive outcome. Contrary to the positive reinforcement of reward learning, the feature of AL is that it is governed by negative reinforcement – the absence of a stimulus motivates behavioural change (Ilango et al., 2012). Like reward learning, AL could also be accounted for by standard theories of reinforcement (Ben et al., 2004; Kim et al., 2006). Kim et al found the common neural mechanism of getting reward and successfully avoiding punishment (Kim et al., 2006). It was proposed that successfully avoiding an aversive outcome itself acts as a reward, but different with the positive reinforcing properties as a real "extrinsic" reward, avoidance of an aversive outcome could be considered to be an "intrinsic reward". Avoidance behaviour thus is positively reinforced on each trial when the aversive outcome is avoided, just as receipt of reward reinforces behaviour during reward conditioning. In summary, a reward reinforces the action that causes its delivery, and a punishment - the negative reward signal, reinforces an action that avoids its delivery (Kenji Doya, 2008).

1.1.4 Summary

Decision making is a complex process that happens every day in our life. When facing multiple choices, humans always have an inherent desire to maximize the prospects of obtaining the resources needed to survive and minimize the probability of encountering situations leading to harm. Thus, the reward and avoidance-related learning forms two important components of decision making. RL theory is widely suggested to solve the problem facing the decision making – learning from the previous experience through trial and

error. RL theory was also suggested to provide a plausible account of reward learning and avoidance-related learning.

1.2 Neurological basis of reinforcement learning

1.2.1 Functional anatomy of dopamine system

As discussed in 1.1.2., the preclinical experiment provides the empirical support for the phasic firing rate of dopaminergic neurons encoding error signals to drive the learning. In humans, the majority of dopamine neurons reside in the midbrain including the substantia nigra (SN) and the ventral tegmental area (VTA) (Glimcher, 2011). Dopamine neurons send widespread projections through these nuclei to regions such as the striatum (caudate and putamen), the amygdala and the cerebral cortex (Glimcher, 2011) (shown in ***Figure 1-4***).

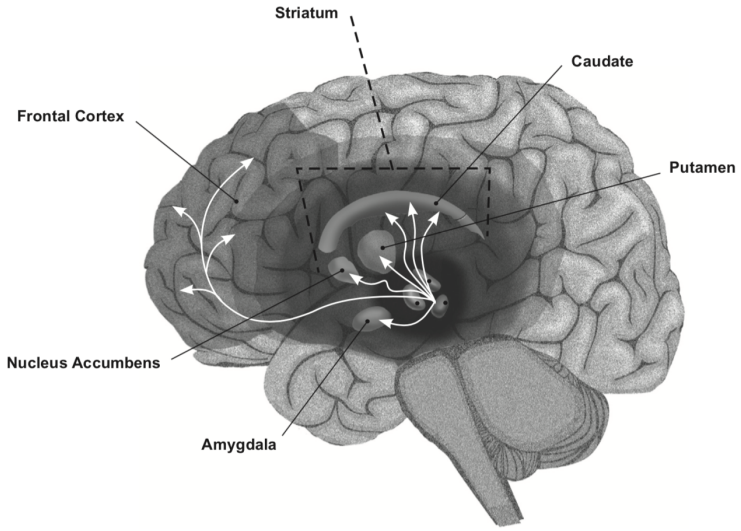

Figure 1-4 *The dopaminergic system of the midbrain and its projection pathways of striatum, the amygdala and the cerebral cortex* (Glimcher, 2011).

It's well accepted that the dopamine system is highly associated with reward. Dopamine neurons generate action potentials when a reward is encountered, and the higher the reward, the stronger the dopamine response (W Schultz et al., 1993). However, the dopamine response to the reward itself will be reduced when the reward is predicted. But if more than the predicted reward occurs, the dopamine neurons will show stronger responses. By contrast, their activity decreases if no, or less than predicted reward occurs. The dopamine response thus reflects the PE signal which affects neuronal activity in brain regions involved

in reward learning, including the striatum (McClure et al., 2003; J. P. O'Doherty et al., 2003; W Schultz et al., 1993; Wolfram Schultz, 2016), frontal cortex (W Schultz et al., 1993; Wolfram Schultz, 2016), and amygdala (Wolfram Schultz, 2016). Strong evidence has highlighted the critical role of the response in targeted area striatum for reward PE signal (Balleine et al., 2007).

Besides the well documented role of reward and reward-associated learning, the dopaminergic system was also reported to modulate the PE signal in aversive conditioning (Menon et al., 2007). Consistently, increased dopamine release was found over baseline during aversive learning from animal studies (Menon et al., 2007; Pezze & Feldon, 2004; Young, 2004). For the aversive PE signal in the brain, converging evidence has suggested that the aversive PE was represented in the dopamine target brain regions including striatum, prefrontal cortex as well as anterior cingulate cortex (Kim et al., 2006; Seymour et al., 2005, 2009; Tom et al., 2014). The amygdala was also suggested to show activity patterns consistent with aversive PE according to previous animals (McHugh et al., 2014), and human study (Yacubian et al., 2006).

1.2.2 Neurobiological basis of reinforcement-based decision making

According to the RL theory, the reward-based decision making and adaptive choice of actions were realized by the following three distinct phases: firstly, evaluation, in which participants estimated the action value and defined how much reward value each action will yield. Secondly, choice selection, in which an action was chosen by comparing the action values of two or more alternative choices. According to RL, there is some randomness existed in the choices and action selection is accomplished by a 'softmax' decision rule biased toward the seemingly richest options (Daw & Doya, 2006); Thirdly, learning, in which participants updated the action values by the error signal of expected action values.

A good match between the computational RL algorithms and neurobiological process in the brain was significantly observed (Dayan & Balleine, 2002b; Samson et al., 2010; Sharp et al., 2017). Striatum and cortical areas are thought to be involved in ***evaluation***. Through measurement of monkeys' performance in a reward-based task, A previous experiment demonstrated the representation of action values in the striatum, which could guide action selection in the ***basal ganglia circuit*** (Samejima et al., 2005). The basal ganglia circuit including striatum has been reported to participate into the RL process and maintain value representations to guide actions (Daw & Doya, 2006; K Doya, 1999; Lau & Glimcher, 2009; Morris et al., 2006). It has been reported that neurons in the monkey caudate nucleus that create a spatially selective response bias depending on the expected gain (Lauwereyns et al., 2002). Besides the basal ganglia circuit, other cortical regions, such as orbitofrontal cortex (OFC), medial prefrontal cortex (mPFC) and lateral intraparietal area (LIP) was found to involved during evaluation process (Barraclough et al., 2004; Daw & Doya, 2006; Eon & Schultz, 2018; Matsumoto et al., 2018; Roesch & Olson, 2018). Neurons in the dorsolateral prefrontal cortex (dlPFC) was found to encode the animal's past decisions and payoffs, as well as the conjunction between the two, providing signals necessary to update the estimates of expected reward, thus PFC plays a key role in optimizing decision making strategies (Barraclough et al., 2004). Also, the experiment on primate has reported that neuronal activity in orbitofrontal cortex (OFC) represents the value of the expected reward (Roesch & Olson, 2018). Via recording of animal response of OFC in a delayed go-nogo task, the study found the OFC neurons could report reinforcers are concerned with the expectation of reward, and also detect reward delivery at trial end (Eon & Schultz, 2018). Further, it has been suggested that the lateral intraparietal area (LIP) might be another brain region involved in the potential action value map (Daw & Doya, 2006).

The purpose of action evaluation is to direct the next step of action *selection* among the multiple choices in decision making. It was suggested that there is overlap of neural substrates for action choice and evaluation. (Daw & Doya, 2006). Through the simultaneous recordings in primate prefrontal cortex and dorsal striatum during a learning task in which the learned associations between stimuli and actions were periodically reversed, the recent study has found that the time course of change in behavioural responses over trials following a reversal was more related to the change in prefrontal responses than to that of striatal response, which suggested the prefrontal regions was more likely to be controlling behaviour (Samejima, 2009). As the expectation value attainment is important for guiding purposeful behaviour, the primate study demonstrated that prefrontal cortex (PFC) is a neuronal substrate for working memory used to guide reward-oriented behaviour (Amemori & Sawaguchi, 2018). Alternatively, considering the choice evaluation in the striatum, it was suggested that the action selection could be in the cortico-basal ganglionic loop (Daw & Doya, 2006).

Learning from the experience is through the PE, the deviation signal between the expectation and the actual outcome. The seminal work of Schultz et al. (Wolfram Schultz, 2018; Wolfram Schultz et al., 1997) suggests that phasic firing of midbrain dopamine neurons correspond to the neural representation of PE, encoding contingency-based reward signals. Also, the dopaminergic projected area striatum along with the diverse connected areas of predominantly anterior cerebral cortex including medial prefrontal cortex and anterior cingulate as well as insula were the key brain area encoding for the PE signal (Garrison et al., 2013a). In a dissociable manner, the PE signal in the reward processing (refer to reward PE) was correlated with the functional activity in ventral striatum and orbitofrontal cortex (Garrison et al., 2013b; Kim et al., 2006), while the aversive PE signal in the

avoidance learning was associated with brain responses in the amygdala-striatal regions (Zhang et al., 2016), and bilateral insula (Garrison et al., 2013b; Kim et al., 2006).

1.2.3 Summary

In summary, according to the literature on animal and human studies, there are specific neurobiological basis involved with the different phases of decision making process: cortico-basal ganglia circuit are responsible for the value estimation, and the prefrontal brain regions are involved with the choice selection, together with the brain regions of midbrain dopamine neurons and its projected areas are encoding the PE signal to drive the learning. A good parallel relationship of the neurobiological processes in the brain and implementation of RL was suggested: the cortico-basal ganglia circuit from the cortex, through striatum, the pallidum and the thalamus is involved into the multiple reinforcement-based decision making stages (Kenji Doya, 2007) (see ***Figure 1-5***).

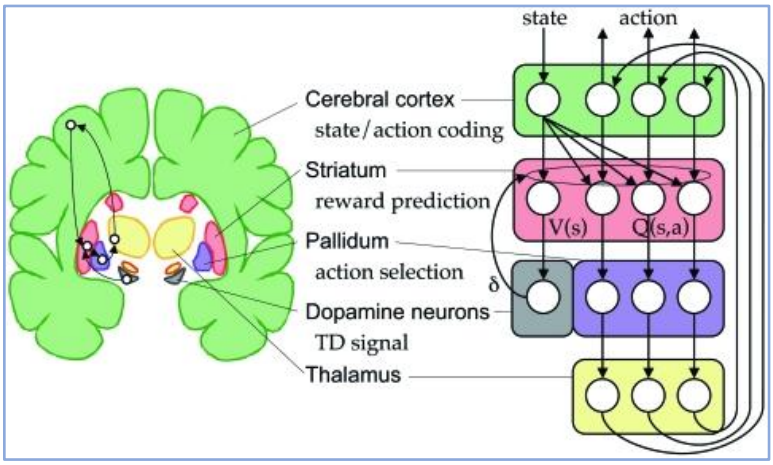

Figure 1-5 *Schematic model of implementation of RL in the cortico-basal ganglia circuit. The striatum learns and action value functions. The action value coding striatal neurons project to dopamine neurons, which sends the temporal difference (TD) signal back to the striatum. The outputs of action value coding striatal neurons channel through the pallidum and the thalamus, where stochastic action selection may be realized (Kenji Doya, 2007).*

1.3 Clinical application

1.3.1 Constructs of impulsivity and compulsivity

Impulsivity represents a multidimensional construct, and it covers a wide range of elements including: 1) decreased sensitivity to negative consequences of behaviour; 2) rapid and unplanned reactions to stimuli before complete processing of information; 3) regardless for long-term undesirable consequences or outcomes (Moeller et al., 2017). The impulsive behaviour could be simply defined as the tendency to act prematurely without foresight and unduly with risk, is associated with most forms of impulse control and addictive behaviours, especially gambling disorder (GD) (Dalley et al., 2011). For decision making, impulsive choice occurs when the individual preferentially chooses an immediately available small reward in preference to experiencing a delay for a larger one, which is probably governed by factors including the decisions about relative value of rewards and the ability to inhibit choices made to the more immediate options (Dalley et al., 2011).

Compulsivity on the other hand is a complicated concept referring to repetitive behaviours that are performed according to certain rules or in a stereotypical fashion (Berlin & Hollander, 2009; Grant & Kim, 2018). And it was involved with three elements: 1) the inability not to perform an act with the unpleasant feeling; 2) experienced loss of control in a habitual and stereotyped way; and 3) feeling one has to repetitive the act even under the perceived negative consequences (Luigjes et al., 2019). In decision making, compulsivity includes several behaviour characteristics, such as self-defeating repetitive behaviours and the diminished ability to stop or divert unwanted ideas suggesting the presence of cognitive and behavioural inflexibility, and also habitual responding and diminished goal-directed control implying excessive habit-learning or impaired reward/punishment processing (Figee et al., 2016). Obsessive compulsive disorder (OCD) is the representative disorder with the compulsive feature such as the compulsions to washing hands.

The two constructs of impulsivity and compulsivity shared some commonalities, for example, Robbins et al.,(Robbins, Gillan, Smith, Wit, et al., 2012) pointed out that both of them may reflect failures of response inhibition or top-down cognitive control (Robbins, Gillan, Smith, Wit, et al., 2012). While, the two constructs of impulsivity and compulsivity are different in nature, in which they differ in aspects of response inhibition: compulsivity relates to an inability to terminate action, whereas impulsivity refers to problems initiating actions (Lai & Ip, 2011). Traditionally, it has been suggested that impulsivity and compulsivity constitute opposite ends of spectrum across dimensional disorders (***Figure 1-6***), in which disorder like GD shares the feature of impulsivity while the OCD are characterized by the compulsivity.

A shift existed from impulsivity to compulsivity with a transmission from initial positive reinforcement to later negative reinforcement has been reported in recent studies (El-Guebaly et al., 2012; Everitt & Robbins, 2005). With the increase of impulsive behaviour, it has found that participants with GD would acquire the compulsivity feature with habitual process (El-Guebaly et al., 2012; Fontenelle et al., 2011). On the other hand, as the conceptualization of a compulsive disorder, it is also suggested that OCD shares behavioural components of impulsivity (Abramovitch & Mckay, 2016; Fontenelle et al., 2011; Grassi et al., 2015).

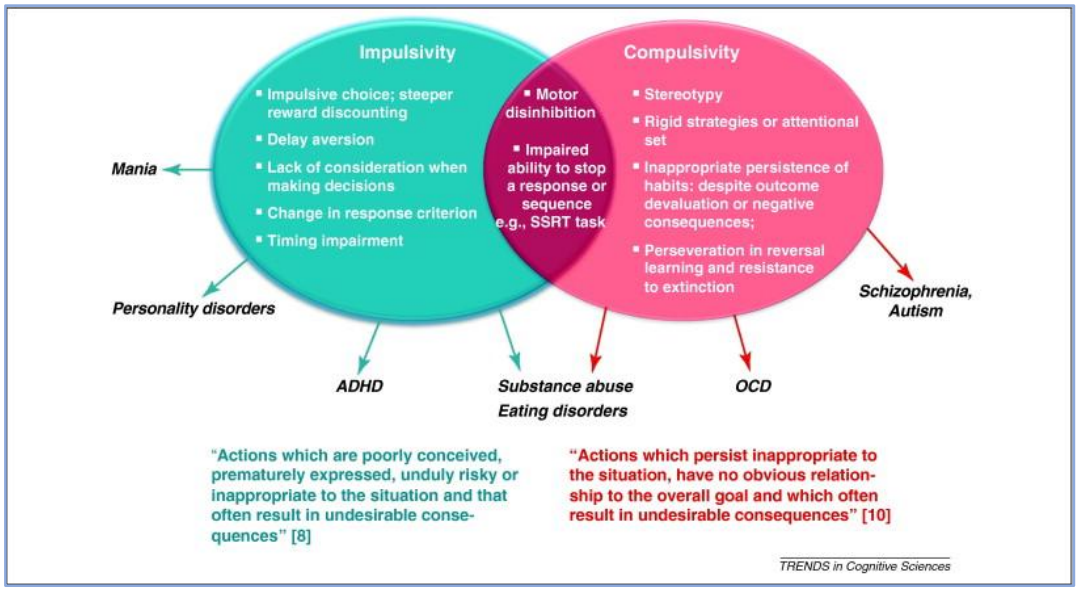

Figure 1-6 *The impulsivity and compulsivity constructs and possible psychological mechanisms underlying the two constructs (Robbins, Gillan, Smith, de Wit, et al., 2012).*

1.3.2 Obsessive compulsive disorder

Obsessive-compulsive disorder (OCD) is a relatively common, chronic and disabling neuropsychiatric disorder with an estimated prevalence between 1% and 3% of populations (Figee et al., 2011). It is characterized by experience of unwanted repetitive thoughts (obsessions) and repetitive behaviours (compulsions) (***Figure 1-7***). As OCD is conceptualized as a compulsive disorder, the portrait of OCD related to risk averse individuals to avoid potential punishment or harm is well received. While recent studies have also linked OCD to impulsivity with risky decision making and dysfunctional reward processing. Further, OCD is suggested to have a poor cognitive flexibility to rapidly change behaviour in the face of changing circumstances. I will review studies on those neurocognitive factors including reward processing, punishment/harm avoidance and behavioural inflexibility conducted on OCD populations, and their relationship with the key symptom of compulsivity or repetitive behaviour as well the potential impulsivity feature.

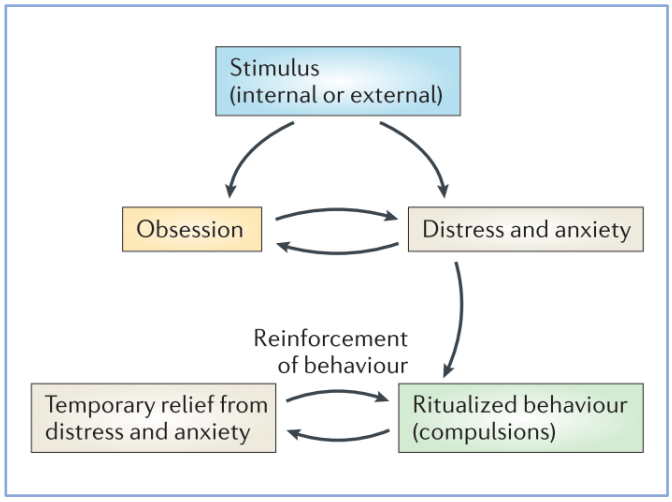

Figure 1-7 *Theoretical basis of obsessive-compulsive behaviour. An individual with obsessive-compulsive disorder experiences exaggerated concerns about danger, harm or loss that result in persistent obsession. Further, Relief from the distress and/or anxiety associated with these obsessions leads to reinforcement of the behaviours, leading to repetitive, compulsive behaviour when obsessions occur (Pauls et al., 2014).*

Reward processing

Compulsivity in OCD may in part be explained by dysfunctional brain reward system, driving the development of a maladaptive behavioural at the cost of healthy rewarding actions and a relative failure to switch to more adaptive, goal-directed behaviours (Figee et al., 2016). As reward processing is critically dependent on the cortico-basal ganglia circuit talked in 1.3.2, participants with OCD have been consistently found the abnormal brain activation within this circuit, and the compulsivity was reported to be related to this impaired reward processing (Figee et al., 2016). Specifically, through the implementation of a monetary incentive delay task, the study has found that participants with OCD displayed attenuated reward anticipation activity in the nucleus accumbens (NAc) compared with healthy controls (Figee et al., 2011).

Besides the NAc, the higher orbitofrontal activation (Lagemann et al., 2012), and blunted responsiveness of the orbitofrontal-striatal loop during reward processing were found in participants with OCD (P. L. Remijnse et al., 2009). The study on dopaminergic dysfunction of reward processing in OCD has reported the abnormally cingulate error signalling during this process, and the exaggerated error signal was related to the trait of self-regulating behaviour difficulty (Murray et al., 2017). While, the application of deep brain stimulation treatment for OCD has reported a normalization of anticipatory reward responses in the ventral striatum and reduced excessive connectivity between the NAc and prefrontal cortex within the brain reward circuits, which provided a further hint for the reward processing impairment (Figee et al., 2014). Along with the aberrant reward processing found in the cortico-basal ganglia circuit for participants with OCD, the increased functional connectivity between NAc and middle frontal gyrus cortex in OCD was found to be correlated with severity of repetitive behaviours (Akkermans et al., 2018). In summary, these studies provide neural evidence for the altered reward processing in the cortico-basal ganglia circuit for participants with OCD, and the correlations showed that the compulsivity feature is the contributing factor to the malfunction.

Punishment or harm avoidance

Avoiding harm/punishment is critical for maintaining physical and mental health. However, excessive harm avoidance can be maladaptive, such as OCD. Also, the avoidance has been included in several diagnostic criteria in latest DSM-V (American Psychiatric Association, 2013).

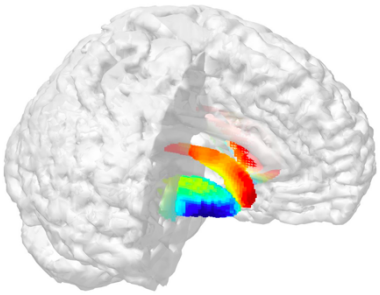

Figure 1-8 Striatum plays a key role in harm-avoidance habits: its structure predicts avoidance learning style (red/yellow: more gray matter in active avoiders, blue/green: less gray matter (Hauser et al., 2020).

Greater avoidance habits or punishment sensitivity in OCD compared to healthy controls was reported in a number of previous studies based on the avoidance/punishment related task (Eldar et al., 2016; Endrass et al., 2011; Figee et al., 2016; Gillan et al., 2014, 2016). And these excessive avoidance habits could come from the OCD individuals' engagement in compulsive behaviours as they may be less capable of noticing its self-damaging consequences (Figee et al., 2016). Interestingly, through a pain shock paired with a conditioned stimulus experiment, Eldar et al. found in their study that OCD individuals had an imbalance of harm avoidance behaviour, which is the better learning from the shocks whereas the poorer learning from success in avoiding shocks. And this imbalance could be predicted by the striatum's function and structure (***Figure 1****-8*). It was concluded that the higher gray matter volume found in the OCD might engender this excessively persistent harm avoidance pattern, and failure to adjust to success in harm (Eldar et al., 2016). Besides from the striatum, the exhibited avoidance behaviour was found associated with the caudate

nucleus (Gillan et al., 2014, 2016), and medial prefrontal cortex (mPFC) hyper-activation in participants with OCD according to the functional neuroimaging studies (Kaufmann et al., 2013). While, the decreased activation was also found in the ventral striatum in the loss outcome for OCD (Wh et al., 2011). Further, the increased brain activation in the anterior insular during the phase of loss anticipation was found in OCD compared to controls. It was concluded that this dysfunction may be involved in the trait of compulsivity with a diminished ability to foresee the negative consequences of compulsive actions (Figee et al., 2016). For the punishment response in OCD, a more impulsive response style was demonstrated in OCD with failure to slow down the response after receiving punishment compared with controls under punishment conditions in a go/no-go task (Morein-Zamir et al., 2013). While this sensitivity to punishment has been found positively correlated with the high scores of disease hoarding dimension of participants with OCD in a questionnaire study (Fullana et al., 2004).

Cognitive and behavioural inflexibility

The contingency related flexibility refers to the adaptation of behaviour or cognitive strategies by evaluating adaptive response in the face of changing environment (Figee et al., 2016). The task like stop-signal task and reversal learning task has been used for the cognitive and behavioural flexibility measurement. Using the stop-signal task, the study has demonstrated the cognitive and behavioural inflexibility was limited to neurocognitive profile of OCD (Samuel et al., 2006). The reversal learning task combining with neuroimaging technique has further reported this cognitive flexibility deficits were related to the fronto-striatal circuit dysfunction (Bechara et al., 2000; Chamberlain et al., 2007; Gu et al., 2008; Peter L Remijnse et al., 2006).

1.3.3 Gambling disorder

Gambling disorder is associated with loss of control and continued gambling in spite of negative consequences. GD is classified as behavioural addiction in the Diagnostic and Statistical Manual of Mental Disorders V (DSM-V), with a lifetime prevalence of 0.5 -1% (Miedl et al., 2012, 2014; Petry et al., 2005; Potenza, 2008). In detail, GD is characterized by persistent and recurrent maladaptive patterns of gambling behaviour, and is associated with impaired functioning, reduced quality of life, and high rates of bankruptcy, divorce, and incarceration. The behaviours that characterize GD are impulsive in that they are often premature, poorly thought out, risky and result in long-term side effects (Lai & Ip, 2011). Broadly speaking, a gamble involves a decision to place a wager on an uncertain event that offers the potential for a larger prize, and gambling can be considered a prototypical example of a risky decision and represents a harmless form of entertainment for most consumers, it has the capacity to become dysfunctional in a minority. As GD has been also proposed to represent a 'behavioural addiction', the impairments in the impulsivity responding and risky decision-making was found in GD as well as alcohol-dependent group (Goudriaan et al., 2006; Lawrence et al., 2009; Ledgerwood et al., 2012).

Reward processing

Dysfunction in reward processing has been found in GD, which is associated with loss of control and continued gambling in spite of negative consequences. According to a recent meta-analysis based on the studies of reward processing in addiction reported that individuals gambling addiction showed decreased striatal activation both at phases of anticipation and outcome compared to healthy controls (Sources et al., 2017). Not only the striatum, there are studies that have reported hypo-activity in other brain regions in the reward circuit in the GD during both the anticipation and receipt of monetary rewards. The study found that GD group

exhibited significantly reduced activity in the reward circuit areas including ventromedial prefrontal cortex (vmPFC) and ventral striatum as well insula during the phases including the prospect and anticipation in the task related studies (Pearlson & Potenza, 2013; Reuter et al., 2005; Ruiter et al., 2009). And, the activity in the ventral striatum was found to be inversely correlated with the levels of impulsivity (Pearlson & Potenza, 2013), and the gambling severity (Reuter et al., 2005). Jan et al. has also reported a reduced activation of fronto-striatal circuit of GD population implying the blunted response to reward stimuli. And the activation was found negatively correlated with gambling severity, thus linking the hypoactivation of the reward processing brain regions to disease severity (Reuter et al., 2005). Further, the ventral striatum of problem gamblers showed an imbalance response to different reward types during reward anticipation and also, imbalance of response existed in the posterior OFC during reward outcome (Sescousse et al., 2018). While there are several studies found hyper-activity within those brain regions during reward processing (Romanczuk-seiferth et al., 2009), implying some sensitization of the reward system. Ruth et al., has found that problem gamblers showed stronger activation in the bilateral ventral striatum to 5 euro than to 1 euro trials together with more activation associated with gain-related expected value in the reward circuit than controls, which showing GD are characterized by abnormally increased reward expectancy, which may render them overoptimistic with regard to gambling outcomes (Holst et al., 2011). GD is also considered as behavioural addiction as sharing the same characteristics with addiction (Janssen et al., 2015). GD's adherence to the disadvantageous decks for receiving higher immediate rewards with suffering higher overall losses were reported before in the gambling task studies (Goudriaan et al., 2006; Sescousse et al., 2018). The neurochemical study reported the speculated gambling behaviour was related to a deficiency of the mesolimbic dopaminergic reward system (Blum et al., 1996).

Punishment or harm avoidance

The punishment of monetary loss was found to be associated with activation of prefrontal cortex including OFC and inferior prefrontal sulcus (J. O'Doherty et al., 2001). GD is reported to be related to response preservation and diminished punishment sensitivity as indicated by hypoactivation of the ventrolateral prefrontal cortex when money is lost (Ruiter et al., 2009). During the phase of loss anticipation, GD group also exhibited significantly reduced activity in the vmPFC, insula and ventral striatum and ventral striatum activation was found to be inversely correlated with levels of impulsivity (Pearlson & Potenza, 2013). Neural activation in the ventromedial caudate nucleus during anticipation of loss decreased in participants with GD compared to OCD and healthy controls, and additionally, reduced activation in the anterior insula during anticipation of loss was observed in GD, which was intermediate between the OCD and the controls. Further, there was a significant positive correlation between anterior insula activity and gambling scores (Choi et al., 2012).

Cognitive and behavioural inflexibility

Problem gamblers were reported a deficiency feedback processing in the Card Playing Task with less likely to change from the deck after experiencing loss, reflecting a response inflexibility (Goudriaan et al., 2005). It was reported that the disadvantages of "Chasing one's losses" might underlie the development of this inflexibility in gamblers (Linnet et al., 2006). Application of reversal learning task in GD found the vmPFC is implicated to be involved with this cognitive process (Clark et al., 2004). Further, the other study provides a demonstration that the impairment of vmPFC would affect this reversal learning performance (Fellows & Farah, 2005).

1.3.4 Summary

Participants with OCD and GD have been found with aberrant decisions (Franken et al., 2008; Sachdev & Malhi, 2005). The OCD and GD are polar opposite representative of compulsive and impulsive disorder on the supervised dimensional model of ***impulsive-compulsive spectrum disorders***, respectively. Findings of altered reward processing (Figee et al., 2011; Wh et al., 2011), and harm avoidance and less sensitivity to punishment (Kaufmann et al., 2013; Wh et al., 2011) were found in OCD. Under the GD condition, previous studies have reported the impaired and risky decision making (Goudriaan et al., 2006; Lawrence et al., 2009; Ledgerwood et al., 2012) and diminished sensitivity in reward and punishment (Ruiter et al., 2009). As a typical conception of compulsive disorder, recent studies also suggested that OCD shares behavioural components of impulsivity (Abramovitch & Mckay, 2016; Fontenelle et al., 2011; Grassi et al., 2015), and also based on the existed portrait of impulsive disorder, a compulsivity feature was suggested to be acquired in participants with GD with the increase of the impulsive behaviour (Fontenelle et al., 2011). The interesting problem is how the constructs of impulsivity/compulsivity are linked to the reward/avoidance processing in participants with OCD and GD.

1.4 Research overview

This dissertation specifically focuses on investigating the differences of reward and avoidance-based decision process at three distinct phases including outcome processing, value expectation and error signal processing. Behaviourally, the RL algorithm was used to interpret the decision making process based on the participants' behavioural data (Dayan & Balleine, 2002a). The model could help specify a set of structural assumptions along with free parameters that can be adjusted to capture a range of behaviours such as learning

efficiency (Dezfouli et al., 2018). Further, the neuroimaging technique could reveal the brain mechanism under these distinct phases.

Using the same paradigm of behavioural modelling and neuroimaging technique, we examined the proposed aberrant reward and avoidance-based decision process in OCD and GD population. Also, we investigated the impulsivity/compulsivity constructs effect on the reward and avoidance-based decision process in those clinical groups. The schedule of the thesis chapters is as follows:

In Chapter 2, we introduced the neuroimaging technique in this study including the physics of functional magnetic resonance imaging (fMRI), task design as well as imaging processing.

In Chapter 3, we examined participants' behavioural response under the reward and avoidance condition in the learning task. the statistical analysis was carried out from the four measurements: 1) the response time of choice making; 2) the number of Correct and Incorrect choice; 3) the learning curve of Correct and Incorrect fractal choice in reward/avoidance condition to model the participants learning of the task; 4) one trial back – the stay ratio on the rewarded/punished versus the non-rewarded/not-punished trials to understand participants responding pattern. A Q-learning model was applied to model the participants' trial-by-trial learning process. The two characteristic parameters were learning rate and inverse temperature parameter which showed the participants' learning and the balance of exploitation versus exploration on the choice respectively.

In Chapter 4, we examined the neural mechanism under the behavioural tendencies based on the chapter 3. Using event-related fMRI and computational modelling, the participants' neural activity was investigated under three distinct phases: 1) outcome delivery and 2) expectation, as well as 3) prediction error (PE) signal processing under conditions of receipt and avoidance of reward and punishment compared to neutral condition. Due to the

probability switch in the learning task, we were supposed to see the enhanced PE signal and its associated brain activation.

In Chapter 5, we investigated the proposed maladaptive and aberrant decision process in participants with OCD (i.e. high compulsivity) and participants with GD (i.e. high impulsivity). At first stage, the statistical analysis of the participants' behavioural performance was carried out under the reward/avoidance condition with comparison to healthy controls, and four measurements were included: 1) the response time; 2) the number of Correct and Incorrect fractal choice; 3) the learning curve; and 4) one trial back of the stay ratio. Next, the RL algorithms were implemented to model the participants' behavioural process in the learning task and extract the learning traits including: 1) learning rate; 2) inverse temperature parameter. Further, to identify the neural substrates supporting aberrant differences compared to healthy controls, imaging regression analysis was carried out to examine the brain activation in OCD and GD group compared to healthy controls at the three phases of decision process: 1) outcome processing; 2) expectation value; and 3) error processing.

In Chapter 6, based on the findings of previous chapters, we provided a summary of whole project findings and suggest potential fruitful avenues forward for future studies.

References

Abramovitch, A., & Mckay, D. (2016). Behavioral Impulsivity in Obsessive – Compulsive Disorder. 5(3), 395–397. https://doi.org/10.1556/2006.5.2016.029

Akkermans, S. E. A., Rheinheimer, N., Bruchhage, M. M. K., Durston, S., Brandeis, D., Banaschewski, T., Boecker-Schlier, R., Wolf, I., Williams, S. C. R., Buitelaar, J. K., van Rooij, D., & Oldehinkel, M. (2018). Frontostriatal functional connectivity correlates with repetitive behaviour across autism spectrum disorder and obsessive–compulsive disorder. Psychological Medicine, 1–9. https://doi.org/10.1017/s0033291718003136

Amemori, K., & Sawaguchi, T. (2018). Contrasting Effects of Reward Expectation on Sensory and Motor Memories in Primate Prefrontal Neurons. July 2006. https://doi.org/10.1093/cercor/bhj042

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. 27(31), 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Barraclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed- strategy game. 7(4), 404–410. https://doi.org/10.1038/nn1209

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. 295–307.

Ben, S., P, O. J., Peter, D., Martin, K., K, ones A., J, D. R., J, F. K., & J, F. R. S. (2004). Temporal difference models describe higher-order learning in humans. Nature, 429(June), 664–667. https://doi.org/10.1038/nature02636.1.

Berlin, H. A., & Hollander, E. (2009). Understanding the differences between the impulsivity and Compulsivity. July 2008, 58–61.

Blum, K., Ph, J. G. C., Ph, P. J. S., Braverman, E. R. M., & Ph, T. J. H. C. (1996). The 02 dopamine receptor gene as a determinant of reward. 396–400.

Chamberlain, S. R., Blackwell, A. D., Fineberg, N. A., Robbins, T. W., & Sahakian, B. J. (2007). Europe PMC Funders Group Strategy implementation in obsessive – compulsive disorder and trichotillomania. 36(1), 91–97. https://doi.org/10.1017/S0033291705006124.Strategy

Choi, J. S., Shin, Y. C., Jung, W. H., Jang, J. H., Kang, D. H., Choi, C. H., Choi, S. W., Lee, J. Y., Hwang, J. Y., & Kwon, J. S. (2012). Altered Brain Activity during Reward Anticipation in Pathological Gambling and Obsessive-Compulsive Disorder. PLoS ONE, 7(9). https://doi.org/10.1371/journal.pone.0045938

Clark, L., Cools, R., & Robbins, T. W. (2004). The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning. 55, 41–53. https://doi.org/10.1016/S0278-2626(03)00284-7

Dalley, J. W., Everitt, B. J., & Robbins, T. W. (2011). Impulsivity, Compulsivity, and Top-Down Cognitive Control. Neuron, 69(4), 680–694. https://doi.org/10.1016/j.neuron.2011.01.020

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. Current Opinion in Neurobiology, 16(2), 199–204. https://doi.org/10.1016/j.conb.2006.03.006

Daw, N. D., & Tobler, P. N. (2013). Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning. Neuroeconomics: Decision Making and the Brain: Second Edition, 283–298. https://doi.org/10.1016/B978-0-12-416008-8.00015-2

Dayan, P., & Balleine, B. W. (2002a). and Reinforcement Learning. 36, 285–298.

Dayan, P., & Balleine, B. W. (2002b). Reward, motivation, and reinforcement learning. Neuron, 36(2), 285–298.

Dezfouli, A., Griffiths, K., Ramos, F., Dayan, P., & Balleine, B. W. (2018). Models that learn how humans learn: the case of decision-making and its disorders. In bioRxiv. https://doi.org/10.1101/285221

Doherty, J. P. O., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. https://doi.org/10.1146/annurev-psych-010416-044216

Doherty, J. P. O., & Pauli, W. M. (2017). Learning, Reward, and Decision Making Learning, Reward, January. https://doi.org/10.1146/annurev-psych-010416-044216

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? 12, 961–974.

Doya, Kenji. (2007). Reinforcement learning: Computational theory and biological mechanisms. HFSP, 1(1), 30–40. https://doi.org/10.2976/1.2732246

Doya, Kenji. (2008). C onnections B etween C omputational and N eurobiological P erspectives on D ecision M aking Decision theory, reinforcement learning, and the brain. 8(4), 429–453. https://doi.org/10.3758/CABN.8.4.429

El-Guebaly, N., Mudry, T., Zohar, J., Tavares, H., & Potenza, M. N. (2012). Compulsive features in behavioural addictions: The case of pathological gambling. Addiction, 107(10), 1726–1734. https://doi.org/10.1111/j.1360-0443.2011.03546.x

Eldar, E., Hauser, T. U., Dayan, P., & Dolan, R. J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. 113(17). https://doi.org/10.1073/pnas.1519829113

Endrass, T., Kloft, L., Kaufmann, C., & Kathmann, N. (2011). Approach and avoidance learning in obsessive-compulsive disorder. 172(August 2010), 166–172. https://doi.org/10.1002/da.20772

Engel, A., & Cáceda, R. (2015). Can Decision Making Research Provide a Better Understanding of Chemical and Behavioral Addictions? 75–85.

Eon, L., & Schultz, W. (2018). Reward-Related Neuronal Activity During Go-Nogo Task Performance in Primate Orbitofrontal Cortex.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nature Neuroscience, 8(11), 1481–1489. https://doi.org/10.1038/nn1579

Fearing, F., Pavlov, I. P., & Anrep, G. V. (1929). Conditioned Reflexes. An Investigation of the Physiological Activity of the Cerebral Cortex. In Journal of the American Institute of Criminal Law and Criminology (Vol. 20, Issue 1, p. 153). https://doi.org/10.2307/1134737

Fellows, L. K., & Farah, M. J. (2005). Different Underlying Impairments in Decision-making Following Ventromedial and Dorsolateral Frontal Lobe Damage in Humans. January. https://doi.org/10.1093/cercor/bhh108

Figee, M., Luigjes, J., Smolders, R., Wingen, G. Van, Kwaasteniet, B. De, Mantione, M., Ooms, P., Koning, P. De, Vulink, N., Levar, N., Droge, L., Munckhof, P. Van Den, Schuurman, P. R., Nederveen, A., Brink, W. Van Den, & Mazaheri, A. (2014). Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. 16(4). https://doi.org/10.1038/nn.3344

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., Brink, W. Van Den, Goudriaan, A., Potenza, M. N., Robbins, T.

W., & Denys, D. (2016). Compulsivity in obsessive – compulsive disorder and addictions. European Neuropsychopharmacology, 26(5), 856–868. https://doi.org/10.1016/j.euroneuro.2015.12.003

Figee, M., Vink, M., De Geus, F., Vulink, N., Veltman, D. J., Westenberg, H., & Denys, D. (2011). Dysfunctional reward circuitry in obsessive-compulsive disorder. Biological Psychiatry, 69(9), 867–874. https://doi.org/10.1016/j.biopsych.2010.12.003

Fontenelle, L. F., Oostermeijer, S., Harrison, B. J., & Pantelis, C. (2011). Obsessive-Compulsive Disorder, Impulse Control Disorders and Drug Addiction Common Features and Potential Treatments. 71(7), 827–840.

Franken, I. H. A., Strien, J. W. Van, Nijs, I., & Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. 158, 155–163. https://doi.org/10.1016/j.psychres.2007.06.002

Fullana, M. A., Mataix-Cols, D., Caseras, X., Alonso, P., Manuel Menchón, J., Vallejo, J., & Torrubia, R. (2004). High sensitivity to punishment and low impulsivity in obsessive-compulsive patients with hoarding symptoms. Psychiatry Research, 129(1), 21–27. https://doi.org/10.1016/j.psychres.2004.02.017

Garrison, J., Erdeniz, B., & Done, J. (2013a). Prediction Error in Reinforcement Learning: A Meta - analysis of Neuroimaging studies. Neuroscience & Biobehavioral Reviews, 1–50.

Garrison, J., Erdeniz, B., & Done, J. (2013b). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neuroscience and Biobehavioral Reviews, 37(7), 1297–1310. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gillan, C. M., Apergis-schoute, A. M., Morein-zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2016). Europe PMC Funders Group Functional neuroimaging of avoidance habits in OCD. 172(3), 284–293. https://doi.org/10.1176/appi.ajp.2014.14040525.Functional

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. Biological Psychiatry, 75(8), 631–638. https://doi.org/10.1016/j.biopsych.2013.02.002

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 108(SUPPL. 3), 15647–15654. https://doi.org/10.1073/pnas.1014269108

Goudriaan, A. E., Oosterlaan, J., Beurs, E. De, & Brink, W. Van Den. (2005). Decision making in pathological gambling: A comparison between pathological gamblers, alcohol dependents, persons with Tourette syndrome , and normal controls. 23, 137–151. https://doi.org/10.1016/j.cogbrainres.2005.01.017

Goudriaan, A. E., Oosterlaan, J., Beurs, E. De, & Brink, W. Van Den. (2006). Psychophysiological determinants and concomitants of deficient decision making in pathological gamblers. 84, 231–239. https://doi.org/10.1016/j.drugalcdep.2006.02.007

Grant, J. E., & Kim, S. W. (2018). Brain circuitry of compulsivity and impulsivity. 2014, 21–27. https://doi.org/10.1017/S109285291300028X

Grassi, G., Pallanti, S., Righi, L., Figee, M., Mantione, M., Denys, D., Piccagliani, D., Rossi, A., & Stratta, P. (2015). Think twice: Impulsivity and decision making in obsessive – compulsive disorder. 4(4), 263–272. https://doi.org/10.1556/2006.4.2015.039

Gu, B. M., Park, J. Y., Kang, D. H., Lee, S. J., Yoo, S. Y., Jo, H. J., Choi, C. H., Lee, J. M., & Kwon, J. S.

(2008). Neural correlates of cognitive inflexibility during task-switching in obsessive-compulsive disorder. Brain, 131(1), 155–164. https://doi.org/10.1093/brain/awm277

Hauser, T. U., Eldar, E., & Dolan, R. J. (2020). Neural Mechanisms of Harm-Avoidance Learning A Model for Obsessive-Compulsive Disorder? 73(11), 1196–1197. https://doi.org/10.1016/j.biopsych

Holst, R. J. Van, Veltman, D. J., Büchel, C., Brink, W. Van Den, & Goudriaan, A. E. (2011). Distorted Expectancy Coding in Problem Gambling: Is the Addictive in the Anticipation? BPS, 71(8), 741–748. https://doi.org/10.1016/j.biopsych.2011.12.030

Ii, L. (1927). Conditioned Reflexes: An Investigation of the. Medicine, 1–15.

Ilango, A., Shumake, J., Wetzel, W., Scheich, H., & Ohl, F. W. (2012). The role of dopamine in the context of aversive stimuli with particular reference to acoustically signaled avoidance learning. 6(September), 1–9. https://doi.org/10.3389/fnins.2012.00132

Janssen, L. K., Sescousse, G., Hashemi, M. M., Harmina, M., Timmer, M., Peter, N., Everdina, D., Geurts, M., & Cools, R. (2015). Abnormal modulation of reward versus punishment learning by a dopamine D2-receptor antagonist in pathological gamblers. 3345–3353. https://doi.org/10.1007/s00213-015-3986-y

Kaufmann, C., Beucke, J. C., Preuße, F., Endrass, T., Schlagenhauf, F., Heinz, A., Juckel, G., & Kathmann, N. (2013). NeuroImage : Clinical Medial prefrontal brain activation to anticipated reward and loss in obsessive – compulsive disorder ☆. YNICL, 2, 212–220. https://doi.org/10.1016/j.nicl.2013.01.005

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), 1453–1461. https://doi.org/10.1371/journal.pbio.0040233

Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How We Learn to Make Decisions: Rapid Propagation of Reinforcement Learning Prediction Errors in Humans. 635–644. https://doi.org/10.1162/jocn

Lagemann, T., Rentzsch, J., Montag, C., Gallinat, J., Jockers-Scherübl, M., Winter, C., & Reischies, F. M. (2012). Early orbitofrontal hyperactivation in obsessive-compulsive disorder. Psychiatry Research - Neuroimaging, 202(3), 257–263. https://doi.org/10.1016/j.pscychresns.2011.10.002

Lai, F. D. M., & Ip, A. K. Y. (2011). Impulsivity and pathological gambling among Chinese: Is it a state or a trait problem? BMC Research Notes, 4(1), 492. https://doi.org/10.1186/1756-0500-4-492

Lau, B., & Glimcher, P. W. (2009). NIH Public Access. 58(3), 451–463.

