

# **The cognitive and neural impact of perceived uncontrollability on reward learning**

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial  
fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Vihang Vaidya



Indian Institute of Science Education and Research Pune  
Dr. Homi Bhabha Road,  
Pashan, Pune 411008, INDIA.

Date: March 16, 2025

Under the guidance of

Supervisor: Marc Guitart-Masip,  
Associate Professor, Karolinska Institute

From May 2024 to Mar 2025

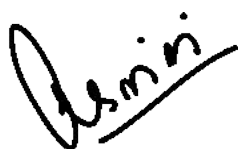
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH  
PUNE

# Certificate

This is to certify that this dissertation entitled **The cognitive and neural impact of perceived uncontrollability on reward learning** submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Vihang Vaidya at the Aging Research Center, Karolinska Institutet under the supervision of Dr. Marc Guitart-Masip, Associate Professor, Aging Research Center, Karolinska Institutet, during the academic year 2024-2025.



Supervisor:  
Dr. Marc Guitart-Masip  
Associate Professor,  
Aging Research Center,  
Karolinska Institutet



Thesis Advisory Committee:  
Dr. Collins Assisi  
Assistant Professor,  
Department of Biology,  
Indian Institute of Science Education and Research, Pune

This thesis is dedicated to my loved ones.

# Declaration

I hereby declare that the matter embodied in the report entitled **The cognitive and neural impact of perceived uncontrollability in reward learning** are the results of the work carried out by me at the Aging Research Center, Karolinska Institutet, Sweden under the supervision of Dr. Marc Guitart-Masip and the same has not been submitted elsewhere for any other degree.

A handwritten signature in black ink that reads "Vaidya". The signature is written in a cursive style with a horizontal line underneath the name.

Vihang Vaidya

Date: 21/03/2025

# Table of Contents

Declaration .....	4
Abstract .....	8
Acknowledgements .....	9
Contributions .....	10
Chapter 1 Introduction.....	11
Reward Learning.....	11
Stress.....	14
Helplessness and Depression.....	17
Controllability .....	19
Chapter 2 Methods .....	22
Task Design .....	22
Collected Data .....	27
Behavioural Data Analyses .....	28
Computational Models .....	29
Neuroimaging Analyses .....	32
Chapter 3 Results.....	38
Lack of control impairs learning after reversal, but only in bonus games .....	38
Subjective perception of control mediates the effect of uncontrollability .....	43
Brain representation does not differ by controllability or valence .....	46
Chapter 4 Discussion .....	50
References.....	53

# List of Tables

Table 1: GLMM results. ....	39
Table 2: Results of the GLMM with controllability rating added as a regressor. ....	45

# List of Figures

Figure 1: Three-armed probabilistic reward learning task. ....	23
Figure 2: Hidden target reversal. ....	24
Figure 3: Controllability manipulation. ....	25
Figure 4: Average learning trajectory of the game. ....	38
Figure 5: Learning curves separated by reversal. ....	40
Figure 6: Learning difference after reversal by controllability. ....	41
Figure 7: Mean accuracy per subject for each condition. ....	42
Figure 8: Subjective ratings of controllability. ....	43
Figure 9: Pairwise reversal difference by rating. ....	46
Figure 10: Group-level activation for GLM1.....	47
Figure 11: Mean Activity for GLM2.....	48
Figure 12: Mean BOLD Activation for GLM3.....	49

# Abstract

A perceived lack of control—the belief that one’s actions don’t determine outcomes—is closely related to uncontrollable stress and helplessness, which may contribute to increased vulnerability to depression. This project examines how perceived uncontrollability influences reward learning and decision-making under uncertainty. Such learning is known to be impaired in stress-related disorders. We manipulated controllability in an fMRI study where 55 participants played a multidimensional probabilistic three-armed bandit task with a hidden target reversal, while simultaneously playing to obtain monetary bonuses or avoid electric shocks. These goal outcomes were determined by performance in controllable games but were unrelated to their actions in uncontrollable games.

We found that a lack of control impaired learning after the hidden target reversal had occurred, but only when participants were playing to gain extra money. We also found that people’s subjective belief in the controllability manipulation mediated the extent of their impairment after the reversal in uncontrollable games but not in controllable games. Participants who perceived higher uncontrollability performed worse after reversal when experiencing uncontrollable conditions, and this was independent of goal outcome valence. Next, we used a hidden Markov model to best capture participants’ trial-by-trial choices, then used its estimates to examine the neural correlates of expected value. Activity associated with stimulus onset and the model-derived expected value of the chosen option was observed in the orbitofrontal cortex and striatum, but mean activation and voxel patterns did not differ by controllability or goal outcome valence. Although neural results are negative, our behavioural findings may further our understanding of mechanisms contributing to learning deficits in stress disorders.

# Acknowledgements

First, I would like to thank my supervisor, Dr. Marc Guitart-Masip, for his mentorship, time, and constant support. His positive and welcoming attitude made for an incredibly enjoyable working environment. I am grateful to members of the lab and the centre, especially Tobias Granwald, Štěpán Wenke, Martin Nilsson, Robin Pedersen, and Javier Oltra, for their advice, discussions, and company.

I would like to thank Dr. Collins Assisi, Dr. Aurnab Ghose, Dr. Leelavati Narlikar, and Dr. Anveshna Srivastava for granting me opportunities to grow and learn as a student and researcher.

I would like to thank my friends Nakul, Aditya, Akilan, Saransh, Vivek, Ujwal, and Parth for their constant support and friendship throughout my time in Pune and Stockholm.

Most importantly, I would like to especially thank my family for their unwavering support and unconditional love in every form possible. All that I am and will be, I owe it to them.

# Contributions

Contributor name	Contributor role
Marc Guitart-Masip, Amy Walsh	Conceptualization Ideas
Marc Guitart-Masip, Amy Walsh	Methodology
-	Software
Vihang Vaidya	Validation
Vihang Vaidya, Amy Walsh, Marc Guitart-Masip	Formal analysis
Amy Walsh	Investigation
Christian Lynghaug	Resources
Vihang Vaidya, Amy Walsh	Data Curation
Vihang Vaidya	Writing - original draft preparation
Vihang Vaidya, Marc Guitart-Masip	Writing - review and editing
Vihang Vaidya	Visualization
Marc Guitart-Masip	Supervision
Marc Guitart-Masip	Project administration
Marc Guitart-Masip	Funding acquisition

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy<sup>1</sup>.

---

<sup>1</sup> <https://journals.biologists.com/jcs/pages/author-contributions>

# Chapter 1 Introduction

Learning from complex and ever-changing environments requires an adaptive reward learning system that continuously integrates feedback to optimize decisions. Stress disrupts reward learning, but it is not yet fully understood how the perceived uncontrollability of stress contributes to this learning. When stress is seen as uncontrollable, it not only hampers learning but also fosters a sense of helplessness, a precursor to depression. In this context, elucidating how uncontrollability impacts reward learning is essential for understanding the neurocognitive underpinnings of stress-related disorders. In the sections below, each concept is explored in depth from the perspective of uncontrollability.

## Reward Learning

Every day, we make decisions such as choosing what to eat, how much effort to put into work, and whether to persist in a difficult task. One of the main factors guiding these decisions is reward learning, a fundamental aspect of both human and animal behaviour. It is the mechanism through which organisms adapt their actions based on previous experiences. In essence, reward learning involves associating certain behaviours with outcomes, thereby changing the likelihood of that behaviour being repeated in future. By reinforcing actions that lead to rewards and discouraging those that do not, reward learning enables individuals to navigate uncertain environments, optimise decisions, and sustain motivation toward goal-directed behaviour.

Broadly, there are two major types of learning, the first of which is classical conditioning (Pavlov, 1927). Classical conditioning is the learning of an association between a stimulus that is neutral (conditioned) and a stimulus that is innately salient (unconditioned) by being exposed to repetitions of the two being paired together. The neutral stimulus alone can then elicit a behavioural response triggered by the unconditioned stimulus. Rescorla and Wagner developed a famous theoretical model of classical conditioning (Rescorla and Wagner, 1972), introducing concepts such as scalar associative strengths and prediction error, which are foundational concepts used in reinforcement learning models to this day. The second type is instrumental learning, first explored by researchers such as Thorndike and Skinner (Thorndike, 1898; Ferster and Skinner, 1957). Unlike Pavlovian learning, where associations form between stimulus and outcomes independent of behaviour, instrumental learning involves a direct action-outcome contingency. This means that an individual's choices influence rewards or punishments, which in turn affect future behaviour.

Within instrumental learning, another distinction between different kinds of learning is model-based versus model-free learning. Model-based learning involves the usage of an explicit internal model of the environment (cognitive map) and learning causal action-outcome contingencies. The choices made are deliberative and goal-directed, optimally performed by thinking ahead and prospectively simulating the results of possible actions. This learning is computationally intensive but offers flexibility in decision-making because one can update their mental map to adapt to changes in the environment. On the other hand, model-free learning does not construct an internal model of the world. The value of an action is learned instead through trial and error by observing the reward associated with it. This type of behaviour is more habitual, relying on cached values of simple stimulus-response associations. It is less computationally intensive and thus is relevant in situations where one does not have sufficient resources such as energy and time. Both decision-making systems are thought to co-exist, and which type is used more depends on the context and constraints such as the amount of uncertainty in the environment (Daw *et al.*, 2005; Drummond and Niv, 2020).

Learning and decision-making are often examined through the lens of reinforcement learning (Sutton and Barto, 1998). The class of algorithms, originally from artificial intelligence research, describe how a human, non-human animal, or agent learns to maximise reward over time via a reward signal. The reason why RL models have found so much success in psychology and neuroscience is because of the seminal findings of Schultz and colleagues (Schultz *et al.*, 1993; Schultz, 1998). They found that the phasic activity of midbrain dopaminergic neurons in monkeys matched the reward prediction error in a temporal difference RL algorithm. This suggested that these neurons 'encode' a TD error signal, and that midbrain dopamine neurons and associated circuits may implement an RL-like algorithm in the brain (Schultz *et al.*, 1997). It is important to note that these algorithms are model-free RL. Recently, there has been some criticism against this standard model that dopamine encodes reward prediction errors (Wang *et al.*, 2018), but the framework is still widely used in literature regardless.

The major target of these midbrain dopamine neurons is the striatum, which has been studied as a candidate region for learning even before the role of dopamine in the reward prediction error was discussed. The numerous spines of striatal neurons also receive glutamatergic input from the cortex. Wickens *et al.* showed that these cortico-striatal synapses are modulated by dopamine input from the midbrain (Wickens *et al.*, 1996). These synapses are strengthened when their activation coincides with increased dopamine but get depressed when their activation is not associated with dopamine release. Dopamine signal thus mediates the potentiation or depression of the same cortico-striatal synapses (Reynolds and Wickens, 2002).

The striatum has been shown to be involved in habit formation (Yin and Knowlton, 2006). Yin *et al.* showed that lesions in the dorsolateral but not the dorsomedial striatum impaired the ability of rats to form habits from feedback interval schedules (Yin *et al.*, 2004). The striatum is associated with reward-orienting behaviour as well (Hikosaka *et al.*, 2006). Activity in the ventral striatum is correlated with the value of rewards (Knutson *et al.*, 2001). The ventral striatum, and in particular the nucleus accumbens, are associated with the value of states, whereas the dorsal striatum is involved more in the value of actions. Moreover, the dorsal medial parts of the striatum are associated with value encoding and goal-directed actions, while the dorsolateral striatum encodes more associative, habitual and associative aspects of action (Burton *et al.*, 2015). Samejima *et al.* showed that the activity of neurons in the striatum is correlated with the probability of selecting an action based on the expected reward (Samejima *et al.*, 2005). In human fMRI studies the activity in the striatum is correlated with reward magnitude (McClure *et al.*, 2004), as well as prediction errors (O'Doherty *et al.*, 2003).

Neural activity predicting reward has also been observed in the cortex, especially in regions such as the prefrontal cortex (Matsumoto *et al.*, 2003; Roesch and Olson, 2004). The ventromedial prefrontal cortex (vmPFC) has been implicated in the representation of the subjective value of choices and outcomes (Chib *et al.*, 2009). A positive correlation between the BOLD signal and the value of chosen options suggests its role in decision-making (Boorman *et al.*, 2009; Amarante and Laubach, 2014; Chung *et al.*, 2020). It is also shown that vmPFC encodes the value of items even when the person is not involved in making any choice (Lebreton *et al.*, 2009), or when observing someone else make a decision (Cooper *et al.*, 2010). Levy *et al.* showed that these value representations are similar across reward types, suggesting that the vmPFC tracks value in a domain-general manner, a common currency value representation (Levy and Glimcher, 2011; McNamee *et al.*, 2013).

The orbitofrontal cortex (OFC) is another region that has been implicated in reward learning (Tremblay and Schultz, 2000; Noonan *et al.*, 2012). The OFC is thought to represent the value of states and outcomes and to update these values based on feedback (Padoa-Schioppa and Assad, 2006; Hare *et al.*, 2008). In a study by O'Doherty *et al.*, participants were asked to choose between two stimuli that were associated with different reward probabilities. The BOLD signal in the OFC was correlated with the expected value of the rewards and punishments (O'Doherty *et al.*, 2001). Plassman *et al.* showed that medial OFC activity is associated with the value of rewarding objects (Plassmann *et al.*, 2007). The OFC is also thought to be involved in representing a cognitive map of hidden states in a task environment. Schuck *et al.* showed that unobservable task states could be decoded from activity in the OFC using pattern-classification techniques (Schuck *et al.*, 2016). In sum, brain regions like the striatum and cortex form a highly interconnected network that is involved in reward learning and decision-making (Neubert *et al.*, 2015; Tanaka *et al.*, 2015).

Much of this knowledge arises from standard principles of measuring cognition and the brain. Reward learning is examined using a variety of tasks like probabilistic learning tasks, reversal learning tasks, effort-based tasks, etc. More recently, computational models have been used on behavioural data obtained from these tasks. Popular models like RL have parameters like learning rate and reward sensitivity that are thought to represent underlying cognitive variables. Reward sensitivity represents the subjective extent to which rewards are valued and how random the decisions are, whereas learning rate is a parameter that represents the magnitude of learning from rewards.

Various neural indices of learning have also been used to measure reward learning. A popular technique is functional magnetic resonance imaging (fMRI), which measures the blood oxygen level-dependent (BOLD) activity as a proxy for neural activity. By regressing model-derived variables like prediction errors and expected value against fMRI activity, researchers pinpoint where in the brain these computations are likely to occur. Another popular technique is electroencephalogram (EEG), which has greater temporal resolution than fMRI. Event-related potential (ERP) signals after observing reward outcomes have been found to be related to prediction errors. For example, positive signals associated with positive prediction errors, better-than-expected outcomes, are termed as feedback-related positivity. Together, these neurocognitive measures help clarify the behavioural and neural mechanisms of reward processing and learning in normal conditions—an understanding that becomes important when examining how disruptions in learning occur due to things like stress.

## Stress

Stress can be defined as a natural response to challenging conditions, causing both psychological and physiological changes. However, a lot of negative situations can only be mildly challenging and do not qualify as stressors. Some stress researchers recommend that the term ‘stress’ should be used only when aversive environmental stimuli impose demands on an organism that exceed its natural regulatory capacity. This is especially relevant in situations that can be characterized as unpredictable or uncontrollable. This is because, from a physiological perspective, stress is characterized by a lack of sufficient and appropriate neuroendocrine responses to deal with such extreme conditions (Koolhaas *et al.*, 2011). Stress involves a complex interplay between the brain and body, activating the hypothalamic-pituitary-adrenal axis and the sympathetic-adrenal medullary system. This response results in the secretion of cortisol and adrenaline, hormones commonly released in stressful situations. These hormones, along with other physiological changes, prepare the biological system to function and respond appropriately by changing things such as increased pumping of blood, mobilizing of energy stores, increased breathing, etc.

Stress is a pervasive factor that can influence significant cognitive and behavioural changes. The most well-known effect is that acute stress induces a shift away from goal-directed, flexible, model-based behaviour towards habitual, inflexible, experience-dependent, model-free behaviour (Hartogsveld *et al.*, 2020). Otto *et al.* showed that acute stress reduces the amount of model-based contribution to behaviour and that this effect is mediated by working memory capacity (Otto *et al.*, 2013). Stress has also been shown to impair learning and decision-making in other ways. Petzold *et al.* showed that although psychosocial stress did not affect overall performance in a probabilistic selection task, it did impair the ability to learn from negative feedback compared to controls (Petzold *et al.*, 2010). A study by de Berker *et al.* showed that stress induced by a standard socially evaluated cold pressor test impaired learning to produce an act, due to Pavlovian associations between punishment and passivity (de Berker *et al.*, 2016). Acute stress has also been shown to reduce cognitive flexibility, an effect that is correlated both with total cortisol increase (Goldfarb *et al.*, 2017) and the time course of hypothalamic-pituitary-adrenal axis activation (Plessow *et al.*, 2011).

