

**Motivation and reward learning: neural mechanisms,
changes in depression and following pharmacological
treatment**

Jacqueline Scholl

**A dissertation in partial fulfilment of the requirements for the
degree of Doctor of Philosophy in the University of Oxford
Christ Church College**

Trinity Term 2015

Abstract

The primary goal of this thesis was the understanding of the cognitive and computational mechanisms of reward and effort guided processing and their change in depression and commonly used/putative treatments. In chapter 2, I developed a new paradigm that allowed me to examine the neural basis of concurrent learning about rewards and effort in a complex, naturalistic environment. Measuring brain activity using functional magnetic resonance imaging, I proposed a prefrontal cortical mechanism for overcoming behavioural biases generated by irrelevant reward features (Scholl et al., 2015). In chapter 3, I used this paradigm to measure the effects of prolonged administration of a commonly used antidepressant, a serotonin re-uptake inhibitor, which has previously been hypothesized to enhance learning and neural plasticity. I found that it increased neural learning signals for both reward and effort in a widespread network of brain areas, suggesting a general learning enhancing effect (Scholl et al., under review). In chapter 4, I measured whether a putative novel treatment for depression, the partial NMDA agonist d-cycloserine, could also influence learning. While I found this not to be the case, it led to a marked improvement in how participants integrated the available information to make decisions (Scholl et al., 2014). In chapter 5, I tested how these different aspects of reward-guided behaviour, some of which were changed in my pharmacological manipulation, are related to changes in depression. We found that dysphoric participants did not differ from healthy controls in how quickly they learnt appetitive or aversive/effortful information. However, when making decisions they used this learnt information less than healthy controls to determine their decisions. Together, these studies revealed that complex decisions require additional cognitive and neural mechanisms that go beyond those seen in the simplest learning tasks. These diverse mechanisms can be computationally teased apart, rely on different neural systems and can thus be influenced separately. Lastly, we found that common treatments of depression did not necessarily target the same mechanisms affected by depression, suggesting potential new avenues for new treatments.

Acknowledgments

First of all, I will always be grateful to both Catherine Harmer's and Matthew Rushworth's guidance and support throughout my DPhil. Their continuous insights and advice allowed me to develop into a far more useful and thoughtful researcher than I could have hoped for. Beyond their personal support they also both created such incredibly supportive lab environments, which always made me feel welcome and showed how research can be a collaborative endeavour.

I also want to thank my husband and closest collaborator Nils Kolling for his enthusiasm and especially also his optimism about our work together. He was always quick with advice whenever needed and he made my breakfast every morning.

I would also like to thank others who have been indispensable for my research over the years. Natalie Nelissen stoically helped me with data collection over many long weekends and was also always there for discussions, when I needed an additional opinion. Magdalena Soukupova was a great help to me during her MSc project, helping with the data collection in the dysphoria study, for which I am very grateful. I am very thankful to Andrea Reinecke and Jan Günthner, with who I collaborated in the d-cycloserine study, for useful discussions and collecting the data.

Other colleagues have also been very important to me over the years. Michael Browning, Marco Wittmann, Bolton Chau, Miriam Klein-Flügge, David Meder and Lennart Verhagen were always supportive and ensured I was never lacking in productive and fun discussions.

Of course, I cannot forget my Friends and Family, who I thank for their support and friendship throughout my thesis. They all know more than they ever thought they would about depression, serotonin, computational modelling and learning.

I also want to thank the Wellcome Trust for their generous financial support of both my DPhil as well as research expenses.

Table of Contents

1. GENERAL INTRODUCTION	5
1.1 BEHAVIOURAL AND NEURAL MECHANISMS OF LEARNING	7
1.1.1 PAVLOVIAN LEARNING	7
1.1.2 MODEL-BASED LEARNING	10
1.2 LEARNING AND DEPRESSION	14
1.2.1 LEARNING	14
1.2.1 PLASTICITY	20
1.2.2 VALUATION / REWARD AND PUNISHMENT	23
1.2.3 MOTIVATION AND EFFORT IN DEPRESSION	26
1.3 SEROTONIN – EMOTIONAL PROCESSING AND PLASTICITY	29
1.3.1 EMOTIONAL PROCESSING AND ANTIDEPRESSANTS	30
1.3.2 PLASTICITY AND SEROTONERGIC ANTIDEPRESSANTS	41
1.3.3 REWARD AND PUNISHMENT LEARNING IN THE LIGHT OF EMOTIONAL BIASES AND PLASTICITY	45
1.4 NMDA RECEPTORS AND THE PLASTICITY HYPOTHESIS OF DEPRESSION	50
1.4.1 D-CYCLOSERINE AND DEPRESSION	50
1.4.2 D-CYCLOSERINE AND LEARNING	52
1.5 SUMMARY AND EXPERIMENTAL QUESTIONS	54
2. NEURAL MECHANISMS OF LEARNING REAL AND HYPOTHETICAL REWARD AND EFFORT	55
2.1 INTRODUCTION	56
2.2 METHODS	58
2.3 RESULTS	82
2.4 DISCUSSION	105
3. SEROTONERGIC PLASTICITY IN CONCURRENT REWARD AND EFFORT LEARNING	111
3.1 INTRODUCTION	112
3.2 METHODS	114
3.3 RESULTS	126
3.4 DISCUSSION	138
4. A ROLE BEYOND LEARNING FOR NMDA RECEPTORS IN REWARD-BASED DECISION-MAKING – A PHARMACOLOGICAL STUDY USING D-CYCLOSERINE	143
4.1 INTRODUCTION	144
4.3 RESULTS	157
4.4 DISCUSSION	165
5. REWARD AND EFFORT LEARNING IN DYSPHORIA	170
5.1 INTRODUCTION	171
5.2 METHODS	174
5.3 RESULTS	187
5.4 DISCUSSION	199
6. GENERAL DISCUSSION	ERROR! BOOKMARK NOT DEFINED.
7. REFERENCES	205

1. General Introduction

Precise quantification is often essential for scientific empirical inquiry. Depression, like many mental disorders, is difficult to categorize and even harder to quantify. Currently, patients are diagnosed based on an interview led by a clinician (e.g. the Patient Health Questionnaire, PHQ-9 (Kroenke et al., 2001)). A diagnosis of depression is made if patients have at least five out of nine possible symptoms, for at least two weeks, of which one needs to be one of two key symptoms. These key symptoms are depressed mood and anhedonia, i.e. loss of interest in or enjoyment of activities previously found enjoyable. The other seven possible symptoms include: weight loss or weight gain, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, diminished ability to think or concentrate and recurrent thoughts of death or suicide. While this approach to diagnosis has good face validity (i.e. psychiatrists can identify depressions with high certainty), it is purely symptomatic rather than mechanistic. In other words, the symptoms are not clustered together in one diagnosis because of an underlying model of cognitive or biological processes that cause them. Addressing this lack of mechanistic as well as quantitative understanding is one aim of computational psychiatry approaches.

Computational psychiatry aims to describe complex behaviours mathematically and hopefully mechanistically in terms of their different components at a both neural and cognitive level. It can thus be considered a computational extension of the cognitive/neural systems approach. There are several advantages to this approach: Firstly, experiments can be designed that can measure particular subprocesses within more complex behaviours, rather than having to rely on simpler tasks or (potentially faulty) introspection from patients. Secondly, it is therefore possible to find out which processes are affected more specifically. Thirdly, psychiatric findings can be more easily connected to knowledge from decades of cognitive or systems neuroscience research as the processes or cognitive mechanisms that are changed by depression can be related to the

equivalent processes in healthy humans or animal models. Fourthly and finally, it allows the use and analysis of more complex, and thus more ecologically valid, experiments.

Applying this approach to the depression key symptom anhedonia, it has been proposed (Romer Thomsen et al., 2015; Treadway and Zald, 2011, 2013) that it can be related to different aspects of reward or punishment processing, including valuation (liking/disliking something or attributing it a reward/incentive value), evaluation (deciding between things you like/dislike), motivation and learning (learning the value of a stimulus from experience). The main focus of this thesis will be on value-based learning.

In the following I will first describe computational and neural mechanisms of learning in healthy humans and animals in order to lay out the basic mechanisms in questions. Then I will discuss learning in depression and possible treatments more specifically.

1.1 Behavioural and neural mechanisms of learning

In real life, choices are only very rarely between options with explicitly indicated values. Instead, more commonly one needs to retrieve, from memory, the learnt value of each option under consideration, making continuous tracking and learning essential throughout. That animals are indeed capable of learning to predict outcomes associated with options was first demonstrated over 100 years ago, leading to the beginnings of behaviourism. Ivan Pavlov found that dogs could learn to salivate to cues predicting foods, while Edward Thorndike discovered at around the same time that cats could learn to pull on a string to escape from a box. These early findings already suggested that learning is potentially not a unitary process, but that different neural and computational learning systems are responsible for learning between stimulus-outcome associations (Pavlovian learning), stimulus-response associations (habitual learning) and model-based (or goal-directed) learning.

1.1.1 Pavlovian learning

In the simplest scenario the human/ animal learns to predict previously unpredicted outcomes, after repeated exposure, estimating the probability of an outcome (e.g. a food reward) following a certain stimulus (e.g. a light). Mathematically, this learning process can be described by a simple equation such as the Rescorla-Wagner learning rule:

$$Prediction_t = Prediction_{t-1} + \alpha * PE_{t-1}$$

with

$$PE_{t-1} = Outcome_{t-1} - Prediction_{t-1}$$

This learning rule conceptualizes that every time the outcome follows the stimulus, one first computes the prediction error (PE), i.e. the difference between the actual outcome and the prior prediction about the outcome. This PE is then used to update the prediction of the outcome. The

size of this update is determined by the learning rate (α , between 0 and 1), the higher the learning rate, the faster the learning (updating).