Lauwereyns, J., Watanabe, K., & Coe, B. (2002). A neural correlate of response bias in monkey caudate nucleus. 418(JULY), 413–417. https://doi.org/10.1038/nature00844.1.

Lawrence, A. J., Luty, J., Bogdan, N. A., Sahakian, B. J., & Clark, L. (2009). Problem gamblers share deficits in impulsive decision-making with alcohol-dependent individuals. 1006–1015. https://doi.org/10.1111/j.1360-0443.2009.02533.x

Ledgerwood, D. M., Orr, E. S., & Kaploun, K. A. (2012). Executive Function in Pathological Gamblers. 89–103. https://doi.org/10.1007/s10899-010-9237-6

Linnet, J., Røjskjær, S., Nygaard, J., & Maher, B. A. (2006). Personality and Social Sciences Episodic chasing in pathological gamblers using the Iowa gambling task. 1987, 43–49.

Luigjes, J., Lorenzetti, V., de Haan, S., Youssef, G. J., Murawski, C., Sjoerds, Z., van den Brink, W., Denys, D., Fontenelle, L. F., & Yücel, M. (2019). Defining Compulsive Behavior. Neuropsychology Review, 29(1), 4–13. https://doi.org/10.1007/s11065-019-09404-9

Matsumoto, K., Suzuki, W., & Tanaka, K. (2018). Neuronal Correlates of Goal-Based Motor Selection in the Prefrontal Cortex. 301(5630), 229–232.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron, 38(2), 339–346. http://www.ncbi.nlm.nih.gov/pubmed/12718866

McHugh, S. B., Barkus, C., Huber, A., Capitao, L., Lima, J., Lowry, J. P., & Bannerman, D. M. (2014). Aversive Prediction Error Signals in the Amygdala. Journal of Neuroscience, 34(27), 9024–9033. https://doi.org/10.1523/JNEUROSCI.4465-13.2014

Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., & Kapur, S. (2007). Temporal Difference Modeling of the Blood-Oxygen Level Dependent Response During Aversive Conditioning in Humans: Effects of Dopaminergic Modulation. Biological Psychiatry, 62(7), 765–772. https://doi.org/10.1016/j.biopsych.2006.10.020

Miedl, S. F., Fehr, T., Herrmann, M., & Meyer, G. (2014). Risk assessment and reward processing in problem gambling investigated by event-related potentials and fMRI-constrained source analysis. 1–11.

Miedl, S. F., Peters, J., & Bu, C. (2012). Altered Neural Reward Representations in Pathological Gamblers Revealed by Delay and Probability Discounting. 69(2), 177–186.

Moeller, F. G., S.Barratt, E., Dougherty, D. M., Schmitz, J. M., & C.Swann, A. (2017). Psychiatric aspects of impulsivity. Yale Companion to Jewish Writing and Thought in German Culture, 1096-1996, November, 101–107. https://doi.org/10.2307/j.ctt1ww3vmm.21

Morein-Zamir, S., Papmeyer, M., Gillan, C. M., Crockett, M. J., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2013). Punishment promotes response control deficits in obsessive-compulsive disorder: Evidence from a motivational go/no-go task. Psychological Medicine, 43(2), 391–400. https://doi.org/10.1017/S0033291712001018

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. 9(8), 1057–1063. https://doi.org/10.1038/nn1743

Murray, G. K., Knolle, F., Ersche, K. D., Craig, K. J., Abbot, S., Shabbir, S. S., Fineberg, N. A., Suckling, J., Sahakian, B. J., Bullmore, E. T., & Robbins, T. W. (2017). Dopaminergic drug treatment remediates exaggerated cingulate prediction error responses in obsessive-compulsive disorder. BioRxiv, 1–34. http://biorxiv.org/content/early/2017/11/27/225938.abstract

Niv, Y. (1997). Reinforcement learning in the brain. 1–38.

Niv, Y., Edlund, J. A., Dayan, P., & Doherty, J. P. O. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. 32(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

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, 4(1), 95–102. https://doi.org/10.1038/82959

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain RID D-9230-2011. Neuron, 38(2), 329–337. https://doi.org/10.1016/S0896-6273(03)00169-7

Pauls, D. L., Abramovitch, A., Rauch, S. L., & Geller, D. A. (2014). Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nature Reviews Neuroscience, 15(6), 410–424.

https://doi.org/10.1038/nrn3746

Pearlson, G. D., & Potenza, M. N. (2013). monetary rewards and losses in pathological gambling. 71(8), 749–757. https://doi.org/10.1016/j.biopsych.2012.01.006.Diminished

Petry, N. M., Stinson, F. S., & Grant, B. F. (2005). Comorbidity of DSM-IV pathological gambling and other psychiatric disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. The Journal of Clinical Psychiatry, 66(5), 564–574.

Pezze, M. A., & Feldon, J. (2004). Mesolimbic dopaminergic pathways in fear conditioning. Progress in Neurobiology, 74(5), 301–320. https://doi.org/10.1016/j.pneurobio.2004.09.004

Potenza, M. N. (2008). The neurobiology of pathological gambling and drug addiction: an overview and new findings. July 3181–3189. https://doi.org/10.1098/rstb.2008.0100

Remijnse, P. L., Nielen, M. M. A., Van Balkom, A. J. L. M., Hendriks, G. J., Hoogendijk, W. J., Uylings, H. B. M., & Veltman, D. J. (2009). Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. Psychological Medicine, 39(9), 1503–1518. https://doi.org/10.1017/S0033291708005072

Remijnse, Peter L, Nielen, M. M. A., van Balkom, A. J. L. M., Cath, D. C., van Oppen, P., Uylings, H. B. M., & Veltman, D. J. (2006). Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. Archives of General Psychiatry, 63(11), 1225–1236. https://doi.org/10.1001/archpsyc.63.11.1225

Reuter, J., Raedler, T., Rose, M., Hand, I., Gläscher, J., & Büchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. 8(2), 147–148. https://doi.org/10.1038/nn1378

Reynolds, J. N. J., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. 67–70.

Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. Trends in Cognitive Sciences, 16(1), 81–91. https://doi.org/10.1016/j.tics.2011.11.009

Robbins, T. W., Gillan, C. M., Smith, D. G., Wit, S. De, & Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. 16(1), 81–91. https://doi.org/10.1016/j.tics.2011.11.009

Roesch, M. R., & Olson, C. R. (2018). Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex Published by: American Association for the Advancement of Science Stable URL: http://www.jstor.org/stable/3836780 digitize, preserve and extend access to Science Neurona. 304(5668), 307–310.

Romanczuk-seiferth, N., Koehler, S., Dreesen, C., Wüstenberg, T., & Heinz, A. (2009). Pathological gambling and alcohol dependence: neural disturbances in reward and loss avoidance processing. 557–569. https://doi.org/10.1111/adb.12144

Ruiter, M. B. De, Veltman, D. J., Goudriaan, A. E., Oosterlaan, J., & Sjoerds, Z. (2009). Response Perseveration and Ventral Prefrontal Sensitivity to Reward and Punishment in Male Problem Gamblers and Smokers. 1027–1038. https://doi.org/10.1038/npp.2008.175

Sachdev, P. S., & Malhi, G. S. (2005). Obsessive – compulsive behaviour: a disorder of decision-making.

Samejima, K. (2009). Representation of Action-Specific Reward Values in the Striatum Representation of

Action-Specific Reward Values in the Striatum. Science, 1337(2005), 1337–1341. https://doi.org/10.1126/science.1115270

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science, 310.

Samson, R. D., Frank, M. J., & Fellous, J.-M. (2010). Computational models of reinforcement learning: the role of dopamine as a reward signal. Cognitive Neurodynamics, 4(2), 91–105. https://doi.org/10.1007/s11571-010-9109-x

Samuel, R., Naomi, A., Andrew, D., Trevor, W., & Barbara, J. (2006). Motor Inhibition and Cognitive Flexibility in Obsessive-Compulsive Disorder and Trichotillomania.

Schultz, W, Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 13(3), 900–913. http://www.ncbi.nlm.nih.gov/pubmed/8441015

Schultz, Wolfram. (2016). Dopamine reward prediction error coding. Dialogues in Clinical Neuroscience, 23–32.

Schultz, Wolfram. (2018). Predictive Reward Signal of Dopamine Neurons.

Schultz, Wolfram, Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. 275(June 1994), 1593–1600.

Sescousse, G., Barbalat, G., Domenech, P., & Dreher, J. (2018). rewards in pathological gambling. April. https://doi.org/10.1093/brain/awt126

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2009). UKPMC Funders Group Author Manuscript Differential Encoding of Losses and Gains in the Human Striatum. Neuroscience, 27(18), 4826–4831. https://doi.org/10.1523/JNEUROSCI.0400-07.2007.Differential

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nature Neuroscience, 8(9), 1234–1240. https://doi.org/10.1038/nn1527

Sharp, M. E., Foerde, K., Daw, N. D., & Shohamy, D. (2017). Dopamine selectively remediates 'model-based' reward learning: a computational approach. November, 355–364. https://doi.org/10.1093/brain/awv347

Sources, D., Selection, S., Extraction, D., & Outcomes, M. (2017). Disruption of Reward Processing in Addiction An Image-Based Meta-analysis of Functional Magnetic Resonance Imaging Studies. 74(4), 387–398. https://doi.org/10.1001/jamapsychiatry.2016.3084

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. IEEE Transactions on Neural Networks, 9(5), 1054. https://doi.org/10.1109/TNN.1998.712192

Tom, S. M., Fox, C. R., Trepel, C., Poldrack, R. a, & Fox, R. (2014). The Neural Basis of Loss Aversion Under in Decision-Making under Risk. Science, 315(5811), 515–518.

Wh, J., D-h, K., Jy, H., Jh, J., B-m, G., J-s, C., & Mh, J. (2011). Aberrant ventral striatal responses during incentive processing in unmedicated patients with obsessive – compulsive disorder. 376–386. https://doi.org/10.1111/j.1600-0447.2010.01659.x

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. Journal of

Neuroscience, 26(37), 9530–9537. https://doi.org/10.1523/JNEUROSCI.2915-06.2006

Young, A. M. J. (2004). Increased extracellular dopamine in nucleus accumbens in response to unconditioned and conditioned aversive stimuli: Studies using 1 min microdialysis in rats. Journal of Neuroscience Methods, 138(1–2), 57–63. https://doi.org/10.1016/j.jneumeth.2004.03.003

Zhang, S., Mano, H., Ganesh, G., Robbins, T., & Seymour, B. (2016). Dissociable Learning Processes Underlie Human Pain Conditioning. Current Biology, 26(1), 52–58. https://doi.org/10.1016/j.cub.2015.10.066

2 A review of methodologies: neuroimaging and modelling

2.1 Methods background

In this chapter, we are going to review the most critical methods in this thesis - Functional magnetic resonance imaging (fMRI), and also computational modelling based on the typical reinforcement learning task introduced in ***Chapter*** 1. FMRI is a non-invasive neuroimaging technique that enables quantification of brain function with balance of temporal and spatial resolution. Over the past decades, fMRI has made substantial contributions to localize and understand normal human brain functions involved with various cognitive processes, including reward/avoidance learning. Also, fMRI has been suggested as a powerful diagnosis tool, which is effective and efficient in the detection and understanding of abnormal brain function under diverse clinical conditions. In this Chapter, we will review many aspects of fMRI including the underlying physics, task design and imaging processing.

The computational modelling is nowadays widely used to interpret the task and extract the feature variables from the behavioural data as well as probe the computational processes underlying the behaviour (Calder et al., 2018). Combining fMRI technique, fitting the model to the experimental data could find the related neural correlates of the calculated computational values. We will also review the modelling application and related fMRI imaging processing.

2.2 Physics of fMRI

2.2.1 MR physics

The fMRI images are obtained using the MRI scanner. During imaging scanning, the participant is exposed to a radiofrequency (RF) electromagnetic field pulse delivered through the head coil surrounding the participant's head. The main magnetic field (B_0) aligns the spins of protons in hydrogen atoms in the participant's brain along its axis. Those protons

absorb the energy at a very specific frequency band that depends on the field strength and become excited. The nuclei of these protons then start the relaxation process and emit the energy. Different tissue types have different relaxation times, thus creating the contrast among gray matter, white matter and cerebrospinal fluid (CSF) in the images. Several kinds of relaxation that are used to create contrast in images: T1, T2, and T2*. T1 is the rate at which spins relax back to the main magnetic field in the direction along the external magnetic field, usually referring to the z-axis. T1-weighted images have values across the image sensitive to the differences in T1 across gray matter, white matter and CSF, and thus providing excellent detailed images of brain anatomy. T2 refers to how quickly the total magnetic component decreases after emitted by the RF pulse dissipates, which also depends on the tissue type in the brain. T2-weighted images also provide excellent anatomical structure and additional detail in some subcortical, brainstem nuclei and many brain pathologies. T2* is the rate of attenuation of the magnetic field stimulated by the RF pulse, and it depends on local inhomogeneity in magnetic susceptibility that are caused by changes in blood flow and oxygenation (Wager & Lindquist, 2011). Such special effect is the fundamental of T2*-weighted imaging method, which is sensitive to blood oxygenation level-depend (BOLD) signal. More details will be provided in 2.1.2 later.

As the spin relaxes, the emitted energy is detected by the receiver coil. This energy is detected as a one-dimensional series of fluctuations over time. To reconstruct a three-dimensional (3D) image from these signals, gradients magnetically changing the strength of the magnetic field in systematic ways across space are applied. So that the frequency and the phase of the signals could be detected by the receiver coil encode the location of the signal in the brain. Pulse sequence is one of the techniques designed to implement particular patterns of RF and gradient manipulations.

2.2.2 BOLD signals

As mentioned above, T2*-weighted functional imaging is used to obtain measures of regional brain activity, thus offering a noninvasively experimental window to observe the human brain. The popular method uses the blood oxygenation level-depend (BOLD) signal based on the difference in T2* between oxygenated and deoxygenated haemoglobin (Voos & Pelphrey, 2013). When haemoglobin (red blood cell) is fully saturated with oxygen (oxyhemoglobin), it behaves as a diamagnetic substance. As neural activity increases, the metabolic demand for oxygen and nutrients also increase. When oxygen is extracted from the blood, the haemoglobin becomes paramagnetic, that creates small distortions in the B0 field that T2* decrease with faster decay of the signal. Increases in deoxyhemoglobin can lead to a decrease in BOLD signal. Such property links the local oxygen supply in the blood to the image contrast generated by the magnetic resonance. So the blood-oxygen-level dependent (BOLD) fMRI is introduced based on this concept, which is further developed into one of the principal imaging methods used to demonstrate regional, time-varying changes in brain activation (Glover, 2012).

Specifically, the BOLD signal changes after a stimulus e.g. visual picture, could be described as the hemodynamic response function (HRF). The first phase of stimulus is accompanied by a transient increase in deoxyhemoglobin concentration, which is called ‘dip’. This phenomenon is because the regional neural activation induced by stimulus consumed more oxygen than the supply in this area. Then, an increase in the oxy/deoxy-hemoglobin ratio leading to a high MR signal. This signal increase is proportional to the underlying neural activity. An “undershoot” is the last phase of HRF before the BOLD signal reaches the baseline (depicted in ***Figure 2-1***), which is again due to the imbalance and delay between the blood supply and consumption. Through the BOLD images, we could indirectly detect the

neuronal activity by subtracting the signal during a particular task (peak of HRF) to that during no task condition (baseline).

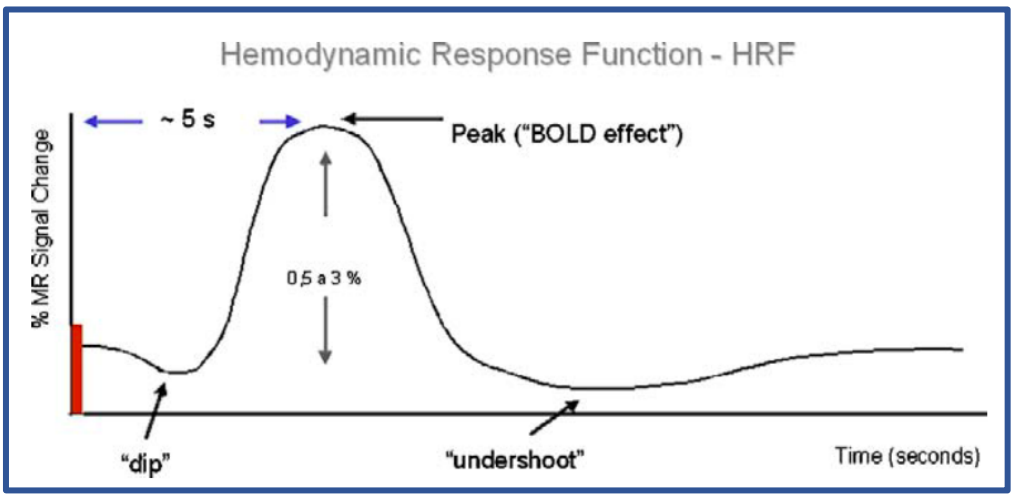

Figure 2-1 *Hemodynamic Response Function (HRF) in the fMRI from a hypothetical short duration stimulus (red bar) (Amaro & Barker, 2006).*

2.2.3 fMRI task design

The typical fMRI task activation experiments utilize visual, auditory or other stimuli to alternately induce two or more different cognitive states in the participants. Task-based fMRI detects neural activity based on comparison between one or more activation conditions relative to one or more control conditions. In another word, brain areas that significantly change (positively or negatively) along with the stimuli are identified as activated by the task. There are two types of fMRI designs: block or event-related design.

2.2.4 Block design and event design

Using a block design, the trials are arranged to alternate between the experimental and control conditions, with each block typically being a few tens of seconds (usually 15-30 seconds) long. In a simple blocked design, one control condition alternates with one activation condition. During each block, the subject either performs a continuous task or

responds to serial stimuli in relatively rapid succession. Activated cortex is identified by block-to-block periodic BOLD signal changes that are correlated with the task paradigm. Advantages of blocked designs include their simplicity and power for detection of an activation response. In particular, blocked paradigms summate the hemodynamic response over multiple neural events within each block, yielding relatively high BOLD contrast-to-noise ratio. Excessively low block frequencies are vulnerable to low-frequency noise (such as scanner drift) while excessively high block frequencies are vulnerable to attenuation of the BOLD response amplitude. A choice between 15 and 30 seconds often represents a reasonable compromise.

Different from the block design (shown in ***Figure 2-2***), the stimuli are presented randomly in event-related designs, which is similar as ERP (Event-Related Potential) in other neuroimaging methods, such as electroencephalography (EEG) or magnetoencephalography (MEG). The responses to trials belonging to each condition are selectively averaged and statistically compared. Responses can also be sorted according to the nature of the response: for example, correct responses can be separated from incorrect responses.

The advantages of event related design: greater control over cognitive stimuli, avoidance of cognitive adaptation that may occur during extended trials, more flexible analysis strategies, and great power to measure the hemodynamic response. Disadvantages, the slow time course of the hemodynamic response complicates the implementation and analysis of event-related paradigms. For block design is optimum for detecting activation and suitable for long-lasting stimulus (e.g., pain), but an event-related design is superior when characterization of the amplitude or timing of the hemodynamic response is desired (Matthews & Jezzard, 2004).

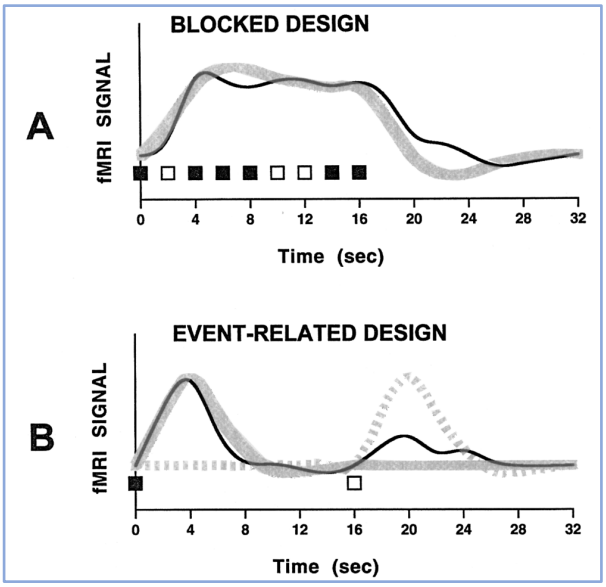

Figure 2-2 *Schematic diagram of the differences in the design of blocked versus event-related fMRI designs (D'Esposio et al., 1999).*

2.2.4.1 Typical cognitive task design example

As introduction in ***Chapter*** 1, decision making is a complex process, and reward/avoidance are two critically important components. Various cognitive tasks have been conducted in the MRI scanner to investigate the underlying brain mechanism. The reinforcement learning task (Kim et al., 2006), is one of the typical tasks designed to investigate the brain mechanism of reward learning and avoidance learning (shown in ***Figure 2-3***). The event-related task consists of three conditions: reward, avoidance and neutral. Under each condition, a specific pair of fractals was displayed, and the participants were required to choose one of them. Then, the chosen action was highlighted. Further, the outcome of the chosen action of getting reward or punishment was depicted. The specific pair of fractals under each condition has a higher and lower probability to get reward/punishment, respectively. This probability difference is to drive the participants' learning of the task. At the time of performing the task, participants underwent fMRI scanning sessions. To investigate the brain mechanism under the distinct phases of reward/avoidance conditioning

compared to the neutral condition, the conditions were pseudorandomly intermixed throughout the sessions.

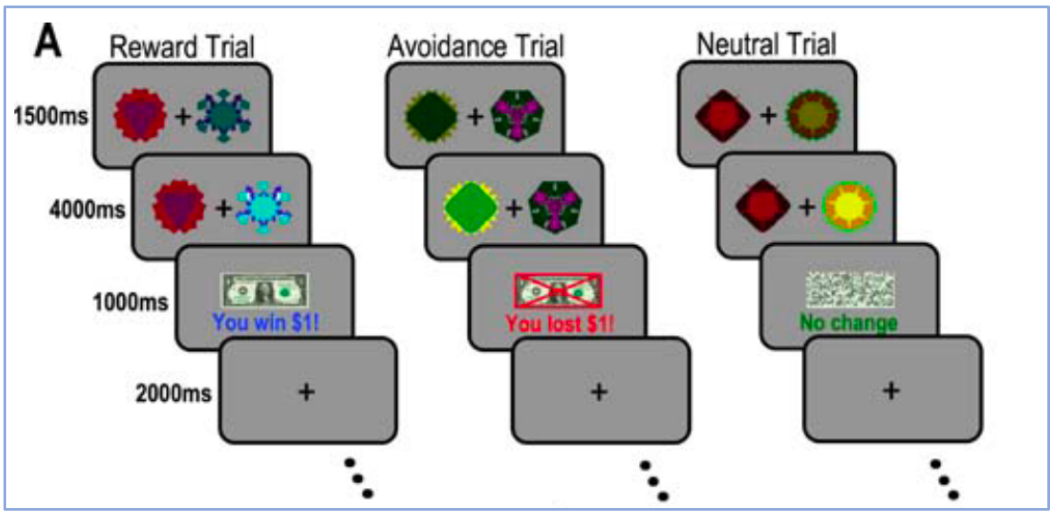

Figure 2-3 *The typical schematic of a RL learning task design included three conditions: Reward, Avoidance and Neutral; each condition has a pair of fractals with different probability of monetary reward, loss or no change (Kim et al., 2006).*

2.2.4.2 Modelling the task

As we have introduced in ***Chapter*** 1, the RL theory has been adopted as the main theoretical framework to interpret the task (Kim et al., 2006). Based on the RL theory, a computational model (advantage learning model) has been used to interpret the reward and avoidance-based decision process in the task (Kim et al., 2006). The detailed description is as follows:

Advantage learning uses a temporal difference (TD) learning rule to learn value prediction of future reward. In temporal difference learning, the prediction $\hat{V}(t)$ of the value $V(t)$ at any time t within a trial is calculated as a linear product of the weights ω_i and the presence or absence of a conditioned stimulus (CS) at time t, coded in the stimulus representation vector $x_i(t)$:

$$\hat{V}(t) = \sum_i \omega_i x_i(t);$$

Learning occurs by updating the predicted value of each time-point in the trial by comparing the value at time t+1 that at time t, leading to a PE or:

$$\delta(t) = r(t) + \gamma\hat{V}(t+1) - \hat{V}(t);$$

Where r(t) is the reward at time t; The parameter γ is a discount factor that determines the extent to which rewards that arrive earlier are more important than rewards that arrive later on. Set $\gamma = 1$; the weights ω_i are then updated on a trial-by-trial basis according to the correlation between PE and the stimulus representation

$\omega_i = \alpha \sum_t \quad x_i(t)\delta(t)$; α is the learning rate.

Assign six time points to each trial, and use each participants' individual event history as input. Setting r(t) as -1, 0 or 1 to denote receipt of a reward outcome, no outcome, or an aversive outcome, respectively. On each trial, the CS was delivered at time point 1, the choice was made at time point 2, and the reward was delivered at time point 6. For the analysis, reward PEs are calculated for the specific CS that was illuminated: at the time of choice, where $\hat{V}(t)$ was generated based on just one of the two stimuli shown. Here, the PE signal is a variant of $\delta(t)$ known as the advantage PE signal $\delta^A(t)$;

$$\delta^A(t) = r(t) + \gamma\hat{V}(t+1) - \hat{Q}(t,a);$$

$\hat{Q}(t,a)$ corresponds to the value of the specific chosen action at time t, and V(t) is the value of the state at the current time t as calculated in Equation 1 above.

And these action values are used to determine the probability of choosing a given action using a logistic sigmoid:

$$p(t,a) = \sigma(\beta(\hat{Q}(t,a) - \hat{Q}(t,b)));$$

where β is an inverse temperature that determines the ferocity of the competition? This probability is then used to define the value of the initial state at t =1 as:

$$\hat{V}(1) = p(1,a)\hat{Q}(1,a) + p(1,b)\hat{Q}(1,b);$$

2.2.4.3 Brain targets under the task

Combining the fMRI technique and modelling, the brain mechanism of reward/avoidance-related decision process has been revealed. At the outcome stage, the activity of the brain area *medial orbitofrontal cortex (OFC)* increases after receiving reward as well as avoiding an aversive outcome. At the expectation stage, the *medial and lateral OFC* was found correlated with expected reward value in both the reward and avoidance

trials. Last, the *ventral striatum* extending from the ventral putamen into the nucleus accumbens proper was found correlating with the reward PE signal derived from the model. While *left and right insula* was found significantly correlated with an aversive PE signal on avoidance trials.

2.3 Task-based fMRI processing

For task-based fMRI processing, it usually includes three steps: i) imaging quality control, ii) imaging pre-processing, and iii) post-processing. All the image processing was realized through the statistical parametric modulation (SPM, version 12) on Matlab v2015a environment on the platform of Massive (https://www.monash.edu/research/infrastructure/platforms-pages/massive).

2.3.1 Imaging quality control

Ensuring the quality of neuroimaging data is becoming crucial as the first step for any imaging analysis workflow, this process includes identify and exclude the low-quality images to increase the reproducibility (Fessler, Michael B.; Rudel, Lawrence L.; Brown, 2008a; O'connor, 2016). In the early days, manual quality control (QC) was used to entail screening every single image of a dataset individually. However, manual QC suffers at least two problems: unreliability and time-consuming nature for large datasets. Automated QC has now attracted great attention with the convergence of machine learning solutions (Gedamu et al., 2008). The automated methods estimate image quality using "image quality metrics" (IQMs) that quantify variably interpretable aspects of image quality (e.g., summary statistics of image intensities, signal-to-noise ratio, coefficient of joint variation, Euler angle, etc.). The web application program interface (web-API) for QC of magnetic resonance imaging data (MRIQC) (https://mriqc.readthedocs.io/en/latest/workflows.html) provides a unique platform

to perform the QC automatically. It was reported that over 50,000 and 60, 000 records of anatomical and functional IQMs has been collected via the platform of MRIQC (shown in ***Figure 2-4***) (Esteban et al., 2019).

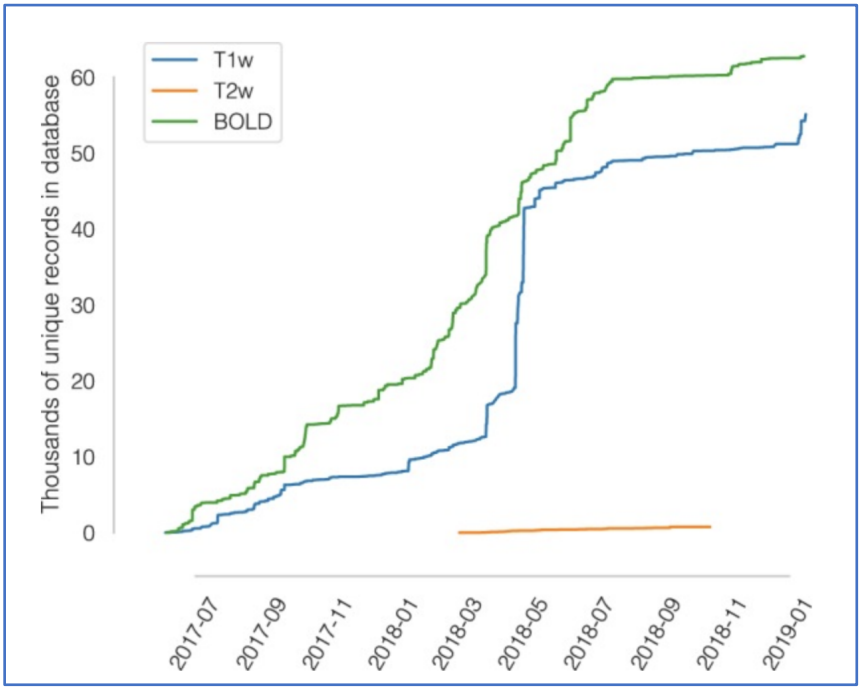

Figure 2-4 *The MRI quality control cases increase rapidly. Database has accumulated over 60,000 records of IQMs for BOLD images. Records are shown after exclusion of duplicated images (Esteban et al., 2019).*

MRIQC is an open-source project, which is compatible with input data formatted according to the Brain Imaging Data Structure (BIDS) standard (Esteban et al., 2017). It extracts the IQMs for each subject's image data and generates the summary report. The tool was developed under several engineering principles: 1) Modularity and integrability with implementation of a nipype workflow to integrate modular sub-workflows that rely upon third party software toolboxes such as FSL and AFNI (Gorgolewski et al., 2011). 2) Minimal preprocessing: the workflow should be as minimal as possible to estimate the IQMs (shown in ***Figure 2-5***). 3) Interoperability and standards: MRIQC is compatible with input data

formatted according to the BIDS standard. 4) Reliability and robustness: the software undergoes frequent vetting sprints by testing its robustness against data variability.

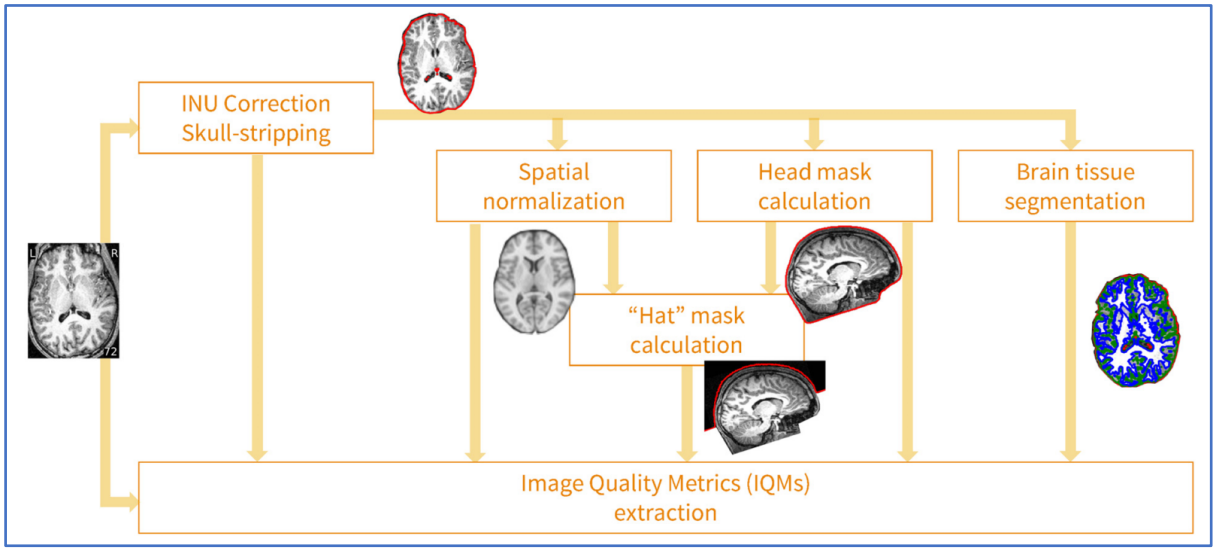

Figure 2-5 *MRIQC's processing data flow. Images undergo a minimal processing pipeline to obtain the necessary corrected images and masks required for the computation of the IQMs (Esteban et al., 2017).*

The IQMs for fMRI could be grouped in four broad categories (shown in ***Table 2-1***), providing a vector of 64 features per bold scan (Esteban et al., 2019). For example, the item DWARS measures the temporal variations by calculating the rate of change of BOLD signal across the entire brain at each frame of data. The FD (framewise displacement) is proposed to regress out instantaneous head-motion in fMRI studies (Fessler, Michael B.; Rudel, Lawrence L.; Brown, 2008b).

Table 2-1 *Summary table of image quality metrics for functional (BOLD) MRI (Esteban et al., 2019). MRIQC produces a vector of 64 image quality metrics (IQMs) per input BOLD scan.*

IQMs measuring temporal variations

tSNR	A simplified interpretation of the original temporal SNR definition by Kruger et al (Krüger & Glover, 2001). We reported the median value of the tSNR map calculated as the average BOLD signal across time over stthe corresponding temporal s.d. map.
GCOR	Summary of time-series correlation as in (Saad et al., 2013) using AFNI @compute_gcor
DVARS	The spatial standard deviation of the data after temporal differencing. Indexes the rate of change of BOLD signal across the entire brain at each frame of data. DVARS is calculated using Nipype, after head-motion correction.
IQMs targeting specific artifacts	
FD	The six-realignment displacement – proposed by Power et al (Fessler, Michael B.; Rudel, Lawrence L.; Brown, 2008b) to indicate instantaneous head-motion in fMRI studies. MRIQC reports the average FD.
GSR	The Ghost to Signal Ratio (Giannelli et al., 2010) estimates the mean signal in the areas of the image that are prone to N/2 ghosts in the phase encoding direction with respect to the mean signal within the brain mask. Lower values are better.
DUMMY	The number of dummy scans – A number of volumes at the beginning of the fMRI time-series identified as nonsteady states.
IQMs from AFNI	
AOR	AFNI's outlier ratio – Mean fraction of outliers per fMRI volumes as given by AFNI's 3dToutcount
AQI	AFNI's quality index – Mean quality index as computed by AFNI's 3dTqual
IQMs measuring spatial information	
EFC	The entropy-focus criterion (Atkinson et al., 1997) uses the Shannon entropy of voxel intensities as an indication of ghosting and blurring induced by head motion. Lower values are better.

FBER	The foreground-background energy ratio (Atkinson et al., 1997) is calculated as the mean energy of image values within the head relative to the mean energy of image values in the air mask. Consequently, higher values are better.
FWHM	The full-width half-maximum (Pezzulo et al., 2013) is an estimation of the blurriness of the image calculated with AFNI's 3d FWHMx. Smaller is better.
SNR	MRIQC includes the signal-to-noise ratio calculation proposed by Dietrich et al. (Dietrich et al., 2007), using the air background as noise reference. Additionally, for images that have undergone some noise reduction processing, or the more complex noise realizations of current parallel acquisitions, a simplified calculation using the within tissue variance is also provided.
SSTATs	Several summary statistics (mean, standard deviation, percentiles 5% and 95%, and kurtosis) are computed within the following regions of interest: background, CSF, WM and GM.

2.3.2 Imaging preprocessing

As the goal of fMRI study is to identify brain areas activated by the task, the preprocessing of fMRI data is needed to perform to remove the variability in raw data. The variability in raw fMRI data could be great to swamp out the small changes in the BOLD response induced by most cognitive tasks. Some unavoidable variability such as thermal or system noise couldn't be controlled, but other sources of variability are measurable. For example, when a subject moves the head, the BOLD response could be sampled from each spatial position within the scanner and suddenly changes in a predictable manner. The preprocessing could remove those artefacts at certain degrees from the data. Generally, preprocessing steps must be performed prior to the statistical analysis of fMRI data. These steps have two primary goals: 1) to reverse displacements of the data in time or space that

may have occurred during acquisition, 2) to enhance the ability to detect spatially extended signals within or across participants.

In our study, we focus on a number of common preprocessing steps. Firstly, the slice timing correction is used to correct for variability in the BOLD responses that are due to the fact that data in different slices are acquired at different times. Secondly, the realignment is to correct for variability due to head movement. Thirdly, the coregistration is to align the structural and functional data, and then the normalization is to warp the subject's functional image to a standard space according to the structural information. Lastly, spatial smooth is done to reduce the nonsystematic high-frequency spatial noise.

SLICE TIMING

The fMRI data are acquired in slices using sequential 2D imaging techniques like single-shot echo planar imaging sequences. As the data analysis is essentially a time course analysis, exact timing with respects to the stimulus presentation paradigm is crucial. As a whole volume can be acquired within typical repetition times (TR) ranging from hundreds of milliseconds to several seconds, a slice acquisition delay between slices is generated. Thus, the delay between slices could add up to significant temporal shifts over the full volume between the expected and actually measured hemodynamic responses (shown in ***Figure 2-6***). In order to compensate for this slice acquisition delay, the slice timing correction has been proposed as a necessary pre-processing step (Calhoun et al., 2000; Henson et al., 2002). To do the slice timing correction, the individual slice is temporally realigned to a reference slice based on its relative timing using an appropriate resampling method. Usually, the linear, sinc and cubic spline interpolation could be used. Slice timing correction was demonstrated to

improve the sensitivity in group statistical analysis, particularly true for event-related designs and task designs (Sladky et al., 2011).

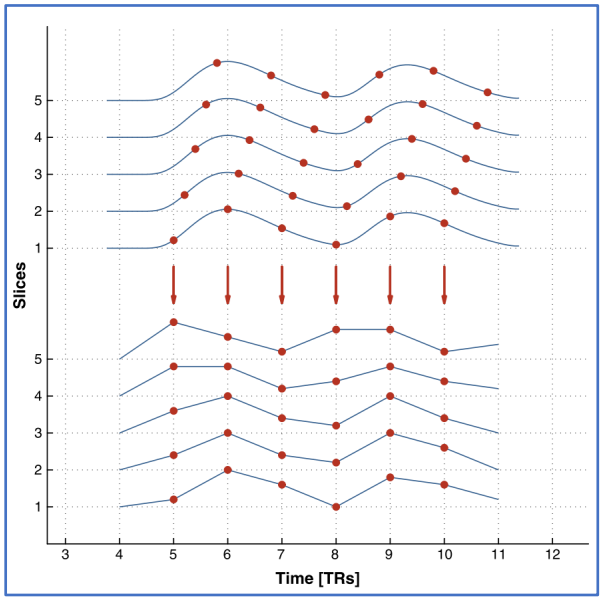

Figure 2-6 *Illustration of the slice-timing problem (Sladky et al., 2011). The hemodynamic responses of the individual slices are acquired at different time points. Without adequate compensation, this will lead to biased estimators in fMRI analysis.*

SPATIAL REALIGNMENT (HEAD MOTION CORRECTION)

Correcting for head motion is the critically important preprocessing step. Correction for head movements based on the assumption that when a subject moves his or her head, the brain does not change shape or size. Under this circumstance, the brain could be treated as a rigid body. Thus, the head movement correction then becomes a problem of rigid body registration. Any movement of a rigid body can be described by six parameters. When a person lies in the MRI scanner, the centre of any voxel could be described as a set of three coordinate values. Based on the coordinate system, the only possible rigid body movements are 1) a translation along the x axis, 2) a translation along the y axis, 3) a translation along the z axis, 4) a rotation about the x axis, 5) a rotation about a single parameter. Each of these

translations could be characterized by the distance moved along that axis and each of these rotations characterized by the angle of rotation. Suppose collecting BOLD responses from the whole brain on two separate TRs only with head movement. In order to correct for this head movement, the two sets of data need to be back into spatial alignment. One of the correction strategies is to take the data from the first TR as the standard and then perform rigid body movements on the data from the second TR until the BOLD responses from the two TRs agree as closely as possible at each coordinate point. The process is called rigid body registration. The goal of rigid body registration is to find the values of the six parameters of the equation that align the two data sets as closely as possible.

With interleaved slice acquisition, a small head movement that moves a point in the brain into a neighbouring slice will cause a significant change in acquisition time. As a result, timing differences will accentuate the effects of head movement. To solve this problem, slice-timing corrections are usually made before head movement corrections when interleaved slice acquisition is used.

To achieve sufficient temporal and spatial resolution, echo-planar imaging (EPI) technique is one of the most used imaging protocols. Multiple phases encoded trajectories will be acquired within a single TR period. Thus, these sequences are sensitive to the main magnetic field (B0) because of the undesired accumulation of signal phase occurring during the relatively long encoding periods. In EPI sequences, this phase accumulation results in spatial distortions in the resulting reconstructed images. The distortions appear as geometric warping of the image, predominantly in the phase-encoding dimension (Elliott et al., 2004). It is also preferable to perform motion realignment before correction for geometric distortion. Applying this approach to real EPI data, greater subject motion was detected, and superior realignment was achieved.

NORMALIZATION

The spatial resolution of the functional data is poor with a common voxel size of 3 mm * 3 mm * 3 mm. Whereas the voxel size in the structural image might be 0.86 mm * 0.86 mm * 0.86 mm. The coregistration is to improve spatial localization of the functional data using the enhanced resolution of the structural data. Different with head movement correction, only the structural image and any one functional image must be aligned in the coregistration. Individual differences existed in the sizes and shapes of individual brains, and these differences make it difficult to assign a task-related activation observed in some cluster of voxels to a specific neuroanatomic brain structure. To address this problem, the structural scan of each subject needs to be registered to some standard brain with identified coordinates in an atlas. This process of registering a structural scan to the structural scan from some standard brain is called normalization.

The MNI atlas produced by the Montreal Neurological Institute (MNI) is commonly used. And the MNI atlas was created by averaging the results of high-resolution structural scans that were taken from 152 different brains. The origin of the coordinate system is set to the midpoint of the anterior commissure. The normalization will include not only the rigid body differences between the standard brain and the brain of typical participants, but also the size and shape differences. Size differences could be accommodated via a linear transformation, but a nonlinear transformation is almost always required to alter the shape of a subject's brain to match the MNI standards. Besides the alignment differences, there will be intensity differences between the subject's image and the reference image.

SPATIAL SMOOTHING

To do the spatial smoothing, the BOLD value in each voxel is replaced by a weighted average of the BOLD responses in neighbouring voxels. This will essentially blur the data at

each TR by smoothing off peaks and filling in valleys. To do the spatial smoothing, a three-dimensional filter is applied to the BOLD responses. For the fMRI data, the common measure of width is the full width at half maximum (FWHM) (shown in ***Figure 2-7***). In practice, a common choice for kernel width is somewhere between 1 and 3 voxel widths. Such as a standard choice for the voxel size 3 mm * 3mm * 3.5 is FWHMx = FWHMy = 6 mm and FWHMz = 7 mm.