Oftentimes, individuals encounter stressful situations that lie beyond their control, where no immediate action can alter the circumstances causing distress. Uncontrollable stress refers to situations where individuals face stressors that they cannot influence or manage. Acute, uncontrollable stress in the form of threat-of-shock paradigms results in poorer outcomes on probabilistic learning tasks by reducing reward responsiveness (Bogdan and Pizzagalli, 2006), smaller feedback-related positivity (Bogdan *et al.*, 2011), and impaired reversal learning (Paret and Bublatzky, 2020). A meta-analysis by Dickerson and Kemeny concluded that uncontrollability in tasks elicited greater cortisol and adrenocorticotrophic hormone responses and took longer to recover, supporting the behavioural deficits with physiological changes (Dickerson and Kemeny, 2004).

Many animal studies have established that acute but mild stressors reliably activate mesocortical dopamine neurons, which project to the prefrontal cortex (PFC) and thus substantially increase dopamine in the medial prefrontal cortex (mPFC) (Abercrombie *et al.*, 1989). If rodents are subjected to larger and more chronic stressors, the mesolimbic dopamine system, in particular the nucleus accumbens (NAc), is also activated, though to a much lesser extent than the mesocortical dopamine system (Chrapusta *et al.*, 1997). This enhanced mesolimbic dopamine is also associated with increased coping and behavioural activation behaviourally (Cabib *et al.*, 2002). Interestingly, the mesolimbic and mesocortical systems react in opposite ways when stress is uncontrollable. When animals face sustained inescapable stressors, behaviourally, they show reduced coping, and neurally, they show reduced dopamine in the nucleus accumbens. However, administration of imipramine, a tricyclic antidepressant, just before the uncontrollable stressor, prevents this dopamine depletion (Rossetti *et al.*, 1993).

Inescapable stressors result in higher medial prefrontal cortex dopamine than when exposed to a similar stressor which is escapable (Cuadra *et al.*, 1999). However,

there is also evidence that this is not as straightforward. Giorgi *et al.* suggest that dopamine in the mPFC is not a function of stress, but rather the coping strategy in reaction to the stress. Rat lines that show proactive coping behaviours did have increased dopamine in the mPFC in response to stress induced by tail pinch, but this increase was not observed in another rat line that shows reactive coping (Giorgi *et al.*, 2003). Also, artificially increasing dopamine in the mPFC also increased active coping in mice (Wilke *et al.*, 2022).

Dopamine released in the mPFC inhibits its function, which includes regulation of the hypothalamic-pituitary-adrenal axis, involved in stress response (Maier *et al.*, 2006). Moreover, prefrontal cortex dopamine release also inhibits the release of dopamine in the nucleus accumbens. This is thought to be a mechanism by which the brain can regulate the balance between the mesocortical and mesolimbic dopamine systems (Del Arco and Mora, 2008). Thus, stress-induced mesocortical dopamine activity suppresses mesolimbic dopamine activity, which in turn leads to reduced coping and passive, helpless behaviour. This is supported by Cabib *et al.*, who showed that exposure to stress to an inbred susceptible strain of mice led to despair-like behaviour. This behaviour was also associated simultaneously with increased mesocortical dopamine system activity and decreased mesolimbic dopamine system activity (Cabib *et al.*, 2002). Ventura *et al.* showed that in the same susceptible strain, this despair-like behaviour was reversed by both activating the mesocortical dopamine system as well as administering an antidepressant clomipramine (Ventura *et al.*, 2002).

Stress, and in particular uncontrollable stress, has been studied in the development of helplessness and subsequent stress-related disorders such as depression and anxiety. In fact, many pre-clinical models of depression and anxiety involve exposing animals to uncontrollable stressors. For example, the learned helplessness model of depression shows that when animals face inescapable stressors, it generalizes to other situations (Overmier and Seligman, 1967). Similarly, the chronic mild stress model of depression involves exposing animals to a series of mild, unpredictable stressors, which leads to anhedonia and other depressive-like symptoms (Katz *et al.*, 1981; Katz, 1982; Willner *et al.*, 1992).

The relation between stress and depression is also supported by clinical studies. For example, individuals with a history of childhood trauma or chronic stress are at increased risk of developing depression later in life. Moreover, stress is a common trigger for depressive episodes, and individuals with depression often report high levels of stress (Hammen, 2005, 2015). Stress has been linked with the development, severity, and relapse of major depression (Pizzagalli, 2014). In addition to severe stressors, chronic stressors, as well as events characterized by a perceived lack of control, inescapability, and humiliation, are linked to the risk of depression (Kendler *et al.*, 2003). Early life stress is also a risk factor for depression and is associated with deficits in reward learning (Min *et al.*, 2024) and usage of value information in decision-making (Smith and Pollak, 2022).

Overall, evidence points to ways in which uncontrollable stress impairs learning, as well as increases the likelihood of helplessness and depression. However, it is unknown to what extent the perceived uncontrollability of this stress contributes to the impairment of reward learning.

## Helplessness and Depression

Helplessness is a state of passivity and resignation that arises when individuals perceive that they have no control over their environment. It is marked by feelings of powerlessness, hopelessness, and a lack of motivation to act. Helplessness can be learned by exposure to uncontrollable stressors. Learned helplessness is a psychological phenomenon in which individuals learn to be passive and helpless in the face of adversity, even when opportunities for escape or change are present. It occurs when an individual repeatedly faces uncontrollable and adverse situations, eventually leading them to believe that their actions have no impact on outcomes. This belief in lack of control makes them less likely to try to change their situation, even in a new and controllable environment. This state of helplessness can have profound effects on mental health and well-being, leading to symptoms of depression, anxiety, and other stress-related disorders.

Learned helplessness was first described by Seligman and Maier in 1967 who observed that dogs exposed to inescapable electric shocks later failed to escape from shocks that they could have avoided (Seligman and Maier, 1967). This phenomenon was later found to exist in humans too, who exhibited similar patterns of passivity and resignation when faced with uncontrollable stressors (Hiroto and Seligman, 1975). The learned helplessness model of depression posits that exposure to uncontrollable stressors can lead to a state of helplessness and anhedonia, which are core symptoms of depression. For example, a study in genetically prone rats found that helplessness reduced preference for sucrose, demonstrating a classic diminished hedonic response (Sanchis-Segura *et al.*, 2005). This model has been widely used in preclinical research to study the neurobiological mechanisms underlying depression because of its high face and predictive validity, as well as decent construct validity. The model has big translational value and has helped with the development of treatments such as new antidepressants (Vollmayr and Gass, 2013).

Given that helplessness is learned, and that intentional mental content is important in mood disorders like depression and anxiety, there has been a lot of interest in cognitive processes that are disturbed in these disorders. Using signal detection analyses, Joormann and Gotlib found negative cognitive biases in individuals with depression, who required a greater intensity of emotional expressions in morphed faces to detect happiness relative to controls (Joormann and Gotlib, 2006). Going beyond just perceptual deficits, another study found impaired working memory in depressed individuals when estimating the probability of fractal stimuli (Rupprechter *et al.*, 2018). This is consistent with another finding that working memory capacity is reduced in individuals with depression (Snyder, 2013). Anhedonia, a cardinal

symptom of depression, is characterized by reduced motivation and willingness to work for rewards. In their now widely used EEfRT task, Treadway and colleagues found that MDD patients were willing to put in less effort to obtain rewards compared to healthy controls (Treadway *et al.*, 2012a). In another study using the same task, they found that dopamine levels in healthy people in the striatum and ventromedial prefrontal cortex correlated with effort expenditure, particularly in low reward trials (Treadway *et al.*, 2012b). Using an apple-gathering task and individually calibrated effort levels, a recent study has also found that both current and remitted depressed patients were willing to expend less effort for rewards compared to healthy controls with and without a family history of depression (Valton *et al.*, 2024).

Reward learning deficits have been widely observed in depression. For example, a meta-analysis by Huys and colleagues using a computational model-based approach found that depressed individuals showed reduced reward sensitivity when performing probabilistic reward learning tasks (Huys *et al.*, 2013). This is consistent with the finding that striatal activity in response to reward is reduced in depression (Steele *et al.*, 2007). Reduced prediction error signals were also observed in the striatum and midbrain, with the signals in the caudate and nucleus accumbens correlating with depression severity (Gradin *et al.*, 2011). However, this seems to be restricted to prediction error signals in learning, as reward prediction error signals in a probabilistic task that did not have a learning component showed no reduction in depressed patients (Rutledge *et al.*, 2017). Furthermore, both unmedicated and medicated MDD patients also had blunted temporal difference error signals in the ventral striatum (Kumar *et al.*, 2008). Another meta-analysis found small to medium reward processing deficits in depression compared to healthy controls across forty-eight studies (Halahakoon *et al.*, 2020). Reduced reward learning has also been found to predict treatment outcomes eight weeks later in MDD patients, even after controlling for initial depression levels (Vrieze *et al.*, 2013). Since depression is heterogeneous, it is important to consider the relation of learning with individual symptoms. Along that route, Brown *et al.* found that anhedonia (as measured from the MASQ subscale) was associated with reduced learning rates and that it mediated the relationship between striatum expected value and prediction error signals (Brown *et al.*, 2021). Moreover, symptom improvement after twelve weeks of cognitive behavioural therapy treatment correlated with learning rate improvement as well.

As described, depression is linked to a variety of cognitive and behavioural deficits, which could arise from various neurocomputational mechanisms. In fact, researchers have talked about general deficits in overall executive function (Bredemeier *et al.*, 2016) as well as cognitive control (Grahek *et al.*, 2018). A possible mechanism to explain some of the impairments is reduced cognitive and behavioural flexibility, an inability to appropriately adapt behaviour when shifts in environmental reward structures give rise to higher-level uncertainty. In a study by Murphy *et al.*, depressed individuals showed reduced cognitive flexibility in a dynamic go/no-go task, especially when the stimuli were emotionally arousing (Murphy *et al.*, 2012). MDD patients showed reduced shifting to maximise rewards in various versions of

the Iowa Gambling Task (Must *et al.*, 2013). Cognitive flexibility was also found to be inversely correlated with emotional regulation in depression (Gao *et al.*, 2025).

Adapting to changing reward contingencies and learning under uncertainty are some key aspects of flexibility. Probabilistic reversal learning tasks are often used to study these aspects, as they require participants to learn under second-order uncertainty. Robinson and colleagues showed that unmedicated MDD patients displayed impaired accuracy after unexpected rewards on reversal trials, which correlated with activity in the striatum (Robinson *et al.*, 2012). Another recent study found that depressed individuals showed reduced learning rates, were slower to adjust to reversals, and also displayed lower sensitivity to both rewards and punishments (Mukherjee *et al.*, 2020). Depression and anxiety are often comorbid and also showed similar deficits in adjusting to volatility regardless of outcome valence (Gagne *et al.*, 2020).

Thus, learned helplessness and depression are closely related, with both conditions characterized by cognitive and behavioural deficits in reward learning and decision-making. One of these deficits common in depression and anxiety is reduced flexibility in adjusting to changing reward structures. However, the extent of these deficits in helplessness is unknown.

## Controllability

Controllability, or perceived control, refers to an individual's subjective belief in their ability to influence events, actions, or outcomes in their environment. The concept has been studied in psychological literature for a long time, with different theories and conceptualizations. For example, Rotter's locus of control theory posits that individuals can be classified as having either an internal or external locus of control, depending on whether they believe that outcomes are determined by their own actions or external forces (Rotter, 1966). Similarly, Bandura's self-efficacy theory suggests that individuals with high self-efficacy are more likely to believe that they can control their environment and achieve their goals (Bandura, 1977). Intuitively, these theories more or less refer to the same underlying concept of perceived control.

Controllability is an important factor in one's everyday life, influencing motivation, well-being, and mental health. The perception of control has been linked to various positive outcomes, such as increased self-esteem, reduced stress, and better mental health (Skinner, 1996). The reasons for this are threefold. Firstly, controllability in itself is rewarding. This reflects White's theory of effectance motivation, which states that individuals have a fundamental psychological need to influence their environment through their actions (White, 1959). This is supported by behavioural studies that show both animals and humans prefer choice over no-choice, even if it requires more effort and does not lead to better outcomes (Leotti *et al.*, 2010). This preference for choice was associated with increased BOLD response in the ventral striatum, in both positive and negative outcomes (Leotti and Delgado, 2014).

Secondly, controllability dampens the negative effect of stressors, generalizable across different contexts. Animal studies show that the ability to exert control over a negative event in the environment not only blunts the impact of that event but also has longer-lasting effects, which blunt the negative impact of aversive events experienced later on. Maier suggests that a circuit from the ventromedial prefrontal cortex to the dorsomedial striatum is involved in detecting the controllability of the environment (Maier, 2015). This circuit then suppresses the activity of the amygdala and dorsal raphe nucleus, which are usually involved in stress response. Lastly, the ability to exert control positively affects subsequent decision-making and learning. Karsh and Eitam showed that participants with higher judgements of self-agency in a task increased both the speed and frequency of actions being performed (Karsh and Eitam, 2015). Animals also show proactive behaviour such as increased social exploration, decreased freezing, and improved learning upon experiencing controllable shocks (Moscarello and Hartley, 2017).

On the flip side, uncontrollability, or the perception of lack of control, can have detrimental effects on both affect and learning. For example, inescapable stressors impair fear extinction learning after fear conditioning in humans and also show increased fear expression afterwards (Hartley *et al.*, 2014). As described before, uncontrollable stressors are a major factor in the development of learned helplessness, which itself is a model of depression. A study showed that in conditions with low controllability, people shift from instrumental learning to Pavlovian, consistent with the idea that situations with low control gain no benefit from the flexibility that instrumental learning offers (Dorfman and Gershman, 2019). This is similar to what is observed in animals, who shift from proactive to reactive behaviour when exposed to uncontrollable stressors (Moscarello and Hartley, 2017). Parallely, it is observed that in depression, there is a shift from goal-directed to habitual behaviour, which is consistent with the idea that depression is a state of learned helplessness. Perceived control is not linked to just depression, though. Perceived control is present across anxiety disorders (Gallagher *et al.*, 2014a). Moreover, anxiety patients demonstrated improvements in perceived control after cognitive behavioural therapy (Gallagher *et al.*, 2014b). A study in our lab demonstrated that perceived uncontrollable stress during the COVID-19 pandemic predicted deficits in reversal learning in the form of greater probabilistic errors after negative feedback, an effect mediated by state anxiety (Guitart-Masip *et al.*, 2023). Lack of control then seems to be linked to depression as well as anxiety.

To summarize, a lack of controllability is known to have negative effects on affect and some forms of learning. Stress, and in particular uncontrollable stress, causes impairments in learning and decision-making, and also leads to helplessness and subsequent stress disorders like depression and anxiety. One of the ways in which learning is impaired in these disorders is behavioural flexibility, the ability to adapt to changing reward contingencies. We do not know whether perceived uncontrollability, an upstream factor in stress disorders, also impairs flexibility in the same way. In the present study, we have aimed to answer precisely this question: what is the impact of perceived lack of control of reward learning? To this end, 55 healthy participants

played a multidimensional probabilistic reversal learning task in an MRI scanner, where they had to learn to choose the correct stimulus based on reward feedback. We manipulated task controllability in the form of controllable and uncontrollable games to see how it affected reward learning. We formulated the following hypotheses on how controllability and outcome valence affect reward learning:

- Lack of control impairs learning after reversal
- Threat of shock impairs learning after reversal
- Threat of shock exacerbates the effect of uncontrollability on learning
- Mean BOLD activity associated with stimulus presentation in the orbitofrontal cortex (OFC) and/or striatum differs by controllability and/or goal outcome valence
- Mean BOLD activity associated with the expected value of choice at the time of stimulus presentation and reward feedback in the OFC and/or striatum is reduced in uncontrollable games relative to controllable ones
- The degree of attenuation of the expected value signal is greater in shock games relative to bonus games
- No differences in the mean BOLD activity associated with reward outcome at the time of reward feedback in the OFC and/or striatum by controllability
- Differences in voxel patterns in the OFC and/or striatum, quantified by voxels able to predict better than chance the condition by training support vector machines on neural voxel patterns.

# Chapter 2 Methods

Fifty-five healthy participants took part in the study, during which they completed a series of reward learning tasks inside an MRI scanner. In this task, we manipulated two key factors: controllability (whether participants' actions directly influenced goal progress) and outcome valence (playing for a monetary bonus versus avoiding an electric shock). This design allowed us to probe how perceived control and motivational context affect reward-based learning under conditions of uncertainty, including periods of unexpected target reversals. The following sections detail the task design, data collection, and analysis pipelines used to disentangle the behavioural, computational, and neural mechanisms underlying these effects.