Neurally, Pavlovian learning has most strongly been associated with the dopaminergic system. In now classic experiments, Schultz et al. (1997) recorded electrical activity from putatively dopaminergic midbrain neurons and found that their patterns of activity corresponded to prediction errors. More specifically, neurons fired more with unexpected rewards (positive PE), did not change firing to expected rewards (no PE) and decreased their baseline firing levels when an expected reward was omitted (negative PE). Interestingly, dopaminergic midbrain neurons send projections to different parts of the brain including the frontal cortex and the striatum. They are therefore in a position to broadcast the learning signals to these cortical and sub-cortical areas. Meta-analyses of human fMRI studies show that prediction error signals are commonly found in the ventral or dorsal striatum and midbrain (Chase et al., 2015; Garrison et al., 2013). Despite the strong emphasis on those regions for learning, the meta-analyses have also revealed other cortical areas as carrying PE signals (e.g. insula, thalamus or posterior cingulate cortex), showing that those signals are more ubiquitous than previously thought. It is of note that these meta-analytic results rely on studies that all used correlations between brain signals and the mathematical PE. However, this technique is prone to confounding areas coding outcome signals with those coding true prediction errors because of mathematical correlations between these constructs. To overcome this problem, Rutledge et al. (2010) used a more rigorous axiomatic analytical approach and tested whether signals from different areas fulfilled all necessary conditions for being a PE signal¹. They found activity in striatum, medial prefrontal cortex, amygdala and posterior cingulate cortex to fulfil conditions of prediction errors. In contrast, they

¹ This study also had increased sensitivity by calculating measurements based on different timings for baseline and peak BOLD time points, thus correcting for area-specific shifts in the haemodynamic response function.

found that the insula (though it was identified in meta-analyses, Chase et al., 2015) did not fulfil conditions for prediction error signals.

The usefulness of the learnt predictions is that they can then be used to drive behaviour. For example, having learnt the predicted values (rewards or punishments) associated with two options, one can then chose between them. While intuitively, it might appear that an animal should simply select the option that promises the greater reward, animals and humans do not in fact behave in this way. Rather, selection appears to be a stochastic process in which either option is chosen with a certain probability depending on the difference in value between the options. Mathematically this is well described in a so-called softmax or logistic function:

$$P(\text{Option}_A) = \frac{e^{\beta * \text{Utility}_A}}{e^{\beta * \text{Utility}_A} + e^{\beta * \text{Utility}_B}}$$

In this equation the parameter β , the inverse temperature (also called the exploitation/exploration parameter), determines to what extent one selects the option with the higher value. If the inverse temperature is low, humans are more prone to selecting the lower value option. While behaving in this stochastic way may not seem like a good strategy in an environment in which the value of each option is known and does not change, it is highly useful in changing environments (Daw et al., 2006): stimuli that did not have the highest value in the past may become more valuable, which can only be discovered through exploration. Neurally, the ventromedial prefrontal cortex (vmPFC) has consistently been identified as the site of the representation of the learnt values (Chase et al., 2015). Studies have also found that it is in vmPFC that the selection between two options take place: Boorman et al. (2009) found, using fMRI, that the vmPFC had information about the value differences between two choice options and Hunt et al. (2012) found using MEG, a human neural imaging technique with high temporal resolution, that vmPFC activity, during choices between explicitly shown valued options, is in agreement with a system in which the values of two options compete with each other to determine choice. These findings suggest

that the vmPFC's crucial role in the learning and decision process is to use the learnt information to decide which option to select. As exploration in itself also has a value that is separate from the learnt value of the available options, studies have also tried to identify its neural correlates. Daw et al. (2006) found that the lateral anterior prefrontal cortex was more active in human subjects when they decided to explore rather than exploit.

1.1.2 Model-based learning

The Pavlovian learning system is a type of model-free learning as all learning only depends stimulus-outcome associations, rather than depending on a model of the world or other relational aspects of the environment. Another type of model-free learning is habitual learning in which animals learn stimulus-response associations. The advantage of such simple learning is that it is not computationally expensive, thus allowing the animal to make choices rapidly and to extract information without needing to understand a causal structure. However, in many situations, this model-free system alone is insufficient for adaptive behaviours. For example, Hershberger (1986) designed a task in which young chicks were put in front of a food cup that moved away from the chick with twice the speed at which the chick was moving towards it. Therefore, to get to the food cup, the chick needed to overcome the behaviour suggested by the Pavlovian system, namely to approach the food cup, but instead walk away from it. Or in other words, it needed to use the model of the world ('food cup moves away from me at twice my speed') to guide behaviour. As it turned out however, chicks were not capable of this.

While the above is a demonstration of the inability in some circumstances in some animals to behave differently than the Pavlovian system, there is also evidence to the contrary. Early evidence for more model-based mechanisms comes from devaluation experiments (reviewed in Dickinson (1985)). In devaluation, the animal is first trained that selecting a certain option (e.g. picking stimulus X) leads to reward A (e.g. apples), while picking stimulus Y leads to reward B

(e.g. bananas). The animal is then fed the reward it prefers (e.g. A) to satiation. Now, when presented again with the two options, based on the model-free system, one would expect the animal to choose again X, as the model-free system has not had the possibility to learn that the option X is now less valuable (as the animal does not want to eat A anymore). However, what is found is that the animal now selects the other option to obtain the reward for which it has not been satiated. In humans, the ability to devalue stimuli predicting different kinds of food items has been linked to the vmPFC (Valentin et al., 2007). Based on these kinds of results, it has been proposed that there is another learning mechanism that has a model of the world (model-based learning). This system is crucial for allowing more complex behaviour. At the same time, model-based learning is also more computationally expensive and slower, which is why with extensive practice, i.e. when the model-free system has been able to learn to perform the task, behaviour switches from model-based to model-free behaviour (Balleine and Dickinson, 1998; Tricomi et al., 2009).

In contrast to model-free learning, which is algorithmically constrained (and described by the Rescorla-Wagner rule), there are many different facets to model-based learning, which is important to note as these may have very different computational/algorithmic and neural implementations. Here I want to highlight how we already know about several mechanistically distinct aspects that all could be considered model-based: 1) Learning environmental statistics in the form of transition probabilities between environmental states. Here, rewards do not simply reinforce a chosen option (which would be Pavlovian), but instead reinforce the choice of an option that is – because of the environmental transition probabilities – most likely to lead to the same outcome as the current reward. This is maybe the most famous example in the literature of a process that has been termed ‘model-based’. VmPFC and striatum show activity correlating with the model-based learning signals (Daw et al., 2011), while activity in parietal sulcus and dorsolateral prefrontal cortex reflects learning of the model itself (Glascher et al., 2010). Studies

have also shown its relevance to the roles of neurotransmitter systems (Worbe et al., 2015; Wunderlich et al., 2012b) and for psychiatric research (Otto et al., 2013). 2) Learning environmental statistics in the form of how quickly the environment changes (volatility). Here, information about how quickly there is a change in whether one option or another option leads to rewards is used to adjust the rate for learning about how likely an option is to lead to a rewarding outcome. This form of model-based learning implicates the dorsal anterior cingulate cortex (dACC, Behrens et al. (2007)). 3) Learning to assign rewards appropriately in the face of several potential causes. For example, in a typical experiment, outcomes should only be assigned to choices made on that trial, and not to choices made on past trials. This has been shown to rely on lateral OFC in macaques (Chau et al., 2015; Noonan et al., 2010; Walton et al., 2010). 4) Learning from an alternative course of action not taken (counterfactual learning). Here, learning is about reward information for an option that was not chosen. The lateral anterior prefrontal cortex has been proposed as important for this process (Boorman et al., 2011). As this short overview shows, there is no unitary model-based process. Rather, model-based behaviour can be found in many situations and relies on different brain mechanisms. Therefore, it may be more beneficial to study one of these mechanisms specifically, rather than aiming to understand ‘model-based’ learning in general.

What these mechanisms described above have in common is that they deal with learning something about the environment based on a model. However, the next question that then arises is, if the brain has both model-based and model-free learning happening at the same time, how does it decide which of this learnt information should guide decisions? In the context of learning the transition probabilities between environmental states (Lee et al., 2014), this control process has been linked to aPFC, keeping track of the reliability of information carried by these two systems, and to vmPFC, integrating the information from the two systems. However, this question has not been addressed yet in other model-based situations. One such situation of

common behavioural relevance is separating the reward information received into informational content (relevant for learning) and reward content (enjoyable, but not relevant for learning) – e.g. after visiting a supermarket that plays music you like, can you learn how much you liked the supermarket based on its assortment or is your memory biased by the music that has nothing to do with the goods on offer? We address this question experimentally in chapter 2.

1.2 Learning and depression

1.2.1 Learning

Changes in learning have long been proposed to be relevant in depression. Experimentally, over 30 years ago, Seligman (1972) observed ‘learned helplessness’ in animals (rats and dogs). He found that after successive experiences of inescapable punishment, animals were much slower at learning to escape from escapable punishment. In other words, the experience of not being able to control their environment or outcomes within it had changed how they acted and learnt about new contingencies. The concept of learned helplessness has been applied to depression in humans based on studies finding that aversive and uncontrollable life events often preceded the onset of a depression and that depressed people reported, on questionnaire measurements, high feelings of helplessness (recently reviewed in Pryce et al. (2011)). While studies from the 1970s have provided some experimental support (Abramson et al., 1978) for learned helplessness specifically, more recent research has focused on studying learning more generally, i.e. not only in the context of learning to avoid punishments, but also learning about rewards and in situations of selection between courses of action, rather than focusing on learning to control negative outcomes (but see also Diener et al. (2009) and Byrom et al. (2015)). These studies have used simple model-free (both Pavlovian and instrumental conditioning) learning tasks, while no tasks have yet looked specifically at more sophisticated model-based learning. I will review these studies, focusing on the questions 1) does depression change learning or plasticity? 2) are changes in learning specific to certain domains such as a reward or punishment?