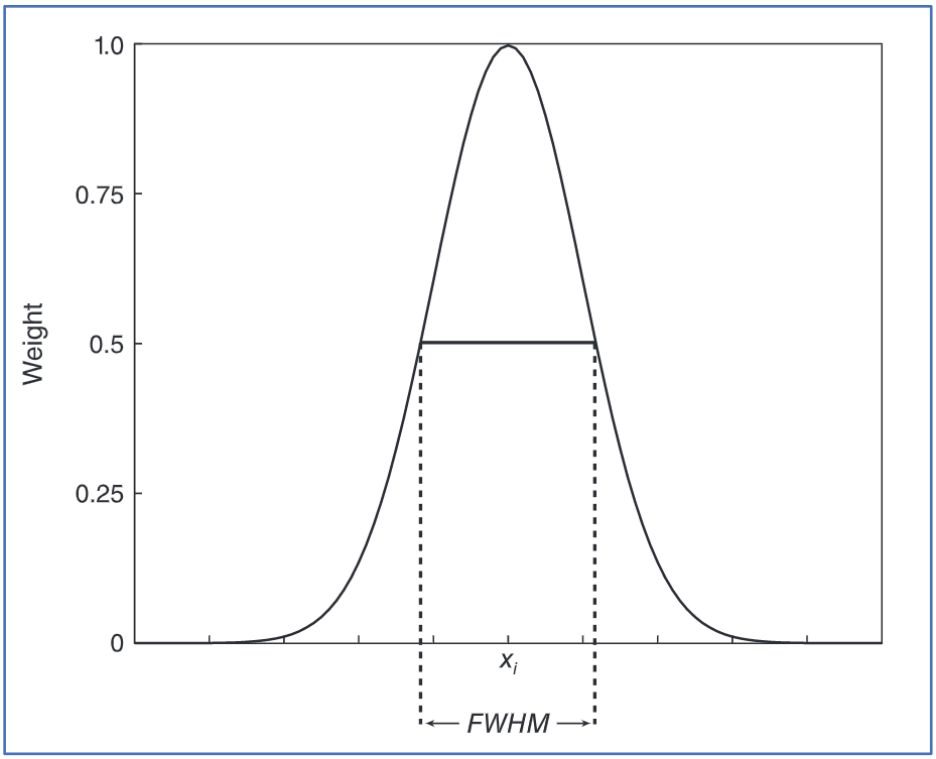

Figure 2-7 *A Gaussian smoothing kernel and an illustration of its width parameter FWHM.*

2.3.3 Imaging post processing

After preprocessing, the next step is to examine the research hypothesis of the designed experiment. The activation maps are aligned and normalized into the same space, which is suitable for the further voxel-based analysis to identify corresponding brain regions activated by the task. The most popular approach is a general linear regression (GLM) based method that is the foundation of the fMRI software packages (Friston et al., 1994). Combining with modelling, the fMRI post processing consists of several steps including

generating the modelling function, regression through the fMRI imaging data and finally statistical analysis.

MODELLING FUNCTION

Predicting the BOLD response to each stimulus event firstly is to make an assumption about how long the neural activation will last in brain regions that process this event. Usually, it is assumed that the neural activation induced by the event onset will persist for as long as the stimulus is visible to the participants. Alternatively, the neural activations persist until the participants respond with a duration equalling subject's response time. Then, all presumed neural activations are modelled via a boxcar function (pulse signal). The function persists for the duration of fMRI data acquisition and equals 1 when neural activation is assumed to be present and 0 when activation is absent (example shown in ***Figure** 2-8*). The correlation method assumes linearity, and next in the analysis is to convolve the boxcar function with the canonical HRF as described in ***chapter*** 2.1.2.

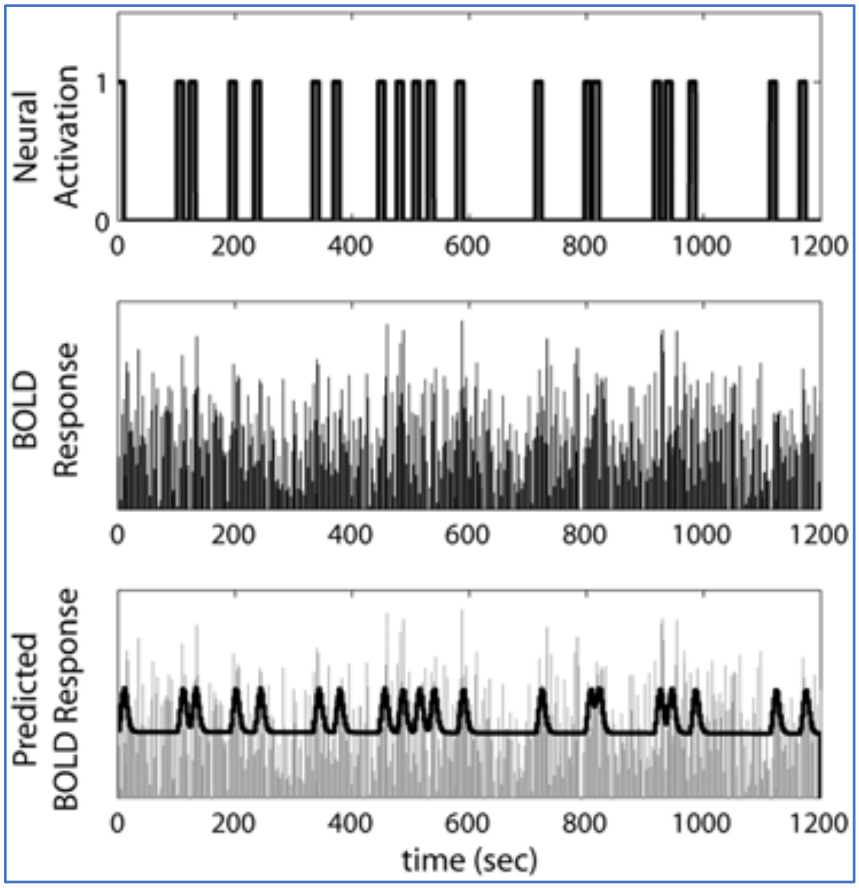

Figure 2-8 *A hypothetical example of the standard correlation-based analysis of fMRI data. The top panel shows the boxcar function that models the presumed neural activation elicited by the presentation of 20 separate stimuli. The middle panel depicts the hypothetical BOLD response in the experiment in a voxel with task-related activity. The bottom panel shows the best-fitting predicted BOLD response that is generated by convolving an HRF with the boxcar function shown in the top panel.*

REGRESSION

The final step is to correlate these predicted BOLD values with the observed BOLD response in every voxel. Voxels where this correlation is high are presumed to show task-related activity. Correlation is typically done within the context of the familiar General Linear Model (GLM) that is the basis of both multiple regression and analysis-of-variance. As the correlation method applied to data from a single voxel at a time, thus if an experiment collects data from the whole brain, this analysis could easily be repeated more than 100,000

times to analyse all of the data collected in the experiment. The result of all these analyses is a value of the test statistic in every voxel that was analysed. The resulting collection of statistics is often called a statistical parametric map, which motivated the name of the well-known fMRI data analysis software package, SPM.

Take the introduced RL task in ***chapter 2.2.2*** for example, the variables derived from the computational model are firstly transferred into the time series, thus generating a regressor. Carefully, the regressor has to be associated with particular time points in the experiment such as a PE signal occurred specifically at the time of the outcome within the trial. Then, this time series was convolved with the HRF (as shown in ***Figure 2-9***) to account for the delay induced by the hemodynamic response. Finally, this newly generated regressor can be then included as a predictor variable in a single-subject fMRI design matrix through multiple linear regression analysis techniques. A statistical contrast on the parameter estimation yields a map with those brain regions related to model-derived variables (Doherty & Gl, 2010) (shown in ***Figure 2-9***).

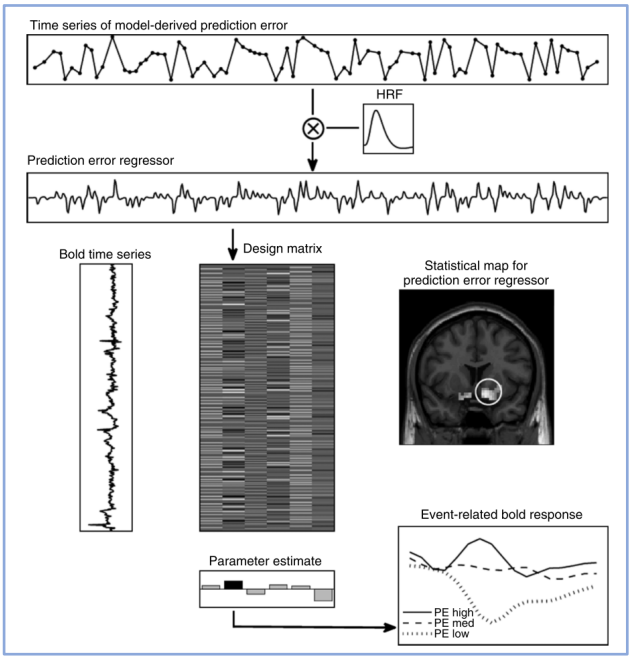

Figure 2-9 *The flow of fMRI image processing combined with the computational model of the introduced RL task. Internal variables derived from the model converted into a time series and convolved with a HRF, thus yielding a regressor in a single-subject fMRI design matrix. This general linear model is fitted at each voxel in the brain, and generated a statistical map describing the degree of correlation between activity in a particular BOLD time series voxel and the internal variable of interest (Doherty & Gl, 2010).*

STATISTICAL ANALYSIS

Similar to the voxel-based regression analysis above, the statistical analysis is also conducted at voxel based. Each voxel is considered as an independent statistical test. There are various GLM based statistical models, including simple two sample t-test, ANOVA, ANCOVA, regression analysis, fully or flexible factorial analysis. In this thesis, the simple two-group t-test is chosen to examine the difference between the control and clinical groups. Similar to the traditional t-test, the statistical significance of the estimated t-value, a p value is defined as the chance of observing a statistical t-value or more extreme results under the null hypothesis. If a voxel's p value is smaller than user defined significance level a, we can hence reject the null hypothesis and classify the voxel as 'active'. To correct for the multiple comparison error, a family wise correction (FWE) is used at brain cluster level. The statistical analysis is realized on xjview (https://www.alivelearn.net/xjview/).

2.4 Research gap

2.4.1 Task design

As introduced in Chapter 1, reward learning in the context of positive and negative one has been successfully described by reinforcement learning models. According to these models, animals learn the uncertain values of positive and negative rewards by updating their subjective valuations of stimuli based on their past experience with those stimuli (Sutton and Barto, 2015). The learning process could be dissected into several signals including outcome,

expected value and critical PE signal. Our understanding of the neural processes associated with learning about reward or avoidance of punishment is still limited by the small number of imaging studies delineating their distinct and/or by introducing a reward/avoidance learning task with probability switch approximately in the middle of the task. To switch the rewarding (or avoidance) probabilities of the two factorials in the task is a plausible practice to increase the task complexity and serve as the novelty of the task. Combined with the aforementioned fMRI technique, we comprehensively investigated brain regions encoding reward and avoidance PE signal, especially in the midbrain and cortical brain areas.

2.4.2 Application to clinical condition

Many psychiatric conditions are associated with participants' aberrant decision processes. For example, people with obsessive compulsive disorder (OCD) repeat endlessly a behaviour such as handwashing; And people with gambling disorder (GD) often seek and engage in risky forms of gambling, despite explicitly acknowledging the harms that may follow. OCD is a relatively chronic and disabling neuropsychiatric disorder with an estimated prevalence between 1-3% (Figee et al., 2011), while GD is classified as a behavioural addiction with a lifetime prevalence of 0.5-1% (Petry et al., 2005). A high burden of individual and socioeconomic cost was caused by the maladaptive decision patterns in both clinical conditions (Fujino et al., 2018; Nestadt et al., 2018). Through the application of proposed reinforcement learning tasks, with a probability switch, we investigated the potential aberrant brain mechanism of reward and avoidance-based decision process in the people with OCD and GD through fMRI technique and computational modelling. Also, we examined how the behavioural constructs of impulsivity and compulsivity affect the reward and avoidance decision processes in OCD and GD."

References

Amaro, E., & Barker, G. J. (2006). *Study design in fMRI : Basic principles*. *60*, 220–232. https://doi.org/10.1016/j.bandc.2005.11.009

Atkinson, D., Hill, D. L. G., Stoyle, P. N. R., Summers, P. E., & Keevil, S. F. (1997). Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Transactions on Medical Imaging*, *16*(6), 903–910. https://doi.org/10.1109/42.650886

Calder, M., Craig, C., Culley, D., De, R., Donnelly, C. A., Douglas, R., Gascoigne, J., Gilbert, N., Hargrove, C., Hinds, D., Lane, D. C., Mitchell, D., Pavey, G., Robertson, D., Rosewell, B., & Sherwin, S. (2018). *Computational modelling for decision-making : where , why , what , who and how Subject Category : Subject Areas :*

Calhoun, V., Golay, X., & Pearlson, G. (2000). Improved fMRI slice timing correction: interpolation errors and wrap-around effects. *Proceedings, ISMRM, 9th Annual Meeting,* 819.

D'Esposio, M., Zarahn, E., & Aguirre, G. K. (1999). *Event-related Functional MRI: Implications for Cognitive Psychology*.

Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F., & Schoenberg, S. O. (2007). Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging*, *26*(2), 375–385. https://doi.org/10.1002/jmri.20969

Doherty, J. P. O., & Gl, J. P. (2010). *Model-based approaches to neuroimaging : combining reinforcement learning theory with fMRI data*. 501–510. https://doi.org/10.1002/wcs.57

Dumais, A., & Bitar, N. (2018). *Loss anticipation and outcome during the Monetary Incentive Delay Task : a neuroimaging systematic review and meta-analysis*. 1–23. https://doi.org/10.7717/peerj.4749

Elliott, M. A., Gualtieri, E. E., Hulvershorn, J., Ragland, J. D., & Gur, R. (2004). The effects of geometric distortion correction on motion realignment in fMRI. *Academic Radiology*, *11*(9), 1005–1010. https://doi.org/10.1016/j.acra.2004.04.022

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS One*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Blair, R. W., Nielson, D. M., Varada, J. C., Marrett, S., Thomas, A. G., Poldrack, R. A., & Gorgolewski, K. J. (2019). Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines. *Scientific Data*, *6*(1), 1–7. https://doi.org/10.1038/s41597-019-0035-4

Fessler, Michael B.; Rudel, Lawrence L.; Brown, M. (2008a). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Fessler, Michael B.; Rudel, Lawrence L.; Brown, M. (2008b). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. https://doi.org/10.1002/hbm.460020402

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1297–1310.

https://doi.org/10.1016/j.neubiorev.2013.03.023

Gedamu, E. L., Collins, D. L., & Arnold, D. L. (2008). Automated quality control of brain MR images. *Journal of Magnetic Resonance Imaging*, *28*(2), 308–319. https://doi.org/10.1002/jmri.21434

Giannelli, M., Diciotti, S., Tessa, C., & Mascalchi, M. (2010). Characterization of Nyquist ghost in EPI-fMRI acquisition sequences implemented on two clinical 1.5 T MR scanner systems: Effect of readout bandwidth and echo spacing. *Journal of Applied Clinical Medical Physics*, *11*(4), 170–180. https://doi.org/10.1120/jacmp.v11i4.3237

Glover, G. H. (2012). *NIH Public Access*. *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001.Overview

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*(August). https://doi.org/10.3389/fninf.2011.00013

Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. *NeuroImage*, *15*(1), 83–97. https://doi.org/10.1006/nimg.2001.0940

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), 1453–1461. https://doi.org/10.1371/journal.pbio.0040233

Kim, S. H., Yoon, H. S., Kim, H., & Hamann, S. (2014). Individual differences in sensitivity to reward and punishment and neural activity during reward and avoidance learning. *Social Cognitive and Affective Neuroscience*, *10*(9), 1219–1227. https://doi.org/10.1093/scan/nsv007

Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). *How We Learn to Make Decisions : Rapid Propagation of Reinforcement Learning Prediction Errors in Humans*. 635–644. https://doi.org/10.1162/jocn

Krüger, G., & Glover, G. H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*(4), 631–637. https://doi.org/10.1002/mrm.1240.abs

Matthews, P. M., & Jezzard, P. (2004). *Functional magnetic resonance imaging*. 6–12.

O'connor. (2016). Subtle In-Scanner Motion Biases Automated Measurement of Brain Anatomy From in Vivo MRI. *Physiology & Behavior*, *176*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Pezzulo, A. a, Tang, X. X., Hoegger, M. J., Alaiwa, M. H. A., Ramachandran, S., Moninger, T. O., Karp, P. H., Wohlford-, C. L., Haagsman, H. P., Eijk, M. Van, Bánfi, B., Horswill, A. R., Hughes, H., Roy, J., & College, L. a C. (2013). Test-restest between-site relaibility in a multicenter fMRI study. *Human*, *487*(7405), 109–113. https://doi.org/10.1038/nature11130.Reduced

Saad, Z. S., Reynolds, R. C., Jo, H. J., Gotts, S. J., Chen, G., Martin, A., & Cox, R. W. (2013). Correcting brain-wide correlation differences in resting-state FMRI. *Brain Connectivity*, *3*(4), 339–352. https://doi.org/10.1089/brain.2013.0156

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage*, *58*(2), 588–594. https://doi.org/10.1016/j.neuroimage.2011.06.078

Sutton, R. S., & Barto, A. G. (2015). *Reinforcement Learning : An Introduction*.

Voos, A., & Pelphrey, K. (2013). Functional Magnetic Resonance Imaging. *Journal of Cognition and Development*, *14*(1), 1–9. https://doi.org/10.1080/15248372.2013.747915

Wager, T. D., & Lindquist, M. A. (2011). Essentials of Functional Magnetic Resonance Imaging. In *The Oxford Handbook of Social Neuroscience* (Issue September). https://doi.org/10.1093/oxfordhb/9780195342161.013.0006

Zhang, S., Mano, H., Ganesh, G., Robbins, T., & Seymour, B. (2016). Dissociable Learning Processes Underlie Human Pain Conditioning. *Current Biology*, *26*(1), 52–58. https://doi.org/10.1016/j.cub.2015.10.066

3 Investigation of reward and avoidance decision processes in healthy young adults through a novel probabilistic reward and avoidance learning task

Aim: In this chapter, we are going to examine participants' behavioural response during the reward/avoidance learning task through the statistical analysis as well as modelling, and the statistical analysis were from the three measurements: 1) the response time of choice making in reward/avoidance condition in comparison with neutral condition; 2) the number of Correct and Incorrect fractal choice as a proxy of reward and avoidance conditioning; and 3) the learning curve of Correct and Incorrect fractal choice in reward/avoidance condition to model the participants' learning of the task. Further, A Q-learning model was applied to model the participants' trial-by-trial learning process. We have tried several ways to estimate the values of model parameters. Firstly, the negative loglikelihood has been used with linear searching a pair of parameters to make the negative loglikelihood sum of each individual participant minimized, and then the Matlab fmincon was used to broad the searching ranges. Both methods were only considering the individual level optimization, and there was some boundary value that existed. Thus, the Bayesian model was finally used for parameter calculation, which considered both individual and group level effects. The two characteristic parameters learning rate and inverse temperature parameter were estimated, which showed the participants' learning and the balance of exploitation versus exploration on the choice, respectively.

Questions: We are interested in 1) the participants' behavioural performance including response time and correct choice ratio under the reward/avoidance condition in the reward/avoidance learning task. 2) the participants' learning characteristics including learning rate and inverse temperature parameter under the reward/avoidance condition. 3)

also, the potential differences of behavioural performance and learning merits between the reward and avoidance condition.

Hypothesis: Based on the previous studies, we hypothesized: 1) Participants will show shorter response time in reward condition and longer in avoidance condition, compared to the neutral condition. 2) Participants will prefer the Correct choices significantly over the Incorrect in reward/avoidance condition, compared to the neutral condition. 3) Participants will show learning of the Correct choice, and a probability switch, as observable in their learning curve. 4) Through the modelling, participants will show a higher learning rate & inverse temperature parameter under the reward condition compared to avoidance condition.

3.1 Introduction

As discussed in previous chapters, decision making can be modelled as a process involved with the choice selection from the available alternatives. In order to optimize this process, the participants would estimate the outcome of the different options, which is based on reward and punishments associated with these alternatives in the past choices. Seeking rewards and avoiding punishments is a propensity to human participants in order to survive, which indicates that reward/avoidance processing and associated learning are important components of the decision making process. Some interesting questions raised, like: 1) how participants carry out the reward/avoidance associated decision making, and 2) what's the similarities and differences of the behavioural sensitivity and performance in those reward/avoidance related learning? For how humans acquire their preferences for different options and outcomes in the decision making process, it is suggested that we always act in a manner that maximize the prospects of obtaining the resources needed to survive and minimize the probability of encountering situations leading to harm (Doherty et al., 2017).

As described in ***chapter 2***, a specific probabilistic learning task (Kim, Shimojo, & Doherty, 2006) including both reward and avoidance types is set up as a typical learning paradigm designed to investigate the detailed decision making process. Such processes can be divided into mainly three distinct stages: i) outcome ii) expected value and iii) error processing. The action value estimation is the process of computing the subjective value of each option, thus to develop a preference for one option over alternatives, which is also referred to as formation of preferences (Verdejo-Garcia et al., 2018). The choice selection is the stage to allocate the response to the preferred choice, and then the learning is realized through the feedback processing of the outcome from the selected choice. Learning from both positive and negative action outcomes introduces reinforcement of obtaining reward and avoiding aversive behaviour, respectively. The capability of learning from successful reward and erroneous punishment were related to D1/D2 receptors at genetic level, respectively (Bravo et al., 2007; Frank & Hutchison, 2013; Haughey et al., 2007). Through the application of the probabilistic learning task, Kim and colleagues (Kim, Shimojo, & Doherty, 2006) demonstrated that successful avoidance of an aversive outcome exhibits the same properties as a reward. Further, they proposed that avoiding an aversive outcome is in itself an intrinsic reward while obtaining a reward is an extrinsic reward, both of the intrinsic and extrinsic reward is serving to reinforce actions during the instrumental reward and avoidance. Thus, punishment-based reinforcement learning processes can be modelled with similar computational methods as reward-based learning. Basically, the reinforcement-based learning is modelled as a set of actions based on trial-and-error learning, while the subject take to maximize rewards or minimize punishments under reward or avoidance condition, respectively (Samson et al., 2010). Several parameters including learning rate and inverse temperature parameter could reflect the learning processes. The learning rate is the parameter to control the velocity of participants' update of the estimation according to the error signal –

which is the difference between the outcome feedback and the expectation value, and the higher learning rate allows for a faster change of the value (Eyal Even-Dar & Yishay Mansour, 2003). The inverse temperature parameter is the balance of the exploration and exploitation to adjust the randomness of action selection (Ishida et al., 2009).

In short, reinforcement learning (RL) (Sutton & Barto, 1998), is an adaptive process in which a subject utilizes its previous experience to improve the outcomes of future choices. It is now widely adopted to address the fundamental question in decision making such as how participants acquire their preference for different actions and outcome, and also how they learn from the previous experience. First, what does the value of an action reflect? RL theory proposes that the value of an action is a rough prediction of the subsequent reward or punishment gained by selecting that action. Secondly, how do we learn from the previous experience? The learning is through the prediction error (PE) signal – the discrepancy between the actual value of the reward and predicted value of the reward. Central to obtaining rewards and avoiding punishments is the ability to represent the value of rewarding and punishing actions, establishing predictions of when and where such rewards and punishments will occur and use those predictions to form the basis of decisions that guide actions. For example, under reward condition, the actions leading to greater predicted reward will produce a positive PE signal, and as the receipt of a rewarding outcome in a given context serves to strength associations between that context and the response performed, thus the afferent reward PE signal will ensure that such a response is more likely to be selected in the future (O'Doherty et al., 2003; Reynolds et al., 2001; Schultz, 2018). Vice versa, the actions leading to the smaller predicted reward will generate a negative PE signal, which would weaken the associations between that context and the choice performed, thus the efferent reward PE signal will indicate that such a choice is less likely to be selected in the future.

To investigate reward/punishment-based learning and decision making, a novel probabilistic reward and avoidance learning task was used in the study. The probabilistic learning task has been widely used to understand the mechanisms of decision making, and the probability difference of getting reward/punishment of the pair of fractals would drive the participants' learning (Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, 2017; Manuscript, 2008). Successfully performing the task, participants' learning was conceptually dissected into learning associates between stimuli and the rewarding or punishing value and specifically, switching to new associations, which implies inhibiting the selection of the previously rewarded/not-punished stimulus and seeking the newly rewarded/non-punished stimulus after contingencies have reversed (Remijnse et al., 2005). The PE signal is critically important for learning, and more complex (or difficult) learning tasks could be designed to generate a greater magnitude and more robust signal under reward and avoidance conditions. With the probabilistic switch implemented in the task, it could alternate the learning processes, and then change the related PE learning signal with more variations. The aim of this chapter was to examine the participants performance in the task through the basic behavioural statistical analysis and behavioural modelling. Then, the behavioural analysis (Correct choice) and behavioural modelling will act as a baseline for the clinical studies.

3.2 Materials and Methods

STIMULI AND TASK

The probabilistic reward and avoidance learning task were derived from the previous learning task used by Kim et al., (2006). With a probability switch happening at the middle stages of the task, we aimed to drive participants' further learning of the task, thus increasing the learning signal. Specifically, on each trial of the Probabilistic reward/avoidance learning task

(***Figure 3-1***), one of three pairs of fractal stimuli were simultaneously presented. Each pair of fractals signified the onset of one of three trial conditions: Reward, Avoidance and Neutral, whose occurrence was semi-random such that each three-trial block contains one of each type, and the order of these three trials were randomized. Participants underwent two ~16 min scanning sessions, each consisting of 90 trials (30 trials per condition). The specific association of fractal pairs to a condition was fully randomized but counterbalanced among participants. Participants' task on each trial was to choose one of the two stimuli by selecting the fractal to the left or right of the fixation cross via a button box (using the right hand). Once a fractal has been selected, depending on the condition, it increased in brightness and was followed by the visual feedback indicating either a reward (a picture of a Myer card with text above saying "you win 1 point!"), an aversive outcome (a red cross overlying a picture of a Myer card with text above saying "You lose 1 point!"), neutral feedback (a scrambled picture of a Myer card with text above saying "No change!"), or nothing (a blank screen with a cross hair in the centre). Participants had 2000 milliseconds to select a fractal. If not selected in time, a screen would be displayed with the text "response omitted", and the trial would be repeated until a response registered for that fractal pair.

In the reward trials, if participants chose the high probability action (also referred to here as the 'Correct' action), they received monetary reward with a 70% probability; on the other 30% of trials they received nothing. In contrast, choosing the low probability action (also referred to here as the Incorrect action), they received monetary reward on only 30% of trials; otherwise, they obtained nothing on the remaining 70% of trials. Similarly, on the avoidance trials, if participants chose the high probability action they received nothing on 70% of trials, on the other 30% they received a monetary loss, whereas choice of the low probability action led to no outcome on only 30% of trials, while the other 70% were associated with receipt of the aversive outcome. A probability switch was introduced at a

time-point between the 11th to 20th trial in the reward/avoidance trials, where the fractal associated with high probability was changed to the low probability and where the fractal associated with low probability changed to the high probability. For the neutral trials, participants had a 70% or 30% probability of obtaining neutral feedback; otherwise, they received nothing.

Prior to the experiment, participants were given instructions that they would be presented with three pairs of fractals and on each trial, they had to select one of these fractals. Participants also had a practice session of the task before going into the MRI. During the task, depending on their choices they would win a point, lose a point, obtain a neutral outcome with no change, or receive nothing. They were not told which fractal pair was associated with a particular outcome neither when the probability switch was to occur. Participants were instructed to try to win as many points as possible and that they would receive a Coles/Myer voucher at the end corresponding to the amount of points they had accumulated.

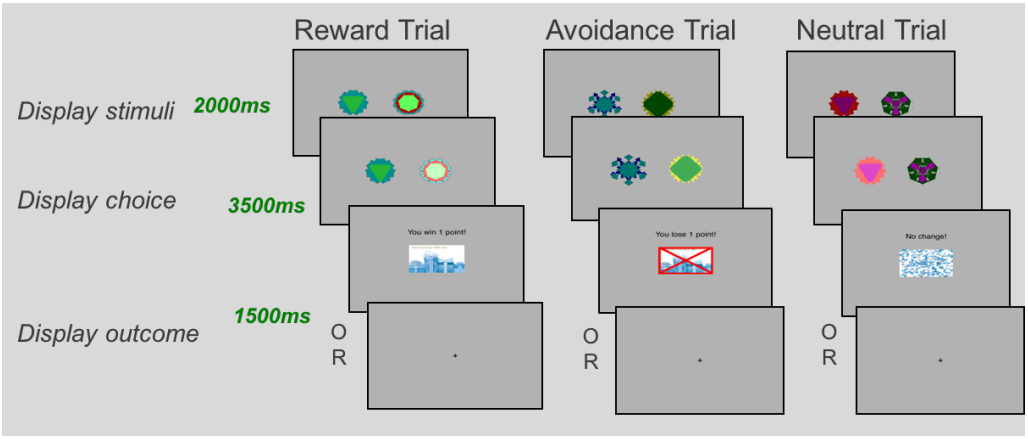

Figure 3-1 *The probabilistic reward and avoidance learning task. Three conditions: reward, avoidance and neutral were included. The pair of fractals under each condition would be displayed for 2000 ms, and then the selected picture would be highlighted for 3500 ms, and the related outcome would be shown for another 1500 ms.*

PARTICIPANTS

As a part of the whole project, 42 healthy controls (21F/21M, 34.03 yrs ± 11.70) have been recruited to perform the task. Several participants were excluded due to incomplete or invalid imaging data, leaving 39 healthy participants (20F/19M, 34 yrs ± 9.47) with complete behavioural and imaging data. Their behavioural data including the fractal choices, outcome and response time of each trial under reward/avoidance/neutral condition was used for this chapter.

METHOD

Statistical analysis

Basic statistical analysis including two-sample t-test was carried out to compare the number of Correct and Incorrect choices as well as response time under each condition. Also, the two-sample t-tests were carried out to compare the response time across three conditions. By realigning the switch to a same point and separation of all trials into eight blocks, a block-based learning curve was drawn based on the number of Correct and Incorrect choices under reward/avoidance condition.

Behavioural modelling

In order to investigate the participants' internal learning traits of the task, a basic Q-learning model was built to describe three components including expectation value, action selection and prediction error. For each pair of fractal stimuli, e.g. A and B, the model estimates the expected value (Q) of choosing A (Qa) or B (Qb) based on the individual sequences of choices and outcomes. The expected value was initially set to zero, and after each trial t > 0, updated according to the chosen stimulus (say choice A): $Q_a(t+1) = Q_a(t) + \alpha * \delta(t)$; while the value for the non-chosen option stays unchanged. $\delta(t)$ – prediction error is the difference

between the actual and expected outcome, α is the learning rate. $R(t)$= \$-1, \$0, \$1 according to the outcome under different conditions. The probability of chosen action was estimated with a soft-max rule, which is the standard stochastic decision rule that calculates the probability of taking one of a set of actions according to their associated values: $P_a(t) = exp\left(\beta Q_a(t)\right)/\{exp\left(\beta Q_a(t)\right) + exp\left(\beta Q_b(t)\right)\}$. The β is the inverse temperature parameter, which indicates how stochastic or exploratory the individual choices are. A low β parameter indicates similar choice probabilities for all choices, which corresponds to low reward/punishment sensitivity; while a high β value indicates that the choice probability is strongly driven by the expected value.

Negative loglikelihood

In order to calculate the parameters of learning rate α and inverse temperature parameter β, one algorithm is to minimize the negative log-likelihood (NLL) of the sum of the observed choices across all trials t given the set of model parameters Θ:

$$arg \sum_{t=1}^{n} \quad - log\ P(a(t)|\Theta)$$

The learning parameter distribution according to the NLL calculation is shown in ***Figure 3-2.*** The NLL calculation is searching a pair of parameters based on each individual. Firstly, the linear searching was realized through self-coded scripts on Matlab (version R2019a), as well as the nonlinear searching method using the Matlab function fmincon. Both estimation methods only consider the individual effects, and some boundary values were appeared.

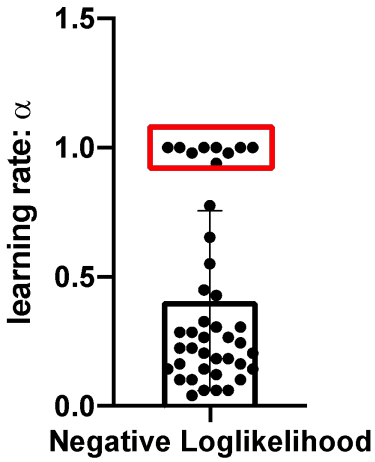

Figure 3-2 *The negative loglikelihood estimation of learning parameters* α*. The red box shows the boundary values.*

Bayesian estimation

In order to solve the problems of boundary values, the hierarchical bayesian method (HBM) was used for the model estimation. In our case, the HBM represents complex and multilevel data structures as shown in ***Figure 3-3***, and it estimates the parameters of the posterior distribution based on a prior knowledge with evidence from data, thus offering a flexible way to specify multilevel structures of parameters. Bayes' theorem is used to integrate the observed data and account for all the uncertainty that is present, and the results of this integration is the posterior distribution, also known as the updated probability estimate, as additional evidence on the prior distribution is acquired. The HBM makes use of two important concepts - hyperparameters and hyperpriors to drive the posterior distribution. The hyperparameters is a parameter following a prior distribution such as the α was suggested to follow a $Norm(\mu_\alpha, \sigma_\alpha)$ distribution. And the related $\boldsymbol{\mu_\alpha}$ and $\boldsymbol{\sigma_\alpha}$ are the hyperpriors of parameter α. In our HBM, both the learning rate α and inverse temperature parameter β was followed to the normal distribution. Application of HBM to the behavioural data has improved the parameter estimation (***Figure 3-4***).

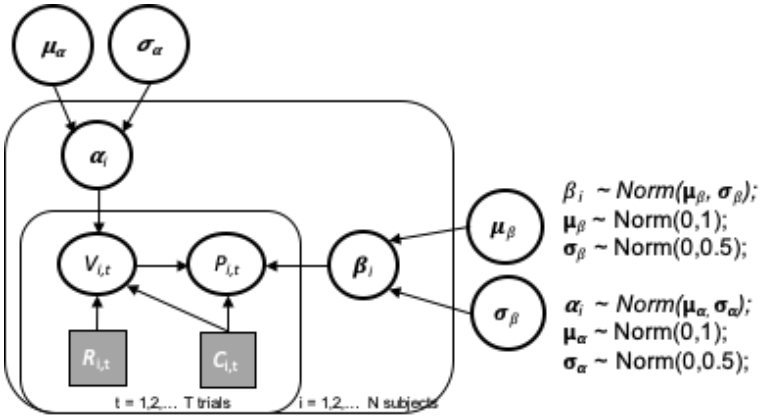

Figure 3-3 *The hierarchical Bayesian model. $R_{i,t}$ is the outcome value of subject i at trial t and $C_{i,t}$ is the choice for the subject i at trial t. The subject related learning rate α_i and inverse temperature parameter β_i was suggested to follow a prior normal distribution with a mean value of μ and a standard deviation value of δ. The group mean value μ_α and μ_β as well as standard deviation value σ_α and σ_β was drawn from a normal distribution, Norm (0,1) and Norm (0,0.5), respectively.*

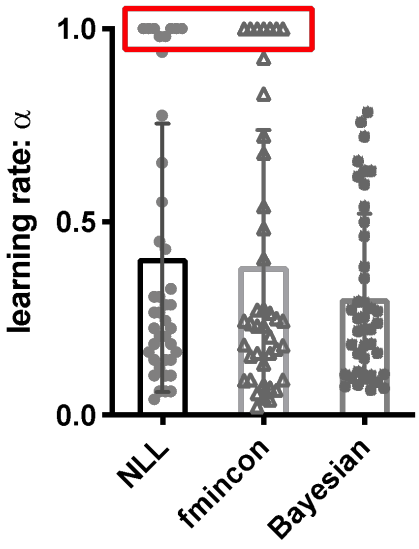

Figure 3-4 *The learning rate estimated performance from three methods. Both the NLL and fmincon had the boundary values, whereas the Bayesian estimation improved the estimation with the consideration of group-level effects.*

Model simulation

In order to demonstrate the feasibility of the model, simulation was carried out. The estimated alpha and beta parameters of the participants under the reward condition were entered into the simulation loop for 20 times (only 5 times were shown for the purpose of convenient display). When comparing the participants' selection probability of the Correct choice with the simulated data, no significant differences were found (shown in ***Figure 3-5***). Also, the selection pattern of Correct and Incorrect choices was shown in ***Figure 3-6***. The simulation analysis was also done under avoidance condition (***Figure 3-7&3-8***). None of these comparisons showed any significant differences of these parameters across, showing the model is validated to simulate similar behaviours as our participants.

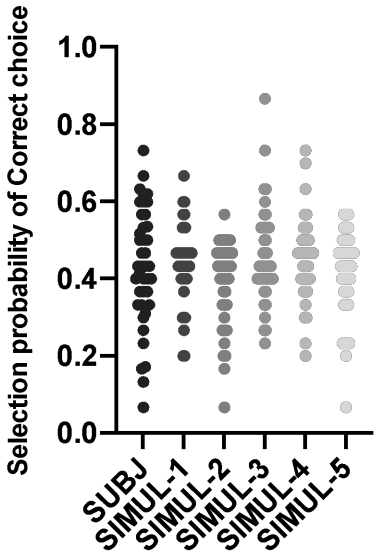

Figure 3-5 *Bayesian model simulation. No significant differences were found between the participants' actual selection probability of the Correct choice and the simulated data under reward condition.*

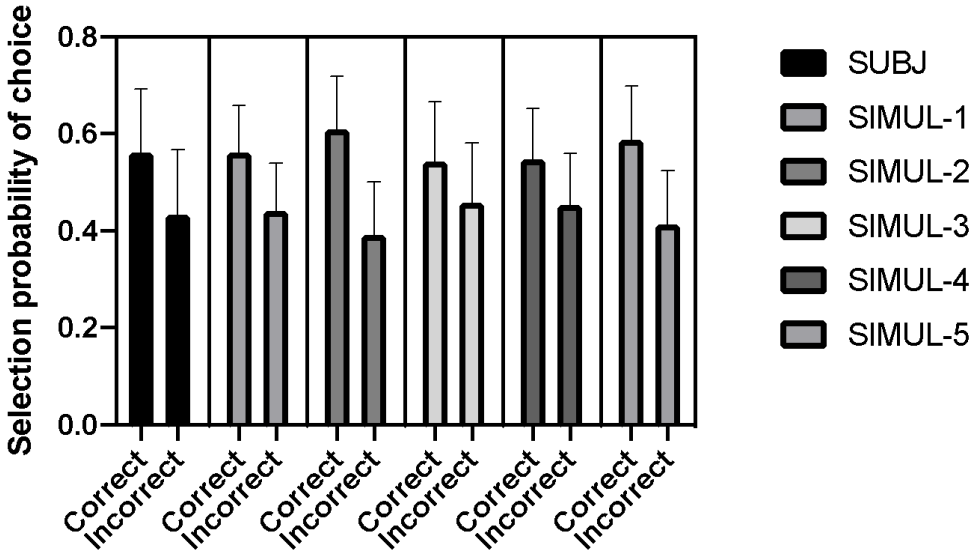

Figure 3-6 *Bayesian model simulation. The selection probability of Correct and Incorrect choice from healthy participants, and the simulated data under reward condition.*

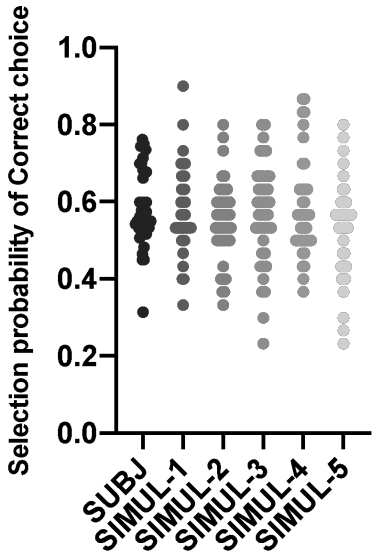

Figure 3-7 *Bayesian model simulation. No significant differences were found between the participants' actual selection probability of the Correct choice and the simulated data under avoidance condition.*

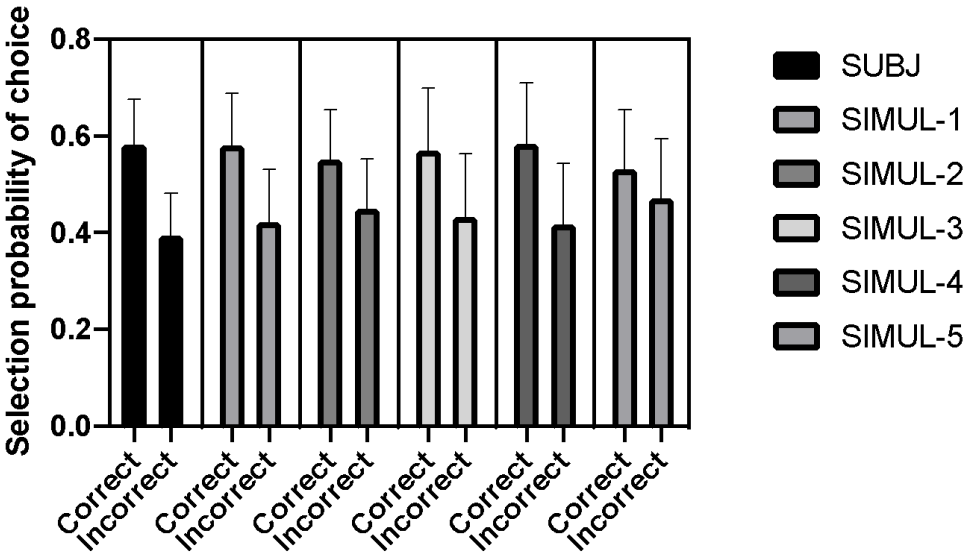

Figure 3-8 *Bayesian model simulation. The selection probability of Correct and Incorrect choice from healthy participants, and the simulated data under avoidance condition.*

3.3 Results

BASIC BEHAVIOURAL RESULTS

Participants showed significant preferences for the Correct choice both in the reward (t = 4.33; $p < .0001$) and avoidance (t = 8.79; $p < .0001$) condition, compared with the Incorrect choice. No significant difference was found in the neutral condition (t = 0.82; p = 0.4138) (***Figure 3-9***); Participants made significantly quicker response to the reward condition (975.3 ms ± 22.84; t = 2.51, p = 0.012) and significantly slower response to the avoidance condition (1133 ms ± 18.35; t = 2.58, p = 0.01). The response time to the neutral condition (1059 ms ± 23.07) is intermediate between the reward and avoidance condition (***Figure 3-10***). The learning curve showed that participants preferred the Correct choice in each block both under the reward/avoidance condition (***Figure 3-11 & 3-12***) before and after the probability switch point.