## Task Design

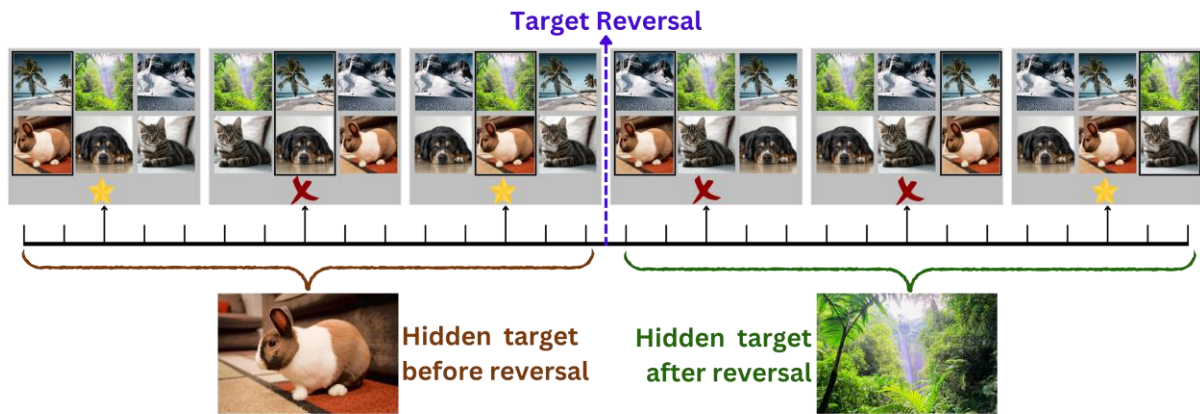
Adapted from Niv and colleagues, participants performed a three-armed bandit task, in which they had to choose one of the three composite stimuli on the left, middle, and right on a screen (Leong *et al.*, 2017). Each composite stimulus consisted of two images, an animal (dog, cat, or rabbit) and a landscape (beach, forest, or mountain). This resulted in a total of six images arranged in a grid of two rows and three columns. The three columns represented the three options, and each row consisted of the same type of image (either animal or landscape). The images in each row were shuffled randomly each trial independent of the other row, ensuring the composite stimuli were different from trial to trial. For example, in one trial the rabbit may be paired with the beach to form a composite stimulus, and in the next trial the rabbit may be paired with the mountain (Figure 2). At one time, only one of the six images was the rewarding image, and the other five were non-rewarding. The participants did not know which of the six images was the rewarding image, and thus it was a hidden target.

Participants had to choose one of the three composite stimuli, and if the chosen option contained the hidden target, they received a reward in the form a gold star on the screen with a probability of 80%. If the chosen option did not contain the hidden target, they received no reward in the form of a red cross with a probability of 80%. At the beginning of each trial, participants saw a fixation cross for a random amount of time (400-700 ms sampled from a uniform distribution). Then, they saw the mentioned 2 by 3 grid of images for 2000 ms, during which they had to make a choice. After selecting one of the three composite stimuli, only their chosen composite stimulus remained on the screen, and they saw the outcome of their choice (feedback) for 1000 ms.



**Figure 1: Three-armed probabilistic reward learning task.** After a fixation cross, 3 composite stimuli were shown consisting of 2 images each. If the chosen composite stimulus contained the hidden target, positive feedback (gold star) was provided with an 80% chance. If the chosen stimulus did not contain the hidden target, negative feedback (red cross) was provided with an 80% chance. After every few trials, the goal progress bar was shown, visually presenting the overall progress for that game.

In the middle of each game, the hidden target randomly switched to another image. This switch was not announced to the participants, and they had to learn the new target by trial and error. This target reversal occurred randomly once per game, between trials 13 and 17. The reward feedback introduces a first order of uncertainty because of its probabilistic nature. 20% of the time, participants received a reward when they chose the non-rewarding option, and 20% of the time they did not receive a reward when they chose the rewarding option. The hidden target reversal was associated with the second order of uncertainty, which occurred randomly and was not signalled to the participants. Thus, along with a multidimensional stimulus space, multiple orders of uncertainty in the task made it a complex learning environment. For every gold star, participants received one point. Within each game, they were playing not only to earn as many points as possible but also to reach a goal of twenty-one points. After every few trials, participants' goal progress for that game was displayed on screen in the form of a goal progress bar. This horizontal progress bar is shown for 1500 ms. The goal progress bar was empty at the start of a game, indicating no progress. As the participants earned points, the bar filled up from left to right with a yellow colour. When participants reached twenty-one points, the bar completely filled with green colour, indicating that they had reached their goal and the game ended.

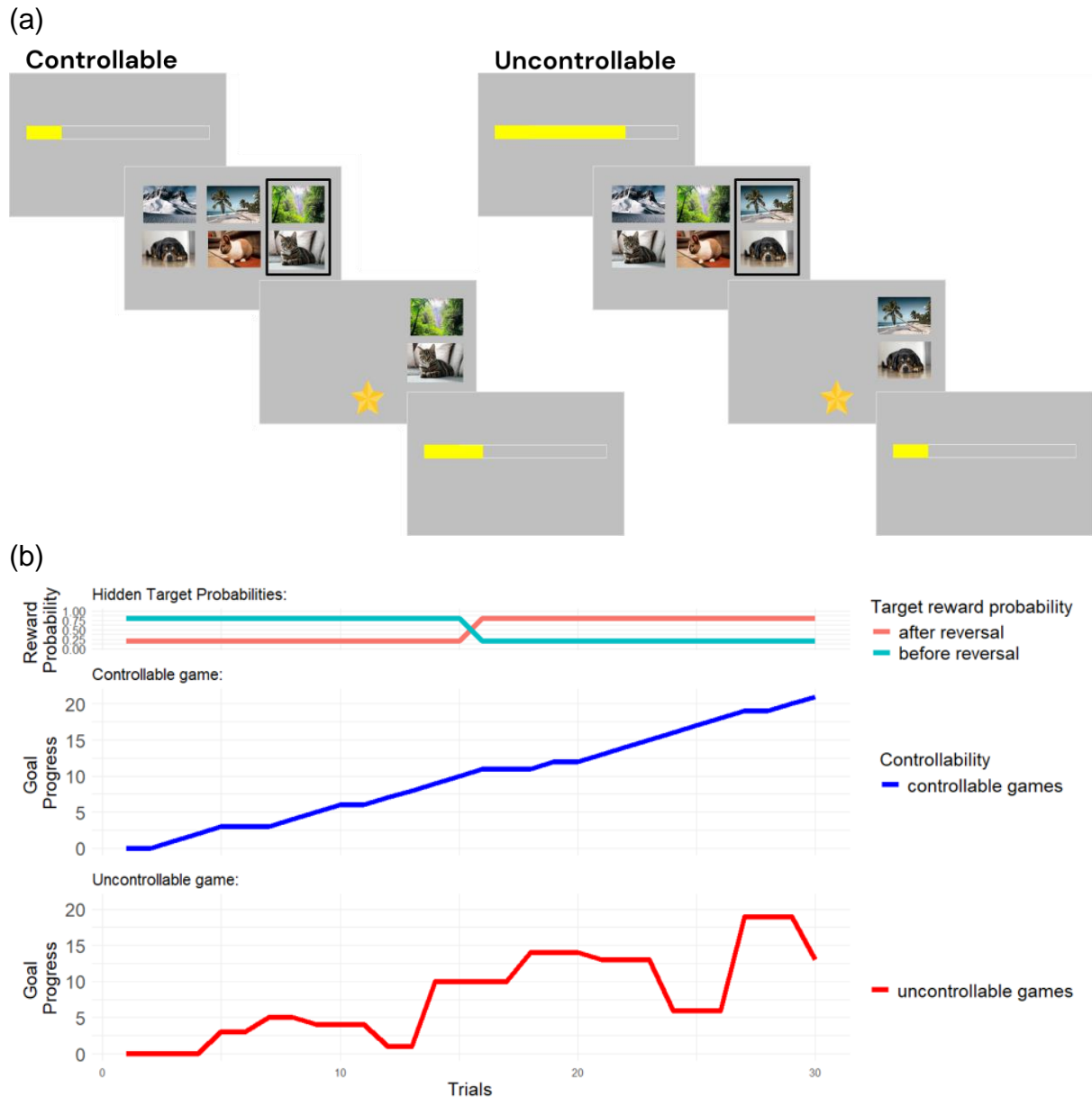


**Figure 2: Hidden target reversal.** In about the middle of each game, the hidden target changed randomly without warning. Participants had to flexibly identify the occurrence of this reversal and learn the new target.

The goal progress bar was shown at specific intervals throughout the game to provide feedback on overall progress of the game. In the initial phase of each game, during the first eight trials, the goal progress bar was displayed every 2-3 trials. As the game progressed and participants accumulated more points, the intervals between progress bar appearances increased slightly. Between trials 9-19, the goal progress bar was shown every 3-4 trials. From trial 20 onwards, the goal progress bar was displayed every 4-5 trials. These structured appearances of the goal progress bar provided periodic but non-continuous feedback, ensuring that participants could monitor their progress without the feedback being overly frequent or distracting.

A key experimental manipulation in the task was controllability, which determined whether participants' performance directly influenced their game success. In controllable games, participants' reward gains directly contributed to goal progress, meaning that their choices had a tangible and predictable impact on whether they would reach the goal of 21 points and complete the game successfully. Each time they received a reward, the goal progress bar would incrementally fill with 1 point, allowing them to actively work toward their goal.

In contrast, in uncontrollable games, participants' reward performance was not directly tied to their goal progress. Instead, the goal progress bar changed in a pseudo-randomized manner, independent of their actions. The first two goal progress updates in uncontrollable games were always set to zero, to quickly establish for participants that their choices were not influencing their goal progress. Subsequently, the next four progress updates varied randomly between values of 1-5, meaning that progress was unpredictable and could even temporarily decrease. The next eight progress updates varied between 6-15, introducing larger fluctuations in progress. Finally, the last four progress updates varied between 16-20. This structure ensured that while participants in uncontrollable games still experienced progress toward their goal, they did not have agency over how quickly or effectively they progressed, introducing an element of uncontrollability.



**Figure 3: Controllability manipulation.** (a) In controllable games (left), the goal progress bar directly reflected the points accumulated till that trial. In uncontrollable games (right), the goal progress was independent of the points accumulated. (b) Example trajectories of goal progress in controllable and uncontrollable games. The top graph shows the switched reward probabilities of the hidden targets before and after reversal. The middle graph (blue) presents an example goal progression in controllable games, where obtaining a gold star adds 1 point to the goal progress, which incrementally fills up the bar. The bottom graph (red) shows an example goal progression in uncontrollable games, where the goal progress is pseudorandom and independent of gold stars accrued. Note that the goal progress bar was shown to participants every few trials only.

Another crucial experimental manipulation involved the valence of goal outcome condition, specifically the presence of a shock in certain games. In shock games, participants were playing to avoid receiving an aversive but not painful electrical shock at the end of the game. The intensity of this electric shock was individually calibrated to be uncomfortable and aversive but not painful. If they successfully

reached the goal within a shock game, they avoided the shock entirely. However, if they failed to reach the goal, they would receive a sequence of three mild shocks during the break between games. In contrast, in bonus games, participants were playing to earn an additional monetary reward. If they successfully reached the goal in a bonus game, they received an extra payment of 10 SEK (approximately 1 Euro). If they failed to reach the goal in a bonus game, they did not receive the additional monetary reward, although they still earned a base amount of 0.75 SEK per rewarded trial regardless of game type. Participants were told that they would earn this base rate for all games regardless of the condition or whether they reached the goal in games. Combined with the fact that the probabilistic feedback was also consistent across controllability and outcome valence, the low-level reward learning and incentive to perform was the same for all conditions because the value and information of obtaining were equal.

To investigate the interplay between controllability and goal outcome, the study employed a fully crossed 2x2 within-subjects design, with four experimental conditions: controllable shock, uncontrollable shock, controllable bonus, and uncontrollable bonus. Inside the scanner, each participant played 20 games in total, with five games per condition. The 20 games were split across 5 runs in the scanner, each run having all 4 conditions in a random order. Participants were not explicitly informed before each game whether it was controllable or uncontrollable, requiring them to infer the nature of the game based on how their goal progress bar changed over time. They were informed before the start of each game whether they were playing to win a bonus or avoid an electric shock. The order of the games was randomized for each participant, ensuring that the sequence of games did not influence the results. The participants played five practice games at the beginning before going inside the scanner that incrementally introduced each element of the task and allowed them to become familiar with the framework.

The task was presented as a scenario in which participants acted as salespeople visiting a different fictional city for each game, where they attempted to sell pets (represented by animal images) and holidays (represented by landscape images). On every trial, they selected one pet and one holiday to sell, with each city favouring a particular pet or holiday (i.e., the target image) that would result in a more successful sale. A yellow star signified a successful sale while a red cross denoted an unsuccessful one. Participants earned a fixed rate of 0.75 SEK per sale, regardless of whether the game was a shock or bonus game, or whether it was controllable or uncontrollable. Controllable games were framed as cities in which the environment was well-organized and predictable, so their sales were reflected in their progress towards the goal. Uncontrollable games were framed as cities which were corrupt, so that even if they successfully made sales, goal progress was erratic and random. They were also informed that the city's preferred item would change unexpectedly at some point during each game. After each game, participants received feedback detailing the target items, their total sales, and whether they had met the criteria to win the bonus or avoid the shock.

The number of trials in each game varied naturally, depending on when the target reversal occurred (randomly between trial 13 to 17), ensuring that pre- and post-reversal phases were balanced. Additionally, the length of some games was modified to ensure an equal number of successful and unsuccessful games per condition. In addition to a fixed base rate of 0.75 SEK per correct sale, a second design strategy was employed to ensure motivation remained consistent across conditions: a balanced win/loss schedule that ensured each condition contained three successful games and two unsuccessful games. In order to ensure this equal number of successes, the number of trials in a game was either increased or decreased. If a game was needed be unsuccessful to equal the success rate, it was cut short just before the participant would have reached the goal. Conversely, to ensure success in games that were needed to be successful, the number of trials was extended beyond the natural stopping point, and reward probabilities were adjusted such that the probability of receiving a reward for choosing the target stimulus increased from 80% to 90%, while the probability for non-target stimuli decreased from 20% to 10%. Piloting indicated that participants did not notice these manipulations.

Uncontrollable games were yoked to a previous controllable game in the same goal condition (shock or bonus) to match them in length and success as closely as possible. The first uncontrollable game was yoked to the controllable practice game, and from then on, each uncontrollable game was yoked to the previous controllable game within the same condition from the previous run. If the yoked controllable game contained an increased reward probability, the uncontrollable game also had its probability increased at the same trial point.

After completing the main experimental phase, participants played two additional shock games outside the scanner—one controllable and one uncontrollable. These additional games were cut short before reaching the goal of 21 points. Following each game, participants rated on a 4-point scale how controllable, frustrating, and motivating they found the game, as well as how unpleasant they found the shock. These post-task ratings provided further insights into the subjective experience of control and its emotional consequences. The task was administered and behavioural data was collected using PsychoPy2 (Peirce *et al.*, 2019).

## Collected Data

At the end of the session, participants also completed questionnaires assessing their anxiety and depression. To assess state and trait anxiety, participants completed the State-Trait Anxiety Inventory (STAI). The STAI consists of two 20-item scales, one measuring state anxiety (how anxious participants feel at the moment) and the other measuring trait anxiety (how anxious participants generally feel). The forty items rated on a 4-point Likert scale, ranging from "not at all" to "very much so". To assess depression, participants completed the Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 is a 9-item scale that assesses the severity of depressive symptoms over the

past two weeks. These items were also on a 4-point Likert scale from "never" to "almost every day".

Other than MRI, the recorded variables included behavioural data, eye-tracking data, and physiological data. Behavioural data consisted of accuracy, coded as 1 for correct choices and 0 for incorrect choices, with non-responses excluded from analyses. Response times were recorded in milliseconds, along with the chosen composite stimulus for computational models. Eye-tracking data was collected using a vpiix dataPiiX (VPiiX Technologies, Canada) eye-tracker at a sampling rate of 2000 Hz, including pupil diameter (measured in pixels) and x and y gaze coordinates (also in pixels). Lastly, physiological data encompassed respiration, which was measured using a belt around the waist, and heart rate, which was monitored using a pulse oximeter attached to the index finger. These measures were obtained using Biopac hardware, and AcqKnowledge software (Biopac Systems Inc., USA).

MRI data were acquired using a Siemens Prisma 3T scanner outfitted with a head-neck 20-channel coil. Structural imaging was performed using a T1-weighted Turbo Flash sequence. The protocol employed a repetition time (TR) of 2.3 seconds and an echo time (TE) of 2.98 milliseconds, with an inversion time set at 0.9 seconds. A flip angle of 9° was used to optimize the signal, and images were acquired with a slice thickness of 1 mm. The image voxels were isotropic, with a resolution of 1 mm<sup>3</sup>. The dimensions were 256 voxels along the j and k axes, and 208 voxels along the i axis. This resulted in a field of view of 256 x 208 x 256 mm<sup>3</sup>.

Functional imaging, designed to capture blood-oxygen-level-dependent (BOLD) contrasts, was conducted with a repetition time of 1.86 seconds and an echo time of 30 milliseconds. A flip angle of 70° was chosen for the BOLD sequence. 62 slices were obtained within a volume, for which a slice thickness of 2.2 mm was maintained. The order of slice acquisition was a simultaneous multi-slice interleaved sequence, in which it used a multiband approach to acquire a pair of slices simultaneously while the slices within each of the two bands are acquired in an interleaved (odd-even) order. The phase encoding direction was set to j-. The image voxels were anisotropic, with a resolution of 2.234 x 2.234 x 2.2 mm<sup>3</sup>. The image was 94 voxels along the i and j axes, and 62 voxels along the k axis. This resulted in a field of view of 210 x 210 x 136.4 mm<sup>3</sup>.