Patients with depression perform worse than healthy controls in simple reward and punishment learning tasks such as the Iowa Gambling Task (Cella et al., 2010; Must et al., 2006), non-reversal learning (Herzallah et al., 2013; Pechtel et al., 2013; Thoma et al., 2015; Vrieze et al., 2013) and reversal learning (Mueller et al., 2015; Remijnse et al., 2009; Taylor Tavares et al., 2008). However, it should be noted that in the above-mentioned studies, no specificity of the

differences between groups could be established. Namely, none of the studies above applied a formal learning model to the data, but rather reported simple aggregate measurements. Instead of breaking the behaviour into its components, e.g. learning rate and exploitation/exploration parameters, they either used measures of overall percentage correct responses and how this changed across trials as a measure of learning. Alternatively, they used error rates after positive or negative, correct or misleading feedback as a measure of ‘sensitivity to misleading feedback’ (e.g. Taylor Tavares et al., 2008). However, the problem with these approaches, as highlighted by results of a model-based meta-analysis (Huys et al., 2013), is that without a computational model, different behavioural components can often not be dissociated. Huys et al. (2013) found in a model-based re-analysis of a reward learning task (Pizzagalli et al., 2005; Tripp and Alsop, 1999) that MDD symptoms (in particular anhedonia) were linked to a problem of reward sensitivity (though it was equivalent to the exploitation/exploration or temperature parameter in this task), rather than a problem in simple reward learning (e.g. reward learning rate) per se. Importantly, similar data from the same task (Pechtel et al., 2013; Vrieze et al., 2013) had been interpreted as showing a change in reward learning in depression, based on analyses not using learning models, which instead only compared the percentage of correct responses early in the experiment and after learning. This highlights the importance of using computational models to analyse behavioural data if there is otherwise the (very common) possibility that different cognitive processes may be affecting participants’ responses to stimuli.

In the following, I will focus on studies employing computational modelling approach, dissecting complex learning behaviour into its components, or on studies that used designs where different components could be identified without the need for modelling. In short, it appears that while there has been some behavioural (Chase et al., 2010; Robinson et al., 2012a) and neural (Gradin et al., 2011; Kumar et al., 2008; Robinson et al., 2012a) evidence for general reinforcement learning deficits in depression, studies so far have not been able to rule out

possible confounds. On the other hand, good evidence exists for more specific reinforcement learning deficits in depression relating to learning about probabilities of negative outcomes for the self (e.g. risk of getting cancer), based on information about other people (Garrett et al., 2014; Korn et al., 2014).

Chase et al. (2010), using a non-reversal probabilistic learning task, found that depression at trend level ($p=0.07$) was associated with decreased learning about monetary reward or punishment. This effect was clearer when instead of the diagnosis of depression, the analysis focused specifically on anhedonia. In this case, the speed of learning from both reward and punishment decreased with increasing anhedonia, or in other words the most recent information was used less to update reward or punishment predictions. However, it should be noted that at the same time anhedonia also strongly correlated with participants' exploitation/exploration parameter – depressed participants had a stronger tendency to 'exploit' (i.e. pick the more advantageous option). The same paradigm was used by Kunisato et al. (2012), with the small modification that instead of using visual stimuli for positive (smiling face) or negative (sad face) feedback, they used monetary reward. In contrast to Chase et al. (2010), they found depressed participants not to differ in their learning rate, but to have a stronger tendency to 'explore' (i.e. picking the less advantageous options more often). It is important to note that this exploitation/exploration parameter in the two studies was the same as what Huys et al. (2013) identified as 'reward sensitivity' – in that study depression was also linked to the opposite effect from Chase, increased exploration. Huys et al. (2013) found this parameter to be very strongly correlated with the learning rate (also see Daw, 2009). This suggests that without control procedures such, as applied in Huys et al. (2013) or suggested by Daw (2009), to control for these correlations it is unclear whether in Chase et al.'s results depression really slowed learning, or whether this result was an artefact of the learning rates correlating with the

exploitation/exploration parameter. In conclusion, the three studies touched on in this section do not provide strong evidence for a learning deficit in depression.

Further support for a lack of straightforward learning effects in depression comes from two studies, in the same participants, looking at Pavlovian (Kumar et al., 2008) and instrumental (Gradin et al., 2011) learning, without reversal, based on primary reinforcers (water). Both studies found no change to the learning rate in depressed patients compared to healthy controls. However, using fMRI they found evidence for decreased reward prediction error signals in reward regions such as the ventral or dorsal striatum and the vmPFC. Furthermore, these effects were stronger in patients with stronger symptoms of depression, such as anhedonia. While these differences between neural and behavioural results may simple be due to the increased sensitivity of fMRI, it is also worth noting that both studies looked for neural prediction error signals by using the difference between the received and the expected outcome, rather than testing these two components separately as has been done in other studies (Boorman et al., 2011). Therefore, it is not clear whether the decreased PE signals they found meant that PE signals are smaller in depression or that outcome signals are smaller (reward and punishment processing in depression without learning are discussed in more detail below) or that paradoxically maybe prediction signals are stronger. In conclusion, while these studies may provide evidence for changed neural learning signals in depression, further studies are needed to address the question whether it is really learning rather than outcome signals that are affected by depression.

Another study (Robinson et al., 2012a) that looked at learning in depression used a design that did not necessarily require computational modelling because it used deterministic rather than probabilistic associations between options and outcomes. In this task, participants had to use the learnt information of two options (namely which one led to visual reward and which one led to

punishment) not to select the better of two options, but to instead make predictions about whether a reward would or would not appear when either of the two options was selected. Robinson et al. found that depressed participants had a specific problem in changing their predictions when an option was surprisingly rewarded after a reversal. Neurally, a corresponding effect was found in part of the ventral striatum (ventral putamen, but not nucleus accumbens) where depressed participants showed less activation to surprising rewards than healthy controls. However, it is also of note that in this task participants showed surprisingly low performance for a deterministic task (healthy controls made ~20% errors on trials without reversal), especially compared with other studies using the same task that had non-reversal error rates of only less than 5% (Robinson et al., 2010; Robinson et al., 2012b). This could suggest that participants might have misunderstood the task instructions, which might confound the interpretation of the results. Another caveat in this task was that it is unclear whether participants perceived ‘reward’ feedback as rewarding and ‘punishing’ feedback as punishing because of the mixed Pavlovian-Instrumental nature of the task and because rewards/punishments only consisted of a smiling/sad smiley faces, rather than e.g. winning or losing money. Then, instead of perceiving the feedback as intended, participants might have re-interpreting the feedback in terms of whether it meant that they made a correct prediction (reward) or a wrong prediction (punishment). If this were the case, it would suggest that the behavioural and neural data should have been analysed in this different framework. In conclusion, based on this study, it is unclear whether depression is associated with a learning deficit.

One might ask whether probabilistic learning in a computerized task translates into differences in more ecological real life learning about actual events. Interestingly, two studies measured updating of real life predictions (such as how likely people think they are to get cancer) in the face of positive or negative information (statistical values of how common cancer is in the population that can either be higher or lower than what participants thought). Korn et al. (2014)

and Garrett et al. (2014) both found in two independent samples of depressed patients that they did not show the ‘optimism bias’ found normally in healthy controls. This bias means that healthy controls update their expectation of aversive life events more in the face of positive than negative information. In contrast, depressed participants updated their expectations similarly for positive and negative information, in other words they lack the optimism bias. Importantly, in their elegant design, they could also show that this was not because depressed participants remembered positive or negative information differently from healthy controls. Thus suggesting that it was really the process of updating that was affected. Previously, the same group had shown that the optimism bias in updating related neurally to a smaller prediction error signal for negative information in the inferior frontal gyrus (IFG, Sharot et al. (2011)). In Garrett (2014), they now found that depressed participants differed from controls in them having larger prediction errors for negative information in the same IFG region, as well as in the inferior parietal cortex. Thus, even though this study did not use a typical learning task in which the value of a small set of stimuli is continuously updated, rather than just once, they provide evidence for changed updating (or learning in other words) in depression. Future studies should look at whether this effect is selective to the updating of personal information (rather than any kind of prediction) and whether the effect is there for updating of predictions about positive life events.

In summary, there is some preliminary evidence that depression affects learning. However, it is still unclear what specific aspects of learning are affected. These components include: the speed of learning, sensitivity to rewards and efforts, being able to use the learnt information appropriately and general ability to make good decisions. We address this question in chapter 5, using a novel learning task in which these different components can be dissociated.

1.2.1 Plasticity

After considering behavioural markers of learning in depression, we will now turn to considering neural markers. It has been shown that behavioural learning relates to diverse phenomena on a molecular level: changes in the strength of synapses (long-term potentiation or depression, LTP/LTD), changes to the numbers of synapses and the growth of new neurons. How these markers of plasticity are affected by depression has mostly been studied in animal models of depression, such as the 'mild chronic stress models' (Willner, 2005). In this model, rats or mice are exposed to mild stressors, such as changes to their social environment, room temperature or wet bedding, for a prolonged time of several weeks. The depressogenic effects of this are most commonly measured as symptoms of anhedonia, namely a reduced preference for sugary foods (though not all studies have found this effect (Willner, 2005)). However, recently, studies have also included other, more diverse symptoms to capture symptoms of depression more comprehensively. Further support for the validity of this model of depression is that while experiencing stress is not the same as feeling depressed, there nevertheless is a consistent predictive relationship between minor chronic stress (as well as major stressful life events) and depression (Hammen, 2005).