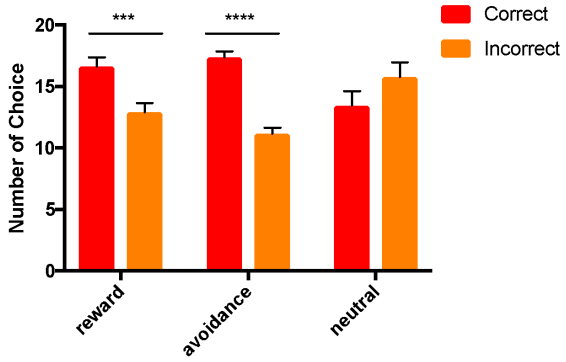

Figure 3-9 *Number of Correct & Incorrect choice under reward, avoidance and neutral condition. participants favoured the Correct choice over the incorrect choice, in both the reward (***p < 0.0001) and avoidance (****p < 0.0001) condition.*

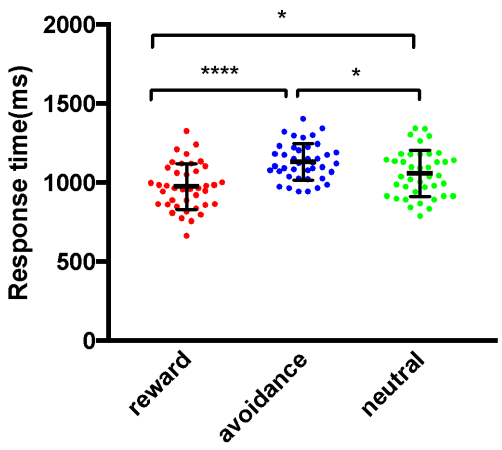

Figure 3-10 *Response time under reward/avoidance/neutral condition. Participants made significantly quicker response to the reward condition (*p < 0.012) and significantly slower response to the avoidance condition (*p < 0.01) compared to the neutral condition.*

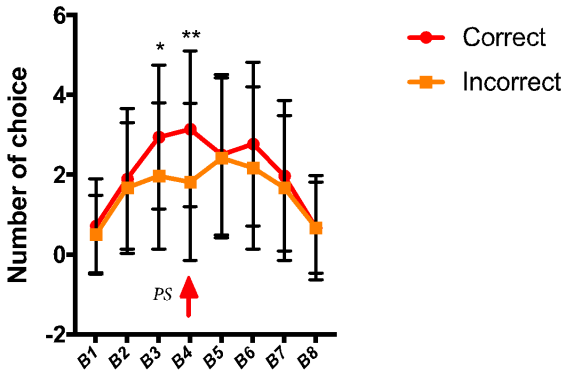

Figure 3-11 *Learning curve under reward condition before and after probability switch. The healthy controls preferred the Correct choice before and after the probability switch (PS).*

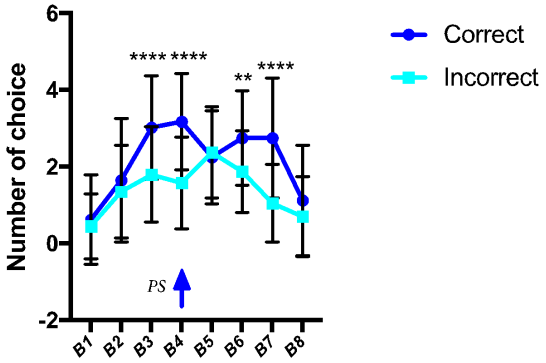

Figure 3-12 *Learning curve under the avoidance condition before and after probability switch. The healthy controls preferred the Correct choice before and after the probability switch (PS).*

MODELLED BEHAVIOURAL RESULTS

The learning rate under reward and avoidance condition was 0.292 +/- 0.212 and 0.754 +/- 0.290, respectively. And the inverse temperature parameter under reward and avoidance condition was 9.00 +/- 2.815 and 1.630 +/- 1.056, respectively. Then, comparing the learning characteristics between the reward and avoidance condition, participants showed a significantly higher learning rate under the avoidance condition ($t = 13.84$, $p < 0.0001$), whereas participants showed a significantly lower inverse temperature parameter under the reward condition ($t = 15.70$, $p < 0.0001$) (***Figure 3-13 & 3-14***). Further, the parameters were entered into the model, and calculated the time series for expected value and prediction error (PE) under reward and avoidance condition (***Figure 3-15, 3-16 & 3-17***).

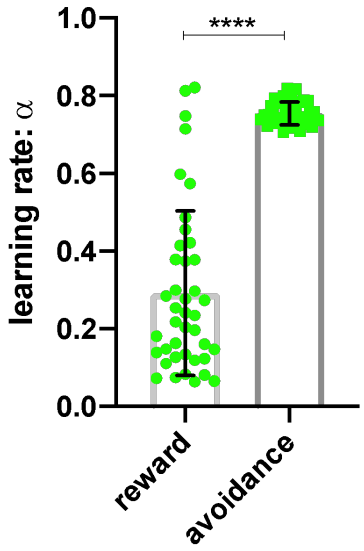

Figure 3-13 *The learning rate under the reward/avoidance condition. The healthy controls showed a significantly higher learning rate under avoidance condition compared to reward condition at ****p < 0.0001.*

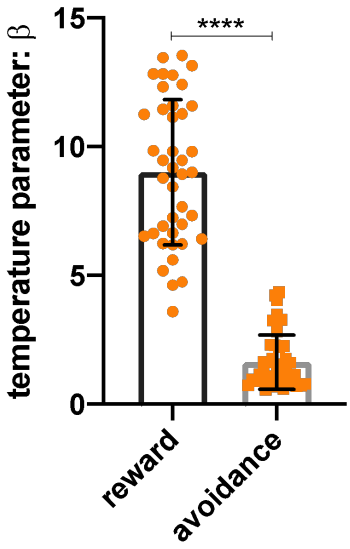

Figure 3-14 *The inverse temperature parameter under the reward/avoidance condition. The healthy controls showed a significantly higher exploitation under reward condition compared to avoidance condition at ****p < 0.0001.*

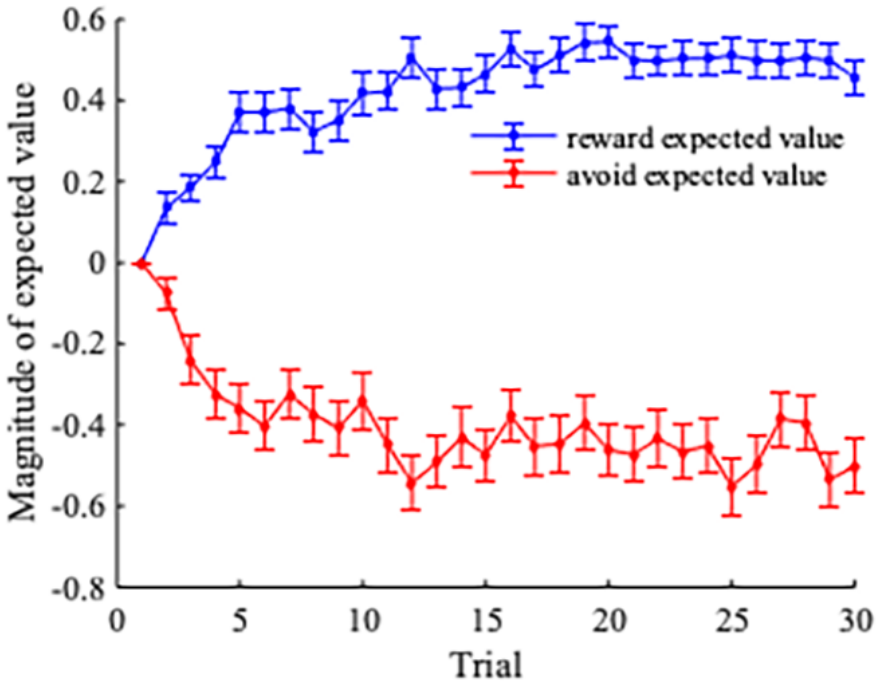

Figure 3-15 *The robust time series of model-derived reward and aversive expected value.*

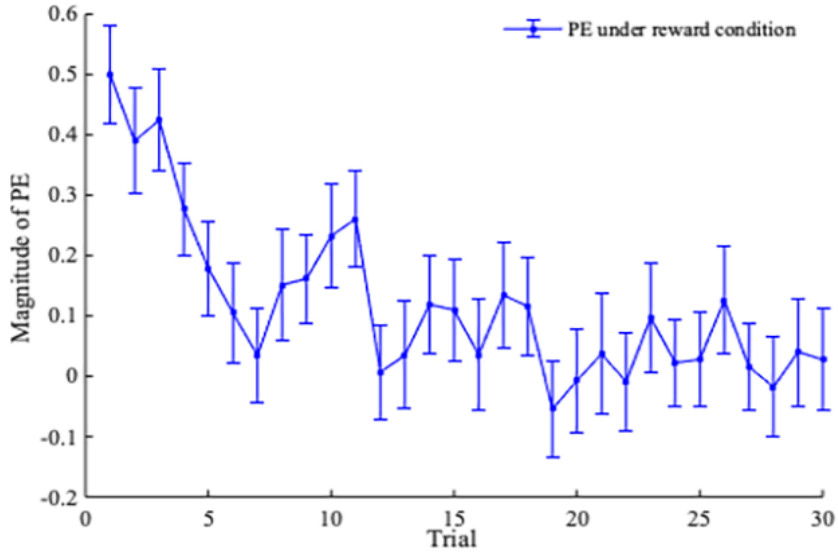

Figure 3-16 *The robust time series of model-derived PE signal under reward condition.*

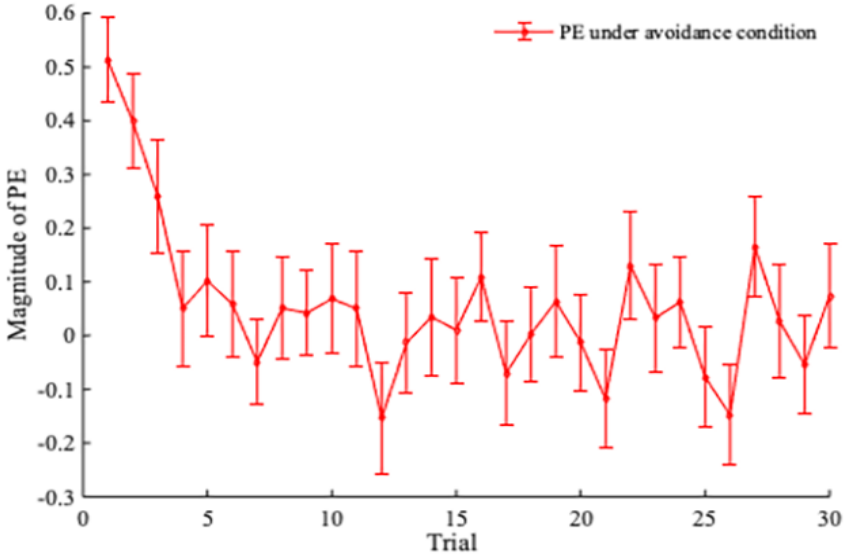

Figure 3-17 *The robust time series of model-derived PE signal under avoidance condition.*

3.4 Discussion

In this chapter, through the application of the probabilistic reward/avoidance learning task, our research found that all healthy participants preferred the Correct choice under reward/avoidance condition to obtain reward and avoid punishments. Also, participants were found to respond more quickly under reward condition while slower in avoidance condition.

Further, the Q-learning algorithm was to model participants' trial-by-trial learning performance. For the model estimation, the comparison with NLL and Matlab fmincon function, the Bayesian model was shown to have better performance. Thus, together with the Q-learning algorithm and Bayesian estimation, participants were found to have a significantly higher learning rate under avoidance condition compared to reward condition, as well as a significantly higher exploitation value under reward condition compared to avoidance condition.

The reward and avoidance learning are similar processes with information acquisition about stimuli, actions and contexts in the environment to get as much reward (or avoid punishments), which are two important components of decision making. In our task, the learning was realized that all participants preferred the Correct choice significantly compared to the Incorrect choice. How were the similarities/differences of the learning traits under both conditions? Previous literature has indicated the participants' difference in the ability to reward and avoid learning and independent processing components for these two types of learning (Carver et al., 1994). In our task, participants responded more quickly to the choice selection under the reward condition compared to the avoidance condition.

Q-learning is the temporal difference learning algorithm which is used to solve the maximization of the rewards or minimization of the punishment in the learning task (Kim, Shimojo, & O'Doherty, 2006; Pavlicek et al., 2011). Starting with the random actions, Q-learning learns the optimization policy during the process of action selection and outcome feedback. The learning rate and inverse temperature parameter are the two key parameters in the Q-learning model to investigate the participants' learning efficiency. Our data showed a higher learning rate under the avoidance condition compared to the reward condition while a higher inverse temperature parameter under reward condition compared to avoidance condition, which means the participants had a relatively lower sensitivity to the outcome

feedback and were highly driven by the expectation value under reward condition. At the same time, participants showed a higher sensitivity to the punishment feedback, thus a higher tendency to explore the two choices. In contrast, study from Kim et al., reported a higher learning rate under reward condition (Kim, Shimojo, & O'Doherty, 2006). Given that the probability switch is the only variance in our paradigm compared to theirs, it could be the reason for the discrepancy.

The preference of Correct choice before and after the probability switch in the learning task could also imply an appropriate cognitive flexibility of healthy participants. The cognitive flexibility – the ability to redirect behaviour to a meeting changing environment plays a crucial role in adaptive decision making (Brusoni, 2018). One of the measurements for the cognitive flexibility is the reversal learning task, in which participants learn an initial response pattern or strategy that must then be adapted when the contingencies or requirements are abruptly changed (Wilson et al., 2018). Usually, the contingency/requirement changes are not cued, so participants must learn that a change has occurred through feedback on obtained outcomes. As adjusting the behaviour to the changing environment is the critical ability, this behavioural flexibility enables one individual to work efficiently to disentangle from a previous paradigm, and reconfigure a new response set to gain a favourable outcome (Dajani and Uddin, 2015). The cognitive flexibility was pointed out to be fundamental for effective decision making, and consequently an important determinant of the organizational ability to learn and adapt to environmental changes (Martinez et al., 2009). The participants with high cognitive flexibility could recognize the value diversity and integrate such diversity in the decision processes to explore the new course of actions (Laureiro-Martínez & Brusoni, 2018).

In summary, the application of reward/avoidance learning task and statistical analysis showed the significant difference of response time existed under reward/avoidance condition.

Further, the modelling of the behavioural data showed the significant learning traits of the two types of learning. In our next chapter, we are going to investigate the computational process, and provide the neural substrates of the computational mechanism under the reward/avoidance-based decision processes.

References

Bravo, A., Crickmore, N., Gould, F., Heckel, D. G., Rie, J. Van, Dennehy, T. J., Yu, L., Wu, Y., Myers, J. H., Weber, E., Ji, I., Ji, T. H., Candas, M., Griko, N. B., Taissing, R., Miranda, R., Bravo, A., Adang, M. J., Dean, D. H., … Wu, L. (2007). *Genetically determined differences in learning from errors*. *318*(December).

Brusoni, D. L. S. (2018). *Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers*. *May 2016*, 1031–1058. https://doi.org/10.1002/smj.2774

Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, C. M. (2017). Differential sensitivity to learning from positive and negative outcomes in cocaine users. *Physiology & Behavior*, *176*, 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Carver, C. S., White, T. L., & The, B. I. S. (1994). *Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: Journal of Personality and Social Psychology*. *doi:10.1037//0022-3514.67.2.319*. *67*(2 SRC-GoogleScholar FG-0), 319–333.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Doherty, J. P. O., Cockburn, J., & Pauli, W. M. (2017). *Learning, Reward, and Decision Making*. https://doi.org/10.1146/annurev-psych-010416-044216

Eyal Even-Dar, & Yishay Mansour. (2003). Learning Rates for Q-learning Eyal Even-Dar Yishay Mansour. *Journal of Machine Learning Research*, *5*, 1–25. http://www.jmlr.org/papers/volume5/evendar03a/evendar03a.pdf

Frank, M. J., & Hutchison, K. (2013). *Genetic Contributions to Avoidance-based decisions: striatal D2 receptor polymorphisms*. *185*(2), 974–981. https://doi.org/10.1038/mp.2011.182.doi

Haughey, H. M., Hutchison, K. E., Curran, T., Frank, M. J., & Moustafa, A. A. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Ishida, F., Sasaki, T., Sakaguchi, Y., & Shimai, H. (2009). Reinforcement-learning agents with different inverse temperature parameters explain the variety of human action-selection behavior in a Markov decision process task. *Neurocomputing*, *72*(7–9), 1979–1984. https://doi.org/10.1016/j.neucom.2008.04.009

Kim, H., Shimojo, S., & Doherty, J. P. O. (2006). Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. *PLoS Biology*, *4*(8). https://doi.org/10.1371/journal.pbio.0040233

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), 1453–1461. https://doi.org/10.1371/journal.pbio.0040233

Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, *39*(4), 1031–1058. https://doi.org/10.1002/smj.2774

Manuscript, A. (2008). Parallel contributions of distinct human memory systems during probabilistic learning. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Martinez, D. L., Brusoni, S., & Zollo, M. (2009). *Cognitive flexibility in decision making: a neurobiological model of learning and change*. *MARCH*.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain RID D-9230-2011. *Neuron*, *38*(2), 329–337. https://doi.org/10.1016/S0896-6273(03)00169-7

Pavlicek, B., Delmaire, C., Palminteri, S., Justo, D., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2011). *Article Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning*. *Bu 2006*. https://doi.org/10.1016/j.neuron.2012.10.017

Remijnse, P. L., Nielen, T. M. M. A., Uylings, H. B. M., & Veltman, D. J. (2005). *Neural correlates of a reversal learning task with an affectively neutral baseline: An event-related fMRI study*. *26*, 609–618. https://doi.org/10.1016/j.neuroimage.2005.02.009

Reynolds, J. N. J., Hyland, B. I., & Wickens, J. R. (2001). *A cellular mechanism of reward-related learning*. 67–70.

Samson, R. D., Frank, M. J., & Fellous, J.-M. (2010). Computational models of reinforcement learning: the role of dopamine as a reward signal. *Cognitive Neurodynamics*, *4*(2), 91–105. https://doi.org/10.1007/s11571-010-9109-x

Schultz, W. (2018). *Predictive Reward Signal of Dopamine Neurons*.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054. https://doi.org/10.1109/TNN.1998.712192

Verdejo-Garcia, A., Chong, T. T. J., Stout, J. C., Yücel, M., & London, E. D. (2018). Stages of dysfunctional decision-making in addiction. *Pharmacology Biochemistry and Behavior*, *164*, 99–105. https://doi.org/10.1016/j.pbb.2017.02.003

Wilson, C. G., Nusbaum, A. T., Whitney, P., & Hinson, J. M. (2018). Trait anxiety impairs cognitive flexibility when overcoming a task acquired response and a preexisting bias. *PLoS ONE*, *13*(9), 1–13. https://doi.org/10.1371/journal.pone.0204694

This chapter has been prepared for submission.

4 A model-based fMRI study of the neural representations of the stages of reward and avoidance decision processes

Xiaoliu Zhang[1], Shinsuke Suzuki[3¶], Amir Dezfouli[2], Leah Braganza[1], Ben Fulcher[1], Linden Parkes[1,4], Leonardo F. Fontenelle[1], Ben J. Harrison[5], Murat Yücel[1], Carsten Murawski[3*¶], Chao Suo[1*¶]

[1] BrainPark, Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging Facility, Monash University, Melbourne, Australia.

[2] Machine Learning Research Group CSIRO's Data61, Sydney, New South Wales, Australia

[3] Brain, Mind & Markets Laboratory, Department of Finance, The University of Melbourne, Melbourne, Victoria, Australia.

[4] Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA, 19104 USA.

[5] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Victoria, Australia.

¶*These are equal corresponding authors*

*Corresponding email: Chao.Suo@monash.edu

This chapter has been prepared as a manuscript for submission.

Connecting: Through the basic behavioural and model-based analysis, the previous chapter has provided evidences of the healthy participants' different response patterns and learning performance at two types of reward and avoidance learning. Specifically, participants achieved the learning with significant preference to the Correct choice compared to the Incorrect choice. When comparing the learning performance, they had quicker response under reward condition and slower under avoidance condition. Also, the participants showed significantly different learning rate and temperature parameter between the reward and avoidance condition. The behavioral findings further motivated to examine the underlying neural mechanisms. In this chapter, we will combine the model fitting and neuroimaging analysis to explore what's the neural activations when getting different outcomes, thus driving participants' learning of the task. Also, what's the neural mechanisms under reward and avoidance conditions, thus causing different learning performance.

Aim: The previous chapter has provided evidences of the healthy participants' different response patterns and learning performance at both types of reward and avoidance learning through both the basic behavioural and model-based analysis. In this chapter, we are going to examine the underlying neural mechanisms through combination of model fitting and neuroimaging analysis. Mainly, we examine the key signals associated with reward and avoidance decision processes: 1) outcome, 2) expected value and 3) PE.

Questions: The reward and avoidance learning include three distinct phases: outcome, expectation and error processing for action selection. In this chapter, we were interested in identifying: the neural substrates of these three distinct signals under reward and avoidance

We included the section of Aim, Questions and Hypothesis to satisfy the consistency of all results chapters.

learning, namely - brain circuits involved in processing the anticipation and receiving of reward or punishment outcomes, as well as the brain circuits that represent the processing of PE signal which drives the subsequent learning from the outcome. Further, how were the enhanced PE signal activation pattern in our learning task due to the probability switch.

Hypothesis: Based on the previous studies on the neural activations of reward/avoidance learning-based decision process, we hypothesised: i) The common brain pattern, e.g. medial orbitofrontal cortex (OFC), was correlated with the rewarded outcome value and punishment avoidance. ii) The activity of medial and lateral OFC will be correlated with expectation value under reward/avoidance condition. iii) As an enhanced PE signal due to the reversal learning component described in Chapter 3, more significant brain activation at the frontostriatal circuit was found associated with the PE. Further, we expected a dissociable manner of PE neural correlates during reward and avoidance learning. e.g. ventral frontostriatal circuit were involved with reward learning whereas dorsal striatum and Insula were associated with punishment learning.

We included the section of Aim, Questions and Hypothesis to satisfy the consistency of all results chapters.

ABSTRACT

Reward and avoidance learning form two critical types of decision making, which involves assigning value to available options, choosing between alternatives based on preferences, assessing consequences for the selected choices, and learning from the outcomes to update the future choices. However, the common or distinct neural representations at separate stages of the two types of decision processes still need clarification. Here, 42 healthy participants were recruited to perform a two-session probabilistic reward and avoidance learning task with fMRI scanning. Together with neuroimaging techniques and computational modelling, we showed that both shared and distinct brain regions are involved at key stages including outcome, expected value and prediction error (PE) in reward and avoidance-based decision processes. At the outcome stage, receiving reward and punishment were both associated with the functional activity in the *insula and cingulum*, whereas the *entire striatum* was selectively active during reward only, and *dorsal striatum* was selectively active during punishment only. The *cingulum* was also found activated for both reward and avoidance expectation. Furthermore, the avoidance expectation recruited broader areas at the cortical and subcortical brain areas including *inferior OFC*, *insula,* and *dorsal striatum*. At the stage of error processing, due to the novel probability switch of the learning task, a robust PE signal and covaried with activity of the *cortical and subcortical brain areas* under the reward condition; Meanwhile the aversive PE signal was found covaried with the activity at the shared frontal-subcortical brain regions and the segregated *dorsal part of striatum*. Our results demonstrate the *dorsal striatum* specific role for differential phases of avoidance processing, and existence of the dissociated computational processes underlying reward and avoidance decision processes.

4.1 Introduction

Decision making is a complex process that involves assigning values to available options, choosing between alternatives based on preferences, assessing consequences of the selected choices, and most importantly learning from the outcome to update future decision processes (Engel and Caceda, 2015). The aim of such decision making processes is to maximise favourable and to minimise unpleasant outcomes (Krigolson et al., 2014). Thus, the two types of learning - namely, reward learning and avoidance learning, form the foundational aspects of decision making. The common goal of most decisions is to maximize reward and minimize punishment, thus, one requires learning which choices are likely to lead to favourable outcomes (Eshel and Steinberg, 2018). This learning process can be explained using reinforcement learning (RL) algorithms (Sutton and Barto, 2015). According to RL, a reward prediction error (PE) is created to reflect the discrepancies between the actual and expected outcome, which is then used to adjust the expected outcome for the next decision to make predictions (Kim et al., 2006; Doherty et al., 2017), and adjust future actions (Schultz, 2017).

RL theory has been useful in providing plausible accounts for animal and human reward-related learning and its neural underpinnings (Reynolds et al., 2001; O'Doherty et al., 2003; Peter and Daw, 2008; Schultz, 2018). For instance, combining the RL model with neural `, animal neurophysiology has shown that dopamine neurons in the midbrain encode reward prediction error (reward PE) during reward-learning tasks (Riaz et al., 2016; Coddington and Dudman, 2018). In humans, functional magnetic resonance imaging (MRI) studies consistently demonstrated the neural encodings of reward PE in the ventral striatum and the encoding of expected value in the orbitofrontal cortex (OFC) (Schultz, 2016; Howard and Kahnt, 2018). Specifically, the expected value signals were found to be correlated with blood oxygen level dependent (BOLD) signal in the medial and lateral OFC at the timing of

decision-making, and the reward PE signals were correlated with BOLD signal in the ventral striatum at the timing of reward delivery (Kim et al., 2006).

Not only for reward learning, the RL theory is also suggested to account for avoidance learning (Ben et al., 2004; Kim et al., 2006). Whether there are common or differential neural mechanisms underlying the distinct stages of reward and avoidance learning processes, remains controversial. For example, during performance of the RL task, and at the stage of outcome processing, the neural activity at medial OFC has been found to increase not only following receipt of reward, but also following successful avoidance of an aversive outcome Kim et al (2006). This is consistent with a recent meta-analysis which also demonstrated the OFC's central role during the outcome of monetary reward and loss (Oldham et al., 2018a). These findings suggest the shared brain mechanism underlying between receiving reward and avoiding punishment (Kim et al., 2006). Using a similar task that focussed on broader patterns of brain activity, it was reported that learning to gain rewards was recruiting striatal brain regions at the outcome stage. In contrast, loss avoidance was associated with the activation of the prefrontal brain regions (Kim et al., 2014). In light of these inconsistent reports, further research is needed to demonstrate the neural mechanism at the outcome stage of obtaining reward and avoiding loss/punishment. At the stage of anticipation, of particular relevance is the study by Kim et al (2006) which reported the medial and lateral OFC were correlated with model-derived expected value under both reward and avoidance condition (Kim et al., 2006). However, a recent neuroimaging meta-analysis has revealed subtle differences exist between the neural processing of reward and punishment. That is, the medial OFC is activated during reward anticipation whereas the loss anticipation recruits the activity of ventro-lateral prefrontal regions (Dumais and Bitar, 2018).

As indicated earlier, PE is an essential neurophysiological signal encoding the discrepancy between the actual and expected outcome, which is subsequently used to guide

predictions and adjust future actions. The neural representations of PE signal is suggested to correspond with the phasic firing of midbrain dopamine neurons in the animal literature (Schultz et al., 1997; Schultz, 2018). Further, striatal dopaminergic projection neurons along with its connected areas, including medial prefrontal cortex and anterior cingulate as well as insula, are all key brain areas involved in encoding the PE signal (Garrison et al., 2013a). Specifically, different to the reward PE signal, the PE signal under avoidance was found to correlate with functional activity in the bilateral insula in the study by Kim et al (Kim et al., 2006). These findings were in line with the literature showing that the reward PE is correlated with the functional activity in ventral striatum and OFC (Kim et al., 2006; Garrison et al., 2013b), whereas the aversive PE signal in avoidance learning is associated with functional activity in the amygdala-striatal regions (Zhang et al., 2016), and bilateral insula (Kim et al., 2006; Garrison et al., 2013b). The PE signal is critically important for learning, and more complex (or difficult) learning tasks could be designed to generate a greater magnitude/more robust signal, thus helping to examine a more complete picture of neural correlates under reward and avoidance conditions. To this end, a previous study has incorporated a probabilistic switch in the task, which serves to increase the difficulty of the task, thereby driving the need for greater learning processes to be engaged and thus increasing PE learning signal (Alexandre Y. Dombrovski et al., 2011). Thus, introducing such a probabilistic switch could be a promising method to investigate the neurocircuitry underlying reward and avoidance learning.

Our understanding of the common or differential pattern of the reward and avoidance decision processes is still limited by the small number of imaging studies utilising modelling and a probabilistic reversal switch, both of which serve to maximise our ability to comprehensively understand the underlying brain mechanism of computational reward and avoidance processing. In the present study, we introduced a probabilistic reward and

avoidance learning task similar to that of Kim's original task, with the addition of a probabilistic switch during the middle stages of the task. The addition of the switch will increase the overall difficulty of the task to drive the need for more learning processes (and thereby increase the PE signal during such learning). We combine this probabilistic switch task with functional MRI and model fitting, to examine the neural correlates of three distinct stages of reward and avoidance learning based decision making - namely, outcome, expected value and PE. With a probabilistic switch occurring at approximately the mid-point of the task, we predicted that both common and distinct brain regions will be associated with distinct stages of reward and avoidance decision processes. Specifically, the mOFC will be recruited for the outcome of receipt of reward and avoidance of punishment. Significant brain activations will be associated with reward PE signals in the fronto-striatal brain regions, and neural circuits associated with aversive PE at brain regions, such as dorsal striatum and insula.

4.2 Materials & methods

PARTICIPANTS

Forty-two healthy controls have been recruited for this task. Study sample has been reported in previous studies (Parkes et al., 2018; Maleki et al., 2020). Inclusion criteria for healthy participants involved the following: age between 18-55 years, having normal to corrected vision, and being fluent in English. Exclusion criteria for all participants included significant head injury or concussion and standard MRI contraindications. Several participants were excluded due to incomplete or invalid imaging data, leaving 39 healthy participants (20F/19M, 34 yrs ± 9.47) with complete behavioural and imaging data. All participants gave informed consent and the study was approved by the Human Research Ethics Committee of Monash University.

PROBABILISTIC REWARD AND AVOIDANCE LEARNING TASK

On each trial of the Probabilistic reward/avoidance learning task (***Figure 4-1 (a)***), one of three pairs of fractal stimuli were simultaneously presented. Each pair of fractals signified the onset of one of three trial conditions: Reward, Avoidance and Neutral, whose occurrence was semi-random such that each three-trial block contains one of each type, and the order of these three trials were randomized. Participants underwent two ~16 min scanning sessions, each consisting of 90 trials (30 trials per condition). The specific association of fractal pairs to a condition was fully randomized but counterbalanced among participants. Participants' task on each trial was to choose one of the two stimuli by selecting the fractal to the left or right of the fixation cross via a button box (using the right hand). Once a fractal has been selected, depending on the condition, it increased in brightness and was followed by the visual feedback indicating either a reward (a picture of a Myer card with text above saying "you win 1 point!"), an aversive outcome (a red cross overlying a picture of a Myer card with text

above saying “You lose 1 point!”), neutral feedback (a scrambled picture of a Myer card with text above saying “No change!”), or nothing (a blank screen with a cross hair in the centre). Participants had 2000 milliseconds to select a fractal. If not selected in time, a screen would be displayed with the text “response omitted”, and the trial would be repeated until a response registered for that fractal pair.

In the reward trials, if participants chose the high probability action (also referred to here as the ‘Correct’ action), they received monetary reward with a 70% probability; on the other 30% of trials they received nothing. In contrast, choosing the low probability action (also referred to here as the Incorrect action), they received monetary reward on only 30% of trials; otherwise, they obtained nothing on the remaining 70% of trials. Similarly, on the avoidance trials, if participants chose the high probability action they received nothing on 70% of trials, on the other 30% they received a monetary loss, whereas choice of the low probability action led to no outcome on only 30% of trials, while the other 70% were associated with receipt of the aversive outcome. A probability switch was introduced at a time-point between the 11^{th} to 20^{th} trial in the reward/avoidance trials, where the fractal associated with high probability was changed to the low probability and where the fractal associated with low probability changed to the high probability. For the neutral trials, participants had a 70% or 30% probability of obtaining neutral feedback; otherwise, they received nothing.

Prior to the experiment, participants were given instructions that they would be presented with three pairs of fractals and on each trial, they had to select one of these fractals. Participants also had a practice session of the task before going into the MRI. During the task, depending on their choices they would win a point, lose a point, obtain a neutral outcome with no change, or receive nothing. They were not told which fractal pair was associated with a particular outcome neither when the probability switch was to occur.

Participants were instructed to try to win as many points as possible and that they would receive a Coles/Myer voucher at the end corresponding to the amount of points they had accumulated.

BASIC BEHAVIOURAL ANALYSIS

Under each condition, the total number of Correct and Incorrect choices were calculated for each participant. Also, the response time was measured at the time of fractal pairs displayed until participants' choice selection at each trial under three conditions.

IMAGING PROCEDURE

All images were acquired with 3.0-T SIEMENS MAGNETOM Skyra syngo MR D13C at Monash Biomedical Imaging. The functional images (fMRI) were acquired through gradient echo T2* weighted echo-planar images (EPI) with BOLD (blood oxygenation level dependent) contrast. The scanning parameters: field of view = 230 mm, 3mm by 3mm in plane resolution, time of repetition = 2000 ms, and time of echo = 30.0 ms. Each volume of fMRI images contains 34 slices with a thickness of 3.0 mm (no gap) in an ascending interleaved way. High resolution T1-weighted (1x1x1 mm^3 resolution) were acquired with a standard MPRAGE sequence (time of echo = 2.07 ms, time of repetition = 2300 ms, flip angle = 9 degree, field of view = 256 mm).

Q-LEARNING MODEL

A basic Q-learning model (Watkins, 1995), was used to characterise participants' behaviour in task. This model estimates the expected value of choosing each stimulus based on the previous history of choices and outcomes. The expected value of each stimulus was

initially set to zero, and after each trial t > 0, was updated according to the chosen stimulus and reward feedback. The expected value of choosing stimulus a was updated as follows,

$$Q_a(t+1) = Q_a(t) + \alpha * \delta(t);$$

while the value for nonchosen stimulus stayed unchanged. α is the learning rate and $\delta(t)$ is the prediction error which is the difference between the actual and expected outcome,

$$\delta(t) = R(t) - Q_a(t);$$

$R(t)$= \$-1, \$0, \$1 is the reward received after choosing the stimulus. The probability of taking each action is based on their values, and according to the softmax rule,

$$P_a(t) = exp\left(\beta Q_a(t)\right) / \{exp\left(\beta Q_a(t)\right) + exp\left(\beta Q_b(t)\right)\};$$

The β is the inverse temperature parameter with a scale from 0 to 20, which indicates how stochastic or exploratory the individual choices are. Lower values of β parameter indicate random action selection, which corresponds to low sensitivity to stimulus values; while a high β value indicates that choices are strongly driven by their expected values.

The hierarchical bayesian method (HBM) was used for the model and parameter estimation. HBM, exploits group-level parameter distributions to inform individual-level estimations, and compared to the individual parameter estimation methods, HBM provides better parameter stability and predictive accuracy (Scheibehenne and Pachur, 2015). The learning rate α and inverse temperature parameter β had a normal prior distribution Norm (0,1) (see **supple. Fig 1** for the details of model structure).

Imaging data analysis

Pre-processing

SPM12 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, United Kingdom) was used to perform the fMRI image analysis. The pre-processing of EPI images commenced with the slice timing correction to the middle slice of each volume. Then, the realignment was applied to remove the motion artefacts. The individual T1-weighted image was co-registered to the mean EPI generated during realignment and then were normalized to the standard Montreal Neurological Institute (MNI) space based on the 6-tissue probability map (TPM) provided by SPM. The motion-corrected and co-registered EPI images were normalized to MNI space using the previously calculated deformation fields and then spatially smoothed with a 8 mm FWHM (full width half maximum) Gaussian kernel (Kim et al., 2006).

1st level analysis

Time series describing expected values and PEs were generated for each participant for each trial in the experiment by entering the participants' trial history into the learning model. These sequences were convolved with a hemodynamic response function and entered into a General Linear Model (GLM) to fit the pre-processed imaging data. The expected value was modelled as a boxcar function beginning at the time of fractals display till the outcome while the PEs modelled as a delta function at the time of outcome display. Separate six regressors were created for different outcomes to model activity at the time of the outcome: rewarded reward trial (R_+), unrewarded reward trial (R_-), punished avoidance trial (P_+), non-punished avoidance trial (P_-), neutral feedback trial (N_+) and neutral trial without feedback (N_-). In addition, the six scan-to-scan motion parameters produced during realignment were included to further remove the nuisance effect of head motion.

Linear contrasts of regressors coefficients were computed at the individual participant level to enable comparison among the Reward, Avoidance and Neutral trials. The simple contrast [R_+ - N_+] was to test the brain response to rewarded outcome and the contrast [P_+ - N_+] was to examine the brain activation related to outcome of getting point loss (referred as aversive outcome). Further, the specific contrast [R_+ + P_-] - [R_- + P_+] was to test those of brain areas showing greater response to obtaining reward and avoidance aversive outcome compared to obtaining aversive outcome and missing reward.

The model-derived expected value and PE were separately parametric modulated from outcome regressor (see **supple. Fig 8-10** for time series for both signals). Then, the contrasts were created to examine the brain areas associated with expected value and PE under reward and avoidance condition. Further, the conjunction analysis of expected value, and PE under reward and avoidance condition were performed to examine the common neural correlates. Moreover, the direct comparison of expected value, and PE under both conditions to examine the differential regional response.

Statistical analysis

For behaviour measurements, independent two-sample t-tests were carried out to compare the number of Correct vs Incorrect choices under each condition. Also, the same test was used to compare the response time between different conditions. The statistical analysis was conducted and visualized using GraphPad Prism (version 8).

For voxel-based group level statistical analysis, the contrast images from each single participant were taken to a one sample t-test design to examine the group effect of various outcome contrasts as well as expected value and PE. SPM12 is used to conduct the analysis. The significant level was initially set at $p < 0.001$ and the threshold for family wise error

(FWE) multiple comparison correction (MCC) was set $q < 0.05$ *with cluster size* $k > 100$ at cluster level.

4.3 Results

BASIC BEHAVIOURAL RESULTS

Participants made significantly quicker responses to the reward condition (975.3 ms ± 22.84; t = 2.51, p = 0.012) and significantly slower response to the avoidance condition (1133 ms ± 18.35; t = 2.58, p = 0.01). The response time of the neutral condition (1059 ms ± 23.07) is intermediate between the reward and avoidance condition (***Figure 4-1. (c)***). Over the course of learning trials, participants showed significant preferences to the choice associated with higher probability obtaining reward points (i.e., Correct choice as shown in ***Figure 4-1. (d)***) in the reward condition (t = 4.3, $p < .0001$; two tailed), and the choice associated with lower probability to get aversive outcome (i.e. in the avoidance condition (t = 8.785, $p < .0001$; two tailed). No significant difference was found in the neutral condition (t = 0.82, p = 0.4138; one tailed).

MODELLED BEHAVIOURAL RESULTS

Before the application of the model to the participants' behavioural data, simulation was carried out to demonstrate the feasibility of the model (see **supple. Fig 2-5**). The learning rate α under reward and avoidance condition was 0.292 +/- 0.212 and 0.754 +/- 0.290, respectively. And the inverse temperature parameter β under reward and avoidance condition was 9.00 +/- 2.815 and 1.630 +/- 1.056, respectively. Then, comparing the learning characteristics between the reward and avoidance condition, participants showed a significantly higher learning rate under the avoidance condition (t = 13.84, p < 0.0001),

whereas participants showed a significantly lower inverse temperature parameter under the reward condition ($t = 15.70$, $p < 0.0001$) (see **supple. Fig 6-7** for parameters distribution).

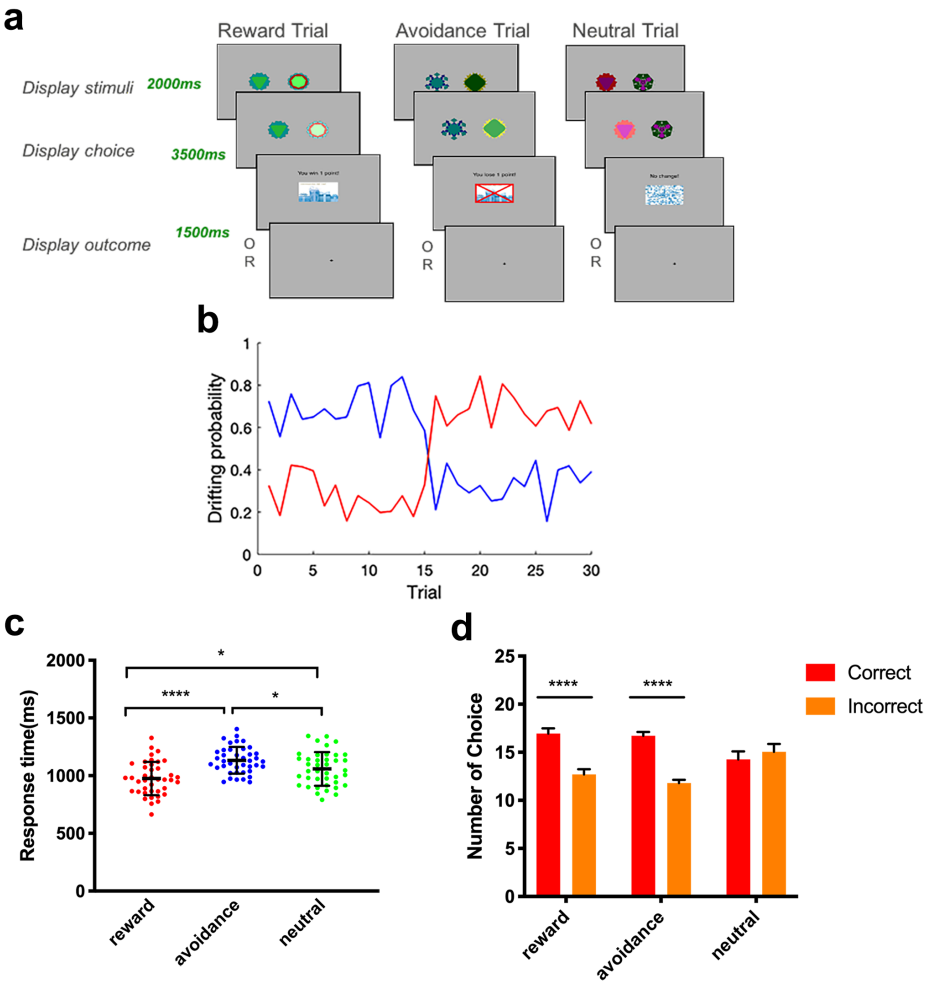

Figure 4-1 *(a) The probabilistic reward and avoidance learning task, which includes three types: reward, avoidance and neutral conditions. The pair of fractals under each condition would be displayed for 2000 ms, then the selected picture would be highlighted for 3500 ms, and the outcome would be shown for another 1500 ms. (b) The drifting probability under reward (in red) and avoidance condition (in blue). (c) The response time under reward, avoidance and neutral condition. Significant quicker response under reward condition and slower under avoidance condition compared to neutral condition were found (* means $p < 0.05$, **** means $p < 0.0001$). (d) The number of Correct and Incorrect choice. The significant preference to the Correct choice were found under reward and avoidance condition (**** means $p < 0.0001$).*

IMAGING RESULTS

Brain regions response to reward receipt and punishment avoidance

When comparing brain responses to rewarded outcome compared to the neutral condition, the largest cluster with the peak at ([8, -4, 2]; t = 9.94, k = 5205) covered *striatum*, *bilateral thalamus*, *bilateral insula* and *left inferior orbitofrontal (OFC)*. The second cluster peaked at the right middle cingulum ([2, 36, 30]; t = 7.22, k = 2494) covered *bilateral ACC* and *bilateral superior medial frontal* and the right middle cingulum ([4, -10, 32]; t = 6.59, k = 860) after correction (see ***Figure 4-2***).

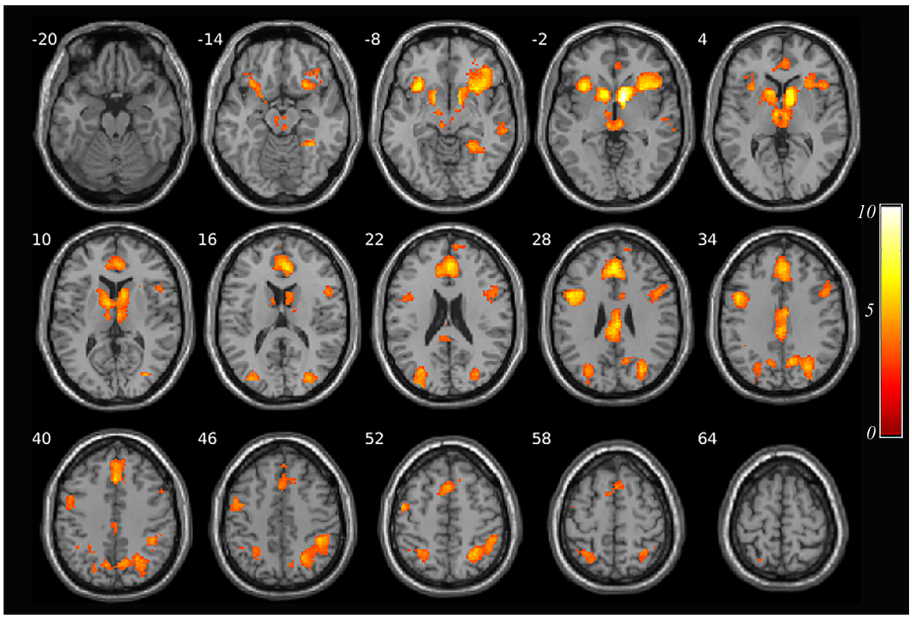

Figure 4-2 *Neural correlates of rewarded outcome. The cluster covering striatum, bilateral thalamus, bilateral insula and left inferior OFC, the right middle cingulum covering bilateral ACC and the right middle cingulum were found to have higher activation in response to the rewarded outcome (see* ***Table 4-1 for details****).*

When comparing brain responses to aversive outcome compared to the neutral condition, we observed significant activation in the right insula ([42, 22, -2]; t = 9.94, k = 1832), left insula ([-32, 18, -10]; t = 9.11, k = 1127), right SMA ([4, 20, 58]; t = 9.08, k =

2374) extending to *right middle cingulum* and the left dorsal striatum ([-10, 2, 14]; t = 4.77, k = 165) after correction (see ***Figure 4-3***).