## Behavioural Data Analyses

Behavioural data were analysed using a general linear mixed-effects model using R (R Core Team, 2024) in RStudio (Posit team, 2024) with the lme4 package (Bates *et al.*, 2015). The model included fixed effects for trial number, controllability, goal outcome valence, reversal, and all possible interactions. Subject was included as a random effect with different intercepts and slopes for trial number to account for individual differences in learning. The dependent variable was accuracy, coded as 0 for incorrect responses and 1 for correct ones. Controllable games were coded as 0.5, while uncontrollable games were coded as -0.5. Bonus games were coded as

0.5, while shock games were coded as -0.5. Similarly, trials before target reversal were coded as 0.5, while those after reversal were coded as -0.5. Trial number was scaled to be centred around 0 with a variance of 1.

The model was fit using maximum likelihood estimation, and the significance of fixed effects was assessed using Bayesian Information Criteria. Post-hoc comparisons were conducted using the emmeans package (Lenth, 2024) with Tukey adjustments for multiple comparisons. To explore interactions, post-hoc linear mixed models on subsets of data were also conducted. Data preprocessing and cleaning was conducted using the tidyverse package (Wickham *et al.*, 2019), and visualizations were created using ggplot2 (Wickham, 2011).

Subjective rating of controllability, obtained after playing the extra games at the end outside the scanner, was also included as a regressor in some models to explore its effect on behaviour. The controllability rating for the uncontrollable game was subtracted from the rating of the controllable game to create a difference score. This difference score was then scaled to have a mean of 0 and a variance of 1, which was then included as a regressor in the model. State anxiety, trait anxiety, and depression scores also included as regressors in some models to explore their mediation on accuracy. Positively worded items were first reverse coded, so that higher scores represented negative affect. These scores were then scaled to have a mean of 0 and a variance of 1. The sum for each scale was then included as regressors in the model.

## Computational Models

We fitted a series of computational models from two main families to the observed choices in the behavioural data: reinforcement learning models and hidden Markov models. The best fitting model was used to derive expected values for the chosen option on each trial. These expected value estimates were then used as regressors in some parts of the fMRI analysis.

Reinforcement learning models were used to capture the learning process in the task. These models update the expected value of an option based on the prediction error, or the difference between the received outcome and the expected outcome. In these models, individuals learn to assign each of the six images with a value, which is updated each trial based on the reward feedback. For each composite stimulus, the values of both images are linearly added to determine the value of an option at each trial. This process is known as feature learning. The basic reinforcement learning model includes a learning rate parameter  $\alpha$ , which determines the extent to which the expected value was updated on each trial based on prediction errors. The support for this parameter is from 0 to 1. an  $\alpha$  of 0 indicates no learning, meaning that there is no updating of feature value based on feedback and prediction error. On the other hand, an  $\alpha$  of 1 means that the model is overly reliant on rewards, and the model wipes out previous value estimates and replaces them with just the current reward outcome. The second parameter is the reward sensitivity parameter  $\beta$ , which

determines the extent to which the model is sensitive to reward feedback and the randomness of choice. The support for this parameter is from 0 to infinity. A  $\beta$  of 0 makes the model not take into account values and makes the choices fully random, while a  $\beta$  approaching infinity indicates that the model is fully deterministic in choices. The observation part of the model includes a simple softmax function, which converts the expected values into choice probabilities. The  $\beta$  is commonly an inverse temperature parameter of the softmax function, but parameterizing it as reward sensitivity outside the softmax in the learning model is equivalent.

Suppose on trial  $t$  a subject is presented with 3 options. Each option  $j$  (with  $j = 1, 2, 3$ ) is composed of 2 stimuli. Let the indices of the stimuli for option  $j$  be given by  $s_{1j}$  and  $s_{2j}$ . The value of option  $j$  is computed as the average of the value of the 2 stimuli:

$$q_j(t) = \frac{1}{2} [Q_{s_{1j}}(t) + Q_{s_{2j}}(t)]$$

The Q-value is updated each trial. For each stimulus  $i$  that is part of the chosen option  $j^*$ , the update rule is:

$$Q_i(t+1) = Q_i(t) + \frac{1}{2} \alpha \delta(t) \quad \forall i \in \{s_{1j^*}, s_{2j^*}\}$$

Where  $\alpha$  is the learning rate, a free parameter.  $\delta(t)$  is the reward prediction error for trial  $t$ . It is computed as:

$$\delta(t) = r_{eff}(t) - q_{j^*}(t)$$

Where  $q_{j^*}(t)$  is the value of the chosen option.  $r_{eff}(t)$  is the effective reward obtained at trial  $t$ . It is calculated as:

$$r_{eff}(t) = \beta \times r(t)$$

Where  $\beta$  is the reward sensitivity parameter, a free parameter. It is equivalent to the inverse temperature parameter used in RL models.  $r(t)$  is the reward obtained on trial  $t$ .

Finally, the decision rule uses a softmax function to convert option values  $q_j(t)$  to choice probabilities. The probability of choosing option  $j$  is:

$$P(j | t) = \frac{\exp(q_j(t))}{\sum_{k=1}^3 \exp(q_k(t))}$$

The first augmentation to this model is adding a second reward sensitivity parameter. This second beta comes into effect only on trials where goal progress bar is shown. If the progress has increased since the previous time it was shown, the second beta increases the value of the effective reward. This, in turn, updates the value of the chosen stimuli. The second augmentation to the simple model is adding a forget parameter. This parameter determines the rate at which the value of unchosen images decay back to 0.5. This parameter is important because it allows the model to forget about the value of unchosen images that are not relevant to the current

choice. Using all combinations of these 2 augmentations, we fit the resultant 4 reinforcement learning models to choice data.

The second class of models that we used are hidden Markov models. The model does not learn the value of each image like the reinforcement learning model. Instead, it estimates the probability of the hidden target being behind each image. Since the target is a hidden state which need to be inferred through reward feedback, hidden Markov models are suitable for this kind of estimation. The basic hidden Markov model works by updating the probability estimate  $\alpha_t(f_i)$  of each image  $f_i$  being the target on observing the feedback. At each trial  $t$ , the model maintains this belief state vector  $\alpha_t \in \mathbb{R}^6$  ( $0 \leq \alpha_t \leq 1 \forall t$ ) over all 6 images. The initial belief state for all images is uniform:

$$\alpha_0(f_i) = \frac{1}{6} \forall i \in \{1, 2, 3, 4, 5, 6\}$$

On each trial, the probability of each  $\alpha(f_i)$  is updated using a Bayesian update rule as follows:

$$\alpha'_t(f_i) = \alpha_t(f_i) \cdot L_i$$

Where  $L_i$  is the likelihood that image  $f_i$  is the hidden target. For the 2 images in the chosen option,  $L_i$  is computed as:

$$L_i = q \cdot r_t + (1 - q) \cdot (1 - r_t)$$

Where  $q$  is a free parameter. It represents the probability of receiving a reward when a chosen option includes the hidden target.

For the remaining 4 images that are not chosen, the likelihood is computed as:

$$L_i = (1 - p) \cdot r_t + p \cdot (1 - r_t)$$

Where  $p$  is also a free parameter. It represents the probability of receiving no reward when the chosen option does not include the hidden target.

Both  $p$  and  $q$  represent probabilities, and thus have support between 0 and 1.  $r_t$  is 1 if a reward is received on trial  $t$ , and 0 otherwise.

After updating with the likelihood, the vector of hidden states is normalized:

$$\alpha''_t(f_i) = \frac{\alpha'_t(f_i)}{\sum_{k=1}^6 \alpha'_t(f_k)} \forall i \in \{1, 2, 3, 4, 5, 6\}$$

The normalized vector is then multiplied by a transition matrix  $T$  that maps the probability of the hidden target changing from one image to another. As with any hidden Markov model, the transition matrix dictates the state evolution:

$$T = \begin{bmatrix} 1 - 5\tau/6 & \tau/6 & \tau/6 & \tau/6 & \tau/6 & \tau/6 \\ \tau/6 & 1 - 5\tau/6 & \tau/6 & \tau/6 & \tau/6 & \tau/6 \\ \tau/6 & \tau/6 & 1 - 5\tau/6 & \tau/6 & \tau/6 & \tau/6 \\ \tau/6 & \tau/6 & \tau/6 & 1 - 5\tau/6 & \tau/6 & \tau/6 \\ \tau/6 & \tau/6 & \tau/6 & \tau/6 & 1 - 5\tau/6 & \tau/6 \\ \tau/6 & \tau/6 & \tau/6 & \tau/6 & \tau/6 & 1 - 5\tau/6 \end{bmatrix}$$

Where  $\tau$  is a free parameter.

The probability then is finally calculated as follows:

$$\alpha_{t+1} = \alpha_t'' \cdot T$$

Then, the probability of a choice is made from these hidden states:

$$P(c) = \frac{\sum_c \alpha_t(f_c)}{\sum_{i=1}^6 \alpha_t(f_i)}$$

Where  $c$  is the composite stimulus consisting of 2 images.

Two additional parameters can be added to this basic model. These are  $p'$  and  $q'$ . These are analogous to  $p$  and  $q$ , but only come into effect on trials where the goal progress bar is shown.  $q'$  represents the probability of goal progress increment since previous update if chosen option includes hidden target, whereas  $p'$  represents the probability of not observing any goal progress increment since previous update if chosen option does not include the hidden target. Thus, adding these parameters gives us a total of 2 hidden Markov models.

These models were fit using the HBI toolbox (Piray *et al.*, 2019) in MATLAB (2022b). The toolbox uses hierarchical Bayesian inference, which estimates model parameters for the entire group along with individual parameter estimates for each subject. These estimates are used to create better priors and then estimated iteratively. The best model is selected by estimating the best-fitting model for each subject and then counting the most frequent model among subjects, giving each model's posterior exceedance probability.

The best-fitting computational model was then used to generate the expected value for the chosen option for each trial. This was used as a regressor in fMRI analysis in some generalized linear models. Note that population averages of each parameter were used to generate the regressor instead of using individual estimates, as this has been suggested to be more robust for fMRI analyses (Daw, 2011). This is because individual estimates can be noisy, and difference in parameters can result in large scaling differences in the beta estimates of individuals, making it difficult to draw group level conclusions.

## Neuroimaging Analyses

MRI data were pre-processed using fMRIPrep (v24.0.0) using the following command:

```
fmriprep ~/test_bids ~/fmriprep_output2 participant --nprocs 24 --omp-nthreads 24 --mem-mb 150000 --level full --output-spaces MNI152NLin2009cAsym:res-2 --return-all-components --verbose --resource-monitor --write-graph --notrack
```

fMRIPrep produces an automated pipeline based on the data, using the most advanced tools and methodologies for each step in the preprocessing. The anatomical T1-weighted image was first corrected for intensity non-uniformity, and then was skull-stripped. Brain tissue segmentation of cerebrospinal fluid, white matter, and grey matter were performed on this T1w image, after which brain surfaces were reconstructed using FreeSurfer. The estimated brain mask was then refined with cortex grey matter segmentations from the surface reconstruction. The T1w image was then normalized using nonlinear registration to the MNI152NLin2009cAsym space with a resolution of 2 mm<sup>3</sup>.

The five functional BOLD files per subject were processed as well. Six head motion parameters were first estimated, consisting of three translational and three rotational parameter estimates for each volume. Then, slice-timing correction was performed on the volume, with the middle slice as reference. The reference volume of each functional run was co-registered to the T1w reference using boundary-based registration, implemented with six degrees of freedom. Various confounding time-series were calculated in addition to the already estimated head-motion parameters. Framewise displacement and global signals within the cerebrospinal fluid, the white matter, and the entire brain were some of the important ones used later in analyses. Additionally, temporal derivatives and quadratic terms of the head motion parameters were also calculated and included as confounds. Finally, all transformations were performed in a single interpolation step to minimize interpolation error. This step consisted of head motion correction, and co-registration to anatomical and standard output space.

Since fMRIPrep does not perform smoothing on the normalized images, the images were smoothed outside of fMRIPrep using SPM12 in MATLAB (R2022b) with an 8mm<sup>3</sup> Gaussian kernel. Our data also had small periods of excessive head motion, particularly due to the administration of an electric shock. To correct this, we used ArtRepair on smoothed images, a toolbox in MATLAB that interpolates frames with excessive motion with neighbouring frames, effectively removing the small number of “bad” frames while keeping the run usable. These frames were de-weighted by a factor of 100 in first-level GLM analyses.

For multivariate analyses an alternative preprocessing pipeline was used because the analyses need co-registered but unnormalized images. Since fMRIPrep performs all resamplings in a single step, it is not possible to obtain it from fMRIPrep’s derivatives. We used SPM12 in MATLAB (R2022b) to perform this preprocessing. First, volumes were slice-time corrected, with the 32<sup>nd</sup> slice being set as the reference slice since it is the middle both spatially in a frame and temporally in the slice acquisition order. The images were then realigned to the mean and the

resliced, producing six realignment parameters consisting of three translational and three rotational parameters. Finally, the functional images were co-registered to the anatomical T1w image. These unsmoothed images in their native space were then fed to first-level generalized linear models (GLMs) for subsequent multivariate analyses.

To test the brain representation of controllability and its interaction with outcome valence, we performed four first-level GLMs, each testing different aspects of controllability. All GLMs were performed using SPM12 in MATLAB (R2022b).

**GLM1:** This GLM was used to test if a lack of control impacts activation in regions of the brain involved in reward processing (striatum) and decision-making (orbitofrontal cortex). We also wanted to see if this representation differed by goal outcome. There were four regressors of interest at the time of presenting the stimuli. These were stimulus onsets in all four types of games: bonus controllable, bonus uncontrollable, shock controllable, and shock uncontrollable. Each of the four regressors had a parametric modulator that indicated trials with increased reward probability. This binary modulator had a value of 1 for trials with increased reward probability, and 0 for trials before reward probability increased. This was done to account for the variance due to reward probability increase at the end of some games, and was not of interest for analyses.

There were other regressors for each event in the game, but were not of interest:

- Reward feedback: presentation of gold star
- No reward feedback: presentation of red cross
- No response feedback: presentation of message that says “too slow”
- Previous goal progress: presentation of the goal progress shown previously
- Goal updating – controllable games: goal progress bar that increases from previous goal incrementally based on performance
- Goal updating – uncontrollable games: goal progress bar that increases pseudorandomly
- End of game feedback: information such as earnings, success/failure, information about the hidden targets for-
  - Bonus games that were failures
  - Bonus games that were successes
  - Shock games that were failures
  - Shock games that were successes
- Anticipation of outcome: waiting for bonus or shock
  - Waiting for shock on lost shock games
  - Waiting for no shock on won shock games
  - Waiting for no bonus on lost bonus games
  - Waiting for bonus on won bonus games
- Actual delivery of outcome: bonus on screen or electric shock
  - Shock on lost shock games
  - No shock on won shock games
  - No bonus on lost bonus games

- Bonus on won bonus games
- Game instructions: presentation of an image of a city with a fake name
- Game instructions specific to bonus games
- Game instructions specific to shock games

Furthermore, nuisance regressors were also added. To account for head motion and other confounds, head motion parameters along with global signals were included. Together with the temporal derivatives and quadratic terms, a total of thirty six regressors were added. Eighteen physiological regressors derived from respiration and pulse data using the PhysIO toolbox (Kasper *et al.*, 2017) were also included. In the first level, contrast images for each regressor of interest for each subject was created. These contrast images were then used in second level analysis, a 2×2 fully crossed within-subjects ANOVA with controllability and outcome valence as the two factors. We then looked for significant activations of main effects of controllability and valence as well as their interaction on the whole brain. We also performed small volume correction for subthreshold activation clusters with the striatum region of interest (ROI) and the orbitofrontal cortex (OFC) ROI. The striatum ROI was defined by combining the bilateral Caudate and Putamen regions in the Automated Anatomical Labelling (AAL) atlas. The OFC ROI included bilateral frontal superior, inferior, middle, and medial OFC regions as well as the rectus.

**GLM2:** The goal of this GLM was to investigate whether the representation of value expectation during decision-making is affected by controllability or valence. For GLM2 (and GLM3), we used the best-fitting computational model on behavioural data to generate regressors for each subject and game. The regressors for GLM2 are an estimate of the value of the chosen option for each trial. These were used as parametric modulators for the same four regressors of interest, except this time we did not include trials with increased reward probability in them. Additionally, we also removed trials in which participants did not respond from the regressors from interest since there was no choice to estimate the value. Each onset of these four regressors had a parametric modulator representing the estimated value of the chosen option in that trial. For trials with increased reward probability, we created an additional 4 event regressors. These were not accompanied by any parametric modulator. We also added another regressor for trials in which there was no response. Otherwise, the remaining event and nuisance regressors were the same as in GLM1. Moreover, all five runs were concatenated into one long run for this analysis.

We created first-level contrast images for each subject and condition by weighing a condition's corresponding parametric modulator by 1. These contrast images were then fed to a 2×2 within-subjects ANOVA with controllability and valence as the factors. The positive effect of all conditions on the contrast estimates was then assessed across the whole brain using family-wise error correction. OFC and striatum ROI masks were used to get the mean activation across conditions within these regions. Binary inclusive masks were created for clusters showing significant activation after family-wise error correction. These masks were then used to obtain the mean activation in contrast images for each subject and condition. These mean

activations were then used in another 2×2 repeated-measures ANOVA to examine the significance of the main effects of controllability and valence and their interaction. Based on behavioural results we also included their subjective rating of controllability in the ANOVA.