It has been found that in these models of depression (reviewed extensively in Duman et al. (2012) and Pittenger and Duman (2008)) striking changes occur to diverse brain areas, including atrophies of brain regions, reduced synaptic connections between neurons and reduced neurogenesis (e.g. Bessa et al. (2009)). Note however that while these changes suggest that the brain has changed, it does not necessarily mean that the brain's ability to change (i.e. plasticity) has changed. This is analogous to a situation in which after a prolonged time of learning a new skill (e.g. juggling, Scholz et al. (2009)), the brain's structure in a certain area has changed; in this situation, one would not conclude that the brain is now more plastic in general. In other words, to measure whether animal models of depression have different plasticity, one should

look at more direct markers of how the brain can change - in other words is plastic - in response to stimuli. This can be measured most directly by measuring changes in synaptic strength – LTP or LTD – in response to stimulation.

LTP or LTD are produced by cells in response to coincidence detection (Hebbian learning), through action of glutamatergic NMDA receptors that allow calcium influx into cells (reviewed in Malenka and Bear (2004)). This has diverse effects on different timescales. Very quickly, it leads to activation of proteins already present in the synaptic terminal (e.g. integration of AMPA channels into the synaptic membrane). At longer timescales, it leads to the transcription of mRNA and to changes in gene expression, release of neurotrophic factors and morphological changes at synapses. Recently (reviewed in Lisman et al. (2011)), it has been proposed that while the early stages of LTP may rely mostly on coincidence detection, later stages (on the time scale of >30min) also need – or are at the least strongly enhanced by - the presence of dopamine. In short, there are different mechanisms of LTP/D that might be differently affected by depression.

Studies have found that chronic stress reduces LTP (Cerqueira et al., 2007; Quan et al., 2011), which is in agreement with a reduction of plasticity. However, other studies (Holderbach et al., 2007) have found that chronic stress increases LTD. This is not in agreement with a reduction in plasticity because recent research suggests that LTD itself actually plays an important role in learning related plasticity (reviewed in Malenka and Bear (2004)) – rather than representing the opposite of LTP and thus a decrease in plasticity. Thus, measurements of short-term plastic change (LTP/LTD) do not provide strong evidence for changes in plasticity with chronic stress. On the other hand, studies have found more consistently that neurotrophic factors are decreased in depression (reviewed in Pittenger and Duman (2008)). How could these divergent findings fit together? As noted above, there are different timescales to LTP/D, with the later ones requiring dopamine. Taking into account that anhedonia, which in turn relates to dysfunction in the

dopamine system, is a key symptom of chronic stress animal models, it could be possible that later phases of LTP/D – which relate more to long-term memory – rather than early phases – which may relate more to the online learning behaviour described in human studies above – may be affected by depression. This is further supported by results from animal studies (reviewed in Conrad (2010)) that find that learning is most strongly impaired by chronic stress in appetitive, but much less or not at all in aversive contexts. This could thus be explained as due to reduced reward motivation (as a result of anhedonia), rather than as a primary learning, i.e. plasticity, deficit. On the other hand, in both rewarding and aversive contexts, long-term memory (i.e. retrieval of learning after >30min) is impaired. Thus, this suggests that chronic stress may not affect plasticity in general, but more specifically at longer timescales.

The studies above were all based on animal models of depression. In post-mortem human studies, there is also evidence for reductions in the size of brain areas, as well as decreases in synapses between neurons (Kang et al., 2012). On the other hand, more direct markers of plasticity such as LTP have not been measured directly in humans, however techniques have been developed to measure proxies of it. One of these techniques (Teyler et al., 2005) involves measuring the change in strength in visually evoked (electrical) brain potentials after a plasticity induction procedure. In this kind of experiment, the electroencephalographic (EEG) response to visually presented stimuli (e.g. checkerboard) is measured before and after induction of plasticity. LTP-related plasticity is induced by rapid presentations of the same visual stimuli. That this procedure measures plasticity has been supported by analogous experiments in animals which found changes in this visually evoked response (VEP) to be NMDA-dependent – a hallmark of synaptic long-term plasticity. Normann et al. (2007) have used this technique to measure effects of depression in humans on LTP. While they claimed their results support that depression decreases plasticity in humans, design choices in their experiment may somewhat weaken this conclusion. Specifically, the frequency used to induce potentiation plasticity here

has previously been described as inducing depression rather than potentiation (Teyler et al., 2005). Furthermore, a combination of presenting stimuli very fast at 2Hz and using only two (rather than the usual 32) electrodes meant that they could not actually know whether the VEP recorded showed potentiation or depression. This is also supported by the fact that their VEPs markedly differed in shape from Teyler et al. (2005), suggesting that they measured other processes than the ones for which it has been established that they relate to increased plasticity.

In conclusion, data from depressed patients and animal models of depression shows that depression leads to changes in the structure of many brain areas, including atrophy and reduced connectivity between neurons. There is also a limited amount of evidence that plasticity may be affected, however, further studies are needed to clarify whether this is specific to different types of plasticity or different brain areas.

1.2.2 Valuation / Reward and punishment

As discussed above, it has been noted (Huys et al., 2013) that what appears as learning difficulties in depression may in fact be due to differences in processing of or sensitivity to rewards or punishments. This might be of particular relevance as it may also relate to specific symptoms of depression such as anhedonia, fatigue or problems making decisions. Reward and punishment processing can be studied more directly in tasks without a learning component. This question has been addressed in recent reviews (Bylsma et al., 2008; Eshel and Roiser, 2010; Groenewold et al., 2013; Hamilton et al., 2012; Zhang et al., 2013).

There are different components to the processing of rewards, including valuation (liking/disliking) of reward/punishment, retrieving and then using reward predictions either in order to decide between them or to more generally to motivate actions in order to obtain such rewards.

For the valuation of rewards and punishments, prominent theories have either stated that depression is characterized by an oversensitivity to punishments and a hyposensitivity to rewards or that depression is linked to a hyposensitivity to both rewards and punishments. Earliest evidence for the first hypothesis was based on results that found that compared to controls, depressed participants showed a “catastrophic response to failure” (Beats et al., 1996; Elliott et al., 1996): they were more likely to make another mistake after having already made a mistake compared to healthy controls. However, another way to look at this finding is that in fact healthy controls are less likely to make a mistake after feedback that they had made an error (e.g. because they now pay more attention as an error signals to them that they were not paying enough attention). Therefore it has been proposed that it is only in comparison to these healthy controls that depressed appear to be over-sensitive to negative feedback and in actual fact they instead show less sensitivity to, i.e. less improvement after, negative feedback (Douglas et al., 2009; Steele et al., 2007). For example, evidence for this idea was found by Steele et al. (2007) in a task in which participants believed that they received feedback about whether they had responded fast enough to a target or not. In actual fact, however, the feedback given did not relate to their performance. This allowed studying changes in performance in response to feedback, de-confounded from performance levels. Steele et al. found that healthy controls tried to improve their responses (by responding faster) after negative compared to positive feedback, while this effect was absent in depressed. This would thus support the view of a generally reduced, in other words blunted, response to positive or negative stimuli, rather than an over-sensitivity to negative ones. Similarly, in a meta-analysis of responses to positive and negative stimuli, Bylsma et al. (2008) found that across a diverse set of measures including self-report, behavioural (facial expression, eye blinks, approach-avoidance behaviour) and physiological (heart rate, blood pressure) measures, depressed patients showed reduced emotional reactivity to both positive and negative stimuli.

On the neural level, the question of changes in reward/ punishment processing in depression has also been addressed in several meta-analyses. However, the results have been somewhat inconsistent. Hamilton et al. (2012) found depression to be characterized by an increase in activation to negative stimuli in amygdala and dorsal anterior cingulate (among other regions) and a decrease in activation in the dorsolateral and the dorsal caudate. These findings were mostly replicated in another meta-analysis (Groenewold et al., 2013), which also found decreased dlPFC activity and increases in amygdala (at least when studies using regions of interest rather than just whole-brain analyses were included). However, this study did not find increases in dorsal anterior cingulate but instead in a more posterior mid-cingulate region. These findings have been interpreted as an increase in the ‘salience’ processing (bottom-up) and the decrease of top-down (model-based) control processing of negative stimuli. This would thus agree with increased sensitivity to punishing information, rather than general blunting. However, it should be noted that it is of course difficult to draw conclusions about behavioural/ psychological effects based on neural data (inverse inference), especially in a meta-analyses; for example, as the studies included in the analyses did not specifically instruct participants to control their negative responses to stimuli, it is unclear whether decreased dlPFC activity can be linked to a failure to regulate responses of ‘lower areas’ to negative stimuli. It should also be noted that in the same study (Groenewold et al. 2013), increased responses in vmPFC were found to positive stimuli, which if one used the same line of argument should suggest stronger responses to positive stimuli in depression, which is not in agreement with behavioural findings.

In conclusion, the above-mentioned studies suggest that the processing of rewards and punishments is affected in depression. While it is not entirely clear whether sensitivity to punishments is increased or decreased, it is nevertheless clear that there are changes in the processing of valenced stimuli. Therefore, studies that want to investigate learning need to control for effects of stimulus valuation. Similarly, beyond effects of pure valuation, learning

studies also need to control for potential confounds of decision-making per se (as highlighted also in section 1.2.1): there is some evidence that depression relates to a deficit in decision making (Mannie et al., 2014; Murphy et al., 2001; Smoski et al., 2009). However, these studies did not use a model-based analysis approach, which makes it more difficult to assess whether these deficits are about decision making per se or are a consequence of changed reward and punishment processing.

1.2.3 Motivation and effort in depression

The last component of value-based processing to consider is motivation to obtain rewards or avoid punishments. This is of particular relevance in depression as decreased interest or engagement in activities is a core symptom relating to decreased motivation. While this motivation to get rewards (or ‘wanting’) is of course strongly related to the ‘liking’ of rewards, a large body of studies has highlighted its mechanistic and neural separation (Berridge et al., 2009).

Experimentally, motivation can be measured as willingness to work for (or exert effort to obtain) rewards. Animal studies have highlighted the anterior cingulate cortex (Walton et al., 2003b) and the dopaminergic system (Denk et al., 2005) as important in the processing of effort. In humans, studies have similarly found the dACC to be important (Croxson et al., 2009), and in particular have highlighted that while many other forms of evaluation (e.g. between different reward magnitudes, probabilities or delays) all involve the vmPFC, evaluation of options with different effort and reward do not and instead rely on a network of dACC and insula (Prevost et al., 2010).