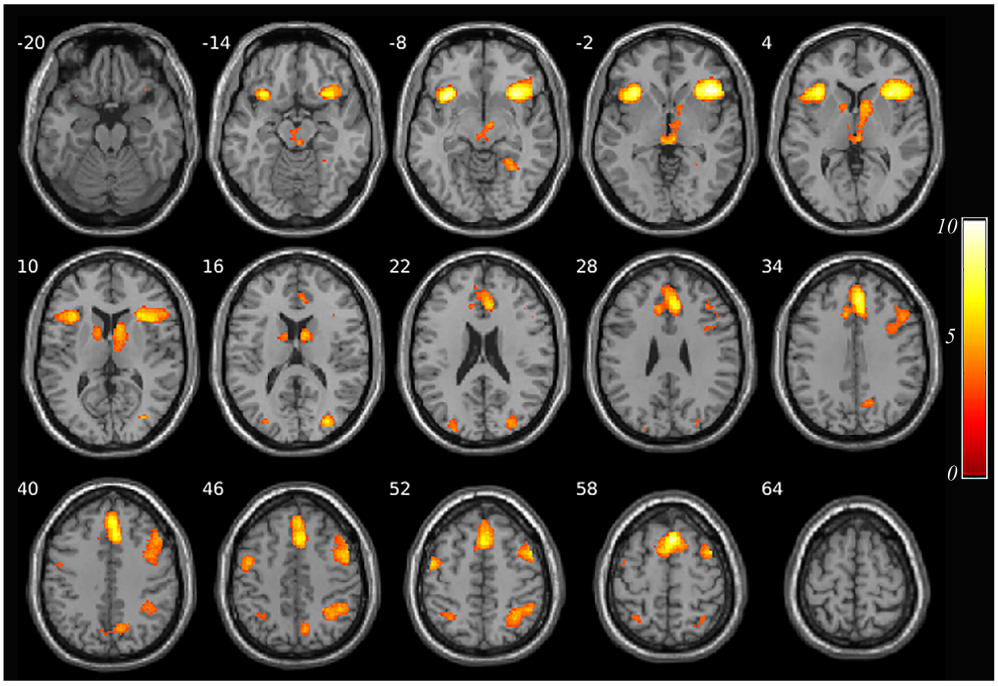

Figure 4-3 Neural correlates of aversive outcome. The bilateral insula, right SMA extending to the right middle cingulum and the left dorsal striatum were found to have higher activation in response to the punished trials (see ***Table 4-1 for details***).

Table 4-1 *Locations of regional response to reward/aversive outcome*

Region	Peak MNI Coordinates			Peak t value	Spatial extent (in contiguous voxels)
	x	y	z		
R_+ - N_+					
R. Caudate	8	4	-2	9.94	5205
L. Insula	-32	20	-6	7.84	
R. Anterior cingulum	2	36	30	7.22	2494
R. Anterior cingulum	8	38	16	6.37	
L. Anterior cingulum	-6	32	26	5.99	

L. Precentral Gyrus	-46	6	30	7.01	864
L. Precentral	-48	-4	52	5.52	
R. Middle cingulum	4	-10	32	6.59	860
R. Middle temporal	62	-32	-6	4.63	152
R. Middle temporal	54	-30	-6	3.80	
P_+ - N_+					
R. Insula	42	22	-2	9.94	1832
R. Insula	34	18	-8	8.84	
L. Insula	-32	18	-10	9.11	1127
L. Insula	-32	20	0	8.59	
R. Supplementary motor area	4	20	58	9.08	2374
R. Middle cingulum	4	36	34	8.44	
R. Middle frontal	42	8	58	6.67	1159
R. Precentral	48	6	50	6.10	
R. Middle frontal	48	20	38	5.28	
L. Precentral	-48	-2	52	6.20	271
R. Caudate	8	6	14	6.03	1075
R. Caudate	12	8	6	5.62	
L. Caudate	-10	2	14	4.77	165

*Note: **R_+ - N_+** –Rewarded outcome – Neutral feedback; **P_+ - N_+** – Punished outcome – Neutral feedback.

Further, the contrast [R_+ + P_-] - [R_- + P_+] showed that the left putamen ([-14, 4, -10]; t = 7.04, k = 409), right putamen ([16, 8, -10]; t = 5.78, k = 576), left medial orbitofrontal (mOFC, [-10, 48, -10]; t = 5.01, k = 412), right posterior cingulate (PCC, [4, -32, 22]; t = 5.31, k = 507), right superior temporal gyrus (sTG, [62, -34, 12], t = 5.12, k =154) and left inferior orbitofrontal (OFC, [-38, 36, -16]; t = 4.90, k =171) were significantly more active in response to receiving reward and avoiding punishment compared to receiving an aversive outcome and missing reward after correction; While the right insula ([32, 20, 10]; t = - 3.22, k

= 223) were shown greater activation in response to missing reward and receiving punishment compared to receiving reward and avoiding punishment (see ***Figure 4-4***).

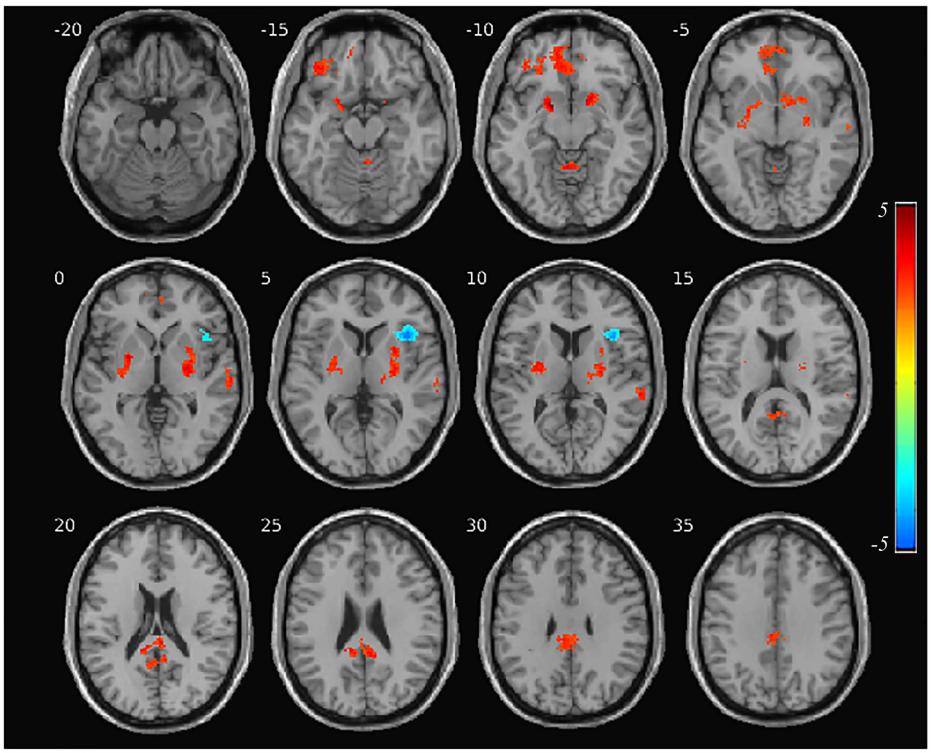

Figure 4-4 *Common neural correlates of outcome under receipt of reward and avoidance of punishment condition. Through the contrast [R_+ + P_-] - [R_- + P_+], the ventral striatum, left mOFC and inferior OFC, right sTG and right PCC were found activated when receiving reward and avoiding punishment. While the right insula was found to have higher activation to missing reward and receiving punishment compared to receiving reward and avoiding punishment (see* ***Table 4-2 for details****).*

Table 4-2 *Locations of common and differential regional response to reward/avoidance outcome.*

Region	Peak MNI Coordinates			Peak t value	Spatial extent (in contiguous voxels)
	x	y	z		
[R_+ + P_-] > [R_- + P_+]					
L. Putamen	-14	4	-10	7.04	409
L. Putamen	-24	0	0	5.25	
R. Putamen	16	8	-10	5.78	576
R. Putamen	30	4	6	4.86	
L. Medial orbitofrontal	-10	48	-10	5.01	412
L. Medial orbitofrontal	-6	40	-8	4.63	
R. Posterior cingulate	4	-32	22	5.31	507
R. Posterior cingulate	8	-44	26	4.79	
R. superior temporal	62	-34	12	5.12	154
R. Superior temporal	64	-22	2	4.47	
L. Medial orbitofrontal	-10	48	-10	5.01	412
L. Medial orbitofrontal	-6	40	-8	4.63	
L. Inferior orbitofrontal	-38	36	-16	4.90	171
L. Middle orbitofrontal	-24	36	-12	4.20	
L. Superior orbitofrontal	-24	46	-8	3.83	
[R_+ + P_-] < [R_- + P_+]					
R. Insula	32	20	10	-3.32	223

*Note: **[R_+ + P_-] > [R_- + P_+]** – [Rewarded outcome + Avoidance outcome] - [Non-rewarded outcome + Punishment outcome].

Brain regions response to reward/avoidance expected value

The reward expected value signal was found positively correlated with the brain activation at the left middle temporal gyrus (mTG, [-56, -10, -12]; t = 4.62, k = 477), left superior medial frontal gyrus ([-2, 58, 34]; t = 3.81, k = 764), right anterior cingulate (ACC, [16, 38, 0]; t = 4.34, k = 539), and also the brain region peaked at ([10, -30, 26]; t = 4.42, k = 1672) covering bilateral middle cingulum and right hippocampus *at $p < 0.01$* after FWE correction (see ***Figure 4- 5***).

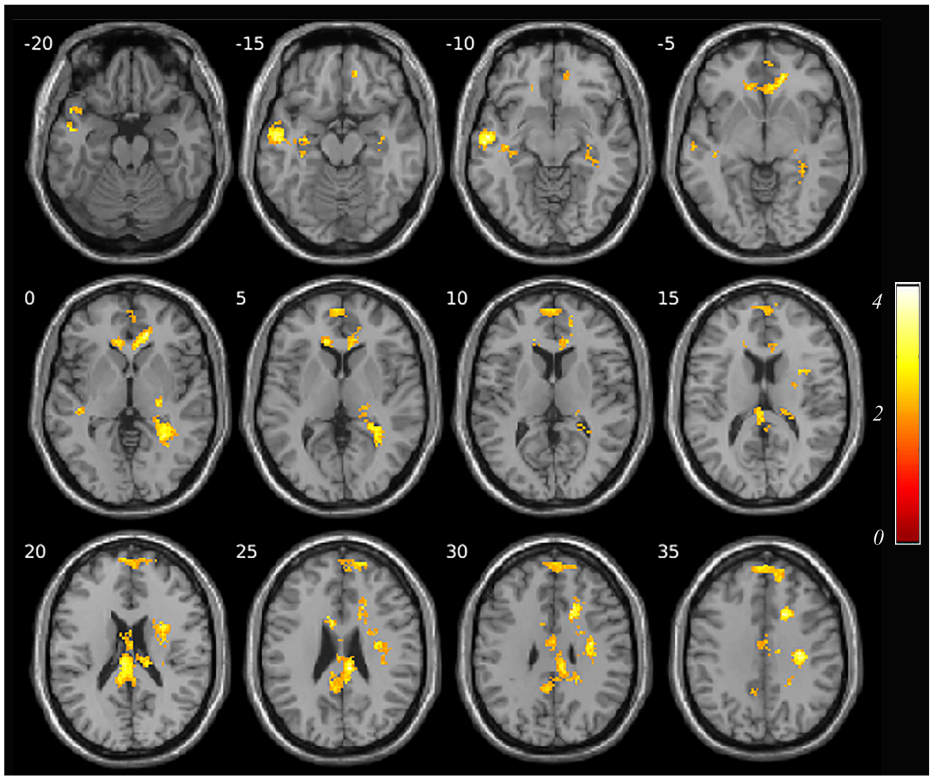

Figure 4-5 *Neural correlates of reward expected value. The reward expected value signal was found positively correlated with the brain activation at the left mTG, left superior medial frontal, right anterior cingulate, and also the brain region including bilateral middle cingulum and right hippocampus (see* ***Table 4-3 for details****).*

Whereas the avoidance expected value was positively correlated with the brain region peaked at ([40, -4, 20]; t = 5.01, k = 241) including *right putamen, left putamen* ([-26, 6, 10]; t = 4.40, k = 108), *left mOFC* ([-2, 46, -14]; t = 4.60, k = 145) extending to the *right mOFC*, as well the *left sTG* ([-54, -4, 6]; t = 4.81, k = 171) (see ***Figure 4-6***). Meanwhile it was found that avoidance expected value was negatively correlated with the *left inferior OFC* peaked at ([-32, 26, -6]; t = 6.47, k = 554), *right insula* ([34, 26, -4]; t = 5.96, k = 765), *right superior medial frontal* ([10, 30, 44]; t = 5.46, k = 1414) including the *right middle cingulum* and *right ACC*, *right caudate* peaked at ([12, 8, 8]; t = 4.86, k = 157) and *right precentral gyrus* ([40, 0, 42]; t = 4.58, k = 110) after correction (see ***Figure 4-7***).

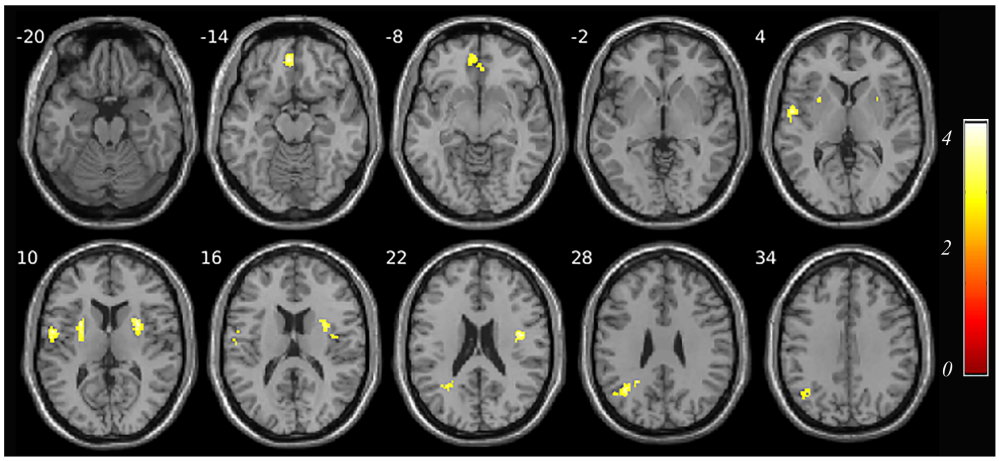

Figure 4-6 *Neural correlates of avoid expected value. The expected value under the avoidance condition was found positively correlated with the brain region covering right putamen, the left putamen, the left mOFC extending to the right mOFC, as well the left sTG (see* ***Table 4-3 for details****).*

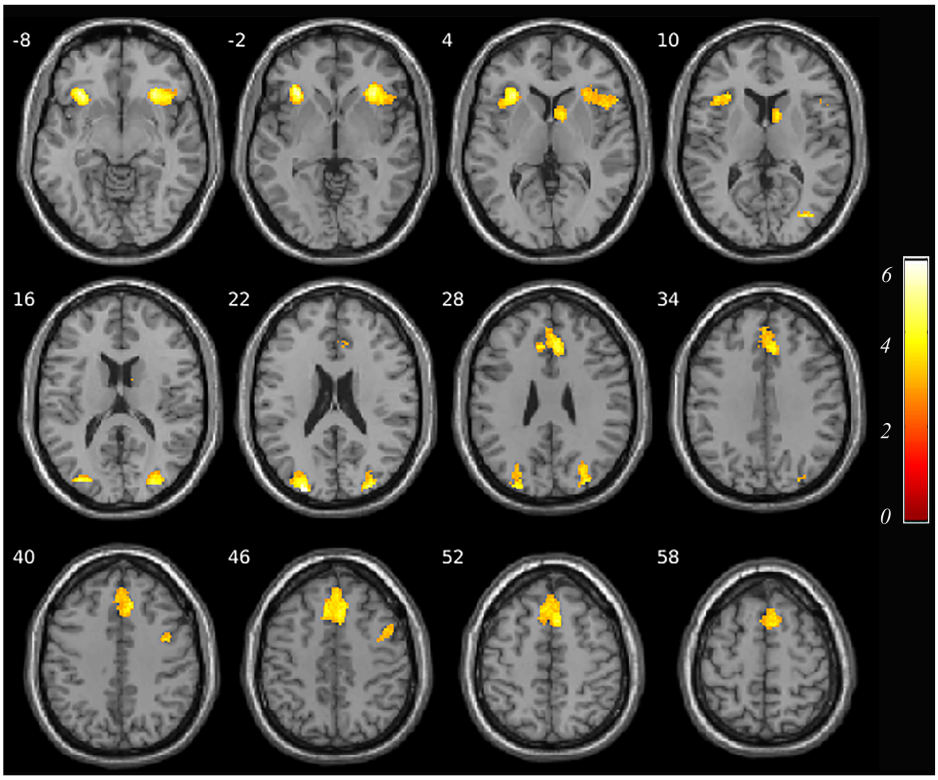

Figure 4-7 *Neural correlates of avoid expected value. The expected value under the avoidance condition was found negatively correlated with the left inferior OFC, right insula, right superior medial frontal including the right middle cingulum and right ACC, right caudate and the right precentral gyrus (see* ***Table 4-3 for details****).*

Table 4-3 *Locations of regional response to reward/avoidance expected value.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward expectation					
L. Middle Temporal	-56	-10	-12	4.62	477
L. Middle Temporal	-50	-6	-18	3.67	
L. Superior Frontal	-2	58	34	3.81	764
R. Middle Frontal	20	60	36	3.69	
L. Superior medial frontal	0	60	8	3.45	

R. Anterior Cingulate	16	38	0	4.34	539
R. Middle Cingulum	10	-30	26	4.42	1672
R. Middle Cingulum	2	-6	30	2.94	
L. Middle Cingulum	-4	-48	26	3.06	
R. Hippocampus	36	-32	-10	2.49	
Avoid expectation					
Positive correlation					
R. Putamen	40	-4	20	5.01	241
L. Superior Temporal	-54	-4	6	4.81	171
L. Postcentral	-56	-12	14	3.56	
L. Medial Orbitofrontal	-2	46	-14	4.60	145
R. Medial Orbitofrontal	4	38	-8	3.92	
L. Angular	-24	-52	28	4.42	186
L. Putamen	-26	6	10	4.40	108
L. Putamen	-24	-8	12	4.17	
Negative correlation					
L. Inferior Orbitofrontal	-32	26	-6	6.47	554
R. Insula	34	26	-4	5.96	765
R. Insula	40	22	-8	4.94	
R. Superior medial frontal	10	30	44	5.46	1414
R. Middle Cingulum	8	26	30	5.36	
R. Anterior Cingulum	8	34	28	5.29	
R. Caudate	12	8	8	4.86	157
R. Precentral	40	0	42	4.58	110

The conjunction analysis showed that the right middle cingulum ([16, -16, 46]; t = 3.63, k = 2225) were commonly activated by the reward and avoided expected value (see ***Figure 4-8***) at $p < 0.05$ with FWE correction at cluster level. Further, the analysis showed

that the right middle frontal ([20, 60, 26]; t = 4.57, k = 292) were found higher activation for reward expected value compared to avoid expected value (see ***Figure 4-9***).

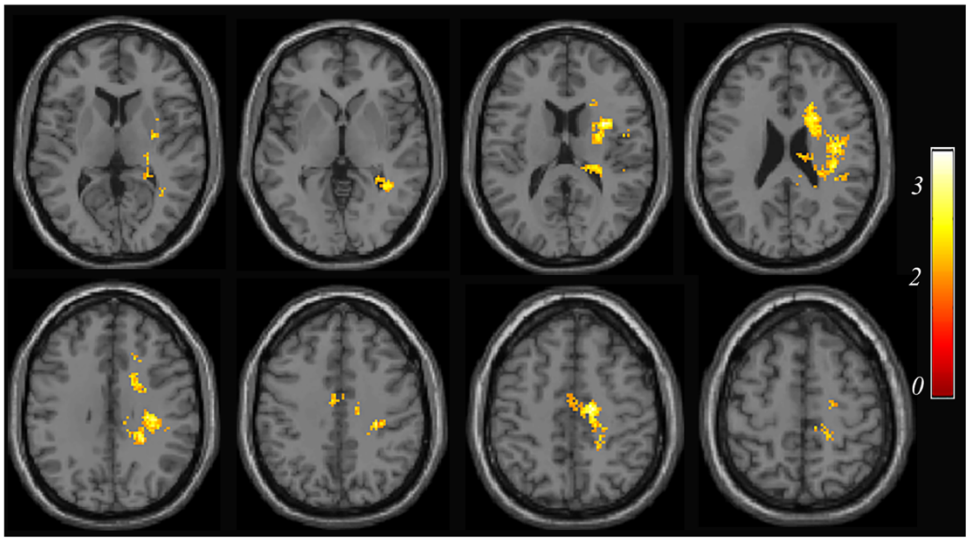

Figure 4-8 *Common neural correlates of positive effects of reward and avoid expected value. The right middle cingulum was found commonly activated by the reward and avoid expected value (see* ***Table 4-4 for details****).*

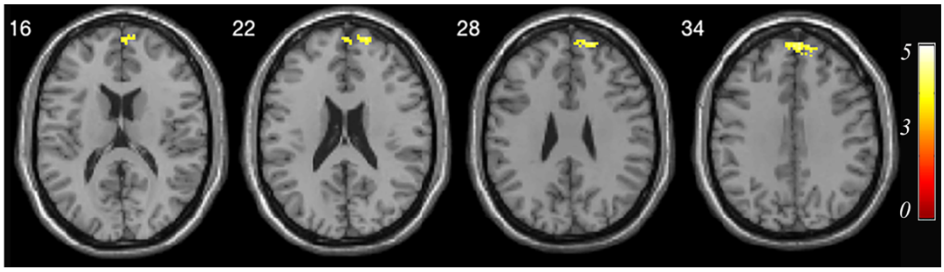

Figure 4-9 *The locations of regional response differences between reward and avoid expected value. The right middle frontal was found showing significantly higher activation to the reward expected value compared to avoid expected value (see* ***Table 4-4 for details****).*

Table 4-4 *Locations of common and differential regional response to expected value under reward and avoidance condition.*

Region	Peak MNI Coordinates			Peak t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward expected value & Avoid expected value					
R. Middle cingulum	16	-16	46	3.53	2225
Reward expected value > Avoid expected value					
R. Middle frontal	20	60	26	4.57	292
R. Medial superior frontal	6	58	34	3.93	
R. Superior frontal	18	54	36	3.85	

Brain regions response to reward/aversive PE

The reward PE signal was found to correlate positively with the activation at several clusters expanding across cortical and subcortical region, the largest cluster peaking at the right fusiform ([30, -50, -12]; t = 12.36; k = 25828) covering the *striatum, cingulate, bilateral insula, hippocampus, thalamus, inferior & middle frontal and sTG*. And the other clusters including left middle occipital ([-24, -86, 18]; t = 9.09, k=2640), right superior frontal gyrus (sFG, [18, 60, 26]; t = 4.53, k = 277) and left supplementary motor area (SMA, [-6, 10, 54]; t = 4.94, k = 234) extending to right SMA (see ***Figure 4-10***).

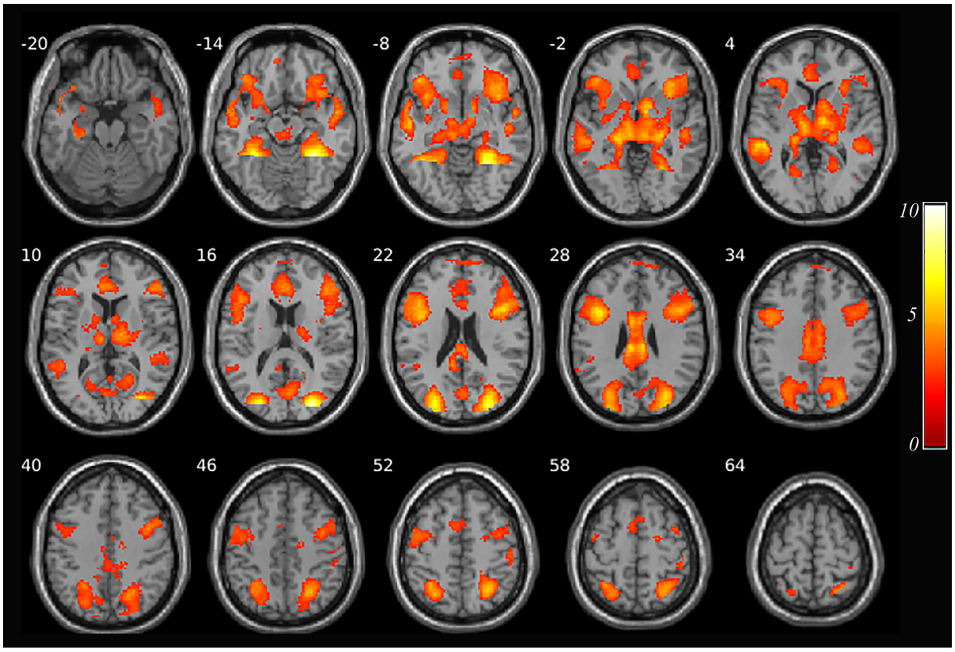

Figure 4-10 *Neural correlates of reward PE. The signal was found positively associated with the activity in the brain regions including right fusiform covering the striatum, cingulate, bilateral insula, hippocampus, thalamus, inferior & middle frontal and sTG. And the other clusters including left middle, right superior frontal and left SMA extending to right SMA (see* ***Table 4-5 for details****).*

The PE signal under avoidance condition which included an aversive outcome was received when unexpected and an aversive outcome was not received when expected, was referred as aversive PE. It was found correlated with the activity at brain regions including

right fusiform ([30, -52, -10]; t = 13.82, k = 3697) expanding to bilateral thalamus and left hippocampus, left insula ([-30, 18, -12]; t = 7.42, k = 2826), right inferior OFC ([40, 28, -27]; t = 7.41, k = 1709) covering right insula, bilateral caudate peaked at left side ([-8, 4, 4]; t = 5.63, k = 183) and right side ([12, 8, 2]; t = 4.99, k = 150) respectively, right SMA peaked at ([2, 8, 60]; t = 6.92, k = 1171) expanding to left SMA and anterior & middle cingulum, right mTG ([56, -40, 4]; t = 5.75, k = 405), right opercular part of the inferior frontal ([38, 10, 30]; t = 6.03, k = 1220) covering the right precentral and right middle frontal, and the left precentral peaked at ([-42, 0, 58]; t = 8.44, k = 631) (see ***Figure 4-11***).

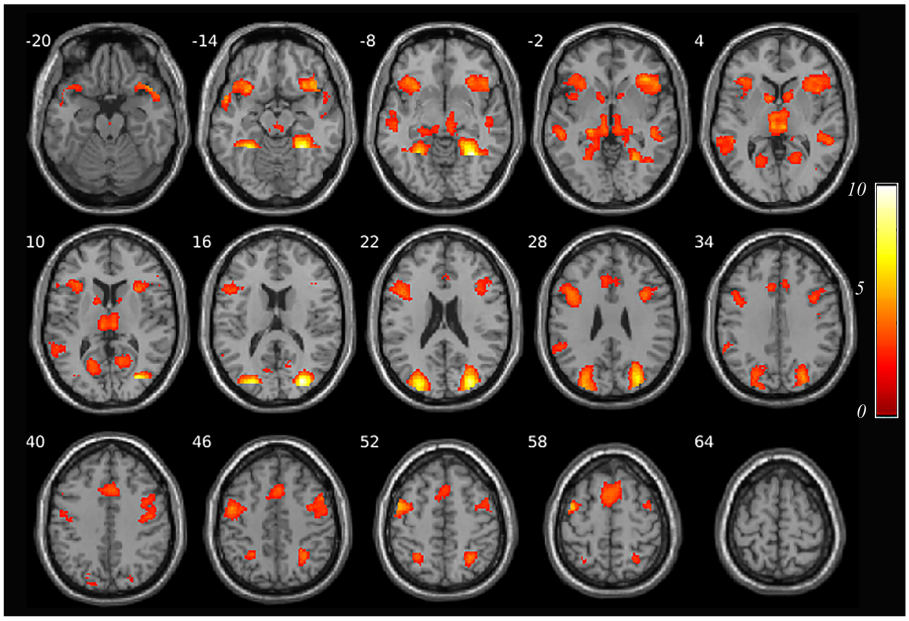

Figure 4-11 *Neural correlates of aversive PE. The activated brain regions were right fusiform, bilateral thalamus and left hippocampus, left insula, right inferior OFC covering right insula, bilateral caudate, right SMA expanding to left SMA and anterior & middle cingulum, right mTG, right opercular part of the inferior frontal covering the right precentral and right middle frontal, and the left precentral (see* ***Table 4-5 for details****).*

Table 4-5 *Locations of regional response to PE under reward/avoidance condition.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward PE					
R. Fusiform	30	-52	-10	12.36	25828
R. Middle occipital	28	-86	16	11.45	
L. Fusiform	-32	-40	-18	9.90	
L. Supplementary motor area	-6	10	54	4.94	234
L. Supplementary motor area	2	14	48	3.47	
R. Postcentral	50	-28	56	5.01	191
R. Postcentral	44	-36	62	4.75	
R. Superior frontal	18	60	26	4.53	277
L. Superior medial frontal	2	62	24	4.15	
Aversive PE					
R. Fusiform	30	-52	-10	13.82	3697
L. Fusiform	-26	-46	-14	10.25	
L. Insula	-30	18	-12	7.42	2026
L. Inferior orbitofrontal	-36	24	-12	6.66	
L. Opercular part of the inferior frontal	-44	16	20	6.42	
R. Inferior orbitofrontal	40	28	-2	7.41	1709
R. Inferior orbitofrontal	40	24	-14	7.17	
R. Insula	32	22	-14	7.02	
R. Supplementary motor area	2	8	60	6.92	1171
R. Supplementary motor area	6	22	56	5.57	
R. Opercular part of the inferior frontal	38	10	30	6.03	1220
R. Precentral	50	8	46	5.31	
R. Middle temporal	56	-40	4	5.75	405

R. Superior temporal	46	-32	-2	5.62	
R. Middle temporal	50	-20	-10	4.25	
L. Caudate	-8	8	4	5.63	183
R. Caudate	12	8	2	4.99	150

The conjunction analysis showed that the right insula ([38, 28, -2]; t = 6.45, k = 2457), left insula ([-32, 18, -10]; t = 5.99, k = 1611), left opercular part of the inferior frontal ([-38, 8, 28]; t = 5.93, k = 761), right caudate ([12, 10, 2]; t = 5.27, k = 11), and left SMA ([-4, 10, 60]; t = 4.37, k = 174) extending to the *right SMA* were found commonly activated for reward and aversive PE (see ***Figure 4-12***). Meanwhile, the right fusiform ([30, -50, -12]; t = 17.76, k = 33802) covering the *bilateral insula*, *bilateral thalamus*, *striatum* and the left SMA ([-4, 10, 60]; t = 6.90, k = 5754) extending to the *right SMA* were found higher activation to reward PE compared with aversive PE (see ***Figure 4-13***).

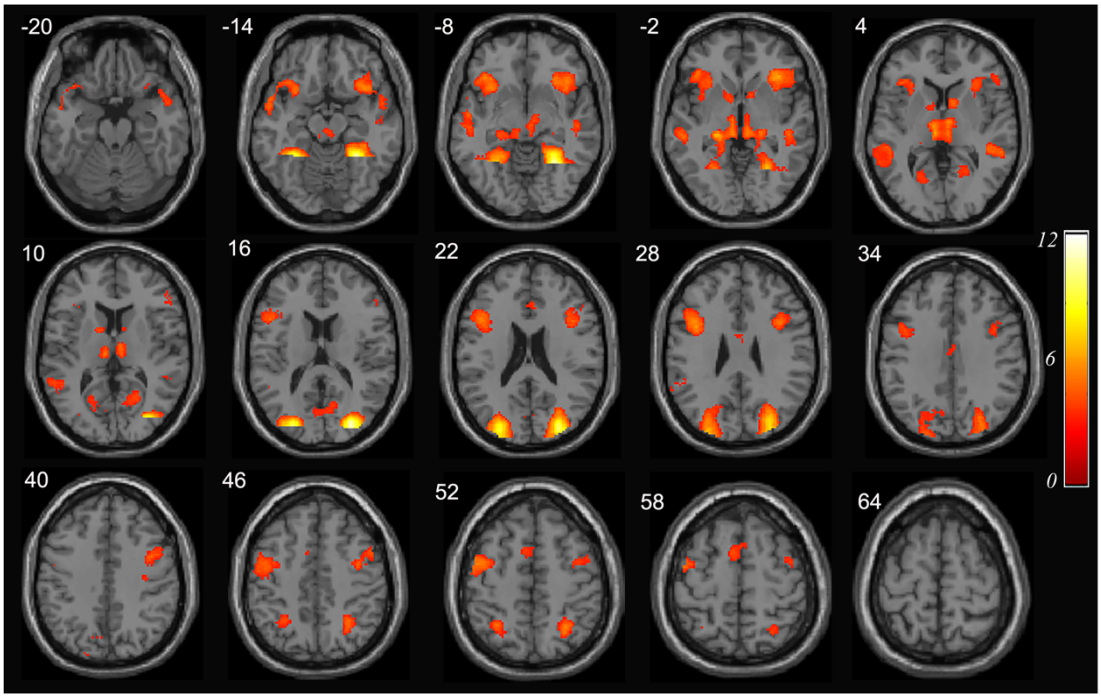

Figure 4-12 *Conjunction analysis of reward and aversive PE. The bilateral insula, left inferior frontal, right dorsal striatum and bilateral SMA were found commonly activated for reward and aversive PE (see* ***Table 4-6*** *for details).*

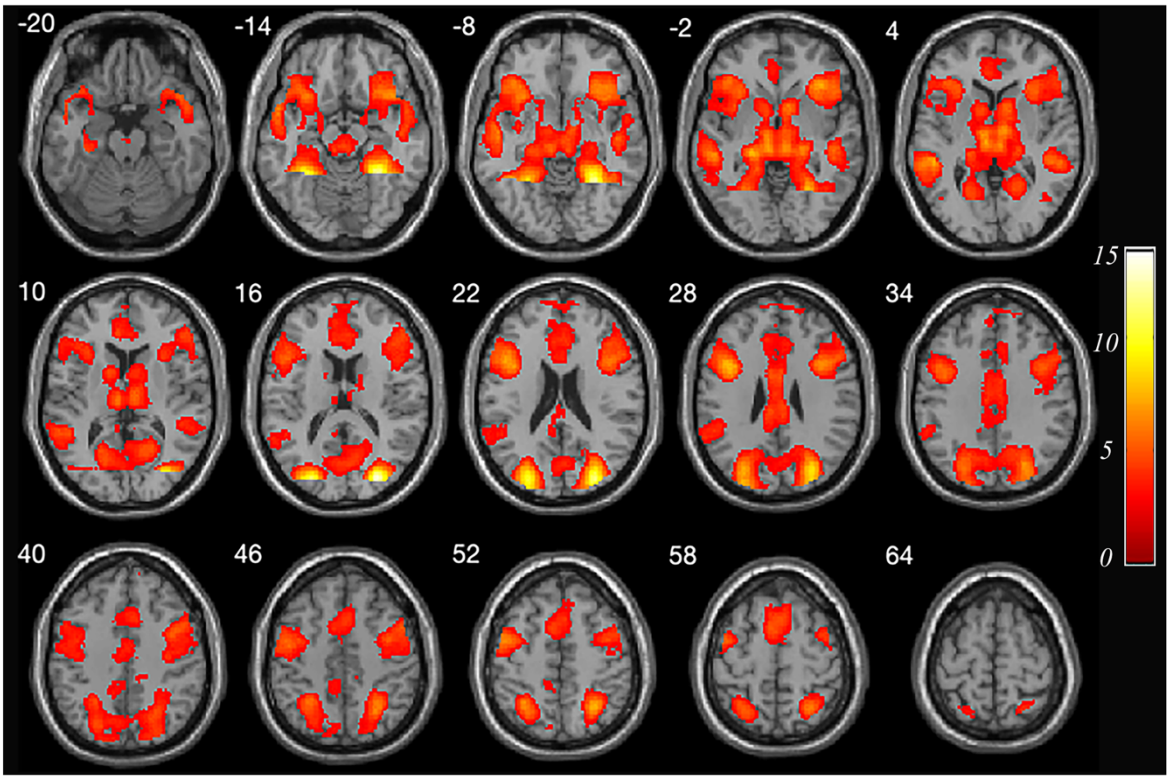

Figure 4-13 *The locations of response differences between reward and aversive PE. The right fusiform covering the bilateral insula, bilateral thalamus, striatum and the left SMA extending to the right SMA were found to have higher activation to reward PE compared with aversive PE (see* ***Table 4-6 for details****).*

Table 4-6 *Locations of common and differential regional responses to PE under reward and avoidance condition.*

Region	Peak MNI Coordinates			Peak t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward PE & Aversive PE					
R. Insula	38	28	-2	6.45	2457
R. Middle temporal	48	-36	2	5.21	
L. Insula	-32	18	-10	5.99	1611
L. Inferior orbitofrontal	-34	26	-10	5.55	
L. Opercular part of inferior frontal	-38	8	28	5.93	761
L. Triangular part of inferior frontal	-44	18	18	5.75	
L. Inferior parietal	30	-56	52	5.67	300
L. Precentral	-42	-2	54	5.53	540
R. Caudate	12	10	2	5.27	115
L. Inferior parietal	-30	-54	50	5.05	206
L. Superior parietal	-22	-64	54	3.37	
L. Supplementary motor area	-4	10	60	4.37	174
R. Supplementary motor area	6	16	58	3.61	
Reward PE > Aversive PE					
R. Fusiform	30	-50	-12	17.76	33802
R. Middle occipital	30	-82	14	16.81	
L. Middle occipital	-26	-84	16	14.09	
L. Supplementary motor area	-4	10	60	6.90	5754
R. Supplementary motor area	4	14	60	6.51	

4.4 Discussion

In this study, through a novel probabilistic reward and avoidance learning task with a switch mid-way through task performance, we examined the full potential neural bases at the outcome and anticipation phases of reward/avoidance learning in healthy controls, and importantly, the neural mechanism of the robust learning PE signal under reward and avoidance conditions. Behaviourally, the healthy participants showed a significant preference of the Correct choice under reward and avoidance condition. Also, the reaction time is significantly faster under reward condition and slower under avoidance condition compared to neutral condition. The results successfully replicated the findings of Kim et al (2016).

The imaging analysis showed that the outcome of receiving reward and getting punishment had a common neural substrate including *cingulum* and *bilateral insula*, but also the distinct neural substrates including *inferior OFC*, *bilateral thalamus* and *whole striatum* as well as *dorsal striatum* involved with the outcome of getting reward and punishment, respectively. The outcome of receiving reward and successfully avoiding punishment was found associated with the consistent *mOFC* implicated in Kim et al. study (Kim et al., 2006), and we also found the new candidates including *posterior cingulum* and *dorsal striatum* involved with this process. While, the outcome of missing reward and receiving punishment was found to be related to higher activations in the common brain region at right *insula*. At the stage of value anticipation, the reward and avoidance expected value had a common neural substrate including *middle cingulum*. And the expected value under avoidance condition recruited broader brain regions including the *inferior OFC, insula and dorsal striatum*.

Specifically, at the phase of error processing, the reward PE signal was found correlated with the activity at the *cortical-basal ganglia* brain areas including the *whole*

striatum, cingulate, insula, hippocampus, thalamus, inferior & middle frontal and SMA. The aversive PE signal was found covaried with the activation at the common brain regions including *cingulate, insula, hippocampus, thalamus, inferior frontal* and *SMA* as well as the specific *dorsal striatum.*

When faced with multiple options, making a decision requires one to compute the values associated with the outcomes of each action. Same with our findings, previous studies have reported that the set of brain regions including OFC, striatum, thalamus, and cingulum are involved in the processing of monetary reward outcome (Oldham et al., 2018b). It has also been previously found that processing loss/aversive outcomes are associated with activation in the dorsal striatum and cingulum (Dumais and Bitar, 2018). The OFC receives information from the object-processing visual stream and could be activated by some primary reinforcers such as pleasant or painful touch (Rolls, 2000). Furthermore, the OFC is suggested to be critical for representing the outcomes of actions and their subsequent impact on the control of behaviour. The medial OFC was consistently found not only activated by rewarding outcomes, there is also evidence for medial OFC activation when successfully avoiding punishment (Kim et al., 2006). An electrophysiological study on non-human primates also found that OFC neurons are involved in both aversive and reward processing, and they encode relative preferences for reward and aversive outcomes (Hosokawa et al., 2007). Besides the OFC, the striatum is reported to be crucial for both learning to approach rewarding outcomes (Lau and Glimcher, 2007), and in avoiding aversive outcomes (Salamone, 1994, 2002). While a distinct anatomical connectivity pattern for the dorsal and ventral striatum has been suggested (Voorn et al., 2004). According to the previous literature, the ventral striatum is reported to be more activated by reward value than loss outcome (Ino

et al., 2010), whereas dorsal striatum was modulated by receiving reward as well as getting punishment (Mattfeld et al., 2011).

Reward-expectation activity has been suggested as an appropriate signal for predicting the occurrence of rewards and thereby provides a suitable mechanism for influencing behaviour that leads to the acquisition of rewards (Schultz, 2000) [37]. Previous studies report that the ACC has strong connections with motor areas and a few direct connections with the sensory cortex, thus, it has been suggested to be responsible for action value calculation to produce a favourable outcome (Jerome et al., 2007; Philiastides et al., 2010). A previous study reported that the activation in rostral ACC/mPFC and amygdala were related to increases in the level of expected reward (Marsh et al., 2007). An overlapping value-related activity within ventromedial prefrontal cortex was reported during anticipation of juice and money reward outcomes (Savage and Ramos, 2009), and the Medial frontal cortex was reported to encode the reward expectation (Silvetti et al., 2014). The anterior insula was found correlated with expected values, occurring when low reward outcomes were expected (Rolls et al., 2008). Further, the Basolateral amygdala projected to OFC was reported to enable the cue-triggered reward expectations that can motivate the execution of specific action plans and allow adaptive conditional responding (Lichtenberg et al., 2017).

In contrast to reward expectation, activation was observed in the inferior OFC, insula and superior medial frontal under the avoidance condition during anticipation phase in our study. The insula activation at the anticipation phases in avoidance condition was reported in a previous study (Kim et al., 2014). The anterior insula is one of the brain structures engaged in emotion processing related to the representation or regulation of an organism's state (Damasio et al., 2000), and its role in the interoception of physiological states elicited during emotional experience has been reported (Palminteri et al., 2012; H and Antoine, 2013). The

activation of the insula suggests that the interoceptive representations of monetary outcomes associated with fractal selections were retrieved during the anticipation phase (Kim et al., 2014). The dorsal striatum activation during anticipation in the avoidance trials was also found in a previous study (Kim et al., 2014), and thus, the dorsal striatum is involved in the evaluation process of the alternative choices to avoid the worst (Palminteri et al., 2012). Besides the previously mentioned role in outcome processing, the OFC has also been implicated in the evaluation of negative outcomes (Kringelbach and Radcliffe, 2005), and the OFC activation during anticipation of outcomes could reflect the potential negative consequences, even if the participants selected the action with a higher expected value.