**GLM3:** In GLM3, we checked whether the representation of the value of the chosen option at the time of reward feedback differed by controllability or valence. We also checked whether the representation of reward at the time of feedback was influenced by controllability or valence. GLM3 was specified similarly to GLM2, but instead of the four regressors of interest being stimulus presentation onsets, they were now reward feedback onset for each condition. The stimulus presentation onsets for each condition were collapsed to one regressor. The feedback onset regressors had a parametric modulator indicating feedback type (1 for reward, 0 for no reward). The regressors also had the same expected value parametric modulator, meaning that these regressors were associated with two parametric modulators. Since reward reception is known to elicit lots of activity, the modulators were orthogonalized so that only the variance unique to the expected value was assigned to the second parametric modulator.

Similar to GLM2, first-level contrasts were estimated for each subject and parametric modulator. These were then used in two 2×2 ANOVAs, one for reward outcome and one for expected value. Significant activations within the striatum and OFC ROIs were used to create binary masks. These masks were then used to extract mean activation of each subject and condition from contrast images, to be fed to a 2×2 ANOVA to investigate the main effects of controllability and valence on expected value and reward outcome representations.

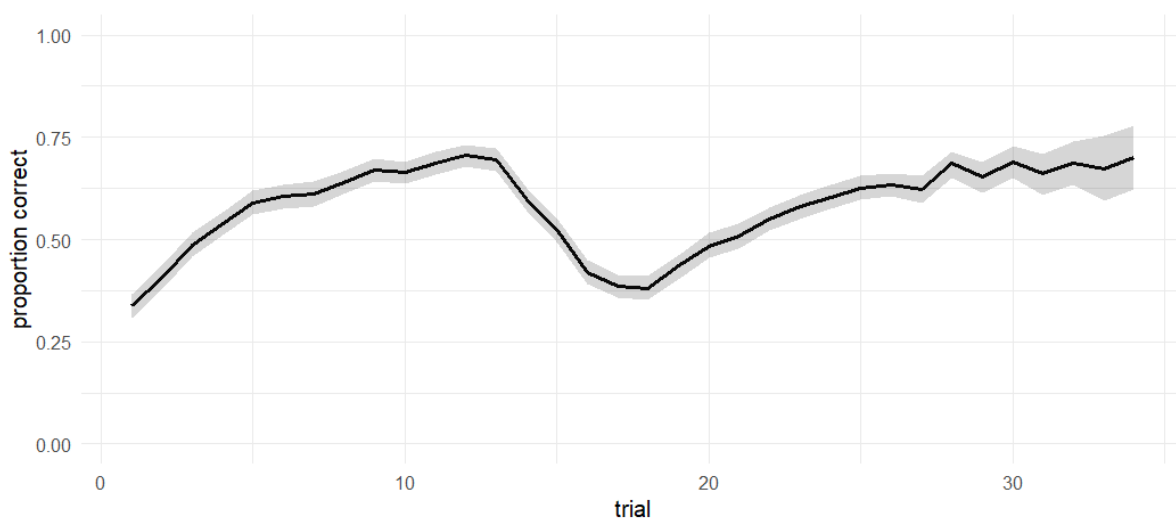
**GLM4:** The first three GLMs performed univariate fMRI analyses, but the last GLM's pipeline incorporated multivariate techniques such as multivariate pattern analysis (MVPA) to understand if there was a difference in the voxel patterns by controllability or valence. The specification for GLM4 was the same as GLM1, but was performed instead on unnormalized and unsmoothed images.

We used The Decoding Toolbox (Hebart *et al.*, 2015) to decode all four conditions from voxel patterns. We trained a linear support vector machine on beta maps generated from the first-level GLM from four of the five instances of each condition and cycled through all for a 5-fold cross validation. We performed searchlight analysis across the whole brain with a sphere with a radius of 3 voxels to generate brain maps, where the value of each voxel represents accuracy minus chance in the model's ability to accurately classify the condition. This map was then normalized and smoothed with an 8mm kernel for group level inferences. A one-sample t-test was then done on these normalized to detect voxels that were able to classify better than chance significantly across participants. Since participants were explicitly told whether they were playing for a bonus or avoid an electric shock, we also did a searchlight analysis to classify controllable from uncontrollable games, regardless of

outcome valence. Similarly, a one-sample t-test was then performed at the second level to see if any voxels were able to classify better than chance.

# Chapter 3 Results

To understand the effect of controllability on learning, we used a reward learning task with a hidden target reversal. As expected, average accuracy dropped after reversal, since participants had to identify that a reversal had occurred and re-learn a new target. Figure 4 shows the average accuracy across participants for each trial number, with a noticeable dip near the middle of the game.



**Figure 4: Average learning trajectory of the game.** Average accuracy for each trial across all conditions and participants. Shaded area is the standard error.

A lack of control is associated with helplessness and passive behaviour. Therefore, an important part of the task design was to provide correct feedback for each trial and match the number of wins across conditions. To ensure that the manipulation did not affect their motivation, we tested the difference in ratings of motivation and frustration in controllable and uncontrollable games. Paired Wilcoxon signed rank tests for both were insignificant ( $p = 0.232$  and  $p = 1$ ), confirming that any differences observed in reward learning were not due to affect or a lack of motivation.

## Lack of control impairs learning after reversal, but only in bonus games

The first question that we asked was whether a perceived lack of control will affect reward learning. Learning is most simply quantified as accuracy across trials. To answer this question, we ran a generalized logistic mixed model (GLMM) on accuracy as the dependent variable with trial number, target reversal, controllability, and outcome valence as fixed effects. We also included subject as a random effect, allowing for different intercepts for each subject and different slopes for trial to

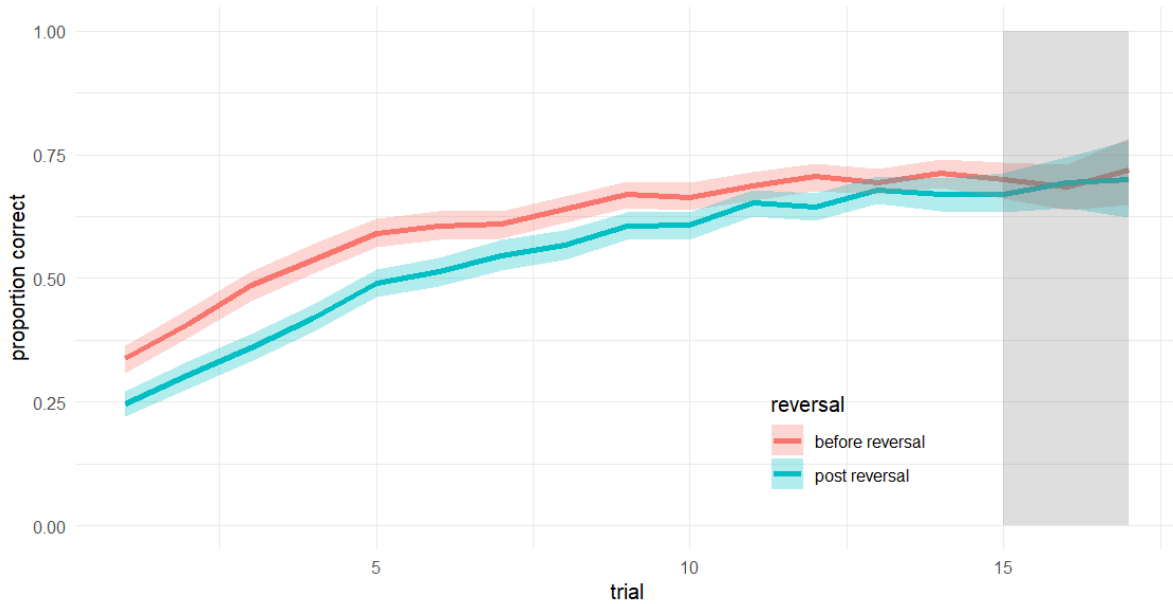
account for individual differences in learning. The results for the model are presented in Table 1.

Predictor	Estimate	SE	z-value	p-value	Lower CI	Upper CI
(Intercept)	0.204	0.044	4.650	3.32e-06	0.118	0.290
trial	0.956	0.031	31.068	6.52e-212	0.895	1.016
reversal	1.938	0.047	41.599	0.00e+00	1.846	2.029
controllability	0.108	0.046	2.324	0.020115	0.017	0.198
valence	-0.039	0.046	-0.849	0.395682	-0.130	0.051
trial:reversal	-0.169	0.046	-3.644	0.000268	-0.261	-0.078
trial:controllability	-0.050	0.046	-1.079	0.280484	-0.141	0.041
reversal:controllability	-0.146	0.093	-1.576	0.114991	-0.328	0.036
trial:valence	-0.051	0.046	-1.106	0.268754	-0.142	0.040
reversal:valence	-0.048	0.093	-0.520	0.603235	-0.230	0.134
controllability:valence	-0.030	0.093	-0.327	0.74372	-0.212	0.151
trial:reversal:controllability	0.198	0.092	2.142	0.032166	0.017	0.379
trial:reversal:valence	-0.094	0.093	-1.016	0.309695	-0.276	0.088
trial:controllability:valence	-0.192	0.093	-2.071	0.038379	-0.373	-0.010
reversal:controllability:valence	-0.749	0.185	-4.046	5.21e-05	-1.112	-0.386
trial:reversal:controllability:valence	-0.177	0.185	-0.959	0.337449	-0.540	0.185

**Table 1: GLMM results.** Table showing all main effects and interactions. The main effect of trial and reversal is highly significant. Two-way interaction of trial and reversal, and three-way interactions of trial  $\times$  reversal  $\times$  controllability and reversal  $\times$  controllability  $\times$  valence is significant.

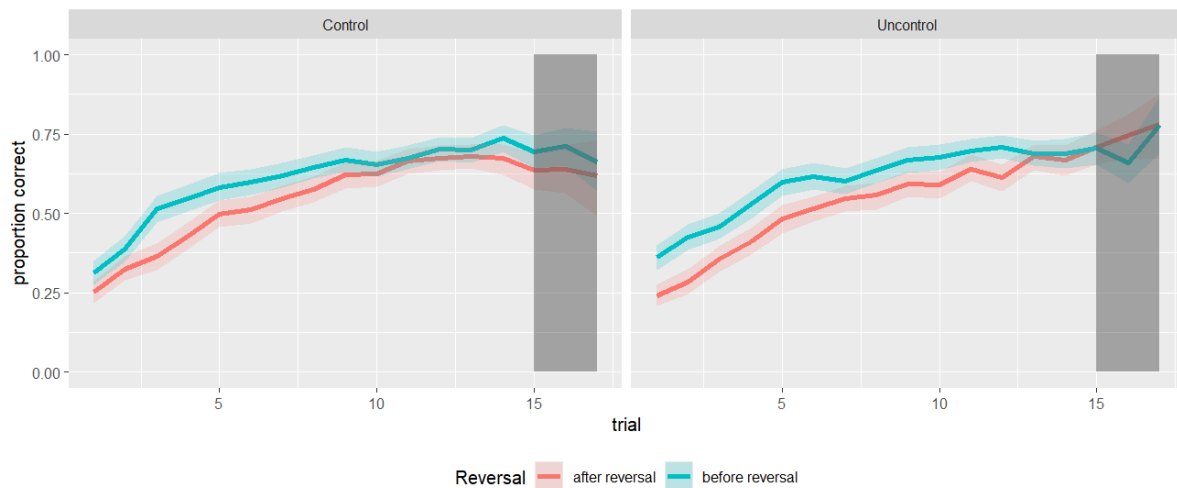
The main effect of reversal was quite strong, with a large effect size of 1.938 and a p-value smaller than machine epsilon. In terms of interactions, we did find a significant reversal  $\times$  trial interaction (estimate = -0.144,  $p = 2.68 \times 10^{-4}$ ). As shown in Figure 5, the negative estimate reflected that participants were slower to learn

after target reversals. This was expected since participants need to identify the occurrence of a reversal and relearn a new target.



**Figure 5: Learning curves separated by reversal.** Participants take some time to find the new hidden target, but reach similar levels of accuracy towards the end of each half of the game. Shaded areas are standard errors, and the shaded rectangle represents trials with smaller sample size due to variable game length.

We predicted that participants' reward learning would be impaired by a perceived lack of control. In the GLMM, this is represented by a negative controllability  $\times$  reversal and a trial  $\times$  controllability  $\times$  reversal interaction. The former represents a difference in accuracy after reversal in uncontrollable over controllable games, while the latter represents uncontrollability affecting learning to a greater degree after reversals. The controllability  $\times$  reversal interaction was not significant ( $p = 0.115$ ), while the controllability  $\times$  reversal  $\times$  trial interaction was significant ( $p = 0.032$ ). Follow-up estimated marginal trend contrasts revealed that in uncontrollable games, the difference in learning before and after reversal was significant ( $p = 3.0 \times 10^{-4}$ ), while in controllable games the difference was not significant ( $p = 0.704$ ). Post-hoc GLMMs on data split by control said the same thing. In uncontrollable games, the trial  $\times$  reversal interaction was significant (estimate =  $-0.270$ ,  $p = 4.5 \times 10^{-5}$ ), while the interaction was not significant in controllable games ( $p = 0.180$ ). As depicted in Figure 6, this indicates that lack of control impaired learning after reversal to a bigger extent.

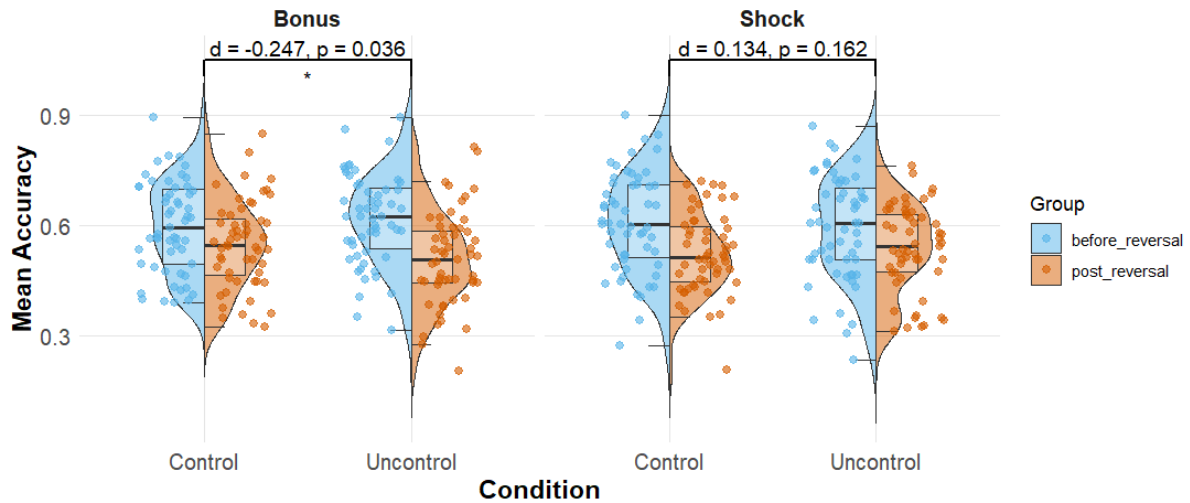


**Figure 6: Learning difference after reversal by controllability.** The learning curves for controllable and uncontrollable games are separated before and after reversal. The learning curve after reversal in uncontrollable games is lower than in controllable games. The shaded areas are standard errors, and the shaded rectangles represent trial numbers for which the sample size is smaller due to variable game lengths.

We also found a highly significant controllability x valence x reversal interaction (estimate =  $-0.748$ ,  $p = 5.21 \times 10^{-5}$ ). To investigate this three-way interaction further, we computed the estimated marginal means for the model. Pairwise contrasts of all possible combinations of conditions were calculated and adjusted for multiple comparisons using the Tukey method. All contrasts with differences across reversal were highly significant ( $p < 0.001$  for all) due to the strong main effect of reversal. However, the contrast of controllability in bonus games after reversal was also significant (estimate =  $-0.353$ ,  $p = 0.0023$ ). Moreover, the contrast of controllability in shock games after reversal was not significant ( $p = 1.000$ ), as well as the contrast of controllability in bonus games before reversal ( $p = 0.636$ ). Post-hoc GLMMs were conducted on data split by outcome valence. In bonus games, the reversal x controllability interaction was significant (estimate =  $-0.515$ ,  $p = 8.38 \times 10^{-5}$ ), while it was insignificant ( $p = 0.085$ ) in shock games. This indicates that there was a negative effect of lack of controllability specifically in bonus games after target reversal.

To confirm this, we conducted paired t-tests on mean accuracy data. First, we calculated the mean difference in mean accuracy before and after reversal in each condition for each subject. Then we split the data into controllable and uncontrollable games and conducted one-sided paired t-tests comparing the two conditions to see its interaction with reversal, quantifying reduction in accuracy after reversal by controllability. These tests were done on bonus and shock games separately to complete the three-way interaction investigation. In bonus games, the effect of controllability on reduction in accuracy after reversal was significant ( $p = 0.036$ ), a lack of control increasing the accuracy drop after reversal. Moreover, this effect was not present in shock games ( $p = 0.162$ ). As shown in Figure 7, the reduction in accuracy after reversal was exacerbated by a lack of control, but only in bonus

games and not shock games. Thus, the overall effect of controllability on learning after reversal, as seen in Figure 6, seemed to be driven by bonus games only.