In depression, several studies have found a reduced willingness to exert effort to obtain monetary rewards (Treadway et al., 2009; Treadway and Zald, 2011; Yang et al., 2014; Yang et al., 2015), although this has not always been found (Sherdell et al., 2012). While some studies (Clery-Melin

et al., 2011; Treadway et al., 2009; Treadway et al., 2012) have suggest that this is due to a changed valuation of the available rewards magnitudes and probabilities, others (Yang et al., 2014; Yang et al., 2015) instead found it not to relate to the valuation of the available rewards, but rather to be due to an unwillingness to exert effort per se. As an alternative hypothesis to a changed appraisal or valuation of reward, it has been suggested (Brinkmann and Gendolla, 2007; Brinkmann and Gendolla, 2008; Brinkmann et al., 2011) that the amount of effort invested depends on the perception how effective one can be in achieving an outcome and how necessary it is to exert effort to achieve it. In this model, depression is thought to generally increase the assessment of how difficult it is to achieve an outcome. Therefore, it was suggested that depressed participants might perceive relatively ‘easy’ tasks, as harder (compared to controls) and therefore exert more effort. At the same time, ‘difficult’ tasks may appear as too hard to them, making them give up and exert less effort. There is some experimental support for this with Brinkmann and Gendolla (2008) finding systolic blood pressure (as a proxy for readiness to exert effort) being changed in the predicted fashion in a task with varying effort levels. It is, however, not clear how these results sit with the other studies measuring effort detailed above as these studies only manipulated how effortful an option was, while not explicitly manipulate how achievable the effort was. It is however of note that for example in Yang et al. (2014), successful completion of effort trials was only 80%, therefore it is quite possible that the decisions participants made were not only ‘should I make the effort?’, but also ‘will I succeed in making the effort?’. This would then suggest that some finding of depressed participants being less willing to exert effort might have actually been due to them being less confident of being able to exert the effort successfully.

In summary, there is good evidence that motivational processes in the form of making efforts to obtain rewards are affected in depression. However, again, in most real-life situations, choices have to be made based on remembered/learnt information, rather than on explicitly shown

possible reward outcomes and effort requirements. There are no studies yet that have addressed how effort is learnt in depression, and we will address this question in chapter 5.

1.3 Serotonin – emotional processing and plasticity

Although we still lack a mechanistic understanding of depression, we do have treatment methods available that offer some success against depressive symptoms. The most commonly used methods are either psychological or pharmacological treatments. The most routinely used pharmacological treatment affect levels of neuromodulators, such as serotonin, dopamine or noradrenaline. Of these, serotonin is the most targeted through the use of a class of drugs called selective serotonin re-uptake inhibitors (SSRI).

Serotonin is produced and released from neurons with cell bodies in the brain stem raphe nucleus, which has a dorsal and a medial part. Neural projections from the raphe nucleus go to most of the brain, in particular also to regions involved with value-guided decision making and learning. In these regions, serotonin acts on many different receptors, which leads to further diversification of its functions (reviewed in Berger et al. (2009)). Once released into the synaptic cleft, serotonin is taken up again by the serotonin transporter (SERT). These transporters are the target side of SSRIs which block them, thus reducing serotonin re-uptake.

While SSRIs can start improving symptoms of depression within the first week, clinically significant effects take several weeks to emerge (Taylor et al., 2006). This has been initially been proposed to be due to serotonin levels not raising immediately with SSRI treatment because of the action of autoreceptors on the serotonergic neurons. This has in turn been suggested to reduce the neurons' activity and thus serotonin release initially. When autoreceptors become re-absorbed after some time, the activity of serotonergic neurons increases while tonic serotonin levels also increase over baseline levels (Artigas et al., 1996). However, later studies in both monkeys (Anderson et al., 2005) and rats (Beyer and Cremers, 2008) have found that when serotonin levels themselves (rather than serotonin neuron activity) are measured, increases are

found immediately with SSRI treatment. Therefore, it seems unlikely that delayed increases in serotonin levels are responsible for the delay in treatment action.

Instead, alternative theories have emerged that explain the delay of treatment effect and the action of serotonin by for example focusing on **emotional biases** that are remedied and unbiased perception taking time to manifest themselves in changed attitudes/mood. Other focus on more basic molecular/cellular processes such as increasing **brain plasticity**. In the following, I will review both of these theories in more depth, with a particular focus on how they relate to value learning. As there have been many studies carried out looking at these phenomena in similar designs, I will try to give an over-view of a large proportion of these studies supporting or not supporting these theories, rather than focusing on details of single studies.

1.3.1 Emotional processing and antidepressants

The main tenet of this hypothesis is that depression is associated with negative biases in the processing of emotional information and that these biases can be reversed by antidepressants such as SSRIs (Harmer and Cowen, 2013). In this framework, the treatment delay effects is explained as follows: SSRIs very rapidly produce changes in emotional processing, but it is only through repeated interactions (over weeks) with their environment that patients *learn* a new, more positive, world view which then leads to an improvement in mood.

Emotional processing in this sense refers most broadly to the processing of stimuli that have either positive or negative valence. Types of emotional stimuli that have been studied in the context of serotonergic antidepressant theories are very varied including emotional words or pictures or faces with emotional expression. These types of studies (see below) have been suggested to show that serotonin is involved with emotional processing in that it is related to a balance between positive and negative processing, with increased serotonin levels – through

administration of an SSRI - producing more positive emotional responses. On the other hand, there exists an extensive animal literature, as well as human studies on serotonin using a procedure called acute tryptophan depletion (ATD) which temporarily lowers serotonin levels, that have instead focused on another kind of emotional stimuli namely rewards or punishments. This line of research initially led to very different conclusions about serotonin, see below, suggesting that it plays a role in the coding of negative stimuli or in linking them to actions and that a decrease in serotonin hampers the processing of negative stimuli.

In the following, I will review studies about emotional processing with serotonergic manipulations. The aim will be to answer the questions: Firstly, do SSRIs influence emotional processing? Secondly, does this differ between different kinds of emotional stimuli? Thirdly, is the shift in emotional processing so that positive emotions are increased and negative ones decreased? Fourthly, do these effects differ between single and prolonged administration?

1.3.1.1 Emotional pictures, words, faces

Many studies have been performed examining the effects of acute SSRI administration or ATD on the processing of emotional pictures, faces or words. The questions they ask are: Do SSRIs bias processing of emotions in general? Is there a more specific effect, namely a shift towards more positive processing?

Acute changes in serotonin

A meta-analysis of 7 fMRI studies (Outhred et al., 2013) revealed that acute administration of SSRIs changed the neural responses to emotional faces and pictures. Specifically, Outhred et al. found that SSRIs decreased responses in amygdala, thalamus and basal ganglia to emotional faces and in amygdala, putamen and insula to emotional stimuli. On the other hand they also

found increases in activity in thalamus and caudate. While this suggests that SSRIs affect emotional processing, it does not speak to the question whether processing is shifted towards enhancing positive and decreasing negative emotional processing. This question however, has been addressed in other studies (see table 1 for an overview of results split into different types of emotional stimuli). Neurally, there is evidence that a single dose of an SSRI enhances processing of positive emotional stimuli and decreases processing of negative emotional stimuli in ‘limbic regions’ (Anderson et al., 2007; Anderson et al., 2011; Murphy et al., 2009; Outhred et al., 2014; Takahashi et al., 2005) such as the amygdala and that it shows the opposite pattern in more prefrontal regions (Bruhl et al., 2010). However, there is also evidence of the opposite pattern² (Anderson et al., 2007; Takahashi et al., 2005), or of no effects (Bruhl et al., 2011; Murphy et al., 2009).

² In the case where a single study had several measurements, it could at the same time provide evidence in agreement and disagreement with the theory.

Acute SSRI					
Faces	Drug	Duration	Emotions used	Task	Result
Anderson (2007)	7.5mg cit	acute, infusion	A, D, F, N	BOLD	No effect for angry ↓BOLD in amygdala to fear ↑left insula, ↓left amygdala to disgust
Anderson (2011)	7.5mg cit	acute, infusion	F, H, Sa		↓amygdala to fear No significant effect for happy or other areas
Bigos (2008)	20mg cit	acute, infusion	A, F, S	Match to sample, BOLD	Amygdala overall more active, no other analysis
Bhagwagar (2004)	10mg cit	acute, infusion	A, D, F, H, Sa, N	Expression recognition	Depressed participants: ↓fear Healthy: ↑fear No effects on other emotions reported
Browning (2007)	20mg cit	acute, oral	A, D, F, H, Sa, Su, N	Expression recognition	↑fear No effect for happy
Capitao (2015)	20mg fluo	acute, oral	A, D, F, H, Sa, Su, N	Expression recognition	↓sad, ↓anger Trend for ↑happy (p=0.08) No effect for fear
			F or H vs N	Attentional dot-probe	No effect
Harmer (2003)	10mg cit	acute, oral	A, D, F, H, S, N	Expression recognition	↑fear ↑happy
Murphy (2009)	20mg cit	acute, oral	A, D, F, H, Sa, Su, N	Expression recognition	↑happy ↓disgust No effect for fear
			F, H, N	BOLD to (un)masked face	↓fear vs baseline in amygdala in unmasked condition No effect in masked condition No effect to happy in either condition
Emotional words					
Browning (2007)	20mg cit	acute, oral	self-referential, +/-	Endorse/ discard	No effect
			Social + or - vs =	Memory	No effect
				Attentional dot-probe	↑bias towards positive words
Emotional pictures					
Browning (2007)	20mg cit	acute, oral	IAPS,+/-/=	Startling to incidental noise	↑startle, not specific to emotion
				BOLD during cue predicting picture	↑BOLD in DLPFC, IFG when predicting negative vs. neutral
Bruhl (2010)	40mg cit	acute, oral	IAPS,+/-/=	Rating of picture valence/ intensity	No effect
				BOLD while viewing picture	No significant effect
Bruhl (2011)	40mg cit	acute, oral	IAPS,+/-/=	Startling to incidental noise	Abolish emotion effect on startle
Capitao (2015)	20mg fluo	acute, oral	IAPS,+/-/=	Rating of picture valence/ intensity	No effect
Kemp (2004)	20mg cit	acute, oral	IAPS,+/-/=	BOLD	↑BOLD in amygdala and IFG to positive ↓BOLD in amygdala and ↑in IFG to negative
Outhred (2014)	20mg cit	acute, oral	IAPS,+/-/=	BOLD	↓amygdala to negative ↓OFC to negative
Takahashi (2005)	50mg fluv	acute, oral	IAPS,+/-/=	BOLD	