The error signal in the reward condition was found to be correlated with activation of cortical-basal ganglia brain regions. In line with these findings, previous studies have suggested that the error signal is computed at midbrain dopamine neurons and a global reinforcement signal is emitted to other neurons in the striatum and prefrontal cortex, which could underlie the learning of appropriate behaviours (Schultz et al., 1997; Schultz, 2000; Glimcher and Bayer, 2005). Previous animal studies have found that dopamine neurons show phasic response to food and liquid rewards (Ljungberg et al., 1991, 1992; Schultz et al., 1993). The striatal dopaminergic system was found to carry distinct messages by different means, which can be integrated differently to shape the basal ganglia responses to reward-related events (Satoh et al., 2003). Ventral striatum is a subdivision of the basal ganglia that includes the nucleus accumbens, parts of the olfactory tubercle, as well as ventral and medial portions of the putamen and caudate nucleus (Holt et al., 1997). Using attractive faces as the visual stimulus, a previous study has shown that the learning process elicits a reward PE in the ventral striatum (Bray and O'Doherty, 2007). Similar results were found in a probabilistic decision task, in which activations in midbrain and ventral striatum were correlated with the

PE signal (Rolls et al., 2008). An overlapping of PE signal was reported during learning with juice and money reward in the dorsal striatum, while the PE signal was significantly stronger during learning with money but not juice reward in the ventral striatum (Valentin and Doherty, 2020). Participants were scanned using event-related fMRI while undergoing appetitive conditioning with a pleasant taste reward. Regression analysis revealed that responses in ventral striatum and OFC were significantly correlated with this error signal (O'Doherty et al., 2003).

The aversive PE is fundamental to avoidance learning, and it was found to be correlated with activity in the frontal-subcortical brain regions including dorsal striatum and insula in our study. Striatal dopamine release is reported to convey a learning signal during both appetitive and aversive conditions (Stelly et al., 2019). Further, striatal structure and function has been reported to predict individual biases in learning to avoid pain (Eldar et al., 2016). A previous study using the Pavlovian conditioning of visual cues to elicit outcomes that simultaneously incorporate the chance of financial reward and loss, the striatal activation, especially the more posterior regions was reported to reflect the PE of loss (Seymour et al., 2007). In a classic fear conditioning paradigm, it was demonstrated that the BOLD signals in the striatum, particularly the head of the caudate nucleus, were correlated with aversive PE (Delgado et al., 2008). The midbrain dopamine system is suggested to be involved in the processing of aversive and reward PE signals (Brooks and Berns, 2013). A previous animal study showed that hemodynamic responses and theta oscillations recorded from the amygdala show activity patterns consistent with aversive PE (McHugh et al., 2014). The basolateral part of the amygdala was reported to encode an aversive PE that quantifies whether cues and outcomes were worse than expected (Michely et al., 2020). As a distinct pattern of PE found for studies using rewarding and aversive reinforcers, reward PE were

observed primarily in the striatum while aversive PE were found more widely including insula and habenula (Garrison et al., 2013c). Another study using juice as a stimulus reported that the activity of a brain network composed of the striatum, anterior insula, and anterior cingulate cortex covaried with the prediction of an aversive taste (Metereau et al., 2013).

In summary, the overlapped and distinct brain regions were found involved in reward and avoidance-based decision processes in the present study. At the outcome stage, receiving reward and punishment was associated with the functional activity in common brain areas including the *insula and cingulum*, whereas there were distinct activations in relation to reward (whole striatum) and punishment (*dorsal striatum*). In addition, the *cingulum* was also activated for both reward and avoidance expectation. Avoidance expectation recruited broader areas at the cortical and subcortical brain areas including *inferior OFC*, *insula,* and *dorsal striatum*. At the stage of error processing, the reward PE signal was found associated with the activity in the *cortical and subcortical areas.* Meanwhile the aversive PE signal was covaried with the activity at the shared frontal-subcortical brain regions and the segregated dorsal part of striatum. The findings show that the specific *dorsal striatum plays a critical* role for differential phases of avoiding decision processes, and also supports the existence of dissocial computational processes in the brain for reward and punishment processing.

ACKNOWLEDGMENTS

The authors thank all healthy participants for participation in this study, and the technical support by the MASSIVE HPC facility (www.massive.org.au).

AUTHOR CONTRIBUTIONS

Xiaoliu, Z., conceived and conducted the data analysis, and drafted the manuscript. Chao, S., Amir, D., Shinsuke S., Ben, F., Leonardo, F. and Ben, J F., contributed to the materials and analysis tools. Leah, B. and Linden, P., performed the experiments and collected the data. Carsten, C. and Murat, Y. conceived and designed the experiments.

FUNDING SUPPORT

Murat Yücel has received funding from Monash University, and Australian Government funding bodies such as the National Health and Medical Research Council (NHMRC; including Fellowship #APP1117188), the Australian Research Council (ARC), and the Department of Industry, Innovation and Science. Also, he has received philanthropic donations from the David Winston Turner Endowment Fund, Wilson Foundations, as well as payment from law firms in relation to court and/or expert witness reports. The funding sources had no role in the design, management, data analysis, or interpretation and writing up of the manuscript. Dr. Fontenelle supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (308237/2014-5), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (203590); the D'Or Institute of Research and Education; and the David Winston Turner Endowment Fund.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

Reference

Alexandre Y. Dombrovski MD, Luke Clark DP, Greg J. Siegle PD (2011) Reward/Punishment reversal learning in older suicide attempters. 167:699–707.

Ben S, P OJ, Peter D, Martin K, K ones A, J DR, J FK, J FRS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664–667 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15190354&dopt=Abstract.

Bray S, O'Doherty J (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. J Neurophysiol 97:3036–3045.

Brooks AM, Berns GS (2013) Aversive stimuli and loss in the mesocorticolimbic dopamine system. Trends Cogn Sci 17:281–286 Available at: http://dx.doi.org/10.1016/j.tics.2013.04.001.

Coddington LT, Dudman JT (2018) The timing of action determines reward prediction signals in identified midbrain dopamine neurons. Nat Neurosci 21:1563–1573.

Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LLB, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. 09:1049–1056.

Delgado MR, Li J, Schiller D, Phelps EA (2008) The role of the striatum in aversive learning and aversive prediction errors. :3787–3800.

Doherty JPO, Cockburn J, Pauli WM (2017) Learning, Reward, and Decision Making.

Dumais A, Bitar N (2018) Loss anticipation and outcome during the Monetary Incentive Delay Task: a neuroimaging systematic review and meta-analysis. :1–23.

Eldar E, Hauser TU, Dayan P, Dolan RJ (2016) Striatal structure and function predict individual biases in learning to avoid pain. Proc Natl Acad Sci 113:4812–4817 Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1519829113.

Engel A, Caceda R (2015) Can Decision Making Research Provide a Better Understanding of Chemical and Behavioral Addictions? Curr Drug Abuse Rev 8:75–85.

Eshel N, Steinberg EE (2018) Learning what to approach. PLoS Biol 16:e3000043.

Garrison J, Erdeniz B, Done J (2013a) Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310 Available at: http://dx.doi.org/10.1016/j.neubiorev.2013.03.023.

Garrison J, Erdeniz B, Done J (2013b) Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310 Available at: http://dx.doi.org/10.1016/j.neubiorev.2013.03.023.

Garrison J, Erdeniz B, Done J (2013c) Prediction Error in Reinforcement Learning: A Meta - analysis of Neuroimaging studies. Neurosci Biobehav Rev:1–50.

Glimcher PW, Bayer HM (2005) Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. Neuron 103:2304–2312.

H NN, Antoine B (2013) The insula and drug addiction: an interoceptive view of pleasure, urges and decision-

making. 214:435–450.

Holt DJ, Graybiel ANNM, Saper CB (1997) Neurochemcial architecture of the human striatum. 25:1–25.

Hosokawa T, Kato K, Inoue M, Mikami A (2007) Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. 57:434–445.

Howard JD, Kahnt T (2018) Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. Nat Commun:1–11 Available at: http://dx.doi.org/10.1038/s41467-018-04055-5.

Ino T, Nakai R, Azuma T, Kimura T, Fukuyama H (2010) Differential activation of the striatum for decision making and outcomes in a monetary task with gain and loss. CORTEX 46:2–14 Available at: http://dx.doi.org/10.1016/j.cortex.2009.02.022.

Jerome S, Rene Q, Marie R, Julien V, Jean-Paul J, Emmanuel P (2007) Expectations, gains, and losses in the anterior cingulate cortex. 7:327–336.

Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:1453–1461.

Kim SH, Yoon HS, Kim H, Hamann S (2014) Individual differences in sensitivity to reward and punishment and neural activity during reward and avoidance learning. Soc Cogn Affect Neurosci 10:1219–1227.

Krigolson OE, Hassall CD, Handy TC (2014) How We Learn to Make Decisions: Rapid Propagation of Reinforcement Learning Prediction Errors in Humans. :635–644.

Kringelbach ML, Radcliffe J (2005) The human orbitofrontal cortex: linking reward to hedonic experience. 6:691–702 Available at: www.nature.com/reviews/neuro.

Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. J Neurosci 27:14502–14514.

Lichtenberg NT, Pennington ZT, Holley SM, Greenfield VY, Cepeda C, Levine MS, Wassum KM (2017) Basolateral amygdala to orbitofrontal cortex projections enable cue-triggered reward expectations. J Neurosci 37:8374–8384.

Ljungberg T, Apicella P, Schultz W (1991) Responses of monkey midbrain dopamine neurons during delayed alter- nation performance. 567:337–341.

Ljungberg T, Physiologie I De, Apicella P (1992) Responses of Monkey Dopamine Neurons During Learning of Behavioral Reactions. 67:145–163.

Maleki S, Chye Y, Zhang X, Parkes L, Chamberlain SR, Fontenelle LF, Braganza L, Youssef G, Lorenzetti V, Harrison BJ, Yücel M, Suo C (2020) Neural correlates of symptom severity in obsessive-compulsive disorder using magnetization transfer and diffusion tensor imaging. Psychiatry Res - Neuroimaging 298:111046 Available at: https://doi.org/10.1016/j.pscychresns.2020.111046.

Marsh AA, Blair KS, Vythilingam M, Busis S, Blair RJR (2007) Response options and expectations of reward in decision-making: The differential roles of dorsal and rostral anterior cingulate cortex. Neuroimage 35:979–988 Available at: http://dx.doi.org/10.1016/j.neuroimage.2006.11.044.

Mattfeld AT, Gluck MA, Stark CEL (2011) Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment. Learn Mem 18:703–711.

McHugh SB, Barkus C, Huber A, Capitao L, Lima J, Lowry JP, Bannerman DM (2014) Aversive Prediction Error Signals in the Amygdala. J Neurosci 34:9024–9033 Available at: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4465-13.2014.

Metereau E, Dreher J, Recherche M De (2013) Cerebral Correlates of Salient Prediction Error for Different Rewards and Punishments. :477–487.

Michely J, Rigoli F, Rutledge RB, Hauser TU, Dolan RJ (2020) Archival Report Distinct Processing of Aversive Experience in Amygdala Subregions. Biol Psychiatry Cogn Neurosci Neuroimaging 5:291–300 Available at: https://doi.org/10.1016/j.bpsc.2019.07.008.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain RID D-9230-2011. Neuron 38:329–337.

Oldham S, Murawski C, Fornito A, Youssef G, Lorenzetti V (2018a) The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. :3398–3418.

Oldham S, Murawski C, Fornito A, Youssef G, Yücel M, Lorenzetti V (2018b) The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. Hum Brain Mapp 39:3398–3418.

Palminteri S, Justo D, Jauffret C, Pavlicek B, Dauta A, Delmaire C, Czernecki V, Karachi C, Capelle L, Durr A, Pessiglione M (2012) Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. Neuron 76:998–1009.

Parkes L, Fulcher B, Yücel M, Fornito A (2018) An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. Neuroimage 171:415–436 Available at: https://doi.org/10.1016/j.neuroimage.2017.12.073.

Peter D, Daw ND (2008) C ONNECTIONS B ETWEEN C OMPUTATIONAL AND N EUROBIOLOGICAL P ERSPECTIVES ON D ECISION M AKING Decision theory, reinforcement learning, and the brain. 8:429–453.

Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR (2010) Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage 53:221–232 Available at: http://dx.doi.org/10.1016/j.neuroimage.2010.05.052.

Reynolds JNJ, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. :67–70.

Riaz N, Wolden SL, Gelblum DY, Eric J (2016) Dopamine neurons share common response function for reward prediction error. Nat Neurosci 118:6072–6078.

Rolls ET (2000) The Orbitofrontal Cortex and Reward. Cereb Cortex 10:284–294.

Rolls ET, McCabe C, Redoute J (2008) Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. Cereb Cortex 18:652–663.

Salamone JD (1994) The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. 61:117–133.

Salamone JD (2002) Motivational views of reinforcement: implications for understanding the behavioural functions of nucleus accumbens dopamine. 137.

Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated Coding of Motivation and Outcome of Decision by

Dopamine Neurons. 23:9913–9923.

Savage LM, Ramos RL (2009) Reward expectation alters learning and memory: The impact of the amygdala on appetitive-driven behaviors. Behav Brain Res 198:1–12.

Scheibehenne B, Pachur T (2015) Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. Psychon Bull Rev 22:391–407.

Schultz W (2000) Multiple Reward Signals in the brain. 1.

Schultz W (2016) Dopamine reward prediction error coding. Dialogues Clin Neurosci:23–32.

Schultz W (2017) Reward prediction error. Curr Biol 27:R369–R371 Available at: http://dx.doi.org/10.1016/j.cub.2017.02.064.

Schultz W (2018) Predictive Reward Signal of Dopamine Neurons.

Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J Neurosci 13:900–913 Available at: http://www.ncbi.nlm.nih.gov/pubmed/8441015.

Schultz W, Dayan P, Montague PR (1997) A Neural Substrate of Prediction and Reward. 275:1593–1599.

Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. J Neurosci 27:4826–4831 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2630024&tool=pmcentrez&rendertype=abstract.

Silvetti M, Nuñez Castellar E, Roger C, Verguts T (2014) Reward expectation and prediction error in human medial frontal cortex: An EEG study. Neuroimage 84:376–382 Available at: http://dx.doi.org/10.1016/j.neuroimage.2013.08.058.

Stelly CE, Haug GC, Fonzi KM, Garcia MA, Tritley SC, Magnon AP, Alicia M, Ramos P, Wanat MJ (2019) Pattern of dopamine signaling during aversive events predicts active avoidance learning. 116:13641–13650.

Sutton RS, Barto AG (2015) Reinforcement Learning: An Introduction.

Valentin V V, Doherty JPO (2020) Overlapping Prediction Errors in Dorsal Striatum During Instrumental Learning with Juice and Money Reward in the Human Brain. :3384–3391.

Voorn P, Vanderschuren LJMJ, Groenewegen HJ, Robbins TW, Pennartz CMA (2004) Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 27:468–474.

Watkins CJCH (1995) Learning from delayed rewards. Rob Auton Syst 15:233–235.

Zhang S, Mano H, Ganesh G, Robbins T, Seymour B (2016) Dissociable Learning Processes Underlie Human Pain Conditioning. Curr Biol 26:52–58 Available at: http://dx.doi.org/10.1016/j.cub.2015.10.066.

This chapter has been prepared as a manuscript for submission.

5 Neural mechanisms of stages of reward/avoidance decision processes in obsessive compulsive disorder and gambling disorder

Connecting: Through the statistical analysis and modelling fitting, the previous Chapter 3 has provided evidences of the healthy participants' behavioural performance of reward and avoidance learning. And through the combination of model fitting and neuroimaging analysis, the Chapter 4 has provided explainable evidences of underlying neural mechanisms associated with keys signals of reward and avoidance learning. These solid findings and well-packed analysis paradigm motivated to explore the potential maladaptive reward and avoidance learning in OCD and GD groups. In this chapter, we will explore the behavioural differences of OCD and GD groups under reward and avoidance learning. Also, how's the neural activation differences underlying the computational processes of both clinical groups compared to healthy participants. Further, we will examine whether the altered neural activations associated with clinical symptoms and behavioural traits.

Aim: In Chapter ***4***, we investigated the brain mechanism of three distinct stages (i.e., outcome, expected value and error processing) in reward and avoidance learning in healthy controls. In this chapter, we extended the investigation to the proposed maladaptive and aberrant decision process in participants with OCD (i.e. high compulsivity) and GD (i.e. high impulsivity). At first stage, the statistical analysis of the participants' behavioural performance was carried out under the reward/avoidance condition with comparison to healthy controls, and four measurements were included: 1) the response time; 2) the number of Correct and Incorrect fractal choice; and 3) the learning curve. Next, the RL algorithms were implemented to model the participants' behavioural process in the learning task and extract the learning traits including: 1) learning rate; 2) inverse temperature parameter. Further, to identify the neural substrates supporting aberrant differences compared to healthy

controls, imaging regression analysis was carried out to examine the brain activation in OCD and GD group compared to healthy controls at the three phases of decision process: 1) outcome processing; 2) expectation value; and 3) error processing.

Questions: Through the application of the reversal learning task, we were interested in the performance pattern of OCD and GD clinical population compared to the healthy controls group at three levels including behavioural, modelling and imaging. At the level of behaviour, how similar or different is performance amongst the compulsive (i.e. OCD), impulsive (i.e. PG) clinical groups and healthy control groups. At the level of modelling, how were the learning merits including learning rate and inverse temperature parameter in the OCD and GD compared to healthy controls? At the level of imaging, how were the brain activation related to outcome processing, expectation value and error processing in participants with OCD and GD?

Hypothesis: According to the previous studies, OCD is reported to be correlated with harm avoidance and exaggeration of aversive events (Choi et al., 2012; Endrass et al., 2011). We hypothesized that: 1) participants with OCD had a significant preference to the Correct choice under avoidance condition. 2) participants with OCD had an aberrant activation in the aversive error signal circuit. Whereas GD is associated with impulsiveness and reward seeking (Won Kim & Grant, 2001). We hypothesized that: 1) participants with GD had a significant preference to the Correct choice under reward condition. 2) participants with GD had an aberrant activation in the reward error signal circuit.

5.1 Introduction

Decision making is the essential skill to live and manage our life, while many psychiatric conditions are associated with participants' aberrant decision making patterns. For example, participants in obsessive compulsive disorder (OCD) commonly repeat a behaviour such as handwashing; And participants in gambling disorder (GD), commonly seek and partake in risky gambling, despite explicitly acknowledging the following harms. OCD is a relatively chronic and disabling neuropsychiatric disorder with an estimated prevalence between 1 and 3% of the world population (Figee et al., 2011), and GD is classified as a behavioural addiction with a lifetime prevalence of 0.5 -1% (Miedl et al., 2012, 2014; Petry et al., 2005; Potenza, 2008). A high burden of individual and socioeconomic cost was caused by the maladaptive DM pattern under both clinical disease (Fujino et al., 2018; Nestadt et al., 2018).

A dimensional model of ***impulsive-compulsive spectrum disorder*** has been previously proposed in which impulsivity and compulsivity represents polar opposite psychiatric spectrum constructs that can be viewed along a continuum of compulsive and impulsive disorders. OCD is recognized as a typical compulsive disorder, characterized by the experience of unwanted repetitive thoughts (obsessions) and repetitive behaviours (compulsions) with **overestimation** of the probability of future harm to carry on the risk avoidance (Pauls et al., 2014). Meanwhile the GD is regarded as an impulsive disorder, characterized by the impulsive choice of persistent and recurrent maladaptive patterns of gambling behaviour with ***underestimation*** of the likelihood or severity of possible harm (Lai & Ip, 2011). As a typical conception of compulsive disorder, recent studies also suggested that OCD shares behavioural components of impulsivity (Abramovitch & Mckay, 2016; Fontenelle et al., 2011; Grassi et al., 2015), and also based on the existed portrait of impulsive disorder, a compulsivity feature was suggested to be acquired in participants with GD with the increase of the impulsive behaviour (Fontenelle et al., 2011).

Studies have been carried out to investigate the impairment in appropriate decision making performance including reward/avoidance learning and its related neural mechanism under those two clinical conditions. The enhancements of harm-avoidance or avoidance habit in OCD with exaggerated anticipation and avoidance of aversive outcomes has been found in previous studies (Gillan et al., 2016; Learning, 2011; Starcevic et al., 2011), and the excessive avoidance behaviour was correlated with the hyper-activation in the orbitofrontal-striatal circuit (Gillan et al., 2016; Remijnse et al., 2006). Based on the latest conception of compulsive disorder mentioned above, it is also suggested that OCD shares behavioural components of impulsivity (Abramovitch & Mckay, 2016; Fontenelle et al., 2011; Grassi et al., 2015). This is supported by recent studies that reported the dysfunctional reward processing with altered neural activity in the brain reward circuit and risk aversion in OCD under the effect of impulsivity trait (Admon et al., 2012; Figee et al., 2010).

The "reward deficiency" in GD has been demonstrated with the hyperactivity in the reward circuitry including striatum and prefrontal brain regions (Brevers et al., 2015; Oberg et al., 2011; Pro et al., 2010). Not only the increased activity found for reward processing, the deactivation to loss aversion in the cortico-striatal circuit has also been reported in GD (Gelskov et al., 2016; Genauck et al., 2017). As the core feature of GD, the levels of impulsivity were found inversely correlated with activity of reward and avoidance processing (Pearlson & Potenza, 2013), and the ability to alter choice behaviour in response to stimulus-reward contingencies (Franken et al., 2008). It was pointed out that compulsivity should be also considered to investigate the deficits in decision making of GD (Ioannidis et al., 2019), and with the increases of the impulsive behaviour, the compulsivity feature would be acquired (Fontenelle et al., 2011).

According to these studies, OCD is usually conceptualized as a compulsive disorder with harm avoidance, but the possible reward processing pattern under the effects of

impulsivity trait remains unclear. While GD is a portrait of impulsive disorder with risky decision making and exaggerated reward processing, the avoidance processing pattern under the compulsivity feature is under investigation. Using computational modelling and neuroimaging techniques, our aim was to investigate ***I)*** the behavioural performance including number of Correct vs Incorrect choices, learning curve and response time of OCD and GD compared with healthy controls. Then, a model was fitted to the behavioural data of both clinical groups to investigate ***II)*** the parameters including the learning rate and inverse temperature parameter under reward and avoidance decision processes. Combining with neuroimaging, we further investigated if there were altered brain activations related to these cognitive processes. Post-hoc analysis was carried out in order to examine ***III)*** how these constructs of the impulsivity and compulsivity affect the reward/avoidance learning performance in OCD and GD. Through the investigation of the neural computational mechanism and brain correlates, the study could help provide a better understanding of the aberrant decision making process in participants with OCD and GD, and also a potential brain area target for treatment intervention. We hypothesized that participants with OCD would be associated with an aberrant activation in the aversive error signal circuit, whereas participants with GD have aberrant activation in the reward error signal circuit. We also predicted that GD would be associated with increased impulsiveness and reward seeking (Won Kim & Grant, 2001).

5.2 Materials & Methods

PARTICIPANTS

Forty-two healthy controls (HC), 40 OCD and 23 GD participants were recruited in our study to complete a probabilistic reward and avoidance learning task while conducting the functional magnetic resonance imaging (fMRI) scanning. Study sample has been reported in previous studies (Maleki et al., 2020; Parkes et al., 2018). Inclusion criteria for all participants involved the following: age between 18-55 years, having normal to corrected vision, and being fluent in English.

Confirmation of OCD diagnosis using the Mini-international Neuropsychiatric Interview (MINI) scale was an inclusion criterion for the OCD group (Lobbestael et al., 2011). Confirmation of GD diagnosis using the Structured Clinical Interview for Axis I DSM-IV Disorders (SCID) scale was an additional inclusion criterion for the GD. Also, the primarily engaged in electronic gaming machine (EGM) gambling determined by clinical services was also an inclusion criteria for GD group.

Exclusion criteria for all participants included a history of neurological diseases or seizures, lifetime history of psychiatric illnesses (apart from participants with OCD and GD), significant head injury or concussion, standard MRI contraindications, significant or sustained steroid use, history of alcohol abuse or dependence, and use of cannabis or other illicit drug use > 50 times. The additional exclusion criteria for clinical groups included primary diagnosis of psychiatric disorders other than OCD and GD (secondary diagnosis of anxiety and depression are not excluded). The diagnosis was corroborated by treatment services and confirmed by the MINI.

All participants were assessed for the severity of obsessive-compulsive symptoms using the OCI-R, an 18-item self-report measure that assess the distress associated with the obsessions and compulsions and six separate dimensions including *obsessing*, *checking*, *neutralizing*, *washing*, *ordering, and hoarding* (Foa et al., 2002). Also, the participants were assessed for the severity of gambling severity by the Problem Gambling Severity Index (PGSI) (Holtgraves, 2009), and the behavioural construct of impulsiveness by the Barratt Impulsiveness Scale (BIS) (Patton et al., 1995). Also, depression and anxiety symptoms were assessed using the Beck Depression Inventory (BID) (Wang & Gorenstein, 2013), and the State and Trait Anxiety Inventory (STAI), respectively (Marteau & Bekker, 1992).

All participants gave informed consent and the study was approved by the Human Research Ethics Committee of Monash University.

PROBABILISTIC REWARD AND AVOIDANCE LEARNING TASK

The probabilistic reward and avoidance learning task paradigm have been referred to in the previous two chapters. Briefly, on each trial of the Probabilistic reward/avoidance learning task (**Fig 1 (a)**), one of three pairs of fractal stimuli were simultaneously presented. Each pair of fractals signified the onset of one of three trial conditions: Reward, Avoidance and Neutral, whose occurrence was semi-random such that each three-trial block contains one of each type, and the order of these three trials were randomized. The specific association of fractal pairs to a condition was fully randomized but counterbalanced among participants. Participants' task on each trial was to choose one of the two stimuli by selecting the fractal to the left or right of the fixation cross via a button box (using the right hand). Once a fractal has been selected, depending on the condition, it increased in brightness and was followed by the visual feedback indicating either a reward (a picture of a Myer card with text above saying

"you win 1 point!"), an aversive outcome (a red cross overlying a picture of a Myer card with text above saying "You lose 1 point!"), neutral feedback (a scrambled picture of a Myer card with text above saying "No change!"), or nothing (a blank screen with a cross hair in the centre). Participants had 2000 milliseconds to select a fractal. If not selected in time, a screen would be displayed with the text "response omitted", and the trial would be repeated until a response registered for that fractal pair.

Participants underwent two ~16 min scanning sessions, each consisting of 90 trials (30 trials per condition). In the reward trials, if participants chose the high probability action (also referred to here as the Correct action), they received monetary reward with a 70% probability; on the other 30% of trials they received nothing. Following choice of the low probability action (also referred to here as the Incorrect action), they received monetary reward on only 30% of trials; otherwise, they obtained nothing on the remaining 70% of trials. Similarly, on the avoidance trials, if participants chose the high probability action they received nothing on 70% of trials, on the other 30% they received a monetary loss, whereas choice of the low probability action led to no outcome on only 30% of trials, while the other 70% were associated with receipt of the aversive outcome. A probability switch was introduced at a time-point between the 11^{th} to 20^{th} trial in the reward/avoidance trials, where the fractal associated with high probability was changed to the low probability and where the fractal associated with low probability changed to the high probability. For the neutral trials, participants had a 70% or 30% probability of obtaining neutral feedback; otherwise, they received nothing.

Prior to the experiment, participants were given instructions that they would be presented with three pairs of fractals and on each trial, they had to select one of these fractals. Participants also had a practice session of the task before going into the MRI. During the task, depending on their choices they would win a point, lose a point, obtain a neutral

outcome with no change, or receive nothing. They were not told which fractal pair was associated with a particular outcome neither when the probability switch was to occur. Participants were instructed to try to win as many points as possible and that they would receive a Coles/Myer voucher at the end corresponding to the amount of points they had accumulated.

IMAGING PROCEDURE

All images were acquired with 3.0-T SIEMENS MAGNETOM Skyra syngo MR D13C at Monash Biomedical Imaging. The functional images (fMRI) were acquired through gradient echo T2* weighted echo-planar images (EPI) with BOLD (blood oxygenation level dependent) contrast. The scanning parameters: field of view = 230 mm, 3mm by 3mm in plane resolution, time of repetition = 2000 ms, and time of echo = 30.0 ms. Each volume of fMRI images contains 34 slices with a thickness of 3.0 mm (no gap) in an ascending interleaved way. High resolution T1-weighted (1x1x1 mm resolution) were acquired with a standard MPRAGE sequence (time of echo = 2.07 ms, time of repetition = 2300 ms, flip angle = 9 degree, field of view = 256 mm).

BEHAVIOURAL DATA ANALYSIS

Basic behavioural analysis

Basic statistical analysis including two-sample t-test was carried out to compare the behaviour outcomes, such as number of Correct and Incorrect choices as well as response time under each condition. By realigning the probability switches of each run to a same point and separation of all trials into eight blocks, a block-based learning curve was drawn based on the number of Correct and Incorrect choices under reward/avoidance condition. The one trial back measurement of the stay ratio on the reward/non-rewarded or punished/non-

punished trials under the reward & avoidance condition respectively to see the participants' choice tendency under these conditions.

Q-learning model

A basic Q-learning model (Watkins, 1995), was used to characterise participants' behaviour in task. This model estimates the expected value of choosing each stimulus based on the previous history of choices and outcomes. The expected value of each stimulus was initially set to zero, and after each trial t > 0, was updated according to the chosen stimulus and reward feedback. The expected value of choosing stimulus a was updated as follows,

$$Q_a(t+1) = Q_a(t) + \alpha * \delta(t);$$

while the value for non chosen stimulus stayed unchanged. α is the learning rate and $\delta(t)$ is the prediction error which is the difference between the actual and expected outcome,

$$\delta(t) = R(t) - Q_a(t);$$

$R(t)$= \$-1, \$0, \$1 is the reward received after choosing the stimulus. The probability of taking each action is based on their values, and according to the softmax rule,

$$P_a(t) = exp\left(\beta Q_a(t)\right) / \{exp\left(\beta Q_a(t)\right) + exp\left(\beta Q_b(t)\right)\};$$

The β is the inverse temperature parameter with a scale from 0 to 20, which indicates how stochastic or exploratory the individual choices are. Lower values of β parameter indicate random action selection, which corresponds to low sensitivity to stimulus values; while a high β value indicates that choices are strongly driven by their expected values.

The hierarchical bayesian method (HBM) was used for the model and parameter estimation. HBM, exploits group-level parameter distributions to inform individual-level estimations, and compared to the individual parameter estimation methods, HBM provides better parameter stability and predictive accuracy [25]. The learning rate α and inverse temperature parameter β had a normal prior distribution Norm (0,1), and at the same time

with group mean value as well as deviation value shrinkage to a normal distribution Norm (0,1) and Norm (0, 0.5), respectively (see supplementary file for details).

IMAGING DATA ANALYSIS

The imaging processing and processing pipeline for OCD and GD followed the mentioned paradigm applied to healthy control groups in Chapter 4. SPM12 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, United Kingdom) was used to perform the fMRI image analysis. The pre-processing of EPI images commenced with the slice timing selecting the middle slice of each volume as reference. Then, the spatial realignment was applied to remove the motion artefacts. The individual T1-weighted image was co-registered to the mean EPI generated during realignment and then were normalized to the MNI space through the 6-tissue probability map (TPM) provided by SPM. The motion-corrected and co-registered EPI images were normalized to MNI space using the previously calculated deformation fields and then spatially smoothed with an 8-mm FWHM (full width half maximum) Gaussian kernel. Time series describing expected values and PEs were generated for each participant for each trial in the experiment by entering the participants' trial history into the learning model. These sequences were convolved with a hemodynamic response function and entered into a regression analysis against the fMRI data. The expected value was modelled as a boxcar function beginning at the time of selection response till the outcome delivered; while the PEs modelled as a delta function at the time of outcome delivered. To map the activation map, a new design was created with other regressors indicating different outcomes to model activity at the time of the outcome: rewarded reward trial (R_+), unrewarded reward trial (R_-), punished avoidance trial (P_+), non-punished avoidance trial (P_-), neutral feedback trial (N_+) and neutral trial without feedback (N_-). In

addition, the six scan-to-scan motion parameters produced during realignment were included to account for residual effects of movement.

Linear contrasts of regressors coefficients were computed at the individual participant level to enable comparison among the Reward, Avoidance and Neutral trials. The results from each individual were taken to a random effects level by including the contrast images from each single participant into a one-way analysis. The simple contrast [R_+ - N_+] was to test the brain response to rewarded outcome and the contrast [P_+ - N_+] was to examine the brain activation related to aversive outcome. Further, the specific contrast [R_+ + P_-] - [R_- + P_+] was to test those of brain areas showing greater response to obtaining reward and avoidance aversive outcome compared to obtaining aversive outcome and missing reward.

The PE and expected value were separately parametric modulation orthogonal to the outcome regressor. Then, the contrasts were created to examine the brain areas associated with expected value and PE under reward and avoidance condition.

STATISTICAL ANALYSIS

The independent two sample t-tests were carried out to compare the number of Correct vs Incorrect choices under each condition for three groups participants. As well, the two-sample t-tests were used to test the Correct vs Incorrect choices of each block under reward and avoidance condition. Also, the two-sample t-test was used to compare the response time between different conditions. Further, the two-sample t-tests were used to compare the modelled parameters including learning rate and inverse temperature parameter under different conditions. The statistical analysis was using GraphPad Prism (version 8).

After the first-level regression analysis through the imaging data, voxel-wise two group t-tests at the second level were conducted to investigate the brain activation differences between groups, with age and gender as covariates using SPM12. For multiple comparison

correction, an initial voxel-wise threshold was set at $p < 0.05$ with subsequent family-wise error corrected p-value < 0.05 at cluster level. The post-hoc analysis was applied to investigate the relationship between the aberrant brain activity and the disease severity, and scales of impulsivity and compulsivity, the mean activation value at the maladaptive brain regions were extracted. Then the Pearson correlation analysis was done to investigate the relationship, and the significant level was set at $p < 0.05$. Also, the correlation analysis was corrected for the multiple comparison correction error.

5.3 Results

Demographics and behavioural statistical analysis

Several participants were excluded due to incomplete or invalid imaging data, leaving 39 healthy participants (20F/19M, 34 yrs ± 9.47)), 28 OCD (14F/14M, 32.11 yrs ± 9.53) and 16 GD (4F/12M, 35.53yrs ± 12.20) with complete behavioural and imaging data. The demographics and characteristics of the participants with OCD or GD is shown as in ***Table 5-1***.

Table 5-1 *Characteristics of participants.*

	GD (n = 16)	**OCD (n = 28)**	**HC (n = 39)**	**Statistics**
Gender, n				
Male	12	14	19	
Female	4	14	20	
Mean, years (S.D.)	35.06(3.11)	32.11(1.80)	34(1.52)	F (2,80) = 0.50, p >0.05
Education (years)				
Pre-Assess				
Barratt Impulsiveness Scale (BIS) (S.D.)	20.06(4.06)	24.44(3.15)	20.54(3.01)	F (2, 79) = 13.90, p < 0.0001
Day-Assess				
OCI-R-Total (S.D.)	12.63(12.45)	32.5(11.25)	5.13(5.54)	F (2, 80) = 71.90, p < 0.0001
PGSI (S.D.)	15.63(7.65)	0.21(0.69)	0(0)	F (2, 80) = 140.1, p < 0.0001

The statistical analysis of behavioural data showed that OCD group significantly preferred the Correct choice both under the reward ($t = 3.23$; $p = 0.0019$) and avoidance ($t = 7.34$; $p < 0.0001$) condition, compared with the Incorrect choice; No significant difference was found in the neutral condition ($t = 0.60$; $p = 0.55$); GD group only showed significant preference to the Correct choice under the avoidance condition ($t = 4.04$, $p = 0.0003$) (***Figure 5-1***).

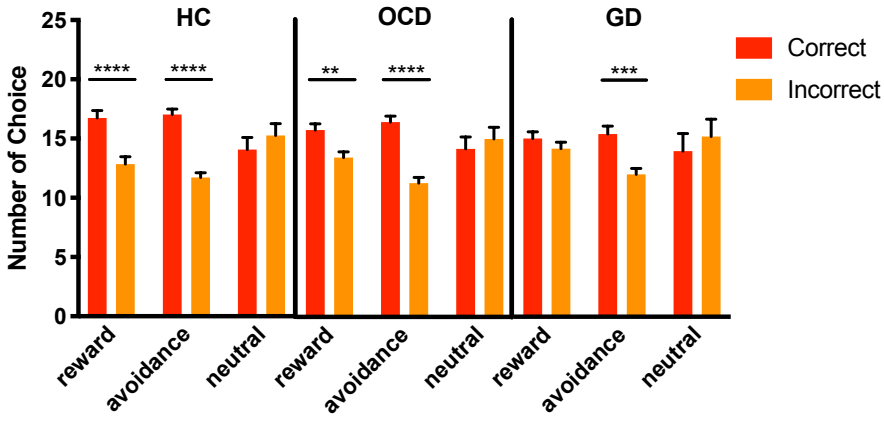

Figure 5-1 *Number of choice under reward/avoidance/neutral condition for controls, OCD and GD groups. OCD group showed significant preference for the Correct choice both in the reward (**p = 0.002) and avoidance condition (****p < 0.0001); while GD* ***group*** *only showed significant difference (***p = 0.0003) in the avoidance condition.*

The OCD group made significantly quicker response to the reward condition (1035 ms ± 29.66; $t = 2.254$, $p = 0.03$) and significantly slower to the avoidance condition (1198 ms ± 21.22; $t = 2.16$, $p = 0.035$). The response time to the neutral condition (1125 ms ± 26.53) is intermediate between the reward and avoidance condition; The GD group made significantly slower response to the avoidance condition (1158 ms ± 30.05; $t = 3.157$, $p = 0.003$) to the avoidance condition compared with the neutral condition (1038 ms ± 23.4). While no significant differences in response time was found in the reward condition (976.4 ms ± 24.65; $t = 1.816$, $p = 0.07$) relative to the neutral condition (***Figure 5-2***).

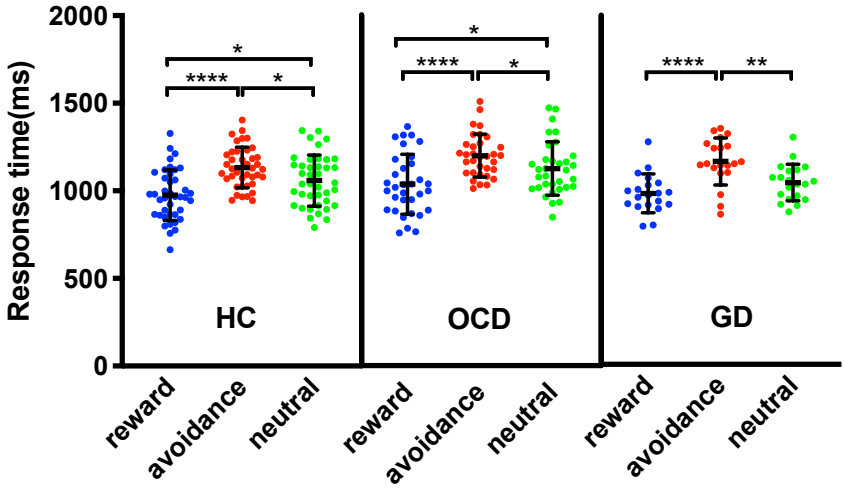

Figure 5-2 *Response time under reward/loss/neutral condition. OCD group made significantly quicker response to the reward condition (*p = 0.03) and slower response (*p = 0.03) to the avoidance condition compared to the neutral condition. GD group showed significantly slower response to the avoidance condition (** p =0.003) compared to the neutral condition.*

The Learning curve showed that the healthy controls did learn the task and made the Correct choice before and after the probability switch, whereas the OCD and GD made the Correct choice only before the probability switch under reward condition (see ***Figure 5-3***). Under the avoidance condition, all three groups of participants preferred the Correct choice before and after the probability switch (see ***Figure 5-4***).

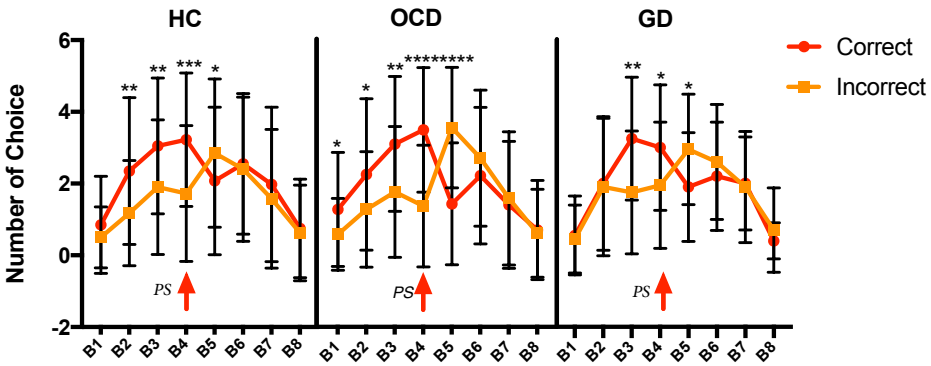

Figure 5-3 *Learning curve under the reward condition.*

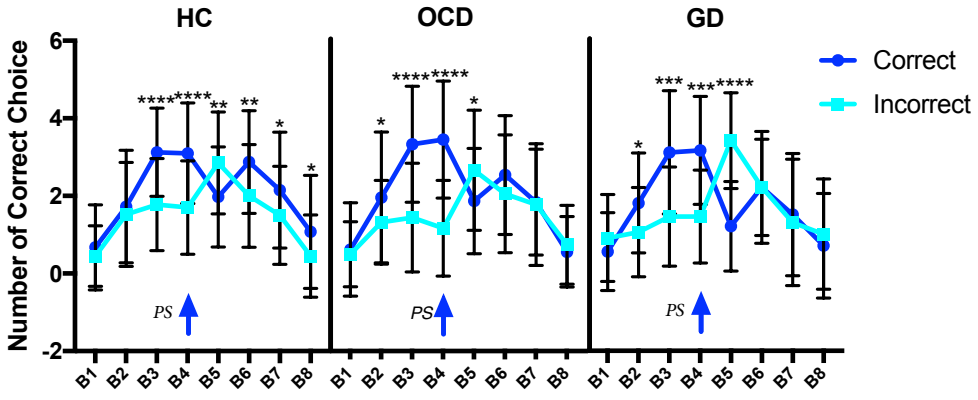

Figure 5-4 Learning curve under avoidance condition.

The Q-learning estimation showed that healthy controls had a significantly higher learning rate compared to GD group ($p = 0.038$, $t = 2.13$) in session 1 **(*Figure 5-5*)**. The healthy control group also had a significantly higher learning rate under avoidance condition in session 1 compared to OCD group ($p < 0.0001$, $t = 6.34$) as well as GD group ($p < 0.0001$, $t = 12.41$) **(*Figure 5-6*)**. For the inverse temperature parameter, the OCD group had a significantly higher inverse temperature parameter compared to healthy controls in session 1 under reward condition ($p < 0.01$, $t = 1.02$) **(*Figure 5-7*)**, and also OCD had a significantly higher inverse temperature parameter compared to healthy controls in session 2 ($p < 0.01$, $t = 1.29$) **(*Figure 5-8*)**.