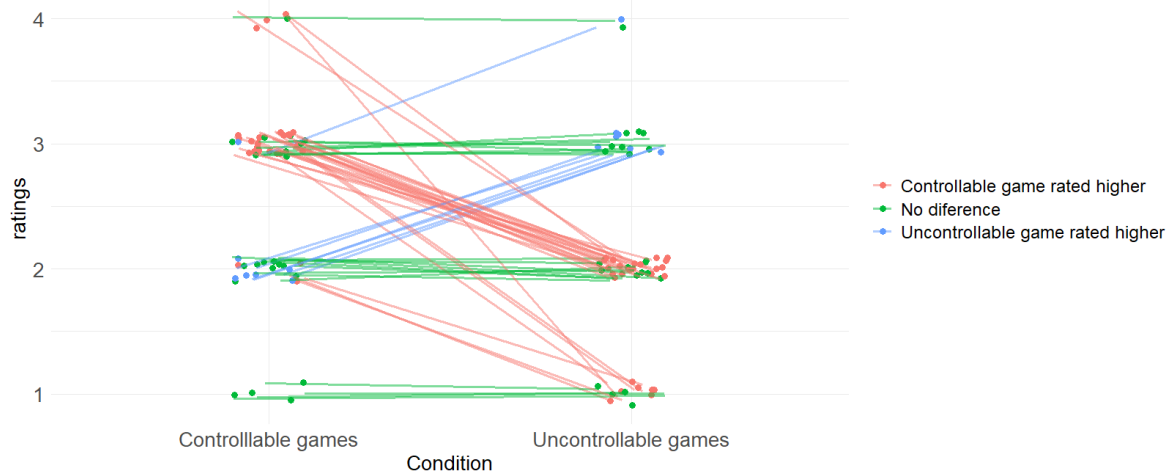
**Figure 7: Mean accuracy per subject for each condition.** Blue is for accuracy before reversal, while orange is for accuracy after reversal. The left two plots are of bonus games, while the right two plots are of shock games. The boxplot represents the median and upper and lower quartiles, and the points represent mean accuracy for each subject. The reduction in accuracy after reversal increases more in bonus uncontrollable games relative to bonus controllable games ( $p = 0.036$ ).

Our second hypothesis was that a threat of shock would impair learning, quantified by a valence  $\times$  reversal interaction with a negative effect size, indicating worse accuracy after a reversal in shock games versus bonus games. This interaction was not significant ( $p = 0.603$ ). We also predicted a positive valence  $\times$  trial interaction, implying worse learning overall due to the threat of shock. This interaction was not significant ( $p = 0.269$ ) too. Finally, we also predicted a three-way valence  $\times$  trial  $\times$  reversal interaction, indicating that the threat of shock would impair learning more after target reversal. This interaction was not significant ( $p = 0.310$ ).

Our next question was whether the effect of controllability differs when the goal outcome is avoidance of an electric shock versus gaining bonus money. We predicted that a threat of shock would exacerbate the effect of uncontrollability on learning, which would be reflected in a negative controllability  $\times$  valence interaction. This interaction was not significant either ( $p = 0.744$ ). We did find a marginally significant controllability  $\times$  valence  $\times$  trial interaction ( $p = 0.038$ ), but this significance disappeared depending on random effects specification. More specifically we also predicted that in shock games, uncontrollable conditions would impair learning after reversal more than in bonus games. This tentative four-way controllability  $\times$  valence  $\times$  reversal  $\times$  trial interaction was not significant ( $p = 0.337$ ).

## Subjective perception of control mediates the effect of uncontrollability

Not everyone perceived the controllability manipulation equally. As shown in Figure 8, there were differences in the subjective perception of controllability. These ratings were obtained in the two extra games (one controllable shock, one uncontrollable shock) that participants played at the end outside the scanner, where they rated how much they felt in control during that game from a scale of 1 to 4. A Paired Wilcoxon signed rank test showed that these controllability ratings were significantly different in the extra controllable game from the extra uncontrollable game ( $p = 0.0005$ ).



**Figure 8: Subjective ratings of controllability.** Ratings of controllability on a 4-point Likert scale, with a higher rating indicating a higher sense of control. Red lines indicate participants who experienced higher controllability in controllable games compared to uncontrollable games, while blue lines indicate participants who experienced higher controllability in uncontrollable games. Green lines indicate participants who felt the same sense of control in both conditions.

To account for this, we subtracted the self-reported score of sense of control in the uncontrollable game from the score in the controllable game, giving us each participant's subjective rating of control. A positive value indicates that a participant felt more in control in the controllable game relative to the uncontrollable game, with the magnitude quantifying the extent of the difference. A value of zero indicates that participants experienced equal controllability in both games, while a negative value indicates that a participant perceived to be in more control in the uncontrollable game. We note that the only negative value observed in the ratings was -1, while the positive values ranged from 1 to the maximum possible difference of 3. This was then added as a regressor to the same GLMM as above with this added main factor of controllability rating and all possible interactions. We note that this GLMM was not included in the main hypotheses and was exploratory.

Predictor	Estimate	SE	z-value	p-value	Lower CI	Upper CI
(Intercept)	0.204	0.044	4.661	3.15e-06	0.118	0.290
trial	0.960	0.031	31.049	1.19e-211	0.900	1.021
reversal	1.946	0.047	41.666	0.00e+00	1.854	2.037
controllability	0.109	0.046	2.348	0.01888	0.018	0.200
valence	-0.040	0.046	-0.865	0.38685	-0.131	0.051
rating	0.048	0.044	1.105	0.269199	-0.038	0.134
trial:reversal	-0.172	0.047	-3.699	0.000217	-0.264	-0.081
trial:controllability	-0.053	0.046	-1.140	0.254174	-0.144	0.038
reversal:controllability	-0.152	0.093	-1.638	0.101344	-0.334	0.030
trial:valence	-0.051	0.047	-1.089	0.276256	-0.142	0.041
reversal:valence	-0.049	0.093	-0.521	0.602341	-0.231	0.134
controllability:valence	-0.032	0.093	-0.346	0.729322	-0.214	0.150
trial:rating	0.091	0.031	2.881	0.003964	0.029	0.152
reversal:rating	0.239	0.048	5.021	5.15e-07	0.146	0.332
controllability:rating	0.115	0.047	2.425	0.015301	0.022	0.207
valence:rating	-0.002	0.047	-0.036	0.971096	-0.095	0.091
trial:reversal:controllability	0.201	0.093	2.172	0.029853	0.020	0.383
trial:reversal:valence	-0.097	0.093	-1.043	0.296974	-0.279	0.085
trial:controllability:valence	-0.190	0.093	-2.046	0.040779	-0.371	-0.008
reversal:controllability:valence	-0.743	0.186	-4.001	6.31e-05	-1.107	-0.379
trial:reversal:rating	0.047	0.048	0.985	0.324815	-0.047	0.142
trial:controllability:rating	-0.057	0.048	-1.192	0.233397	-0.151	0.037
reversal:controllability:rating	-0.250	0.095	-2.639	0.008321	-0.435	-0.064

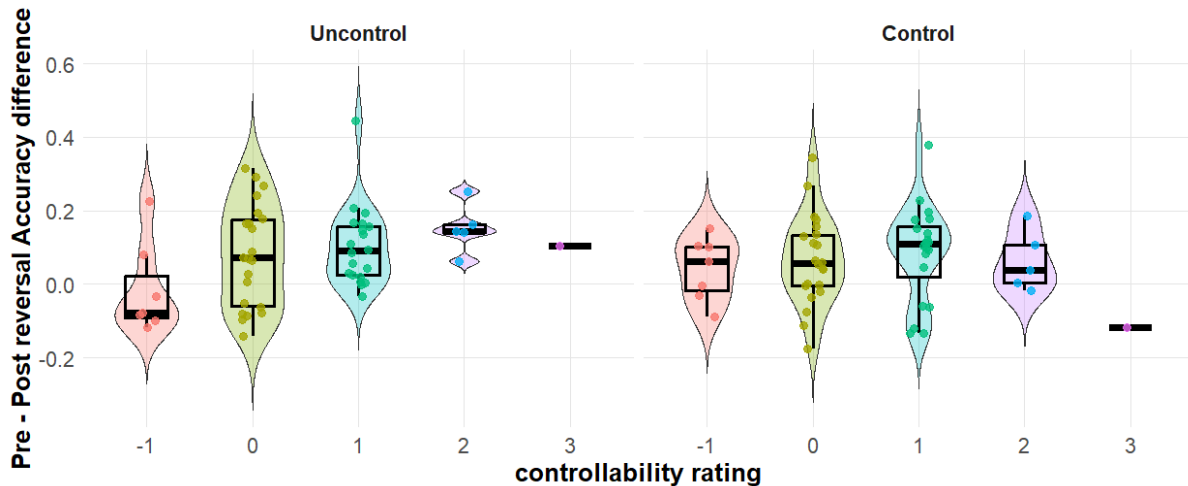
Predictor	Estimate	SE	z-value	p-value	Lower CI	Upper CI
trial:valence:rating	0.064	0.048	1.329	0.183721	-0.030	0.158
reversal:valence:rating	-0.083	0.095	-0.872	0.383232	-0.269	0.103
controllability:valence:rating	0.055	0.095	0.581	0.561234	-0.130	0.240
trial:reversal:controllability:valence	-0.179	0.185	-0.967	0.333578	-0.543	0.184
trial:reversal:controllability:rating	0.250	0.096	2.608	0.009109	0.062	0.437
trial:reversal:valence:rating	-0.081	0.096	-0.844	0.398921	-0.269	0.107
trial:controllability:valence:rating	-0.026	0.096	-0.271	0.786261	-0.214	0.162
reversal:controllability:valence:rating	-0.069	0.189	-0.365	0.715387	-0.440	0.302
trial:reversal:controllability:valence:rating	-0.027	0.191	-0.140	0.888989	-0.402	0.349

**Table 2: Results of the GLMM with controllability rating added as a regressor.** In addition to the significant main effects and interactions already present in the previous model, we also see interactions of controllability rating with trial, reversal, and control. Importantly, we also see a significant three-way reversal  $\times$  controllability  $\times$  rating and a four-way trial  $\times$  reversal  $\times$  controllability  $\times$  rating interaction.

As shown in Table 2, in addition to the effects seen in the original model, we observed a three-way reversal  $\times$  controllability  $\times$  rating and a four-way trial  $\times$  reversal  $\times$  controllability  $\times$  rating interaction. Estimated marginal trends for the subjective rating of controllability for each combination of conditions were calculated from the model. Pairwise contrasts of the estimated slopes—which quantify how subjective ratings moderate the drop in accuracy following reversal—revealed that in uncontrollable games, the relationship between ratings and accuracy loss differed significantly before versus after reversal ( $p < 0.0001$ ). In other words, in uncontrollable games, lower subjective ratings of controllability were associated with a greater drop in accuracy after reversal, an effect that was not evident in controllable games ( $p = 0.331$ ).

Similar effects were seen in separate post-hoc models on data split by controllability. In uncontrollable games, a reversal  $\times$  rating interaction was significant ( $p = 1.25 \times 10^{-7}$ ), indicating that subjective rating of controllability mediated the reduction in

accuracy after reversal. Moreover, this interaction was insignificant in controllable games ( $p = 0.068$ ). Kendall's correlation tests confirmed the same finding. After taking pairwise differences in accuracy after reversal, we correlated this accuracy drop with the subjective rating of control. This correlation was significant in uncontrollable games ( $p = 0.031$ ), but not in controllable games ( $p = 0.648$ ).

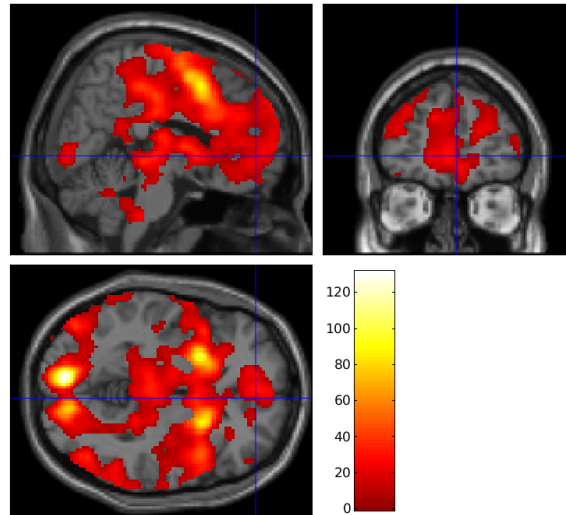


**Figure 9: Pairwise reversal difference by rating.** For each subject, the pairwise difference in accuracy before and after reversal was taken and was separated by controllable and uncontrollable games. Within each condition, the data was split by controllability rating. In uncontrollable games, this accuracy difference was correlated with the rating, but was not significantly correlated in controllable games.

### Brain representation does not differ by controllability or valence

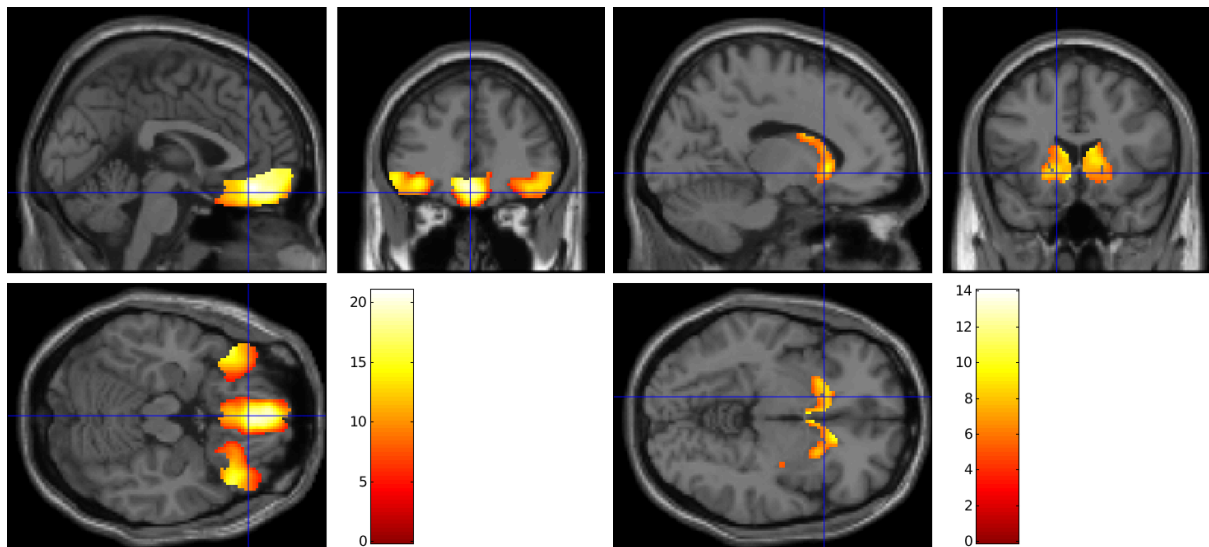
To find out whether activity in the brain reflected observed behavioural differences, we ran four GLMs as detailed in the methods. In the first GLM, we predicted differences in activity due to the main effects of controllability and valence as well as their interaction within the orbitofrontal cortex and striatum. After correcting for family-wise errors, the mean activity for all conditions was present throughout many areas of the brain, such as the visual cortex, frontal cortex, medial prefrontal cortex, striatum, and others.

Activity across conditions in the brain, however, was quite similar. Even within prespecified OFC and striatum ROI masks, there were no voxels in the OFC or striatum that differed significantly by main effects of controllability/valence or their interaction.



**Figure 10: Group-level activation for GLM1.** Mean activity across conditions associated with stimulus presentation in the first GLM. Activation was observed across multiple areas in the brain, but no difference in activity across conditions.