Table 1. Studies looking at effects of acute SSRI administration. Abbreviations: Emotions: anger (A), disgust (D), fear (F), neutral (N), sad (Sa), happy (H), surprise (Su), International affective picture systems (IAPS, photographs with positive (+), negative (-) or neutral (=) valence), dorsolateral prefrontal cortex (DLPFC), inferior frontal gyrus (IFG), orbitofrontal cortex (OFC), citalopram (cit), fluoxetine (fluo), fluvoxamine (fluv). Colours: in agreement with (green), in disagreement with (orange), not as expected from (yellow) and not informative about (white) emotional processing theory.

For tryptophan depletion (see table 2), studies have found that it enhances the processing of negative emotions, especially of fear, in the amygdala (Cools et al., 2005b; van der Veen et al., 2007) or of fear and sadness in dorsal anterior cingulate cortex/ dmPFC and its connection to the amygdala (Robinson et al., 2013). But studies have also found that ATD increases (Robinson et al., 2013; van der Veen et al., 2007) or decreases (Fusar-Poli et al., 2007) BOLD in the amygdala, *independent* of whether the emotion is positive or negative.

ATD			
Faces	Emotions used	Task	Result
aan het Rot (2010)	A, D, F, H, Sa, Su, N	Expression recognition	↓surprise No effect for other emotions
Cools (2005)	H, F, N	BOLD amygdala	↑BOLD in amygdala to fear vs happy in threat sensitive volunteers
Daly (2010)	H, Sa, D, F	BOLD	Interpretation unclear as comparisons relative to fixation cross rather than neutral face
Fusar-Poli (2007)	H, N, Sa	BOLD	↑BOLD in dmPFC, ↓ in occipital and parietal to sad ↑BOLD in IFG and left middle temporal to happy ↓BOLD in amygdala to emotional vs. neutral, but not to happy vs. sad
Harmer (2003)	A, D, F, H, Sa, Su, N	Expression recognition	In females only, ↓fear recognition No effect on other emotions or in male volunteers
Hayward (2005)	A, D, F, H, Sa, Su, N	Expression recognition	Recovered depressed participants: ↓happy Healthy: ↑happy No effect on other emotions
Robinson (2013)	H, F	BOLD	↑BOLD in dACC/dmPFC to fearful faces and ↑connectivity with amygdala
van der Veen (2007)	D, F, H, S, N	BOLD	↑BOLD especially to fear vs. happy in people with high threat sensitivity or with ATD mood effect ↑BOLD to fear vs. happy in dmPFC/vmPFC in people with ATD mood effect
		Rate emotion strength	No effect
		Rate face gender	ATD ↓mood in subjects with depression family history. This correlates with decreased accuracy for sad, fearful and disgusted faces
Emotional words			
Coull (1995)	Panic, depressive, neutral, anxious	Distractor in focused attention task	No effect
		Distractor in attentional search	No effect of emotional distractor
Hayward (2005)	self-referential +/- words -/=	Endorse/ discard Memory	No effect
		Emotional stroop	No effect ↑Interference of negative words
Klaassen (2002)	words +/-/=	List learning	No immediate recall effect, after 30 min: ↓recall of neutral words, next day: ↓recall of positive words
Munafo (2006)	word -/=	Emotional stroop	↑Interference with negative words in group of depressed participants No effect in a healthy control group
Murphy (2002)	words +/-	Affective go/no-go	↓RT to positive words, no change to negative words
Emotional pictures			
Hayward (2005)	IAPS, +/-/=	Startling to incidental noise	Recovered depressed: ↑startle, not specific to emotion
			Healthy: no effect

Table 2. Studies looking at effects of tryptophan depletion. Abbreviations: Emotions: anger (A), disgust (D), fear (F), neutral (N), sad (Sa), happy (H), surprise (Su), International affective picture systems (IAPS, photographs with positive (+), negative (-) or neutral (=) valence), dorsomedial prefrontal cortex (dmPFC), inferior frontal cortex (IFG). Colours: in agreement with (green), in disagreement with (orange), not as expected from (yellow) and not informative about (white) emotional processing theory.

Thus, while there is some evidence that acute manipulations of serotonin (either increases or decreases) affect emotional processing neurally, evidence that it is indeed a shift from negative towards more positive processing is less clear. Furthermore, as many of the studies above did not require participants to use the emotional information for behaviour, it is also important to look at studies investigating the behavioural relevance of these changes.

Based on the emotional theory, it would be expected that increases in serotonin should lead to increased behavioural responsiveness to positive and decreased responses to negative stimuli.

Behavioural studies (see table 1) have found that a single dose of SSRI can increase the recognition of (Harmer et al., 2003a; Murphy et al., 2009) or attention to (Browning et al., 2007) positive facial emotions and decrease the recognition of negative emotions (Bhagwagar et al., 2004; Capitao et al., 2015; Murphy et al., 2009). But also that it can increase the recognition of negative facial expressions (Bhagwagar et al., 2004; Browning et al., 2007; Harmer et al., 2003a) or have no effect on categorization or memory of emotional words (Browning et al., 2007) , attention to emotional faces (Capitao et al., 2015) or rating of emotional pictures (Kemp et al., 2004). For tryptophan depletion, it would be expected based on the emotional theory that ATD would lead to an increased responsiveness to negative emotional stimuli. This has been partly supported (see table 2) with studies finding that ATD increases the interference of negative emotional stimuli in healthy (Hayward et al., 2005) or depressed (Munafò et al., 2006) participants and decreased the recognition of positive facial expression in recovered depressed patients (Hayward et al., 2005). On the other hand, other studies have found no change in interference by negative emotional stimuli in healthy participants (Coull et al., 1995; Munafò et al., 2006; van der Veen et al., 2007), no effect on emotional memory (Hayward et al., 2005), decreased recognition of negative facial expression such as fear (Harmer et al., 2003b), increased (Hayward et al., 2005) or faster recognition of positive facial expressions or faster responses to positive words (Murphy et al., 2002) in healthy volunteers.

In summary, the studies reviewed here suggest that acute SSRI administration or tryptophan depletion affects the neural processing and behavioural responses to of emotional faces, pictures or words. However, the studies also suggest that there is not always a clear shift away from negative towards more positive emotional processing with increases in serotonin or vice versa with decreases. It has been suggested that this pattern may be related to clinically observed effects where acute SSRI can increase feelings of anxiety (Harmer and Cowen, 2013). It is also of note that effects seem to differ between healthy and depressed or at risk participants.

Prolonged changes in serotonin

Beyond this acute timescale, some studies have also looked at the effects of longer-term administration of SSRIs (between 1 and 4 weeks). These studies therefore should give a clearer indication of how increased levels in serotonin affect emotional processing. However with the caveat that after a prolonged administration, SSRIs also affect other neurotransmitter systems such as dopamine (Dremencov et al., 2009) or different aspects of neural plasticity (discussed later). Overall, studies (table 3) at longer timescale have produced more homogeneous results, finding that prolonged administration of SSRIs either decrease amygdala responses to negative facial expressions, especially fear (Arce et al., 2008; Harmer et al., 2006) or increase in activity to positive facial expressions (Norbury et al., 2009). Behaviourally, studies suggest a decreased processing of negative stimuli: they found a decreased negative emotional startle response (Grillon et al., 2009; Harmer et al., 2004), decreased recognition of negative facial expressions (Harmer et al., 2004; Harmer et al., 2006), increased memory for positive words (Harmer et al., 2004) and a shift in endorsement of negative/positive adjectives about the self and others towards more positive (Matthews et al., 2010). However, studies have also found no effect on categorization of emotional words (Harmer et al., 2004), emotional word memory (Harmer et al., 2004), rating of emotional pictures or of brain responses to emotional pictures (Simmons et al., 2009).

In summary, there is strong evidence to support the view that prolonged administration of SSRIs shifts the processing of emotional, and in particular of social, cues towards more positive.

Prolonged SSRI					
Faces	Drug	Dur.	Emotions used	Task	Result
Arce (2008)	10mg esc	21d	A, F, H	BOLD	↓BOLD in amygdala to fear in participant who show urine markers of medication
Hamer (2004)	20mg cit	7d	A, D, F, H, Sa, Su, N	Expression recognition	↓fear, angry and disgust recognition ↓surprise No effect on happy
Harmer (2006)	20mg cit	7d	F, H	BOLD	↓BOLD in amygdala and vmPFC to fear No effect on happy
			A, D, F, H, Sa, Su, N	Expression recognition	↓fear recognition No effect on happy
Norbury (2009)	20mg cit	7-10d	F, H	BOLD	↑happy in amygdala No effect on fearful
Windischberger (2010)	10mg esc	10d	A, D, F, H, Sa, Su, N	Match to sample, BOLD	↓amygdala to emotional faces vs. neutral shapes
Emotional words					
Harmer (2004)	20mg cit	7d	self-referential, +/-	Endorse/ discard	No effect
				Memory	↑memory for positive words
Matthews (2010)	10mg esc	3w	self and other referential words	Endorse/ discard	↓negative to self, ↑positive to other
Emotional pictures					
Harmer (2004)	20mg cit	7d	IAPS, +/-/=	Startling to incidental noise	↓startle to negative pictures
Simmons (2009)	10mg esc	21d	IAPS +/-/= distractors	Continuous performance interference	No effect
				BOLD	No significant effects at statistical threshold

Table 3. Studies looking at effects of prolonged SSRI administration. Abbreviations: Duration (Dur.): days (d), weeks (w); Emotions: anger (A), disgust (D), fear (F), neutral (N), sad (Sa), happy (H), surprise (Su), International affective picture systems (IAPS, photographs with positive (+), negative (-) or neutral (=) valence); citalopram (cit), escitalopram (esc); Colours: in agreement with (green), not as expected from (yellow) and not informative about (white) emotional processing theory.