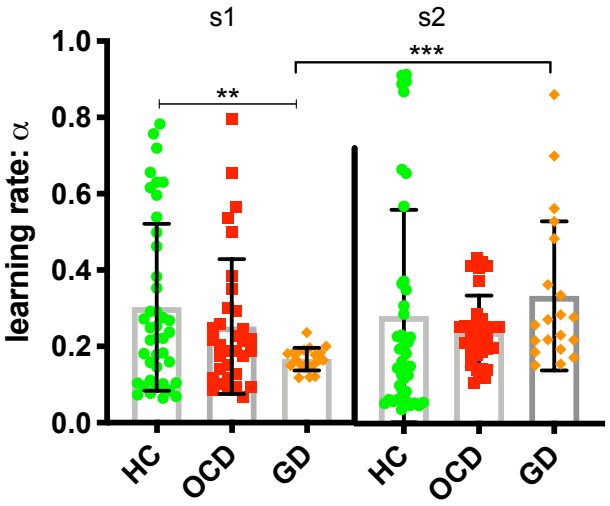

Figure 5-5 *The learning rate of healthy controls, OCD and GD under reward condition.*

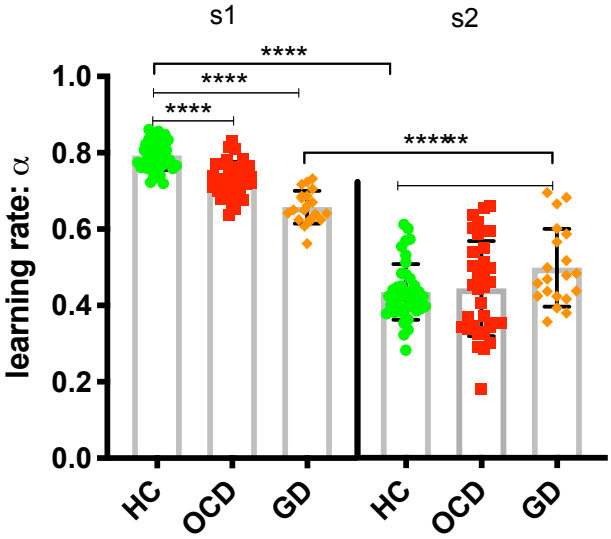

Figure 5-6 *The learning rate of healthy controls, OCD and GD under avoidance condition.*

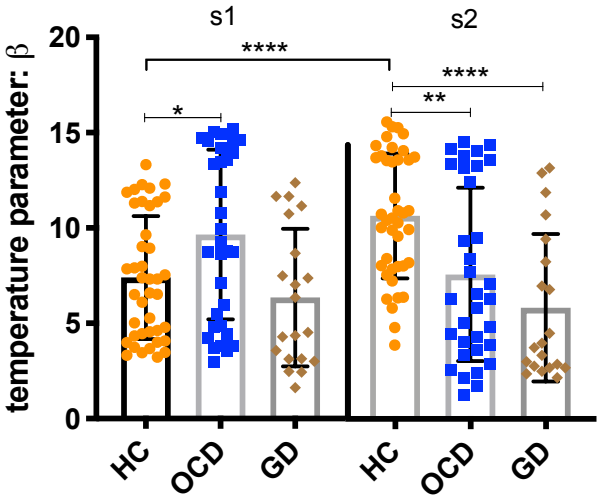

Figure 5-7 *The inverse temperature parameter of healthy controls, OCD and GD under reward condition.*

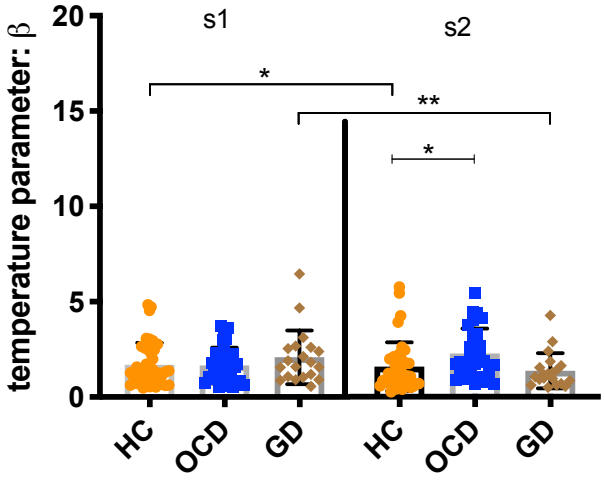

Figure 5-8 The inverse temperature parameter of healthy controls, OCD and GD under avoidance condition.

Relationship between the learning parameters and the clinical measurement of disease severity, and the scales of impulsivity and compulsivity

We have conducted the correlation analysis between the learning parameters and clinical measurement of disease severity, and the scales of impulsivity and compulsivity, no significant correlations were found.

Clinical groups\Learning performance	***Learning rate α***	***temperature parameter β***
OCD		
Reward condition		
OCI-R-Total	r = -0.087; p = 0.66; N = 28	r = 0.19; p = 0.33; N = 28
PGSI	r = 0.08; p = 0.68; N = 28	r = -0.32; p = 0.10; N = 28
BIS	r = -0.25; p = 0.21; N = 28	r = -0.21; p = 0.30; N = 28
Avoid condition		
OCI-R-Total	r = 0.28; p = 0.15; N = 28	r = 0.096; p = 0.63; N =28
PGSI	r = -0.11; p = 0.58; N =28	r = -0.06; p = 0.77; N = 28
BIS	r = -0.20; p = 0.32; N = 28	r = -0.20; p = 0.34; N = 28
GD		
Reward condition		
OCI-R-Total	r = -0.33; p = 0.21; N = 16	r = -0.15; p = 0.58; N =16

PGSI	r = -0.27; p = 0.32; N = 16	r = 0.0007; p = 1.00; N = 16
BIS	r = -0.58; p = 0.018; N = 16	r = 0.31; p = 0.24; N = 16
Avoid condition		
OCI-R-Total	r = -0.075; p = 0.78; N =16	r = -0.26; p = 0.34; N = 16
PGSI	r = -0.37; p = 0.16; N = 16	r = -0.46; p = 0.075; N = 16
BIS	r = -0.42, p = 0.11; N = 16	r = -0.37, p = 0.16; N = 16

Imaging results

Neural response to reward receipt and punishment avoidance

Compared to healthy controls, participants with OCD showed the decreased activity at left Opercula part of the inferior frontal ([-46, 10, 28]; t = 4.80, k = 7385), right Opercula part of the inferior frontal ([48, 14, 30]; t = 4.39, k = 5613) and right Thalamus ([6, -22, 0]; t = 3.80, k = 1482) at the outcome of getting reward after family wise error correction (FWE) (shown in ***Table 5-2 & Figure 5-9***). And participants with OCD also showed the decreased activity at the brain region peaked at ([-40, -16, -18]; t = 4.12, k = 18492) at the outcome of missing reward after correction (shown in ***Figure 5-10***).

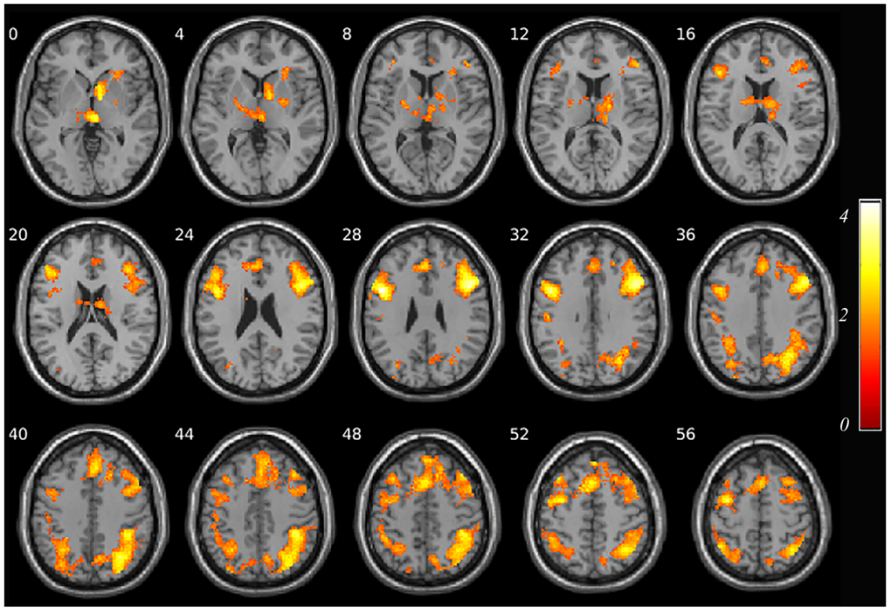

Figure 5-9 *Brain activation for receiving reward relative to neutral condition in participants with OCD compared to controls. The participants with OCD showed decreased activations at bilateral inferior frontal and right thalamus extending to right caudate (see* ***Table 5-2 for details****).*

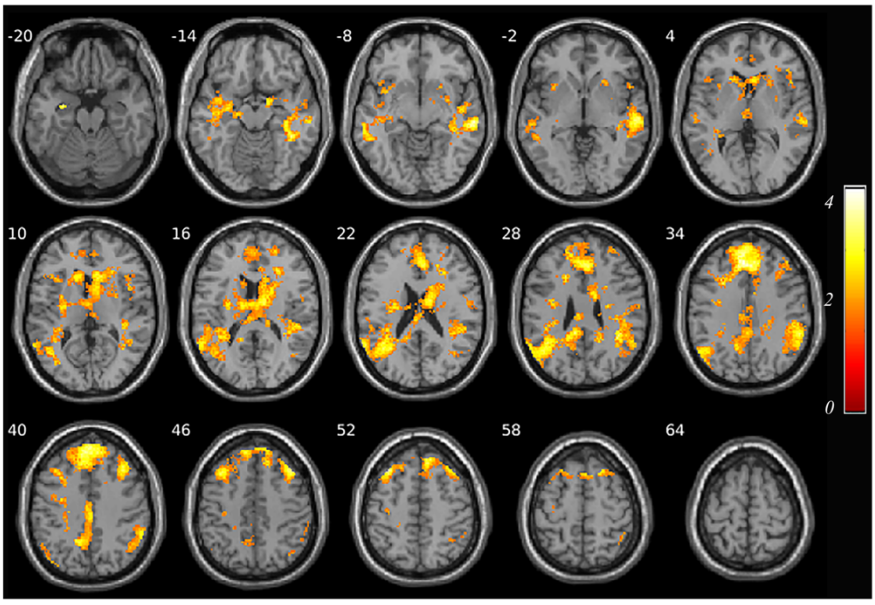

Figure 5-10 *Brain activation when missing reward relative to neutral condition in participants with OCD compared to controls. The participants with OCD showed decreased activations at left medial superior frontal and right middle temporal (see* ***Table 5-2 for details****).*

Table 5-2 *Brain regional response difference to reward/aversive outcome among participants with OCD compared to healthy controls.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward success – neutral success					
OCD < Control					
L. Opercula part of the Inferior Frontal	-46	10	28	4.80	7385
R. Inferior Parietal	40	-40	46	4.05	
L. Precentral	-34	-2	56	3.94	
R. Opercula part of the Inferior Frontal	48	14	30	4.39	5613
R. Opercula part of the Inferior Frontal	52	18	36	3.77	
R. Opercula part of the Inferior Frontal	58	20	28	3.71	
R. Thalamus	6	-22	0	3.80	1482
R. Caudate	10	10	-2	3.39	
Reward unsuccess – neutral unsuccess					
OCD < ***Control***					
L. Superior Temporal	-40	-16	-18	4.12	18492
L. Medial Superior Frontal	-4	36	34	4.00	1115
R. Middle Temporal	58	-20	-10	3.96	490

Compared to healthy controls, participants with GD showed the decreased activity at right Cuneus ([14, -86, 30]; t = 3.81, k = 1650), right Rolandic operculum ([48, -2 ,20]; t = 3.73, k = 7509) and left Postcentral ([-34, -16 ,38]; t = 3.16, k = 1650) at the outcome of getting reward after correction. The decreased activity at left superior frontal ([-20, 26, 46]; t = 3.87, k = 4137), right middle cingulum ([2, -12, 40]; t = 3.59, k = 2636) for missing reward were found in GD compared to healthy controls (shown in ***Figure 5-11***). Successful avoidance aversive outcome was found in higher activation in participants with GD compared to healthy controls at the left caudate ([-22, -2, 24]; t = 4.50, k = 11203). When receiving punishment, the decreased activity was found in participants with GD compared to healthy controls at right middle frontal ([24, -22, 46]; t = 3.84, k = 2359) (see ***Table 5-3*** for details).

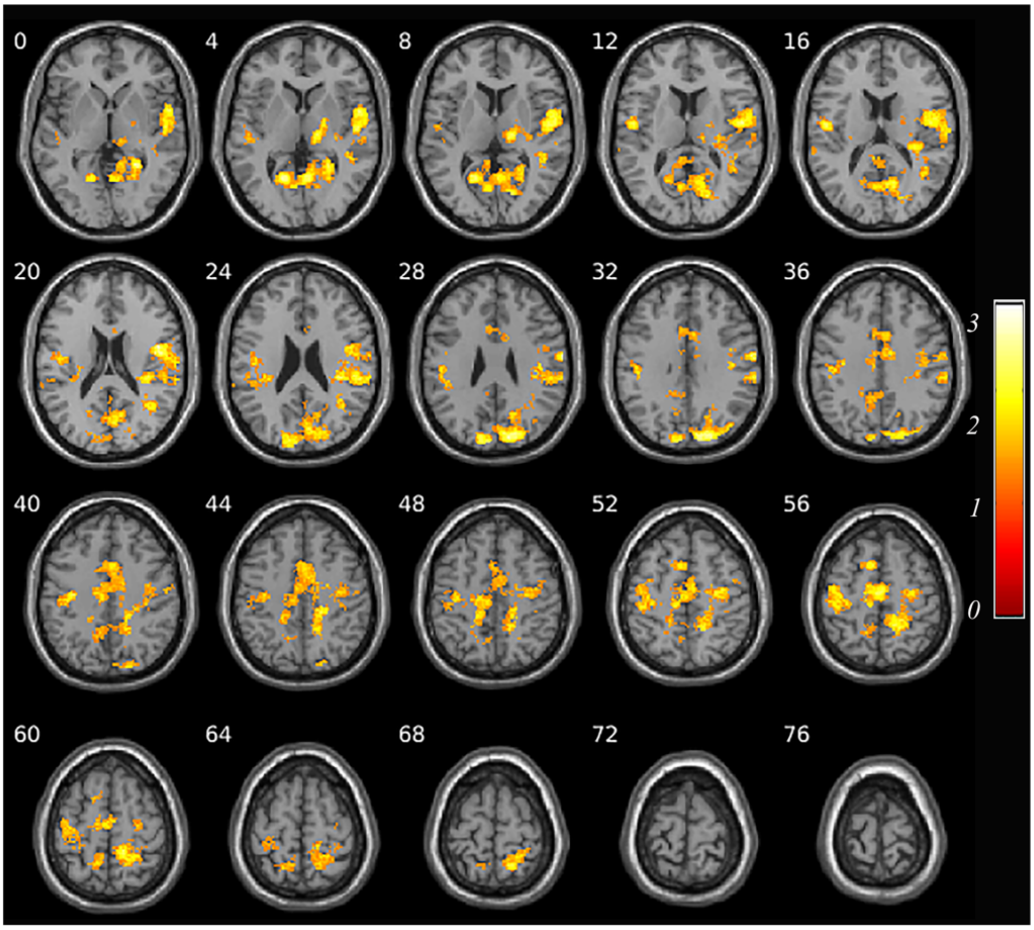

Figure 5-11 *Brain activation of receiving reward relative to neutral condition in participants with GD compared to controls. The participants with GD showed decreased activations at right cuneus, left precentral and left postcentral (see* ***Table 5-3 for details****).*

Table 5-3 *Brain regional response difference to reward/aversive outcome among participants with GD compared to healthy controls.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward success – neutral success					
GD < Control					
R. Cuneus	14	-86	30	3.81	1650
R. Rolandic Operculum	48	-2	20	3.73	7509
R. Postcentral	64	-10	28	3.44	
R. Rolandic Operculum	52	-10	10	3.33	
L. Postcentral	-34	-16	38	3.16	1650
L. Postcentral	-52	-8	14	3.10	
L. Precentral	-42	-10	54	2.71	
Reward unsuccess – neutral unsuccess					
GD < Control					
L. superior frontal	-20	26	46	3.87	4137
L. Superior Frontal	-14	42	38	3.77	
L Superior Medial Frontal	-6	52	18	3.65	
R. Middle cingulum	2	-12	40	3.59	2636
L. Middle Cingulum	-10	-44	34	3.44	
L. Posterior Cingulum	-4	-40	30	3.43	
Avoidance success – neutral unsuccess					
GD > Control					
L. Caudate	-22	-2	24	4.50	11203
Avoidance unsuccess – neutral success					
GD < Control					
R. Middle Frontal	24	-22	46	3.84	2359

When comparing OCD and GD at the phase of outcome processing, participants with OCD showed the increased activity at left Postcentral ([48, -4, 18]; t = 4.28, k = 12690) for receiving reward (***Figure 5-12***), and decreased activations at left Middle occipital ([-34, -72, 10]; t = 4.14, k = 4988) for missing reward. Participants with GD showed the increased activity at left Middle temporal ([-34, -66, 4]; t = 3.99, k = 3276) and left Thalamus ([14, -26, 28]; t = 3.90, k = 2871) for avoiding loss. While participants with OCD showed the increased activity in right Hippocampus ([24, 6, 32]; t = 3.49, k = 5227) for receiving loss (***Figure 5-13***).

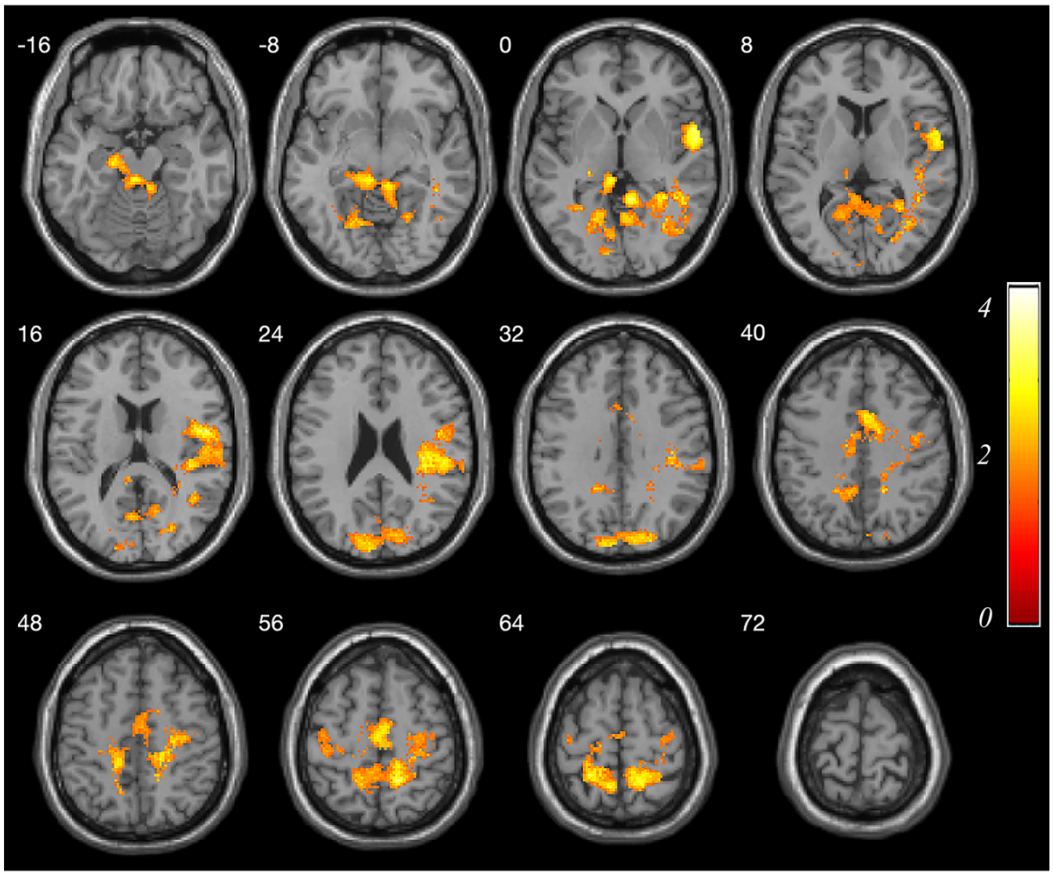

Figure 5-12 *Brain activation for receiving reward in participants with OCD compared to participants with GD. The participants with OCD showed increased activations at the left postcentral (see* ***Table 5-4 for details****).*

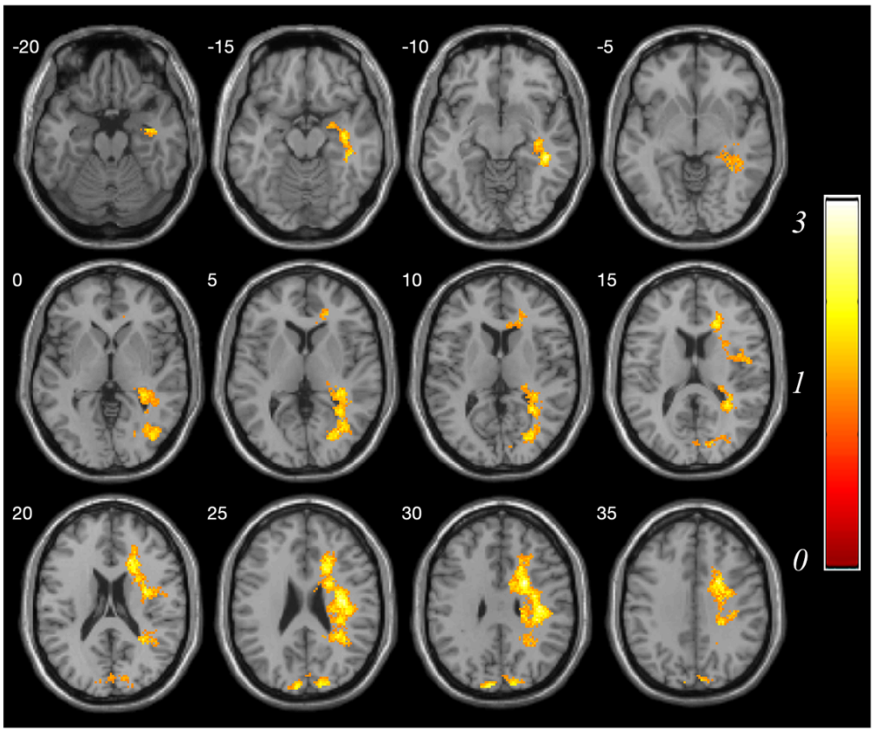

Figure 5-13 *Brain activation of receiving loss in participants with OCD compared to participants with GD. The participants with OCD showed increased activations at the brain region including right pallidum and left superior temporal (see* ***Table 5-4 for details****).*

Table 5-4 *Brain regional response difference to reward/aversive outcome between participants with OCD and GD.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward success – neutral success					
OCD > GD					
L. Postcentral	48	-4	18	4.28	12690
Reward unsuccess – neutral unsuccess					
GD > OCD					
L. Middle Occipital	-34	-72	10	4.14	4988
Avoidance success – neutral unsuccess					
GD > OCD					
L. Middle Temporal	-34	-66	4	3.99	3276
L. Thalamus	14	-26	28	3.90	2871
Avoidance unsuccess – neutral success					
OCD > GD					
R. Hippocampus	24	6	32	3.49	5227

Neural response to reward/aversive expected value

Compared to healthy controls, OCD showed significantly lower activation at the brain region ([-18, 32]; t = 3.94, k = 3416), and the other region ([-8, 6, 26]; t = 3.90, k = 3980) for the reward expected value after correction. At the same statistical level, participants with OCD showed the increased activity at left anterior cingulum ([-8, 44, 10]; t = 3.89, k = 5320) at the phase of value expectation under avoidance condition (shown in ***Figure 5-14***). No significance was found in participants with GD at the phase of value expectation under both reward and avoidance condition compared to healthy controls.

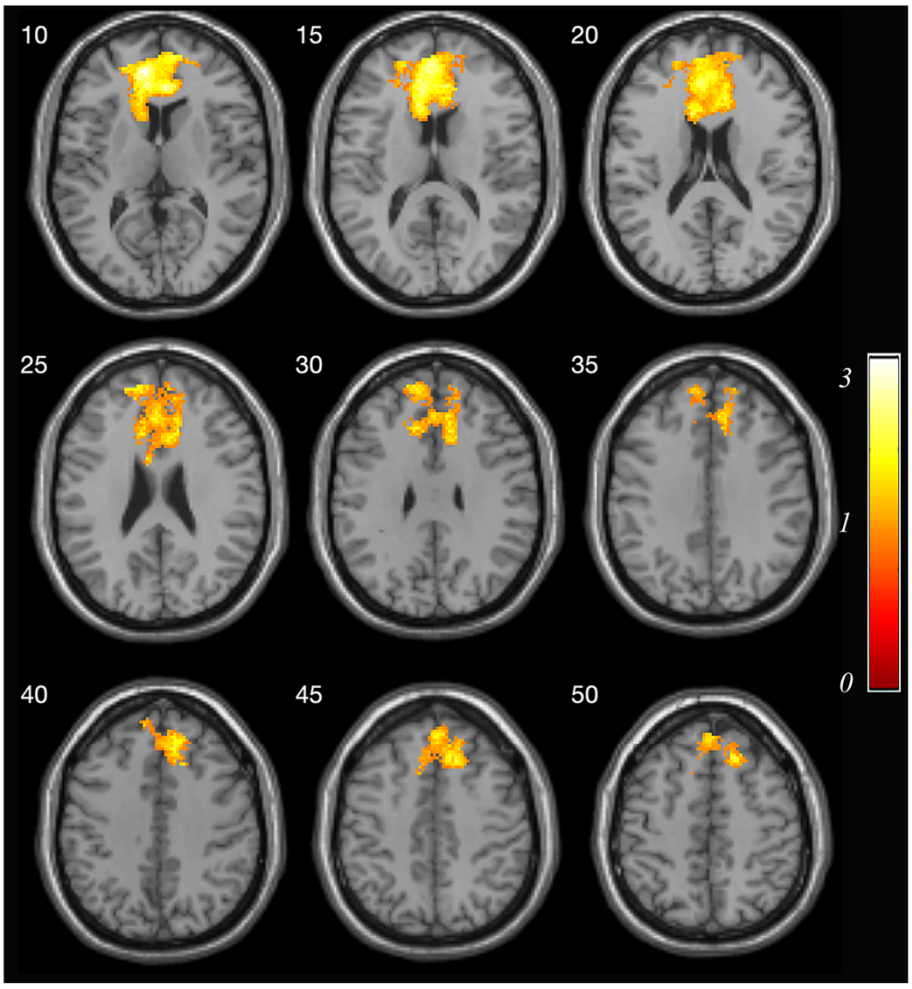

Figure 5-14 *Brain activation of avoid expected value in participants with OCD compared to controls. The left Anterior cingulum extending to the right anterior cingulum was found to have higher brain activation in OCD compared to controls at the phase of expectation under reward condition (see* ***Table 5-5 for details****).*

Table 5-5 *Brain regional response difference to reward/avoidance expected value among participants with OCD compared with healthy controls.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward expected value					
OCD < Control					
R. Precuneus	34	-18	32	3.94	3416
L. Middle Cingulum	-8	6	26	3.90	3980
Avoid expected value					
OCD > ***Control***					
L. Anterior Cingulum	-8	44	10	3.89	5320
R. Anterior Cingulum	8	34	8	3.80	
L. Anterior Cingulum	0	32	-2	3.54	

Compared to participants with OCD, GD showed significantly higher activation at the right Middle cingulum ([4, -12, 28]; t = 3.87, k = 9569) for the reward expected value after correction (see ***Figure 5-15***). No significant differences were found between OCD and GD for the avoidance expected value.

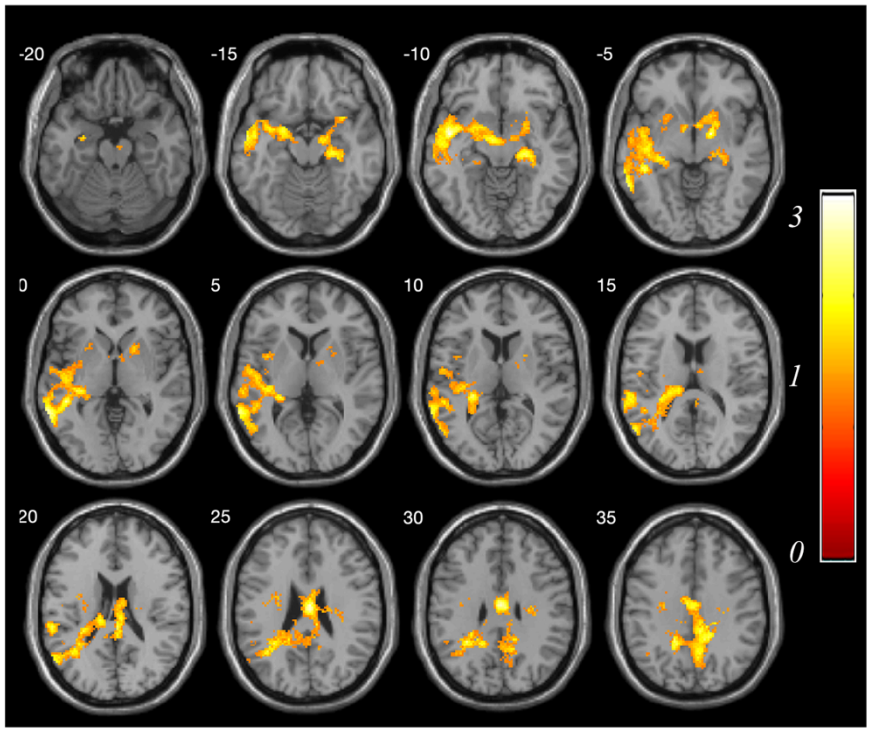

Figure 5-15 *The brain activation of reward expected value in participants with GD compared to participants with OCD. The participants with GD showed the increased activity at the brain region including the right pallidum and left superior temporal to the reward expected value compared to participants with OCD (see* ***Table 5-6 for details****).*

Table 5-6 *Brain regional response difference to reward/aversive expected value between participants with OCD and GD.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward expected value					
GD > OCD					
R. Middle Cingulum	4	-12	28	3.87	9569
R. Pallidum	18	-4	-4	3.86	
L. Superior Temporal	-46	-4	-10	3.75	

Neural response to reward/aversive prediction error

Compared to healthy controls, participants with GD showed the decreased activity at the brain region right posterior cingulate peaked at ([20, -36, 40]; t = 4.09, k = 6388) and the brain region left precuneus peaked at ([-30, -16, 24]; t = 3.09, k = 2030) after FWE correction. At the meantime, the increased activations were found at right thalamus ([28, -40, 10]; t = 4.43, k = 1042), and right middle frontal ([34, 26, 20]; t = 3.93, k = 641) for the aversive PE in GD compared to healthy controls (shown in ***Figure 5-16***) at $p < 0.005$ with FWE correction at cluster level. No significant differences were found at the phase of error signal processing both under the reward and avoidance condition in OCD group compared to healthy controls.

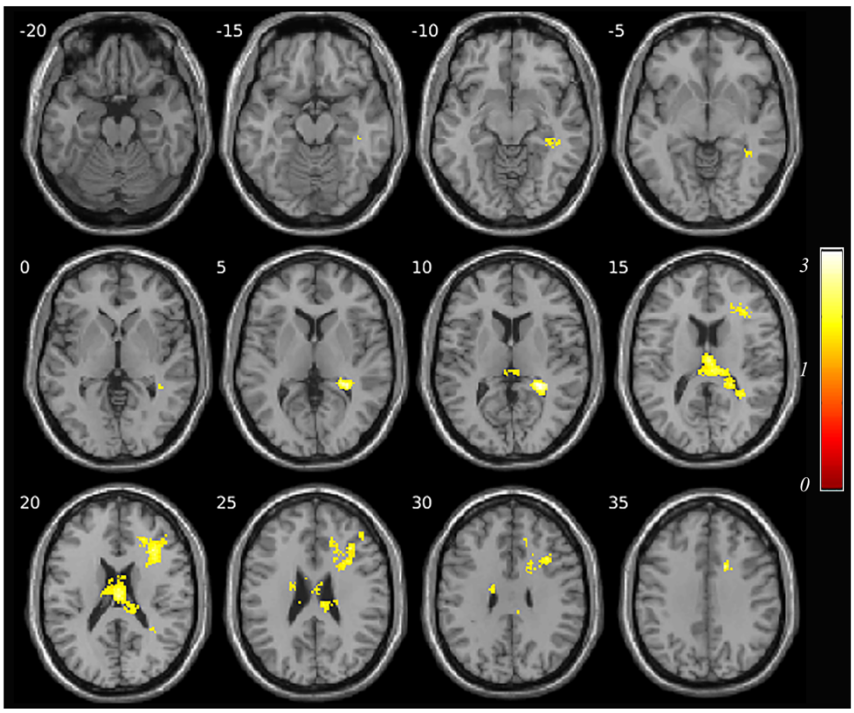

Figure 5-16 *The brain activation of aversive PE in participants with GD compared to controls. The brain region including the right triangular part of the inferior frontal was found to have higher activation in GD compared to controls at the phase of error signal processing under avoidance condition (see* ***Table 5-7 for details****).*

Table 5-7 *Brain regional response difference to reward/avoidance prediction error among participants with GD compared with healthy controls.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward PE					
Healthy controls > GD					
R. Posterior Cingulate	20	-36	40	4.09	6388
L. Precuneus	-30	-16	24	3.09	2030
Aversive PE					
GD > ***Healthy controls***					
R. Thalamus	28	-40	10	4.43	1042
R. Middle Frontal	34	26	20	3.93	641
R. Triangular part of the Inferior Frontal	36	18	22	3.61	-

Compared to GD, participants with OCD showed the increased activity at the brain region left Hippocampus peaked at ([-22, -44, 0]; t = 4.23, k = 5388) for reward PE. Whereas GD showed the increased activity at right Middle frontal peaked at ([-26, 12, 24]; t = 4.77, k = 13560) for the aversive PE compared to participants with OCD (see ***Figure 5-17***).

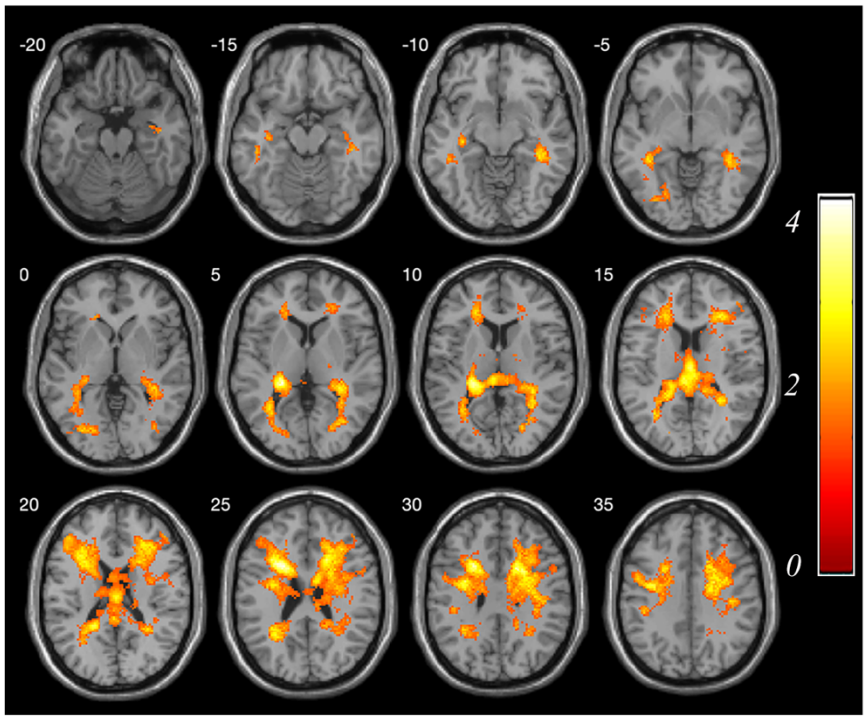

Figure 5-17 *The brain activation of aversive PE in participants with GD compared to participants with OCD. The brain region at the right middle frontal was found to have higher activation in GD compared to participants with OCD at the phase of error signal processing under avoidance condition (see* ***Table 5-8 for details****).*

Table 5-8 *Brain regional response difference to reward/aversive PE between participants with OCD and GD.*

Region	MNI Coordinates			t value	Spatial extent (in contiguous voxels)
	x	y	z		
Reward PE					
OCD > GD					
L. Hippocampus	-22	-44	0	4.23	5388
L. Lingual	16	-34	-10	3.66	
Aversive PE					
GD > OCD					
R. Middle Frontal	-26	12	24	4.77	13560

Relationship between the brain activation and the clinical measurement of disease severity, and the scales of impulsivity and compulsivity

The Pearson correlation analysis showed that the increased activity at left anterior cingulum for the expected value under the avoidance condition in OCD were negatively correlated with the personality traits of compulsivity measured by the total OCI-R scores ($r = -0.36$, $p = 0.06$). Further, the decreased activity at right precuneus and left middle cingulum for reward expected value were negatively correlated with the personality traits of impulsivity measured by the BIS scores at ($r = -0.56$, $p = 0.02$), and ($r = 0.61$, $p = 0.001$) respectively in OCD participants' group. While the decreased activation of getting reward outcome at the left postcentral was positively correlated with the total OCI-R scores ($r = 0.47$, $p = 0.06$) in GD participants' group. Further, the decreased activity of reward PE in GD at the right posterior cingulate was found negatively correlated with the BIS scores ($r = -0.47$, $p = 0.07$). Also, the decreased actives of reward PE in GD at the left precuneus was found negatively correlated with the total severity scores measured by PGSI ($r = -0.48$, $p = 0.06$) (see ***Table 5-9*** for details). After c Bonferroni correction for multiple comparisons, the negative correlation between the decreased activity in the left middle cingulum for reward expected value and the impulsivity scores measured by the BIS scores remained significant in OCD participants ($r = 0.61$, $*p = 0.039 < 0.05$).

Table 5-9 *Relationship between the brain activation and the disease clinical measurement, and the personality measures of impulsivity and compulsivity without correction.*

Brain regions	OCI-R	PGSI	BIS
Reward success – neutral success			
OCD (N = 28) < Control (N = 39)			
L. Inferior frontal	r = -0.23; p = 0.23; N = 28	r = 0.11; p = 0.58; N = 28	r = 0.08; p = 0.70; N = 28
R. Thalamus	r = -0.24; p = 0.22; N = 28	r = -0.10; p = 0.60; N = 28	r = 0.13; p = 0.51; N = 28
R. Inferior frontal	r = -0.09; p = 0.64; N = 28	r = 0.003; p = 0.99; N = 28	r = 0.18; p = 0.38; N = 28
GD (N = 16) > Control (N = 39)			
R. Cuneus	r = 0.37; p = 0.16; N = 16	r = 0.16; p = 0.55; N = 16	r = -0.31; p = 0.25, N = 16
R. Rolandic operculum	r = 0.31; p = 0.25; N = 16	r = -0.08; p = 0.76; N = 26	r = -0.31; p = 0.25; N = 16
L. Postcentral	**r = 0.47; *p =0.06; N = 16**	r = -0.25; p = 0.36; N = 16	r = -0.24; p = 0.36; N = 16
Reward expected value			
OCD (N = 28) > Control (N = 39)			
R. Precuneus	r = -0.11; p = 0.57; N = 28	r = 0.12; p = 0.55; N = 28	**r = -0.56; *p = 0.02; N = 28**
L. Middle cingulum	r = -0.10; p = 0.63; N = 28	r = 0.08; p = 0.68; N = 28	**r = -0.61; **p = 0.001; N = 28**
Avoid expected value			
OCD (N = 28) > Control (N = 39)			
L. Anterior Cingulum	**r = -0.36; *p = 0.06; N = 28**	r = -0.32; p = 0.10; N = 28	r = -0.16; p = 0.43; N = 28
Reward PE			
GD (N =16) < Control (N = 39)			
R. Posterior cingulate	r = -0.38; p = 0.14; N = 16	r = -0.17; p = 0.54; N = 16	**r = -0.47; *p = 0.07; N = 16**
L. Precuneus	r = -0.23, p = 0.40; N = 16	**r = -0.48, *p = 0.06; N = 16**	r = -0.40, p = 0.12; N = 16
Aversive PE			
GD (N = 16) > Control (N = 39)			
R. Thalamus	r = -0.37; p = 0.17; N = 16	r = 0.27; p = 0.31; N = 16	r = 0.29; p = 0.28; N = 16
R. Middle frontal	r = -0.42; p = 0.11; N = 16	r = 0.15; p = 0.58; N = 16	r = 0.22; p = 0.41; N = 16

5.4 Discussion

In this chapter, we carried out the statistical analysis to investigate the learning performance including choice and response time among three groups of participants in the reward and avoidance learning task. Then, applying the Q-learning model with Bayesian estimation to the participants' behavioural data, we extracted the learning characteristics including the learning rate and inverse temperature parameter. Further, time series of expected value and PE under reward and avoidance conditions derived from the model were regressed through the fMRI data to investigate the underlying neural mechanism. In addition, correlation analysis of brain activity with clinical measurements (OCI-R and BIS) was carried out to examine effects of impulsivity and compulsivity behavioural traits.

Obsessive compulsive disorder

As with healthy controls, OCD participants showed a significant preference towards the Correct choice under both reward and avoidance condition. No significant behavioural differences were found between the OCD and healthy controls. The OCD group showed the decreased activity at left and right Opercula region of the inferior frontal and right thalamus at the outcome of receiving reward. OCD participants also showed the increased activity in the left anterior cingulum during the phase of value expectation under avoidance condition. No significant difference of brain activity was found for error signal processing under both reward and avoidance condition in OCD participants compared to healthy controls.

The conceptualization of the pathophysiology of OCD has been within the cortico-striato-thalamo-cortico circuit. According to the model, the projection from the frontal regions to the striatum travels through direct and indirect pathways to the thalamus, and project back to the frontal regions (Moreira et al., 2017). In line with the literature, the decreased activity at left and right Opercular part of the inferior frontal and right thalamus

were found at the outcome of receiving reward in our study. The inferior frontal region has been reported to be associated with reward sensitivity, and higher reward sensitivity was associated with increased activity in the bilateral inferior frontal according to a recent fMRI study (Fuentes-Claramonte et al., 2016). While the inferior frontal junction was reported to be involved in cognitive control (Cole & Schneider, 2007), and may modulate the striatal response that is essential to reward-related behaviours (Delgado, 2007). The dorsal striatum is suggested to carry out an important role in specific action-outcome associations, and the activation could be modulated by the value provided by the reward feedback (Balleine et al., 2007; Delgado, 2007). The monetary reward was found to increase the thalamic activation (Thut et al., 1997), and the thalamus nucleus has projections to the frontal cortex forming the final link in the reward circuit (Haber & Knutson, 2010).

The available data suggest that compulsivity in OCD and addictions are related to impaired reward and punishment processing in the ventral striatum and associated attenuated dopamine release, and with negative reinforcement in limbic and anti-reward systems, which may at least partly explain the presence of repetitive self-defeating behaviours. Also, the habitual responding regardless of its consequences is an aspect of compulsivity that might be related to imbalances between ventral and dorsal frontostriatal recruitment (Figee et al., 2016). The ACC is essential to the reinforcement-based decision process and implied in performance monitoring and cognitive control. With strong connections with motor areas but few direct connections with sensory cortex, the ACC was suggested to be responsible for action value calculation to produce a favourable outcome (Inserm & Cell, 2007; Philiastides et al., 2010). A previous study reported the activation in rostral ACC was related to increases in the level of expected reward (Marsh et al., 2008). Further, the decreased activity at left middle cingulum for reward expected value found negatively associated with the BIS scores was in line with the previously reported dysfunctional reward processing with altered brain

activity at the brain reward areas under the effects of impulsivity trait (Admon et al., 2012; Figee et al., 2010).

Gambling disorder

Different with healthy controls, GD participants were only found showing a significant difference to the Correct choice under the avoidance condition. At the three distinct phases of reward/avoidance-based decision process, GD participants showed the decreased activity at right Cuneus at the outcome of obtaining reward compared to healthy controls; No significant differences of brain activity were found at the phase of value expectation under both reward and avoidance condition. As well the increased activity in brain regions including the right triangular part of the inferior frontal for the error processing under avoidance condition was found for GD participants compared to healthy controls.

The increased activity at the inferior frontal for aversive PE was found in GD participants compared to healthy controls in our study. The inferior frontal has been thought to play a crucial role in response inhibition and behavioural impulse control (Aron et al., 2004, 2014). A previous study found a relationship between the risk PE and activity in the inferior frontal, and the activation was more pronounced in risk aversive individuals during decision-making under uncertainty (Lu et al., 2009). Thus, the increased activity might suggest the improved response inhibition to the aversive events. The thalamus was suggested to play an important role in performance monitoring (Bellebaum et al., 2005), and mediate the error-related cognitive control (Hendrick et al., 2010; S.lde & Li, 2012). The increased activity for the aversive PE in GD participants might suggest the aberrant error-related cognitive control. The cingulum was one of the brain regions involved with reward processing.