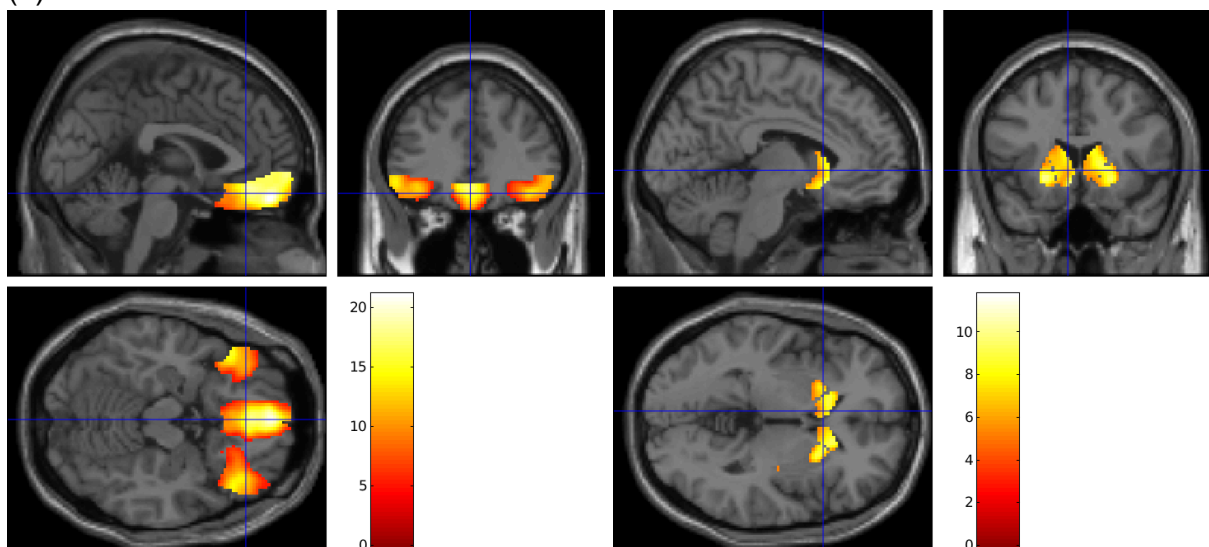
The second GLM associated the expected value of each choice at the time of stimulus presentation with BOLD activity. The expected value was generated from the best-fitting computational model, the basic hidden Markov model with three free parameters  $p$ ,  $q$ , and  $\tau$ . The expected value for the chosen option at each trial was generated by summing the probabilities of each of the two images being the hidden target. After adjusting for family-wise error at the group level, a large cluster of 2700 voxels located in the ventromedial prefrontal cortex was found. Additionally, two clusters of 1053 and 1282 voxels were found to be significant in the left and right lateral OFC respectively. The left and right striatum were active too, with significant clusters of 661 and 791 voxels respectively. These clusters were used to create binary masks, and the mean activity within these clusters was extracted for each subject and condition's contrast image. The ANOVA performed on mean activity in the vmPFC or bilateral OFC revealed no significant predictors. The ANOVA performed on the striatum showed controllability as a significant predictor ( $p = 0.03$ ), and the follow-up paired t-test was also significant ( $p = 0.047$ ). There was also a significant controllability  $\times$  valence  $\times$  rating interaction ( $p = 0.034$ ), but follow-up estimated marginal trends and post-hoc ANOVAs uncovered no differences.



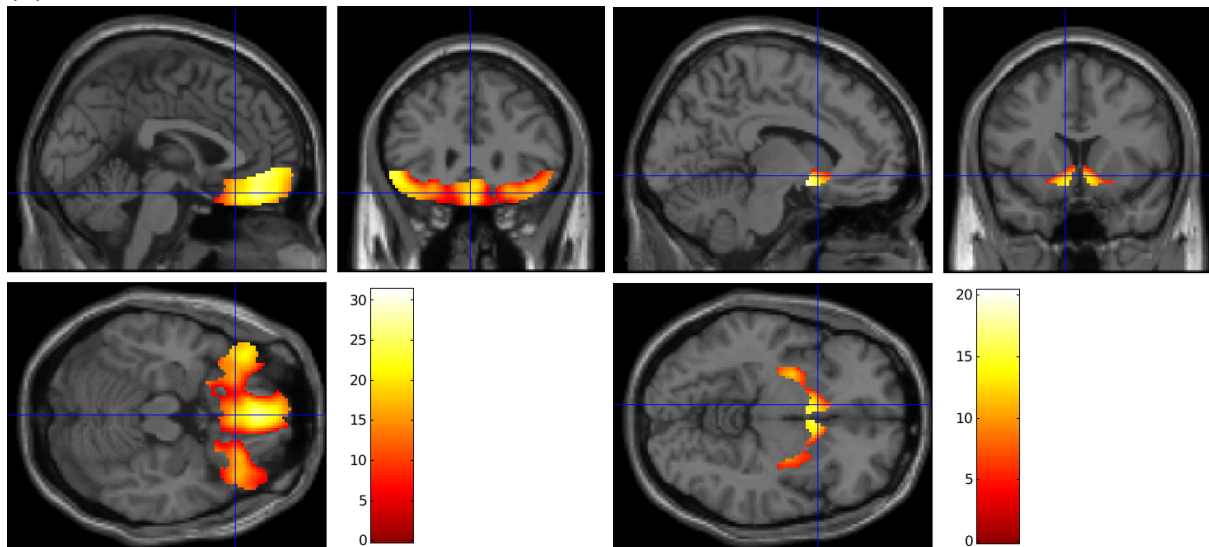
**Figure 11: Mean Activity for GLM2.** Activation in the orbitofrontal cortex (left) and the striatum (right) associated with the expected value of the chosen option at the time of stimulus presentation in each trial. The orbitofrontal cortex contained three clusters, one in the ventromedial prefrontal cortex and two in the lateral orbitofrontal cortex. The striatum was also bilaterally activated, especially the caudate nucleus.

In GLM3, the activity associated with the expected value of choice at the time of reward feedback was similar to GLM2. The vmPFC cluster of size 2603 voxels did not differ by any condition. The lateral OFC clusters of size 1222 and 899 voxels did not differ by conditions, either. The ANOVA performed on the striatum clusters of size 747 and 778 voxels did not show significant effects for controllability ( $p = 0.064$ ) and controllability  $\times$  valence  $\times$  rating ( $p = 0.094$ ). The reward outcome signal at the time of reward feedback presentation was present in the OFC in a singular cluster of 6526 voxels; meanwhile, in the striatum, it was present bilaterally in voxels of sizes 704 and 705. In both regions, mean activity associated with reward feedback did not differ by controllability, valence, controllability ratings, or any of their interactions.

(a)



(b)



**Figure 12: Mean BOLD Activation for GLM3.** (a) The mean activity associated with the expected value of the chosen option at the time of reward feedback in the orbitofrontal cortex (left) and the striatum (right). The pattern of significant clusters is very similar to the one in GLM2, implying that the representation does not differ much from the time of stimulus presentation. The mean activity also does not differ significantly across conditions. (b) The mean activity associated with reward outcome at the time of reward feedback in the orbitofrontal cortex (left) and the striatum (right), which also does not differ across conditions.

The final GLM was run to assess if patterns of voxel activity in the brain could predict the condition being experienced. Whole brain searchlight analysis using beta maps from the first-level GLM did not uncover any voxels that could predict a condition above chance (25%) significantly across participants. Searchlight analysis that trained on binary classification, i.e., to be able to classify controllable from uncontrollable games also did not give any voxels above chance (50%) that could predict controllability.

# Chapter 4 Discussion

In this study, we examined whether a lack of control affects how people learn from rewards using a probabilistic reversal learning task. We found that uncontrollability significantly impaired learning following hidden target reversals, but notably, this effect was restricted to games played for monetary bonuses rather than avoidance of electric shocks. Moreover, participants' subjective perceptions of controllability mediated this impairment across both bonus and shock games, with stronger beliefs in uncontrollability leading to more pronounced deficits in adapting after reversal. Despite these clear behavioural impacts, our neuroimaging analyses revealed no significant differences in brain activation patterns related to controllability or goal outcome valence within regions of interest, namely the orbitofrontal cortex and the striatum.

As expected, the effect of reversal was quite strong, with a drop in mean accuracy after reversal in most conditions. Reduction in accuracy after reversal was more pronounced in uncontrollable games over controllable ones, but this effect was driven by bonus games. This is because we found that in bonus games, uncontrollability exacerbated the impact of reversal, but this wasn't present in shock games. This was contrary to our prediction that a threat of shock would increase the effect of uncontrollability on learning. We had also predicted that the threat of shock would impair learning after reversal in general, but neither hypothesis was seen. Although the threat of shock has been shown to impair learning (Ballard *et al.*, 2019), it is also possible that it increases motivation to perform. This has some support in the literature, as a study found that Parkinson's disease patients not on dopaminergic medication showed increased vigour in response to avoiding a shock compared to gaining money (Shiner *et al.*, 2012). Another study in healthy subjects found increased vigour when faced with an instantaneous threat of punishment in the form of potential monetary loss (Griffiths and Beierholm, 2017).

It is not just the lack of objective control, but the subjective perception of control that affects decision-making too (Wang and Delgado, 2019). Furthermore, there was a lot of variation in the perception of controllability, we also performed another model with participants' ratings of how much they felt in control. These controllability ratings mediated the effect of the control manipulation. The more participants believed in the manipulation, the worse they performed after the reversal in uncontrollable games. Their belief did not impact their performance after the target reversal in controllable games, suggesting that it was not just a lack of control but the subjective belief in

control that affected learning. Taken together, the results point to a clear effect of perceived controllability on flexibility in learning.

To explore the neural representations of controllability, we ran four GLMs to test different facets of brain activity underlying altered reward learning. Because we had behavioural evidence that the subjective belief about controllability affected the ability to learn after reversal, we expected to see differences in activity across controllability. As shown in the results, we observed a lot of activity associated with events in the task. Areas like the visual cortex, frontal cortex, medial and lateral prefrontal cortex, striatum, etc. were significantly active during stimulus presentation. Multidimensional features, probabilistic feedback, hidden target reversal, and controllability manipulations made for a rich and dynamic task environment, so it is not surprising that we observed widespread neural activity. However, we did not detect any brain regions with significant differences in activation when comparing controllable versus uncontrollable games or bonus versus shock games.

The expected value of the choice signal (as predicted by the hidden Markov model) at the moment of both stimulus presentation and reward feedback was associated with a lot of activity, including the orbitofrontal cortex and the striatum. The ventromedial prefrontal cortex within the OFC is thought to represent the value of choices as well as outcomes, so it was reassuring to observe a significant activation cluster in the vmPFC that represented the expected value both at the time of presentation of the options and feedback. Reception of reward is also known to elicit lots of activity in the brain, so it was expected to see activity associated with reward outcome in the OFC and the striatum.

However, no significant differences in conditions were observed in all three univariate analyses, especially in the OFC. This could be due to more subtle or distributed effects of controllability despite the clear behavioural differences. As stated before, the task is complex and quite demanding, which could result in neural responses dominating over any representations of controllability.

Moreover, it could be a specification issue, since one beta was estimated for all trials in a game. Behaviourally, we observed that the effect of reversal on accuracy was quite strong, so perhaps specifying reversal explicitly or estimating when each participant detects reversal could uncover differences in the neural activity that drives differences in behaviour.

There were no differences observed in the representation of expected value by controllability either at the time of stimulus presentation or reward feedback, which indicates that the actual value representation is not affected. The differences in learning could then be because these values are used to a lesser extent in the computations underlying decision-making. This is interesting because the motivation to learn was kept constant across conditions, as participants were told that they would be rewarded with money for each gold star they received, irrespective of the

condition and independently of what the goal progress is. The value and information of each trial were thus designed to be kept constant, in concordance with our finding that the value representation does not differ by condition. Participants performed worse when they perceived a lack of control, even if the representation of value was unaffected. This could suggest that if one believes that their choices do not matter (due to perceived uncontrollability), one may care less about the values, and use them to guide learning.

The beta maps generated from the fourth GLM were used to train support vector machines to detect differences in voxel pattern activity, going beyond simple activity means of univariate analyses to multivariate techniques. Here, too, no voxels were able to predict and differentiate between conditions above chance at a group level. Despite using cross-validation, the problem here was that there was too little data—only four games per condition for training and one for testing. We tried to estimate separate beta maps for each trial to increase data points, but the inter-stimulus interval was only 2.36 seconds on average. Since the peak latency of the canonical haemodynamic response function is about 4-6 seconds, it was difficult for the algorithm to deconvolve individual trials from the BOLD activity. This resulted in highly correlated estimates, an issue not solved even by estimating alternate trials. Preliminary representational similarity analyses (not reported in results) using correlation-based similarity matrices also revealed that the patterns of each of the four conditions were extremely similar to each other. Perhaps if the trials were dissociable or if there were more games per condition, MVPA and RSA would be able to detect any differences in patterns of voxel activity (if at all they actually exist).

In the future, functional connectivity analyses might provide more insight into the neural activity underlying learning impairments observed. Graph-theory-based functional connectivity toolboxes (Mijalkov *et al.*, 2017) offer a variety of network parameters to quantify. We note that these analyses would be purely exploratory, as no predetermined hypotheses were set a priori. On the behavioural side, eye-tracking data collected during the fMRI scans could be analysed to find the effect of controllability on attention and arousal and their interaction with learning. And finally on the computational side, model comparison could be used to test whether there is evidence for separate model parameters for controllable and uncontrollable games. Overall, we found support for the hypothesis that a lack of control, a pre-clinical factor, was enough to impair flexible reward learning in healthy people. More importantly, it was not just the existence of objective uncontrollability, but their subjective belief in controllability that mediated this impairment. This has implications for our understanding of uncontrollability as a contributor to learning deficits observed in stress and stress-related disorders and perhaps as a potential therapeutic target for behavioural treatment strategies.

# References

- Abercrombie, ED, Keefe, KA, DiFrischia, DS, and Zigmond, MJ (1989). Differential Effect of Stress on In Vivo Dopamine Release in Striatum, Nucleus Accumbens, and Medial Frontal Cortex. *J Neurochem* 52, 1655–1658.
- Amarante, LM, and Laubach, M (2014). For Better or Worse: Reward Comparison by the Ventromedial Prefrontal Cortex. *Neuron* 82, 1191–1193.
- Ballard, T, Sewell, DK, Cosgrove, D, and Neal, A (2019). Information Processing Under Reward Versus Under Punishment. *Psychol Sci* 30, 757–764.
- Bandura, A (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev* 84, 191–215.
- Bates, D, Mächler, M, Bolker, B, and Walker, S (2015). Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67, 1–48.
- de Berker, AO, Tirole, M, Rutledge, RB, Cross, GF, Dolan, RJ, and Bestmann, S (2016). Acute stress selectively impairs learning to act. *Sci Rep* 6, 29816.
- Bogdan, R, and Pizzagalli, DA (2006). Acute Stress Reduces Reward Responsiveness: Implications for Depression. *Biol Psychiatry* 60, 1147–1154.
- Bogdan, R, Santesso, DL, Fagerness, J, Perlis, RH, and Pizzagalli, DA (2011). Corticotropin-Releasing Hormone Receptor Type 1 (CRHR1) Genetic Variation and Stress Interact to Influence Reward Learning. *J Neurosci* 31, 13246–13254.
- Boorman, ED, Behrens, TEJ, Woolrich, MW, and Rushworth, MFS (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron* 62, 733–743.
- Bredemeier, K, Warren, SL, Berenbaum, H, Miller, GA, and Heller, W (2016). Executive function deficits associated with current and past major depressive symptoms. *J Affect Disord* 204, 226–233.
- Brown, VM, Zhu, L, Solway, A, Wang, JM, McCurry, KL, King-Casas, B, and Chiu, PH (2021). Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA Psychiatry* 78, 1113–1122.
- Burton, AC, Nakamura, K, and Roesch, MR (2015). From ventral-medial to dorsal-lateral striatum: Neural correlates of reward-guided decision-making. *Neurobiol Learn Mem* 117, 51–59.
- Cabib, S, Ventura, R, and Puglisi-Allegra, S (2002). Opposite imbalances between mesocortical and mesoaccumbens dopamine responses to stress by the same genotype depending on living conditions. *Behav Brain Res* 129, 179–185.
- Chib, VS, Rangel, A, Shimojo, S, and O'Doherty, JP (2009). Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. *J Neurosci* 29, 12315–12320.

- Chrapusta, SJ, Wyatt, RJ, and Masserano, JM (1997). Effects of Single and Repeated Footshock on Dopamine Release and Metabolism in the Brains of Fischer Rats. *J Neurochem* 68, 2024–2031.
- Chung, D, Orloff, MA, Lauharatanahirun, N, Chiu, PH, and King-Casas, B (2020). Valuation of peers' safe choices is associated with substance-naïveté in adolescents. *Proc Natl Acad Sci* 117, 31729–31737.
- Cooper, JC, Kreps, TA, Wiebe, T, Pirkl, T, and Knutson, B (2010). When Giving Is Good: Ventromedial Prefrontal Cortex Activation for Others' Intentions. *Neuron* 67, 511–521.
- Cuadra, G, Zurita, A, Lacerra, C, and Molina, V (1999). Chronic stress sensitizes frontal cortex dopamine release in response to a subsequent novel stressor: reversal by naloxone. *Brain Res Bull* 48, 303–308.
- Daw, ND (2011). Trial-by-trial data analysis using computational models. In: *Decision Making, Affect, and Learning*, Oxford University Press.
- Daw, ND, Niv, Y, and Dayan, P (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8, 1704–1711.
- Del Arco, A, and Mora, F (2008). Prefrontal cortex–nucleus accumbens interaction: *In vivo* modulation by dopamine and glutamate in the prefrontal cortex. *Pharmacol Biochem Behav* 90, 226–235.
- Dickerson, SS, and Kemeny, ME (2004). Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychol Bull* 130, 355–391.
- Dorfman, HM, and Gershman, SJ (2019). Controllability governs the balance between Pavlovian and instrumental action selection. *Nat Commun* 10, 5826.
- Drummond, N, and Niv, Y (2020). Model-based decision making and model-free learning. *Curr Biol* 30, R860–R865.
- Ferster, CB, and Skinner, BF (1957). *Schedules of reinforcement*, East Norwalk, CT, US: Appleton-Century-Crofts.
- Gagne, C, Zika, O, Dayan, P, and Bishop, SJ (2020). Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife* 9, e61387.
- Gallagher, MW, Bentley, KH, and Barlow, DH (2014a). Perceived Control and Vulnerability to Anxiety Disorders: A Meta-analytic Review. *Cogn Ther Res* 38, 571–584.
- Gallagher, MW, Naragon-Gainey, K, and Brown, TA (2014b). Perceived Control is a Transdiagnostic Predictor of Cognitive–Behavior Therapy Outcome for Anxiety Disorders. *Cogn Ther Res* 38, 10–22.