1.3.1.2 Primary or secondary rewards and punishments and motivation

Most of the types of emotional stimuli discussed above (facial expression, emotional words referring to the self or others) have a strong social component. However, there are also simpler emotional stimuli without these social components, i.e. primary or secondary rewards and punishments (such as money, food, sexual rewards, effort). Research in both humans and animals has looked at how serotonin plays a role in their processing. Based on studies about these simpler reinforcers or punishers, theories have been proposed that serotonin codes punishments (Dayan and Huys, 2009) or more specifically, inhibition in the context of punishments (Boureau and Dayan, 2011). According to this theory, decreases in serotonin (by e.g. ATD) should reduce the processing of punishments (and mediate inhibition related to that punishment) while increases (by e.g. SSRIs) should enhance it. In addition, this theory predicts

that changes in serotonin should have no effects on the processing of rewards as these are instead signalled by dopamine.

Studies using ATD depletion broadly support the theory of punishment induced inhibition. It has been found (reviewed in Faulkner and Deakin (2014)) that ATD reduces punishment induced inhibition in the context of learning (Cools et al., 2008; Robinson et al., 2012b) and also without learning (Crockett, 2009; Crockett et al., 2012b; Geurts et al., 2013b), while not affecting punishment/risk processing per se (Anderson et al. (2003); Macoveanu et al. (2013); but see also Talbot et al. (2006) and Rogers et al. (1999)) or inhibition (Cools et al., 2005a; Evers et al., 2006b) per se. For short-term SSRI treatments, the evidence is less clear because of design issues - e.g. lack of delays between different phases of the experiment, confounds with reaction time, or unsuitable control conditions - in some of the FMRI studies (Del-Ben et al., 2005; Macoveanu et al., 2014; Marutani et al., 2011). For prolonged SSRI treatment, McCabe et al. (2010) found evidence that, contrary to the punishment-based theory, prolonged SSRI administration reduced the neural responses to primary food punishments.

Despite some promising results on punishment processing and behaviour, the punishment-inhibition theory also predicts that serotonin should have no role in the coding of rewards. However, studies using ATD found that it impaired reward delay discounting (Schweighofer et al., 2008) or the neural activity in ventral striatum associated with it (Tanaka et al., 2007), as well as reducing the neural and behavioural impact of rewards, but not punishments in a reward-pain trade-off learning task (Seymour et al., 2012). Furthermore, studies using prolonged SSRI administration found that increased serotonin decreased the neural responses to food (McCabe et al., 2010) and sexual (Abler et al., 2011; Abler et al., 2012) rewards, again suggesting that serotonin is also involved in the coding of reward value.

In conclusion, while there is some evidence supporting the punishment-inhibition theory of serotonin in humans, studies also clearly show that the functions of serotonin are not limited to connecting punishments with inhibition. Most recently, it has been possible to directly record from or stimulate serotonergic neurons specifically in the raphe nucleus using a combination of optogenetics and electrophysiological recordings which most importantly allows identification of the recorded neurons as being serotonergic or not. Furthermore, primary rewards and punishments can also easily be studied in animals. These studies can therefore shed light on the basic question whether the information coded in serotonergic neurons in the raphe nucleus is about punishments, punishment induced inhibition or rewards. Using these techniques, Liu et al. (2014) found that stimulation of the raphe nucleus is rewarding to rats and that raphe neurons activate to rewards. However, as a caveat they also found that this effect is not actually dependent on serotonin, but rather on glutamate, which is co-released by the dorsal raphe neurons. Instead, they found that the serotonin release itself only played a role in motivating effortful behaviour. In contrast, Fonseca et al. (2015), using a very similar optogenetic protocol, and Miyazaki et al. (2014), using a different optogenetic stimulation frequency, did not find that stimulation of the raphe nucleus had rewarding properties, instead they found that it increased rats' ability to wait for rewards, but without having a general effect on behavioural inhibition. They noted that differences in results compared to Liu et al. might have been related to differences in the light sensitive ion channels used or the precise location of the stimulation site. Lastly, Cohen et al. (2015) very recently found evidence for a large amount of complexity in the behaviour of serotonergic raphe neurons: using a Pavlovian conditioning task, they found that on short (ms) timescales, serotonergic neurons responded to the receipt of punishment, but more to the predictions of rewards than punishments. Interestingly, they also found that on a longer (s) timescale, serotonergic neurons' tonic rate was modulated by whether overall more rewards or punishments were presented recently.

In conclusion, I have reviewed evidence with respect to whether serotonin plays a role in shifting emotional experiences among the appetitive- aversive axis (emotional processing theory) or whether serotonin plays a more specific role in linking punishment to a specific behavioural strategy i.e. inhibition. Studies found that the emotional processing theory is most strongly supported by studies using social cues (such as faces). This could suggest that – independent of other effects serotonin may have on reward and punishment coding or plasticity (see below) - serotonin may play a special role in social processing. This idea is further supported by studies that found that social behaviour could be shifted towards more pro-social behaviour by increasing serotonin (Knutson et al., 1998; Raleigh et al., 1991; Tse and Bond, 2002), a mechanism that could play a role in treatment success. In contrast, an emotional effect, more specifically a shift from more negative to more positive processing, independent of social aspects does not seem well supported: assuming a shift from negative to positive across all emotions presupposes that all emotions share a common valence axis (positive to negative) that is then, with some further processing, translated into specific emotions. This valence axis should however, given that it is so general as to affect all emotions, also be the same valence axis that processes more simple valenced stimuli (e.g. rewards or punishments such as rewards or punishments). But as the studies reviewed above showed, there is no clear evidence that serotonin levels shift how positively or negatively these kinds of stimuli are processed.

With respect to the punishment-inhibition theory, while there is evidence, mainly based on tryptophan depletion studies, that serotonin plays a role in linking punishment to inhibition, other studies show that the role of serotonin goes beyond that, with serotonin also being involved in responses to rewards.

Based on the studies above, it seems clear that serotonin is involved with the coding of some rewarding or punishing aspects of outcomes. Therefore, when serotonin levels are changed for a

prolonged time, as with SSRI treatment, one would expect this to lead to re-learning of the associations between stimuli and outcomes. For example, if outcomes are perceived as more rewarding, this should lead to increased learning about the stimuli/options predicting these outcomes until it is learnt that these stimuli lead to more reward. In this way, learning is a key process in the emotional theory of antidepressant action, based on the changes in stimulus values. This learning (similarly to increased reward sensitivity as discussed above (Huys et al., 2013)), however would not actually be due to a change in the learning process (learning rate) per se. In the next section we will now look at the plasticity theory of serotonergic antidepressants, which on the contrary posits that changes in the learning process per se are key for antidepressant actions.

1.3.2 Plasticity and serotonergic antidepressants

The second theory of serotonergic antidepressant action that we will consider is the **plasticity theory** (Castren, 2013). According to this theory serotonergic antidepressants, over the course of several weeks of administration, increase the plasticity of the brain, i.e. how able it is to change its structural and functional properties in response to stimuli. Or in other words how able the brain is to learn from experience. These changes in plasticity are thought to then lead to changes in depressed mood. The exact link between plasticity and mood is not clearly determined in this theory; it could either be a direct read-out ('a plastic brain is a happy brain'). Or increased plasticity could change how humans learn about value in the world which leads to a **re-learning** about how good or bad the world is, which eventually leads to a more positive view of the world and then improved mood. In the following, I will review studies that support that serotonin plays a role in neural plasticity in animals, evidence whether this may also be relevant for humans and finally evidence how this may relate to value learning.

1.3.2.1 SSRI induced plasticity in the animal brain

Although it had been held as a dogma among neuroscientist for a long time that the adult brain is very un-plastic and unchangeable, the most recent decades of research have shown how ubiquitous plasticity is and how essential it is for many brain functions, from skill acquisition to potentially even working memory. In fact, neural plasticity is not one unitary mechanisms but can happen on many different levels: new neurons can grow, connections between neurons can grow or change, dendritic spines can move, synapses can get lost or be created or their properties, such as ion channel composition can change. There is a wealth of studies that have found that increases in serotonin can influence these different forms of plasticity. Prolonged SSRI administration has been found to lead to morphological changes in hippocampus, more specifically, it has been shown to lead to either an increase in the number of new neurons born in the hippocampus (Bessa et al., 2009; David et al., 2009) or to a reversal of maturity, making existing neurons in the hippocampus again more immature and thus plastic (Kobayashi et al., 2010). There has been evidence that this SSRI-driven hippocampal neurogenesis is relevant for some depression related behaviours (Bessa et al., 2009; David et al., 2009) such as novelty suppressed feeding, a measure of anxiety. However, the same studies also found that hippocampal neurogenesis was not needed for many other beneficial effects of SSRIs, such as sucrose preference (related to anhedonia), the forced swim test (related to learnt helplessness) or the open field test (measuring anxiety).