Furthermore, the altered brain activity was found at distinct phases of reward and avoidance decision processes between the OCD and GD clinical groups. The participants with OCD were found higher activation at the outcome of receiving loss at right hippocampus compared to participants with GD. Also, the higher activation was found for the PE signal under reward condition in OCD participants compared to GD participants. The hippocampus plays an important role in memory and learning (Tamnes et al., 2014), thus, the altered activity found at hippocampus implied the deficits of reward and punishment processing in OCD. The participants with GD were found higher activation of reward expected value at right middle cingulum, and of aversive PE at right middle frontal compared to participants with OCD.

In summary, the present study found the altered neural activity at distinct stages of reward and avoidance-related decision processes OCD and GD conditions, together with the associations of psychopathology such as impulsivity, which provided a better understanding of the underlying mechanisms of those two clinical conditions.

References

Abramovitch, A., & Mckay, D. (2016). Behavioral Impulsivity in Obsessive – Compulsive Disorder. 5(3), 395–397. https://doi.org/10.1556/2006.5.2016.029

Admon, R., Bleich-cohen, M., Weizmant, R., Poyurovsky, M., Faragian, S., & Hendler, T. (2012). Psychiatry Research: Neuroimaging Functional and structural neural indices of risk aversion in obsessive – compulsive disorder (OCD). Psychiatry Research: Neuroimaging, 203(2–3), 207–213. https://doi.org/10.1016/j.pscychresns.2012.02.002

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. Trends in Cognitive Sciences, 8(4), 170–177. https://doi.org/10.1016/j.tics.2004.02.010

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. Trends in Cognitive Sciences, 18(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. 27(31), 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Bellebaum, C., Daum, I., Koch, B., Schwarz, M., & Hoffmann, K. P. (2005). The role of the human thalamus in processing corollary discharge. Brain, 128(5), 1139–1154. https://doi.org/10.1093/brain/awh474

Brevers, D., Koritzky, G., Bechara, A., & Noel, X. (2015). Cognitive processes underlying impaired decision-making under uncertainty in gambling disorder. 39(10), 1533–1536. https://doi.org/10.1016/j.addbeh.2014.06.004.Cognitive

Choi, J. S., Shin, Y. C., Jung, W. H., Jang, J. H., Kang, D. H., Choi, C. H., Choi, S. W., Lee, J. Y., Hwang, J. Y., & Kwon, J. S. (2012). Altered Brain Activity during Reward Anticipation in Pathological Gambling and Obsessive-Compulsive Disorder. PLoS ONE, 7(9). https://doi.org/10.1371/journal.pone.0045938

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. NeuroImage, 37(1), 343–360. https://doi.org/10.1016/j.neuroimage.2007.03.071

Delgado, M. R. (2007). Reward-Related Responses in the Human Striatum. 88, 70–88. https://doi.org/10.1196/annals.1390.002

Endrass, T., Kloft, L., Kaufmann, C., & Kathmann, N. (2011). Approach and avoidance learning in obsessive-compulsive disorder. Depression and Anxiety, 28(2), 166–172. https://doi.org/10.1002/da.20772

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., Potenza, M. N., Robbins, T. W., & Denys, D. (2016). Compulsivity in obsessive-compulsive disorder and addictions. European Neuropsychopharmacology, 26(5), 856–868. https://doi.org/10.1016/j.euroneuro.2015.12.003

Figee, M., Vink, M., De Geus, F., Vulink, N., Veltman, D. J., Westenberg, H., & Denys, D. (2011). Dysfunctional reward circuitry in obsessive-compulsive disorder. Biological Psychiatry, 69(9), 867–874. https://doi.org/10.1016/j.biopsych.2010.12.003

Figee, M., Vink, M., Geus, F. De, Vulink, N., Veltman, D. J., & Westenberg, H. (2010). dysfunctional reward circuitry in obsessive-compulsive disorder. BPS, 69(9), 867–874. https://doi.org/10.1016/j.biopsych.2010.12.003

Fineberg, N. A., Chamberlain, S. R., Goudriaan, A. E., & Stein, D. J. (2013). New Developments in Human Neurocognition: Clinical, Genetic and Brian Imaging Correlates of Impulsivity and Compulsivity. In

CNS Spectrums (Vol. 23, Issue 1). https://doi.org/10.1038/jid.2014.371

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. Psychological Assessment, 14(4), 485–496. https://doi.org/10.1037/1040-3590.14.4.485

Fontenelle, L. F., Oostermeijer, S., Harrison, B. J., & Pantelis, C. (2011). Obsessive-Compulsive Disorder, Impulse Control Disorders and Drug Addiction Common Features and Potential Treatments. 71(7), 827–840.

Franken, I. H. A., Strien, J. W. Van, Nijs, I., & Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. 158, 155–163. https://doi.org/10.1016/j.psychres.2007.06.002

Fuentes-Claramonte, P., Ávila, C., Rodríguez-Pujadas, A., Costumero, V., Ventura-Campos, N., Bustamante, J. C., Rosell-Negre, P., & Barrós-Loscertales, A. (2016). Inferior frontal cortex activity is modulated by reward sensitivity and performance variability. Biological Psychology, 114, 127–137. https://doi.org/10.1016/j.biopsycho.2016.01.001

Fujino, J., Kawada, R., Tsurumi, K., Takeuchi, H., Murao, T., Takemura, A., Tei, S., Murai, T., & Takahashi, H. (2018). An fMRI study of decision-making under sunk costs in gambling disorder. European Neuropsychopharmacology, 28(12), 1371–1381. https://doi.org/10.1016/j.euroneuro.2018.09.006

Gelskov, V., Madsen, K. H., Ramsøy, T. Z., & Siebner, H. R. (2016). NeuroImage Aberrant neural signatures of decision-making: Pathological gamblers display cortico-striatal hypersensitivity to extreme gambles. 128, 342–352. https://doi.org/10.1016/j.neuroimage.2016.01.002

Genauck, A., Quester, S., Wüstenberg, T., Mörsen, C., & Romanczuk-seiferth, N. (2017). Reduced loss aversion in pathological gambling and alcohol dependence is associated with differential alterations in amygdala and prefrontal functioning. November 1–11. https://doi.org/10.1038/s41598-017-16433-y

Gillan, C. M., Apergis-schoute, A. M., Morein-zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2016). Europe PMC Funders Group Functional neuroimaging of avoidance habits in OCD. 172(3), 284–293. https://doi.org/10.1176/appi.ajp.2014.14040525.Functional

Grassi, G., Pallanti, S., Righi, L., Figee, M., Mantione, M., Denys, D., Piccagliani, D., Rossi, A., & Stratta, P. (2015). Think twice: Impulsivity and decision making in obsessive – compulsive disorder. 4(4), 263–272. https://doi.org/10.1556/2006.4.2015.039

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology, 35(1), 4–26. https://doi.org/10.1038/npp.2009.129

Hendrick, O. M., Ide, J. S., Luo, X., & Li, C. S. (2010). Dissociable processes of cognitive control during error and non-error conflicts: A study of the stop signal task. PLoS ONE, 5(10), 10–11. https://doi.org/10.1371/journal.pone.0013155

Holtgraves, T. (2009). Evaluating the problem gambling severity index. Journal of Gambling Studies, 25(1), 105–120. https://doi.org/10.1007/s10899-008-9107-7

Ioannidis, K., Hook, R., Wickham, K., & Grant, J. E. (2019). Impulsivity in Gambling Disorder and Problem Gambling: A. 43(1), 1–17. https://doi.org/10.1080/00952990.2016.1206113.Impulsivity

Jerome, S., Rene, Q., Marie, R., Julien, V., Jean-Paul, J., & Emmanuel, P. (2007). Expectations, gains, and losses in the anterior cingulate cortex. 7(4), 327–336.

Lai, F. D. M., & Ip, A. K. Y. (2011). Impulsivity and pathological gambling among Chinese: Is it a state or a

trait problem? BMC Research Notes, 4(1), 492. https://doi.org/10.1186/1756-0500-4-492

Learning, A. (2011). Approach and avoidance learning in obsessive- compulsive disorder. 172(October 2010), 166–172. https://doi.org/10.1002/da.20772

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). Clinical Psychology and Psychotherapy, 18(1), 75–79. https://doi.org/10.1002/cpp.693

Lu, Z., Li, X., Linden, M. Van Der, & Bechara, A. (2009). NeuroImage Neural correlates of risk prediction error during reinforcement learning in humans. NeuroImage, 47(4), 1929–1939. https://doi.org/10.1016/j.neuroimage.2009.04.096

Maleki, S., Chye, Y., Zhang, X., Parkes, L., Chamberlain, S. R., Fontenelle, L. F., Braganza, L., Youssef, G., Lorenzetti, V., Harrison, B. J., Yücel, M., & Suo, C. (2020). Neural correlates of symptom severity in obsessive-compulsive disorder using magnetization transfer and diffusion tensor imaging. Psychiatry Research - Neuroimaging, 298(February), 111046. https://doi.org/10.1016/j.pscychresns.2020.111046

Marsh, A. A., Blair, K. S., Vythilingam, M., Busis, S., & Blair, R. J. R. (2008). Response options and expectations of reward in decision-making: the differential roles of dorsal and rostral anterior cingulate cortex. 35(2), 979–988.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). British Journal of Clinical Psychology, 31(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Miedl, S. F., Fehr, T., Herrmann, M., & Meyer, G. (2014). Risk assessment and reward processing in problem gambling investigated by event-related potentials and fMRI-constrained source analysis. 1–11.

Miedl, S. F., Peters, J., & Bu, C. (2012). Altered Neural Reward Representations in Pathological Gamblers Revealed by Delay and Probability Discounting. 69(2), 177–186.

Moreira, P. S., Marques, P., Soriano-Mas, C., Magalhães, R., Sousa, N., Soares, J. M., & Morgado, P. (2017). The neural correlates of obsessive-compulsive disorder: a multimodal perspective. Translational Psychiatry, 7(8), e1224. https://doi.org/10.1038/tp.2017.189

Nestadt, P., Wang, Y., Bakker, A., & Samuels, J. (2018). Doubt and the decision-making process in obsessive-compulsive disorder. 1–4. https://doi.org/10.1016/j.mehy.2016.09.010.Doubt

Oberg, S. A. K., Christie, G. J., & Tata, M. S. (2011). Neuropsychologia Problem gamblers exhibit reward hypersensitivity in medial frontal cortex during gambling. Neuropsychologia, 49(13), 3768–3775. https://doi.org/10.1016/j.neuropsychologia.2011.09.037

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. NeuroImage, 171(December 2017), 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. Journal of Clinical Psychology, 51(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768:AID-JCLP2270510607>3.0.CO;2-1

Pauls, D. L., Abramovitch, A., Rauch, S. L., & Geller, D. A. (2014). Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nature Reviews Neuroscience, 15(6), 410–424. https://doi.org/10.1038/nrn3746

Pearlson, G. D., & Potenza, M. N. (2013). monetary rewards and losses in pathological gambling. 71(8), 749–757. https://doi.org/10.1016/j.biopsych.2012.01.006.Diminished

Petry, N. M., Stinson, F. S., & Grant, B. F. (2005). Comorbidity of DSM-IV pathological gambling and other psychiatric disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. The Journal of Clinical Psychiatry, 66(5), 564–574.

Philiastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. NeuroImage, 53(1), 221–232. https://doi.org/10.1016/j.neuroimage.2010.05.052

Potenza, M. N. (2008). The neurobiology of pathological gambling and drug addiction: an overview and new findings. July 3181–3189. https://doi.org/10.1098/rstb.2008.0100

Pro, U., Gantman, A., Greck, M. De, Tempelmann, C., Northoff, G., & See, A. (2010). Decreased Neuronal Activity in Reward Circuitry of Pathological Gamblers During Processing of Personal Relevant Stimuli. 1812(March 2009), 1802–1812. https://doi.org/10.1002/hbm.20981

Remijnse, P. L., Nielen, M. M. A., van Balkom, A. J. L. M., Cath, D. C., van Oppen, P., Uylings, H. B. M., & Veltman, D. J. (2006). Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. Archives of General Psychiatry, 63(11), 1225–1236. https://doi.org/10.1001/archpsyc.63.11.1225

S.lde, J., & Li, C. R. (2012). A cerebellar thalamic cortical circuit for error-related cognitive control. 54(1), 455–464. https://doi.org/10.1016/j.neuroimage.2010.07.042.A

Starcevic, V., Berle, D., Brakoulias, V., Sammut, P., Moses, K., Milicevic, D., & Hannan, A. (2011). The nature and correlates of avoidance in obsessive – compulsive disorder. https://doi.org/10.3109/00048674.2011.607632

Thut, G., Maguire, R. P., Leenders, K. L., Roelcke, U., Nienhusmeier, M., Schultz, W., & Missimer, J. (1997). Activation of the human brain by monetary reward. Neuroreport, 8(5), 1225–1228. http://www.ncbi.nlm.nih.gov/pubmed/9175118

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. Revista Brasileira de Psiquiatria, 35(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Watkins, C. J. C. H. (1995). Learning from delayed rewards. Robotics and Autonomous Systems, 15(4), 233–235. https://doi.org/10.1016/0921-8890(95)00026-C

Won Kim, S., & Grant, J. E. (2001). Personality dimensions in pathological gambling disorder and obsessive-compulsive disorder. Psychiatry Research, 104(3), 205–212. https://doi.org/10.1016/S0165-1781(01)00327-4

6 General discussion

To achieve the goal of maximizing the rewards or minimizing the punishments, one must adapt their behaviours to learn from the outcome. The learning performance and brain mechanisms underlying those two types of learning is still under investigation. Through a novel probabilistic reward and avoidance learning task, we used the reinforcers of gaining points or losing points to guide participants' learning processes. The basic behavioural analysis mapped the normal trajectory of learning in healthy participants' during task performance. Then a model was used to capture the participants' behaviour and model the computational processes. Combined with neuroimaging, we found the shared or distinct brain regions were involved with key stages of reward and avoidance-based decision processes. At the outcome stage, we replicated the Kim et al., (2016) findings that *medial orbitofrontal cortex (mOFC)* was activated by the outcome of obtaining reward as well as avoiding punishments. Other structures within the reward circuit including *posterior cingulum* and *dorsal striatum* were also found to be activated. Then at the stage of anticipation, the fronto-cortical and fronto-striatal circuits were found to be activated by the reward and avoidance expected value, respectively. The error processing is crucial for learning, and the reward and aversive PE was found to be associated with the activity at the cortical-basal ganglia and temporo-parietal circuit, *insula* and *dorsal striatum*, respectively.

Clinical conditions such as OCD and GD are characterized by maladaptive reward and avoidance-based decision-making processes. As a typical compulsive disorder, the OCD-related behaviours driven by impulsive processes could increase with the progression and chronicity of the disease. On the other side, the GD-related behaviours driven by compulsive processes could increase. Thus, we investigated the aberrant brain mechanisms of reward and avoidance-based decision processes, and how the orthogonal pair of the impulsivity and compulsivity affect the processes. The imaging analysis showed that the OCD group was

found to have decreased activity in the *Opercula part of the inferior frontal (bilaterally)* and *right thalamus* during the outcome of obtaining reward. OCD participants also showed the increased activity in the left *anterior cingulum* at the phase of value expectation under avoidance condition. GD participants showed decreased activity in the *right Cuneus* at the outcome of obtaining reward compared to healthy controls. Also, the increased activity in the *right triangular part of the inferior frontal* for the error processing under avoidance condition was found for GD participants compared to healthy controls. The decreased activity of the *left middle cingulum* during reward expected value was to be found negatively associated with the BIS scores, which showed the aberrant reward processing under the effects of impulsivity.

6.1 Reward and avoidance-based decision performance in healthy participants

The first question is to understand the learning performance under both learning types. The human participants' learning capabilities under both types were demonstrated equally well (Gross, 2006; Palminteri et al., 2015). Through a novel reinforcement learning task, the basic learning model predicts better performance on the reward learning (Kim et al., 2006a). In ***chapter 3***, the analysis of the response time showed that participants showed significantly quicker response under the reward condition while slower under the avoidance condition. Further, the RL model fitted to the behavioural data has found a significantly higher learning rate under avoidance condition compared to reward condition. And the inverse temperature parameter was significantly higher under reward condition compared to avoidance condition. The learning rate showed how quickly the participant's changed their choices, and the inverse temperature parameter showed the participants' balance of exploration and exploitation. The significantly higher inverse temperature parameter found under reward condition showed that participants tended to exploit the trials.

Through a simple choice task to examine the behavioural effects of the magnitudes of reward and punishment in single trials, an asymmetric effect of reward and punishment on the choice behaviour was found in a previous study (Kubanek et al., 2017). The larger a reward outcome, the tendency to repeat a choice is higher. While there is no modulation of the effect by the magnitude of a penalty, a loss drove a uniform avoidance of the choice that led to the loss. One mechanistic explanation could be the reward prediction error that drives learning in computational models of choice behaviour, like in the RL model here, may not be symmetric to corresponding punishment PE terms. Moreover, the recently found dopaminergic neurons in monkey ventral midbrain known encoding for reward PE do not encode the corresponding term for punishments (Fiorillo, 2013), which suggest that there are different neural mechanisms underlying the error processing for reward and punishment.

6.2 Shared and separate neural representations of distinct stages of reward and avoidance-based decision performance in healthy participants

The other question facing the reward and avoidance-based decision process is the neural mechanism underlying the learning processes. The same brain areas located in the *medial orbitofrontal cortex* are reported to be activated when participants received a reward or avoid an aversive outcome (Gross, 2006; Kim et al., 2006a). The PE signal (i.e., encoding the discrepancy between the expectation and actual outcome) in reward processing (refer to reward PE) is reported to be correlated with the functional activity in *ventral striatum* and *OFC* (Kim et al., 2006b; Garrison et al., 2013). The aversive PE signal in avoidance learning is associated with the brain activity in amygdala-striatal regions (Zhang et al., 2016), and *bilateral insula*, according to fMRI studies (Kim et al., 2006b; Garrison et al., 2013).

Through combination of neuroimaging and modelling, we investigated the neural bases at separate stages of reward/avoidance-based decision processes in healthy controls

including outcome, expected value and PE in ***chapter 4*** (see ***Table 6-1 & Figure 6-1*** for summary). At the outcome stage, receiving reward was associated with the activity in the *fronto-cortical circuit*, whereas receiving punishment was correlated with the activity in the *cortical and subcortical* brain regions. Thus, distinct brain regions were recruited during the reward and punishment outcome processing. According to Kim et al., (2016), the common *mOFC* was involved with the outcome of receiving reward and successfully avoiding punishment. We replicated these findings in our study. The *OFC* is suggested to be critical for representing the outcomes of actions, and subsequently impact on the control of behaviour.

At the stage of anticipation, the reward expected value was positively correlated with the activity at fronto-cortical circuit whereas the avoid expected value was negatively associated with the activity at the cortical and subcortical brain regions including *inferior OFC*, *insula, cingulum and dorsal striatum*.

Finally, at the stage of error processing, the enhanced reward PE signal in the novel reversal learning task was found to be correlated with the activity in the *cortical-basal ganglia circuit*; while the aversive PE signal was covaried with the activation in the *temporo-parietal circuit*, dorsal *striatum* and *insula*. Dopamine neurons (DA) is suggested to signal the PE signal. Specifically, the DA neurons show a rapid phasic firing increase only for unpredicted reward outcomes, and suppress firing when reward is omitted. Conversely, many DA neurons display decreased firing rates in response to aversive stimuli. Thus, DA neurons code the discrepancy of reward and its prediction bidirectionally. Excited by rewarding stimuli, it was suggested that DA neurons are also excited by aversive experience. An investigation of same set of DA neurons to both rewarding and aversive conditions in nonhuman primates found that DA neurons can be divided into two categories: a) a population excited by reward and inhibited by aversive stimuli, b) another population by both

reward and aversive events in a similar manner (Hu, 2016). Here, we found the dorsal striatum was activated by both the reward and aversive PE whereas ventral striatum was only activated by reward PE. According to the literature, the dopamine projection mesolimbic pathway begins in the VTA (ventral tegmental area) and projects to the ventral striatum while the nigrostriatal pathway begins in the SN (substantia nigra) and projects to the dorsal striatum (Watanabe and Narita, 2018). It might suggest that different dopamine projection pathways have been involved with the reward and aversive PE encoding. Also, the striatal dopaminergic systems were found to carry distinct messages by different means, which can be integrated differently to shape the basal ganglia responses to reward-related events (Morris et al., 2004).

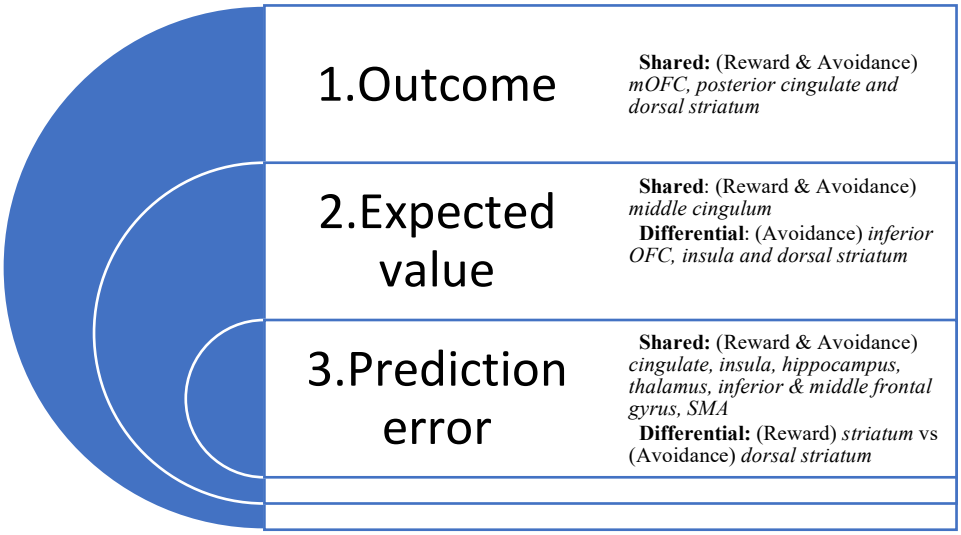

Table 6-1*. The summary of shared and differential neural mechanisms under the three distinct stages of reward and avoidance decision processes (see Figure 6-1 for image view).*

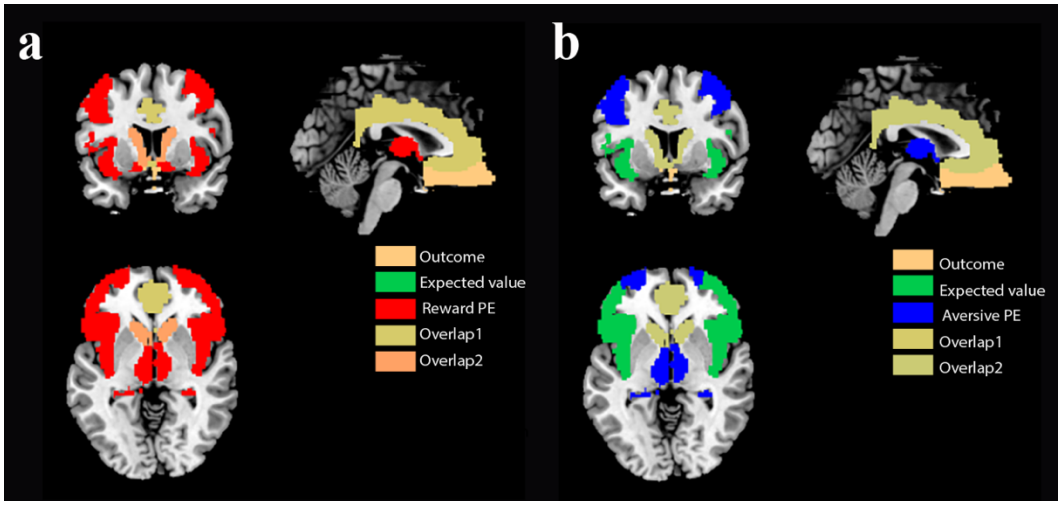

Figure 6-1*. The summary of shared and differential neural mechanisms under the three distinct stages of reward and avoidance decision processes. (a) The brain regions associated with reward decision processes: outcome was coloured in yellow, expected value in green and PE in red. The overlap of outcome and expected value was shown in Overlap1, and the overlap of expected value and PE was in Overlap2. (b) The brain regions associated with avoidance decision processes: outcome was coloured in yellow, expected value in green and PE in blue. The overlap of outcome and expected value was shown in Overlap1, and the overlap of expected value and PE was shown in Overlap2.*

6.3 Maladaptive brain activations underlying distinct stages of reward and avoidance-based decision performance in obsessive-compulsive disorder and gambling disorder

Obsessive compulsive disorder (OCD) and gambling disorder (GD) were usually suggested along the dimensional model of *impulsive-compulsive spectrum disorder* in which impulsivity and compulsivity represents polar opposite psychiatric spectrum constructs that can be viewed along a continuum of compulsive and impulsive disorders (Robbins et al., 2012) (see ***Figure 6-2).*** OCD is characterized by the experience of unwanted repetitive thoughts (obsessions) and repetitive behaviours (compulsions) with overestimation of the probability of future harm to carry on the risk avoidance (Pauls et al., 2014). Whereas the GD was recognised by the impulsive choices of persistent and recurrent maladaptive patterns of

gambling behaviour with underestimation of the likelihood or severity of possible harm (Lai and Ip, 2011).

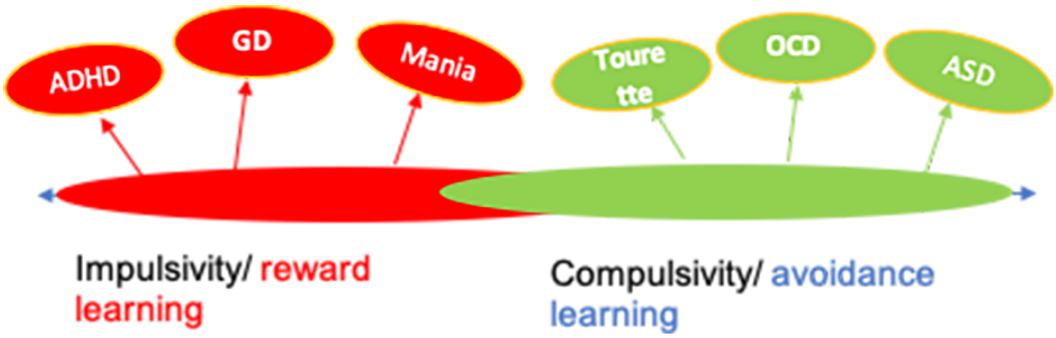

Figure 6-2 *The schematic explanation of dimensional impulsive-compulsive spectrum disorders. The GD is one of the representative impulsive disorders whereas the OCD is one of the representative compulsive disorders (Robbins et al., 2012).*

Instead of the one-dimensional model, compulsivity and impulsivity were recently suggested orthogonal factors that each contribute in varying degrees toward the development of OCD and GD (see the ***Figure 6-3*** for the schematic explanation) (Fontenelle et al., 2011; Fineberg et al., 2013). According to previous literature, the enhancements of harm-avoidance or avoidance habit in OCD with exaggerated anticipation and avoidance of aversive outcomes has been found in previous studies (Learning, 2011; Starcevic et al., 2011; Gillan et al., 2016), and the excessive avoidance behaviour was correlated with the hyper-activation in the orbitofrontal-striatal circuit (Remijnse et al., 2006; Gillan et al., 2016). As the conceptualization of a compulsive disorder, it is also suggested that OCD shares behavioural components of impulsivity (Fontenelle et al., 2011; Grassi et al., 2015; Abramovitch and Mckay, 2016). Recent studies have reported the dysfunctional reward processing with altered neural activity in the brain reward circuit and risk aversion in OCD under the effect of impulsivity trait (Figee et al., 2010; Admon et al., 2012).

According to neuroimaging studies, the “reward deficiency” in GD has been demonstrated with the hyperactivity in the reward circuitry including striatum and prefrontal

brain regions (Pro et al., 2010; Oberg et al., 2011; Brevers et al., 2015). Not only the increased activity found for reward processing, but the deactivation to loss aversion in the cortico-striatal circuit has also been reported in GD (Gelskov et al., 2016; Genauck et al., 2017). As the core feature of GD, the levels of impulsivity were found inversely correlated with activity of reward and avoidance processing (Pearlson and Potenza, 2013), and the ability to alter choice behaviour in response to stimulus-reward contingencies (Franken et al., 2008). It was pointed out that compulsivity should be also considered to investigate the deficits in decision making of GD (Ioannidis et al., 2019), and with the increases of the impulsive behaviour, the compulsivity feature would be acquired (Fontenelle et al., 2011).

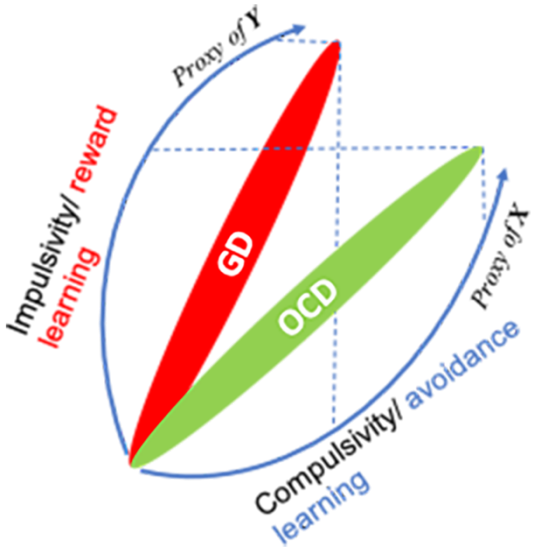

Figure 6-3 *The schematic explanation of reward and avoidance-based learning and decision processes under the effect of impulsivity and compulsivity constructs. The impulsivity and compulsivity are an orthogonalized pair of constructs to contribute certain degrees to affect the reward and avoidance learning in GD and OCD (Fineberg et al., 2013).*

In ***chapter 5***, we examined the potential maladaptive reward and avoidance decision processes and the underlying neural mechanisms in OCD and GD. At the three distinct phases of reward/avoidance-based decision process, OCD showed the decreased activity at *left Opercula part of the inferior frontal*, *right Opercula part of the inferior frontal* and *right*

thalamus at the outcome of getting reward. The *inferior frontal* was reported to be associated with reward sensitivity, and higher reward sensitivity was observed increased activity in the *bilateral inferior front*al according to a recent fMRI study (Fuentes-Claramonte et al., 2016). The thalamus nucleus has projections to the frontal cortex forming the final link in the reward circuit (Haber and Knutson, 2010). Also, OCD showed decreased activity in the *left middle cingulum* for reward expected value, and the decreased activity were negatively correlated with the personality traits of impulsivity measured by the BIS scores. The findings were in line with the aforementioned dysfunctional reward processing with altered neural activity in the brain reward circuit under the effect of impulsivity trait (Figee et al., 2010; Admon et al., 2012). Participants with OCD showed the increased activity in the *left anterior cingulum* at the phase of value expectation under avoidance condition. With strong connections with motor areas but few direct connections with sensory cortex, the *anterior cingulum* was suggested to be responsible for action value calculation to produce a favourable outcome (Jerome et al., 2007; Philiastides et al., 2010).

Participants with GD showed the decreased activity in the *right Cuneus* during the outcome stage of obtaining reward compared to healthy controls; Also, participants with GD showed increased activity in the brain region including the *right triangular part of the inferior frontal cortex* for error processing under avoidance condition. The *inferior frontal* has been thought to play a crucial role in response inhibition behavioural impulse control (Aron et al., 2004, 2014). The increased activity in the *inferior frontal* for PE under avoidance condition might suggest the improved response inhibition to aversive event.

6.4 Future direction

Through the findings of ***chapter 4***, we found several sets of brain regions associated with distinct stages of reward and avoidance decision processes, however, the coupling

nature of those brain areas and how the coupling influenced by the changes of experimental manipulations remain unclear. One potential future project would be to apply dynamic causal modeling (DCM) to establish the experimentally induced coupling changes among these regions (Friston et al., 2003). Secondly, to provide more evidence of the dopamine functioning underlying the reward and avoidance decision process, molecular imaging techniques such as PET/fMRI could be a potential powerful tool (Heiss, 2009). Through our study, the striatum was found involved into the distinct stages of decision processes. Targeting the striatum, the simultaneous PET/fMRI could measure the dopamine release directly (Zürcher et al., 2021). Besides the Dopamine (DA) neurons, the GABAergic (gamma-Aminobutyric acid) neurons are also suggested to be involved with the error signal encoding (Hu, 2016) as a potential inhibitor. Magnetic Resonance Spectroscopy (MRS) is a quantitative method in the family of MRI to assess the concentration of metabolites (such as GABA, glutamate, glutamine, Choline, Creatine) at a certain brain region. Using such technology, a study reported decreased levels of GABA and Glutamine (Glx) in vmPFC has been found during the rewarding information than the aversive events (Padulo et al., 2016). Also, the Chol (choline) was suggested to be involved with the learning processes, and could inform when to learn (Morris et al., 2004). Recently, Li et al., demonstrated the feasibility of rapid, high-resolution, near whole-brain 3D MR spectroscopic imaging (MRSI), which is more powerful to explore metabolites concentration across the whole brain while concurrently collecting fMRI signals (Li et al., 2020). In other words, such sequences combine fMRI and MRSI together to record metabolites concentration (such as GABA and Glu) and neural activation (BOLD signal) simultaneously while the participant is performing tasks in the scanner. It will facilitate the fundamentally mechanistic understanding of these imaging signals we observed at neuronal or biochemical level.

Apart from the current paradigm and model we applied to investigate our research questions in ***chapter 4*** and ***5***, more modified paradigms can be developed to answer more specific questions. For example, the probabilistic switch could be tailored according to the performance of the participant and an increased probabilistic switch may be introduced for an extra level of re-learning process. More complex models can be introduced to enrich the estimation at the behaviour level (e.g., one could add randomness factor or stickiness factor to account for such effects in the decision-making process).

Moreover, the main goal of the thesis is to examine the neural associates of computational signals based on the estimated learning parameters. While, another interesting question was the brain architecture for predicting learning performance. It was observed that individuals' resting baseline activity in motor and visual regions could predict the future learning rate (Tamnes et al., 2014). As we have found the different learning performance and neural mechanisms of reward and avoidance learning, future studies could be carried out to examine the brain regions for prediction both types of learning rate.

In summary, the novel probabilistic reward and avoidance learning task offered an innovative way to examine the reward and avoidance-based decision processes in healthy participants. The learning model was then used to investigate the underlying computational processes. Together with the modelling and neuroimaging, the thesis found shared and distinct neural representations at various stages of reward/avoidance-based decision processes. Specifically, at the outcome stage, the outcome of receiving reward and successfully avoiding punishment was found associated with the consistent *mOFC* implicated in the previous study (Kim et al., 2006). But also, the new candidates including *posterior cingulum* and *dorsal striatum were* found activated. Receiving reward and punishment was associated with the functional activity at common brain areas of *insula* and *cingulum*. And the *cingulum* was also found activated for both reward and avoidance expectation. Whereas

avoidance expectation recruited broader areas at the cortical and subcortical brain areas including *inferior OFC*, *insula* and *dorsal striatum*. At the stage of error processing, reward and aversive PE was found covaried with the activity at the shared frontal-subcortical brain regions including *cingulate, insula, hippocampus, thalamus, inferior & middle frontal and SMA.* Different from the whole *striatum* activation by reward PE, the *dorsal striatum* was specifically activated by the aversive PE. Further, application of the modelling and neuroimaging to the clinical populations of OCD and GD, the aberrant brain activations were found during these processes. Based on the newly suggested orthogonal pairs of impulsivity and compulsivity behavioural traits, we then examined how the impulsivity and compulsivity constructs affect the reward and avoidance decision processes. The whole study offered a clarification of the neural mechanisms underlying the reward and avoidance processes, and also provided a better understanding of the pathology of the aberrant reward and avoidance-based decision making processes in OCD and GD.

References

Abramovitch A, Mckay D (2016) Behavioral Impulsivity in Obsessive – Compulsive Disorder. 5:395–397.

Admon R, Bleich-cohen M, Weizmant R, Poyurovsky M, Faragian S, Hendler T (2012) Psychiatry Research: Neuroimaging Functional and structural neural indices of risk aversion in obsessive – compulsive disorder (OCD). Psychiatry Res Neuroimaging 203:207–213 Available at: http://dx.doi.org/10.1016/j.pscychresns.2012.02.002.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8:170–177.

Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: One decade on. Trends Cogn Sci 18:177–185 Available at: http://dx.doi.org/10.1016/j.tics.2013.12.003.

Brevers D, Koritzky G, Bechara A, Noel X (2015) Cognitive processes underlying impaired decision-making under uncertainty in gambling disorder. 39:1533–1536.

Figee M, Vink M, Geus F De, Vulink N, Veltman DJ, Westenberg H (2010) dysfunctional reward circuitry in obsessive-compulsive disorder. BPS 69:867–874 Available at: http://dx.doi.org/10.1016/j.biopsych.2010.12.003.

Fineberg NA, Chamberlain SR, Goudriaan AE, Stein DJ (2013) New Developments in Human Neurocognition: Clinical, Genetic and Brian Imaging Correlates of Impulsivity and Compulsivity.

Fiorillo CD (2013) Two dimensions of value: Dopamine neurons represent reward but not aversiveness. Science (80-) 341:546–549.

Fontenelle LF, Oostermeijer S, Harrison BJ, Pantelis C (2011) Obsessive-Compulsive Disorder, Impulse Control Disorders and Drug Addiction Common Features and Potential Treatments. 71:827–840.

Franken IHA, Strien JW Van, Nijs I, Muris P (2008) Impulsivity is associated with behavioral decision-making deficits. 158:155–163.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302.

Fuentes-Claramonte P, Ávila C, Rodríguez-Pujadas A, Costumero V, Ventura-Campos N, Bustamante JC, Rosell-Negre P, Barrós-Loscertales A (2016) Inferior frontal cortex activity is modulated by reward sensitivity and performance variability. Biol Psychol 114:127–137 Available at: http://dx.doi.org/10.1016/j.biopsycho.2016.01.001.

Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310 Available at: http://dx.doi.org/10.1016/j.neubiorev.2013.03.023.

Gelskov V, Madsen KH, Ramsøy TZ, Siebner HR (2016) NeuroImage Aberrant neural signatures of decision-making: Pathological gamblers display cortico-striatal hypersensitivity to extreme gambles. 128:342–352.

Genauck A, Quester S, Wüstenberg T, Mörsen C, Romanczuk-seiferth N (2017) Reduced loss aversion in pathological gambling and alcohol dependence is associated with differential alterations in amygdala and prefrontal functioning. :1–11.

Gillan CM, Apergis-schoute AM, Morein-zamir S, Urcelay GP, Sule A, Fineberg NA, Sahakian BJ, Robbins TW (2016) Europe PMC Funders Group Functional neuroimaging of avoidance habits in OCD. 172:284–

293.

Grassi G, Pallanti S, Righi L, Figee M, Mantione M, Denys D, Piccagliani D, Rossi A, Stratta P (2015) Think twice: Impulsivity and decision making in obsessive – compulsive disorder. 4:263–272.

Gross L (2006) Avoiding punishment is its own reward. 4.

Haber SN, Knutson B (2010) The reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26 Available at: http://dx.doi.org/10.1038/npp.2009.129.

Heiss WD (2009) The potential of PET/MR for brain imaging. Eur J Nucl Med Mol Imaging 36:105–112.

Hu H (2016) Reward and Aversion. Annu Rev Neurosci 39:297–324.

Ioannidis K, Hook R, Wickham K, Grant JE (2019) Impulsivity in Gambling Disorder and Problem Gambling: A. 43:1–17.

Jerome S, Rene Q, Marie R, Julien V, Jean-Paul J, Emmanuel P (2007) Expectations, gains, and losses in the anterior cingulate cortex. 7:327–336.

Kim H, Shimojo S, O'Doherty JP (2006a) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:e233.

Kim H, Shimojo S, O'Doherty JP (2006b) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:1453–1461.

Kubanek J, Snyder LH, Abrams; RA (2017) Reward and punishment act as distinct factors in guiding behaviour. Physiol Behav 176:139–148.

Lai FDM, Ip AKY (2011) Impulsivity and pathological gambling among Chinese: Is it a state or a trait problem? BMC Res Notes 4:492 Available at: http://www.biomedcentral.com/1756-0500/4/492.

Learning A (2011) Approach and avoidance learning in obsessive- compulsive disorder. 172:166–172.

Li Y, Wang T, Zhang T, Lin Z, Li Y, Guo R, Zhao Y, Meng Z, Liu J, Yu X, Liang Z-P, Nachev P (2020) Fast high-resolution metabolic imaging of acute stroke with 3D magnetic resonance spectroscopy. Brain:3225–3233.

Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but Distinct Messages of Midbrain Dopamine and Striatal Tonically Active Neurons. 43:133–143.

Oberg SAK, Christie GJ, Tata MS (2011) Neuropsychologia Problem gamblers exhibit reward hypersensitivity in medial frontal cortex during gambling. Neuropsychologia 49:3768–3775 Available at: http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.037.

Padulo C, Delli Pizzi S, Bonanni L, Edden RAE, Ferretti A, Marzoli D, Franciotti R, Manippa V, Onofrj M, Sepede G, Tartaro A, Tommasi L, Puglisi-Allegra S, Brancucci A (2016) GABA levels in the ventromedial prefrontal cortex during the viewing of appetitive and disgusting food images. Neuroscience 333:114–122 Available at: http://dx.doi.org/10.1016/j.neuroscience.2016.07.010.

Palminteri S, Khamassi M, Joffily M, Coricelli G (2015) Contextual modulation of value signals in reward and punishment learning. Nat Commun 6.

Pauls DL, Abramovitch A, Rauch SL, Geller DA (2014) Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nat Rev Neurosci 15:410–424 Available at: http://www.nature.com/nrn/journal/v15/n6/full/nrn3746.html%5Cnhttp://www.nature.com/nrn/journal/v15/n6/pdf/nrn3746.pdf.

Pearlson GD, Potenza MN (2013) monetary rewards and losses in pathological gambling. 71:749–757.

Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR (2010) Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage 53:221–232 Available at: http://dx.doi.org/10.1016/j.neuroimage.2010.05.052.

Pro U, Gantman A, Greck M De, Tempelmann C, Northoff G, See A (2010) Decreased Neuronal Activity in Reward Circuitry of Pathological Gamblers During Processing of Personal Relevant Stimuli. 1812:1802–1812.

Remijnse PL, Nielen MMA, van Balkom AJLM, Cath DC, van Oppen P, Uylings HBM, Veltman DJ (2006) Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. Arch Gen Psychiatry 63:1225–1236.

Starcevic V, Berle D, Brakoulias V, Sammut P, Moses K, Milicevic D, Hannan A (2011) The nature and correlates of avoidance in obsessive – compulsive disorder.

Tamnes, C. K., Walhovd, K. B., Engvig, A., Grydeland, H., Krogsrud, S. K., Østby, Y., Holland, D., Dale, A. M., & Fjell, A. M. (2014). Regional hippocampal volumes and development predict learning and memory. *Developmental Neuroscience*, *36*(3–4), 161–174. https://doi.org/10.1159/000362445

Watanabe M, Narita M (2018) Brain reward circuit and pain. Adv Exp Med Biol 1099:201–210.

Zhang S, Mano H, Ganesh G, Robbins T, Seymour B (2016) Dissociable Learning Processes Underlie Human Pain Conditioning. Curr Biol 26:52–58 Available at: http://dx.doi.org/10.1016/j.cub.2015.10.066.

Zürcher NR, Walsh EC, Phillips RD, Cernasov PM, Tseng CEJ, Dharanikota A, Smith E, Li Z, Kinard JL, Bizzell JC, Greene RK, Dillon D, Pizzagalli DA, Izquierdo-Garcia D, Truong K, Lalush D, Hooker JM, Dichter GS (2021) A simultaneous [11C]raclopride positron emission tomography and functional magnetic resonance imaging investigation of striatal dopamine binding in autism. Transl Psychiatry 11 Available at: http://dx.doi.org/10.1038/s41398-020-01170-0.