- Gao, W, Yan, X, Chen, Y, Yang, J, and Yuan, J (2025). Situation covariation and goal adaptiveness? The promoting effect of cognitive flexibility on emotion regulation in depression. *Emotion* 25, 18–32.
- Giorgi, O, Lecca, D, Piras, G, Driscoll, P, and Corda, MG (2003). Dissociation between mesocortical dopamine release and fear-related behaviours in two psychogenetically selected lines of rats that differ in coping strategies to aversive conditions. *Eur J Neurosci* 17, 2716–2726.
- Goldfarb, EV, Froböse, MI, Cools, R, and Phelps, EA (2017). Stress and Cognitive Flexibility: Cortisol Increases Are Associated with Enhanced Updating but Impaired Switching. *J Cogn Neurosci* 29, 14–24.
- Gradin, VB, Kumar, P, Waiter, G, Ahearn, T, Stickle, C, Milders, M, Reid, I, Hall, J, and Steele, JD (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* 134, 1751–1764.
- Grahek, I, Everaert, J, Krebs, RM, and Koster, EHW (2018). Cognitive Control in Depression: Toward Clinical Models Informed by Cognitive Neuroscience. *Clin Psychol Sci* 6, 464–480.
- Griffiths, B, and Beierholm, UR (2017). Opposing effects of reward and punishment on human vigor. *Sci Rep* 7, 42287.
- Guitart-Masip, M, Walsh, A, Dayan, P, and Olsson, A (2023). Anxiety associated with perceived uncontrollable stress enhances expectations of environmental volatility and impairs reward learning. *Sci Rep* 13, 18451.
- Halahakoon, DC, Kieslich, K, O'Driscoll, C, Nair, A, Lewis, G, and Roiser, JP (2020). Reward-processing behavior in depressed participants relative to healthy volunteers: A systematic review and meta-analysis. *JAMA Psychiatry* 77, 1286–1295.
- Hammen, C (2005). Stress and Depression. *Annu Rev Clin Psychol* 1, 293–319.
- Hammen, CL (2015). Stress and depression: old questions, new approaches. *Curr Opin Psychol* 4, 80–85.
- Hare, TA, O'Doherty, J, Camerer, CF, Schultz, W, and Rangel, A (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *J Neurosci* 28, 5623–5630.
- Hartley, CA, Gorun, A, Reddan, MC, Ramirez, F, and Phelps, EA (2014). Stressor controllability modulates fear extinction in humans. *Neurobiol Learn Mem* 113, 149–156.
- Hartogsveld, B, van Ruitenbeek, P, Quaedflieg, CWEM, and Smeets, T (2020). Balancing Between Goal-Directed and Habitual Responding Following Acute Stress. *Exp Psychol* 67, 99–111.
- Hebart, MN, Görden, K, and Haynes, J-D (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinformatics* 8.

- Hikosaka, O, Nakamura, K, and Nakahara, H (2006). Basal ganglia orient eyes to reward. *J Neurophysiol* 95, 567–584.
- Hiroto, DS, and Seligman, ME (1975). Generality of learned helplessness in man. *J Pers Soc Psychol* 31, 311–327.
- Huys, QJ, Pizzagalli, DA, Bogdan, R, and Dayan, P (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord* 3, 12.
- Joormann, J, and Gotlib, IH (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *J Abnorm Psychol* 115, 705–714.
- Karsh, N, and Eitam, B (2015). I control therefore I do: Judgments of agency influence action selection. *Cognition* 138, 122–131.
- Kasper, L, Bollmann, S, Diaconescu, AO, Hutton, C, Heinzle, J, Iglesias, S, Hauser, TU, Sebold, M, Manjaly, Z-M, Pruessmann, KP, *et al.* (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *J Neurosci Methods* 276, 56–72.
- Katz, RJ (1982). Animal model of depression: Pharmacological sensitivity of a hedonic deficit. *Pharmacol Biochem Behav* 16, 965–968.
- Katz, RJ, Roth, KA, and Carroll, BJ (1981). Acute and chronic stress effects on open field activity in the rat: Implications for a model of depression. *Neurosci Biobehav Rev* 5, 247–251.
- Kendler, KS, Hettema, JM, Butera, F, Gardner, CO, and Prescott, CA (2003). Life Event Dimensions of Loss, Humiliation, Entrapment, and Danger in the Prediction of Onsets of Major Depression and Generalized Anxiety. *Arch Gen Psychiatry* 60, 789–796.
- Knutson, B, Fong, GW, Adams, CM, Varner, JL, and Hommer, D (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport* 12, 3683.
- Koolhaas, JM, Bartolomucci, A, Buwalda, B, de Boer, SF, Flügge, G, Korte, SM, Meerlo, P, Murison, R, Olivier, B, Palanza, P, *et al.* (2011). Stress revisited: A critical evaluation of the stress concept. *Neurosci Biobehav Rev* 35, 1291–1301.
- Kumar, P, Waiter, G, Ahearn, T, Milders, M, Reid, I, and Steele, JD (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain* 131, 2084–2093.
- Lebreton, M, Jorge, S, Michel, V, Thirion, B, and Pessiglione, M (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron* 64, 431–439.
- Lenth, RV (2024). emmeans: Estimated marginal means, aka least-squares means.
- Leong, YC, Radulescu, A, Daniel, R, DeWoskin, V, and Niv, Y (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron* 93, 451–463.

- Leotti, LA, and Delgado, MR (2014). The Value of Exercising Control Over Monetary Gains and Losses. *Psychol Sci* 25, 596–604.
- Leotti, LA, Iyengar, SS, and Ochsner, KN (2010). Born to choose: the origins and value of the need for control. *Trends Cogn Sci* 14, 457–463.
- Levy, DJ, and Glimcher, PW (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *J Neurosci Off J Soc Neurosci* 31, 14693–14707.
- Maier, SF (2015). Behavioral control blunts reactions to contemporaneous and future adverse events: Medial prefrontal cortex plasticity and a corticostriatal network. *Neurobiol Stress* 1, 12–22.
- Maier, SF, Amat, J, Baratta, MV, Paul, E, and Watkins, LR (2006). Behavioral control, the medial prefrontal cortex, and resilience. *Dialogues Clin Neurosci* 8, 397–406.
- Matsumoto, K, Suzuki, W, and Tanaka, K (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301, 229–232.
- McClure, SM, Laibson, DI, Loewenstein, G, and Cohen, JD (2004). Separate Neural Systems Value Immediate and Delayed Monetary Rewards. *Science* 306, 503–507.
- McNamee, D, Rangel, A, and O'Doherty, JP (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat Neurosci* 16, 479–485.
- Mijalkov, M, Kakaie, E, Pereira, JB, Westman, E, Volpe, G, and Initiative, for the ADN (2017). BRAPH: A graph theory software for the analysis of brain connectivity. *PLOS ONE* 12, e0178798.
- Min, S, Mazurka, R, Pizzagalli, DA, Whitton, AE, Milev, RV, Bagby, RM, Kennedy, SH, and Harkness, KL (2024). Stressful Life Events and Reward Processing in Adults: Moderation by Depression and Anhedonia. *Depress Anxiety* 2024, 8853631.
- Moscarello, JM, and Hartley, CA (2017). Agency and the Calibration of Motivated Behavior. *Trends Cogn Sci* 21, 725–735.
- Mukherjee, D, Filipowicz, ALS, Vo, K, Satterthwaite, TD, and Kable, JW (2020). Reward and punishment reversal-learning in major depressive disorder. *J Abnorm Psychol* 129, 810–823.
- Murphy, FC, Michael, A, and Sahakian, BJ (2012). Emotion modulates cognitive flexibility in patients with major depression. *Psychol Med* 42, 1373–1382.
- Must, A, Horvath, S, Nemeth, VL, and Janka, Z (2013). The Iowa Gambling Task in depression – what have we learned about sub-optimal decision-making strategies? *Front Psychol* 4.
- Neubert, F-X, Mars, RB, Sallet, J, and Rushworth, MFS (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc Natl Acad Sci* 112, E2695–E2704.

Noonan, MP, Kolling, N, Walton, ME, and Rushworth, MFS (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *Eur J Neurosci* 35, 997–1010.

O'Doherty, J, Kringelbach, ML, Rolls, ET, Hornak, J, and Andrews, C (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 4, 95–102.

O'Doherty, JP, Dayan, P, Friston, K, Critchley, H, and Dolan, RJ (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron* 38, 329–337.

Otto, AR, Raio, CM, Chiang, A, Phelps, EA, and Daw, ND (2013). Working-memory capacity protects model-based learning from stress. *Proc Natl Acad Sci* 110, 20941–20946.

Overmier, JB, and Seligman, ME (1967). Effects of inescapable shock upon subsequent escape and avoidance responding. *J Comp Physiol Psychol* 63, 28–33.

Padoa-Schioppa, C, and Assad, JA (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226.

Paret, C, and Bublatzky, F (2020). Threat rapidly disrupts reward reversal learning. *Behav Res Ther* 131, 103636.

Pavlov, IP (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*, Oxford, England: Oxford Univ. Press.

Peirce, J, Gray, JR, Simpson, S, MacAskill, M, Höchenberger, R, Sogo, H, Kastman, E, and Lindeløv, JK (2019). PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* 51, 195–203.

Petzold, A, Plessow, F, Goschke, T, and Kirschbaum, C (2010). Stress reduces use of negative feedback in a feedback-based learning task. *Behav Neurosci* 124, 248–255.

Piray, P, Dezfouli, A, Heskes, T, Frank, MJ, and Daw, ND (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Comput Biol* 15, e1007043.

Pizzagalli, DA (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annu Rev Clin Psychol* 10, 393–423.

Plassmann, H, O'Doherty, J, and Rangel, A (2007). Orbitofrontal Cortex Encodes Willingness to Pay in Everyday Economic Transactions. *J Neurosci* 27, 9984–9988.

Plessow, F, Fischer, R, Kirschbaum, C, and Goschke, T (2011). Inflexibly focused under stress: Acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. *J Cogn Neurosci* 23, 3218–3227.

Posit team (2024). *RStudio: Integrated development environment for R*, Boston, MA: Posit Software, PBC.

- R Core Team (2024). R: a language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing.
- Rescorla, RA, and Wagner, AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory*.
- Reynolds, JNJ, and Wickens, JR (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw* 15, 507–521.
- Robinson, OJ, Cools, R, Carlisi, CO, Sahakian, BJ, and Drevets, WC (2012). Ventral Striatum Response During Reward and Punishment Reversal Learning in Unmedicated Major Depressive Disorder. *Am J Psychiatry* 169, 152–159.
- Roesch, MR, and Olson, CR (2004). Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex. *Science* 304, 307–310.
- Rossetti, ZL, Lai, M, Hmaidan, Y, and Gessa, GL (1993). Depletion of mesolimbic dopamine during behavioral despair: Partial reversal by chronic imipramine. *Eur J Pharmacol* 242, 313–315.
- Rotter, JB (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychol Monogr Gen Appl* 80, 1–28.
- Rupprechter, S, Stankevicius, A, Huys, QJM, Steele, JD, and Seriès, P (2018). Major Depression Impairs the Use of Reward Values for Decision-Making. *Sci Rep* 8, 13798.
- Rutledge, RB, Moutoussis, M, Smittenaar, P, Zeidman, P, Taylor, T, Hryniewicz, L, Lam, J, Skandali, N, Siegel, JZ, Ousdal, OT, *et al.* (2017). Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry* 74, 790–797.
- Samejima, K, Ueda, Y, Doya, K, and Kimura, M (2005). Representation of Action-Specific Reward Values in the Striatum. *Science* 310, 1337–1340.
- Sanchis-Segura, C, Spanagel, R, Henn, FA, and Vollmayr, B (2005). Reduced sensitivity to sucrose in rats bred for helplessness: a study using the matching law. *Behav Pharmacol* 16, 267.
- Schuck, NW, Cai, MB, Wilson, RC, and Niv, Y (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* 91, 1402–1412.
- Schultz, W (1998). Predictive Reward Signal of Dopamine Neurons. *J Neurophysiol* 80, 1–27.
- Schultz, W, Apicella, P, and Ljungberg, T (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci* 13, 900–913.
- Schultz, W, Dayan, P, and Montague, PR (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.

- Seligman, ME, and Maier, SF (1967). Failure to escape traumatic shock. *J Exp Psychol* 74, 1–9.
- Shiner, T, Seymour, B, Symmonds, M, Dayan, P, Bhatia, KP, and Dolan, RJ (2012). The Effect of Motivation on Movement: A Study of Bradykinesia in Parkinson's Disease. *PLOS ONE* 7, e47138.
- Skinner, EA (1996). A guide to constructs of control. *J Pers Soc Psychol* 71, 549–570.
- Smith, KE, and Pollak, SD (2022). Early life stress and perceived social isolation influence how children use value information to guide behavior. *Child Dev* 93, 804–814.
- Snyder, HR (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychol Bull* 139, 81–132.
- Steele, JD, Kumar, P, and Ebmeier, KP (2007). Blunted response to feedback information in depressive illness. *Brain* 130, 2367–2374.
- Sutton, RS, and Barto, AG (1998). Reinforcement learning: An introduction, MIT press Cambridge.
- Tanaka, S, Pan, X, Oguchi, M, Taylor, JE, and Sakagami, M (2015). Dissociable functions of reward inference in the lateral prefrontal cortex and the striatum. *Front Psychol* 6.
- Thorndike, E (1898). Some Experiments on Animal Intelligence. *Science* 7, 818–824.
- Treadway, MT, Bossaller, NA, Shelton, RC, and Zald, DH (2012a). Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. *J Abnorm Psychol* 121, 553–558.
- Treadway, MT, Buckholtz, JW, Cowan, RL, Woodward, ND, Li, R, Ansari, MS, Baldwin, RM, Schwartzman, AN, Kessler, RM, and Zald, DH (2012b). Dopaminergic Mechanisms of Individual Differences in Human Effort-Based Decision-Making. *J Neurosci* 32, 6170–6176.
- Tremblay, L, and Schultz, W (2000). Reward-Related Neuronal Activity During Go-Nogo Task Performance in Primate Orbitofrontal Cortex. *J Neurophysiol* 83, 1864–1876.
- Valton, V, Mkrtchian, A, Moses-Payne, M, Gray, A, Kieslich, K, VanUrck, S, Samborska, V, Halahakoon, D, Manohar, SG, and Dayan, P (2024). A computational approach to understanding effort-based decision-making in depression. *bioRxiv*, 2024–06.
- Ventura, R, Cabib, S, and Puglisi-Allegra, S (2002). Genetic susceptibility of mesocortical dopamine to stress determines liability to inhibition of mesoaccumbens dopamine and to behavioral 'despair' in a mouse model of depression. *Neuroscience* 115, 999–1007.

- Vollmayr, B, and Gass, P (2013). Learned helplessness: unique features and translational value of a cognitive depression model. *Cell Tissue Res* 354, 171–178.
- Vrieze, E, Pizzagalli, DA, Demyttenaere, K, Hompes, T, Sienaert, P, de Boer, P, Schmidt, M, and Claes, S (2013). Reduced Reward Learning Predicts Outcome in Major Depressive Disorder. *Biol Psychiatry* 73, 639–645.
- Wang, JX, Kurth-Nelson, Z, Kumaran, D, Tirumala, D, Soyer, H, Leibo, JZ, Hassabis, D, and Botvinick, M (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nat Neurosci* 21, 860–868.
- Wang, KS, and Delgado, MR (2019). Corticostriatal circuits encode the subjective value of perceived control. *Cereb Cortex* 29, 5049–5060.
- White, RW (1959). Motivation reconsidered: The concept of competence. *Psychol Rev* 66, 297–333.
- Wickens, JR, Begg, AJ, and Arbuthnott, GW (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex *In vitro*. *Neuroscience* 70, 1–5.
- Wickham, H (2011). ggplot2. *WIREs Comput Stat* 3, 180–185.
- Wickham, H, Averick, M, Bryan, J, Chang, W, McGowan, LD, François, R, Grolemond, G, Hayes, A, Henry, L, Hester, J, *et al.* (2019). Welcome to the Tidyverse. *J Open Source Softw* 4, 1686.
- Wilke, SA, Lavi, K, Byeon, S, Donohue, KC, and Sohal, VS (2022). Convergence of Clinically Relevant Manipulations on Dopamine-Regulated Prefrontal Activity Underlying Stress Coping Responses. *Biol Psychiatry* 91, 810–820.
- Willner, P, Muscat, R, and Papp, M (1992). Chronic mild stress-induced anhedonia: A realistic animal model of depression. *Neurosci Biobehav Rev* 16, 525–534.
- Yin, HH, and Knowlton, BJ (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci* 7, 464–476.
- Yin, HH, Knowlton, BJ, and Balleine, BW (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* 19, 181–189.