Beyond effects on neurogenesis, SSRIs have also been shown to affect the structure of dendrites, i.e. the parts of the neuron with which it receives inputs from other neurons and of synapses. Specifically, SSRIs have been found to increase how easily neurons can undergo long-term potentiation (Jitsuki et al., 2011; Karpova et al., 2011), i.e. strengthening of synapses or connections between neurons following appropriate stimulation. Whether this effect is mainly driven by direct action of serotonin on glutamatergic receptors (Jitsuki et al., 2011) or via

indirect actions on either levels of GABA and glutamate that regulate neural excitability or neurotrophic factors such as BDNF (Vetencourt et al., 2008) has not been established completely yet.

Importantly, when using SSRIs to change serotonin levels artificially, we might be tapping into a neural mechanism that the brain naturally uses for example in order to increasing plasticity in response to environmental enrichment (Baroncelli et al., 2010) or increased reliance on tactile sense (Jitsuki et al., 2011). Behaviourally, artificial and natural changes in plasticity have been found to relate to visual learning (Bachatene et al., 2013; Baroncelli et al., 2010; Vetencourt et al., 2008) and fear extinction (Karpova et al., 2011).

In conclusion, based on research in animals, mostly in rats, there is strong evidence that serotonin is naturally involved with regulating plasticity and that prolonged increases in serotonin levels by SSRIs can modulate plasticity. One can speculate that an advantage of having this serotonin system to control plasticity is that it could allow animals to adapt more quickly when change to their environment takes place on a large scale (e.g. being placed in a new environment, the environment getting enriched, or changed sensory input after losing one's sense of sight). Tonic serotonin level increases could then make the neural system more plastic in general, to learn a new model of the world. This could be in contrast to dopaminergic PEs that signal when a specific outcome for a specific prediction is unexpected, which would allow fine-tuning of an already existing model of the world. With respect to mood, improved mood could be directly linked to increased plasticity: increased plasticity is only useful if the agent interacts with the world, allowing it to learn a new model. One way to achieve this increased interaction would be to improve mood as depression is characterized by symptoms such as withdrawal from interactions. Therefore, it would not be surprising if improved mood – and thus increased interaction with the world - was directly linked to serotonin levels.

1.3.2.2 SSRI induced plasticity in the human brain

While many animal studies have looked at the effects of SSRIs on plasticity, evidence of these effects in humans is sparser. The reason for this is that it is not possible in humans to have direct markers of plasticity (such as neuron dendrites in living tissue or how easily LTP can be introduced in neurons). However, there are different more indirect markers that have been proposed and used to look at humans with somewhat less clear results.

As suggested by some of the above animal studies, plasticity can be controlled by the level of excitation and inhibition, which relates to the neurotransmitters glutamate and GABA respectively (reviewed in Bavelier et al. (2010)). As an example, in the visual system, the postnatal period of increased plasticity is ended by a shift towards inhibition by increased GABA levels. GABA and glutamate can be measured in the human brain using a technique called magnetic resonance spectroscopy. If SSRIs in humans increase plasticity by affecting the excitation-inhibition balance, this should show either as a decrease in GABA levels or as an increase in glutamate levels. Results of these studies have broadly not been in agreement with these predictions: studies found that prolonged administration of SSRIs either had no effect on GABA levels (Godlewska et al., 2015; Selvaraj et al., 2006; Taylor et al., 2008) or even increased GABA levels (Sanacora et al., 2002). Similarly for glutamate, while one study found an increase (Taylor et al., 2008), others found no effect (Taylor et al., 2009; Taylor et al., 2012). It should be noted that all of these studies had been performed at MRI field strengths of 3T and clearer results may be obtained in future studies using improved acquisition protocols at 7T. Another, more indirect, way of measuring GABA levels is using transcranial magnetic stimulation (TMS) wherein the effect of single or paired (at different inter-pulse intervals) magnetic pulses on the evoked muscle (motor) activity is measured. Using this technique, Gerdelat-Mas et al. (2005) found prolonged SSRI treatments to increase interneuronal glutamate activity, but not to affect GABA activity.

As described earlier (section 1.2.1), another way of measuring plasticity in humans that has been proposed is measuring proxies of LTP using EEG (Teyler et al., 2005). Normann et al. (2007) have used this technique to measure SSRI effects in humans. While they claimed their results support that SSRIs enhance plasticity in humans, design choices in their experiment may somewhat weaken this conclusion, as discussed before. Another approach to measuring a proxy of LTP was chosen by Batsikadze et al. (2013). They measured the effect of prolonged SSRI administration on paired-associate stimulation (PAS)-induced plasticity. The NMDA-dependent LTP-proxy in this paradigm is the extent to which pairing between peripheral nerve stimulation and TMS over the corresponding motor cortex area enhances or decreases the ability of later TMS pulses to this area to produce muscle responses. They found that citalopram decreased the inhibitory effects of PAS, while at trend-level increasing the excitatory effects, suggesting that increased serotonin increased plasticity.

In conclusion, there is strong evidence that prolonged SSRI administration enhances plasticity in animals and some preliminary evidence of related effects in humans. In the next section, we will look in more detail at whether and how these neural effects of plasticity relate to behavioural effects in learning.

1.3.3 Reward and punishment learning in the light of emotional biases and plasticity

In both the different emotional and the plasticity theories of serotonergic antidepressant, learning plays a crucial role. However, they differ in how exactly learning is important. According to the emotional (or the punishment based) theory, changes in learning are due to a change in the coding of outcome valence or salience. In contrast, the plasticity theory predicts that learning

itself, independent from any changes to outcome coding, should be modified. In the following, I will review the relevant literature, focusing on the following critical tests that should tell the theories apart: 1) Are reward and punishment learning affected equally or differently? 2) Can learning be affected independent from outcome processing? 3) Are the effects dependent on the duration of serotonin level changes? This is important because plasticity changes are thus far only predicted to happen later, whereas biases could potentially emerge early.

1.3.3.1 Acute modulations of serotonin levels

Several studies have been performed in both humans and animals looking at the effects of short-term manipulations of serotonin levels on learning. These studies are not able to answer the question whether serotonin-related plasticity can affect behavioural learning, as these plasticity changes need prolonged administration. However, if changes in learning are driven by changes in valence processing, then they should also be seen at this point. Several studies have found that changes in serotonin affect learning about punishments (Chamberlain et al., 2006a; Cools et al., 2008; Robinson et al., 2012b). However, the direction of effects was not consistent across studies: Bari et al. (2010) found that acute SSRI administration improved learning by reducing over-shifting after negative feedback. Even this was only at high SSRI doses, at low doses the opposite effect was seen. Chamberlain et al. (2006a) on the other hand found the opposite effect of the high doses, namely that a single dose of an SSRI led to over-shifting after negative feedback. Furthermore, in line with Chamberlain et al.'s findings, Cools et al. (2008) and Robinson et al. (2012b) both found that ATD led to enhanced punishment prediction in a learning task. However, they later (Geurts et al., 2013b) re-interpreted their data as maybe rather being due to a release of punishment-induced behavioural inhibition than an actual learning effect (in line with e.g. Crockett et al. (2009) discussed above). In the same paper, they also found that ATD had a specific effect on learning to prevent a response to receive a reward. However, Guitart-Masip et al. (2014), using a task in which they could also measure learning to

go/no-go to get reward or avoid punishment, failed to find similar effects. Similarly, several other studies (Evers et al., 2005; Murphy et al., 2002; Schweighofer et al., 2008; Seymour et al., 2012; Worbe et al., 2015) all failed to find effects of serotonin ATD on learning from rewards or punishments.

In conclusion, while there is some evidence that acute manipulations of serotonin might affect learning, overall the evidence for this effect is weak, beyond serotonin's putative role in more general affective processing. However, it is of note that none of the studies above used social cues; as discussed above, social processing appears to be particularly affected by serotonin levels. Therefore it would be worthwhile for future research to examine the effects of acute serotonin on learning in a social context.

1.3.3.2. Prolonged modulations of serotonin levels

It is only after prolonged changes of serotonin levels, that one would expect to see changes in plasticity and related learning. Animal studies have provided strong evidence that learning without a clear value component is affected by chronic serotonin levels, including visual learning (Baroncelli et al., 2010; Vetencourt et al., 2008), tactile learning (Jitsuki et al., 2011) and barrier reaching (Walker et al., 2006). More specifically, they found that chronic increases in serotonin levels enhance learning, while decreases hinder it and that it was specifically the changes in serotonin levels that led to learning (Baroncelli et al., 2010; Jitsuki et al., 2011; Vetencourt et al., 2008). In terms of reward-based learning, a series of studies from the Roberts lab, using strong and prolonged depletions of serotonin levels, have suggested that 5-HT is needed for reward learning (Clarke et al., 2004; Clarke et al., 2005) and have more specifically hypothesized that serotonin in the OFC is important for learning to flexibly stop responding to stimuli that are very salient, e.g. because of visual attributes or because they were rewarded in the past (Clarke et al., 2007; Walker et al., 2009). Related results were found by Karpova et al. (2011) who found that

prolonged treatment with SSRI facilitated fear extinction and that this effect was related to molecular changes in the amygdala. However, another study from the Roberts lab (Rygula et al., 2014) on the other hand found that serotonin in amygdala and OFC is important for both reward and punishment learning³. In accordance with this finding, Bari et al. (2010) found that prolonged systemic increases in serotonin levels in rats improved reversal learning with no difference between changes to reward or punishment learning. Lastly, in humans only one study has addressed this question: Palminteri et al. (2012) found that in OCD patients, who happened to be treated with an SSRI or without, both reward and punishment learning behaviour was improved. However, because of the naturalistic study design (patients were not randomised to treatment), it is unclear whether these differences were really because of the SSRI or other differences between these groups.

In conclusion, there is strong evidence that long-term serotonin levels are important for different kinds of learning and that this relates to neural markers of plasticity. Furthermore, there is evidence that it can affect both learning about rewards and punishment similarly in animals, which is more in agreement with the plasticity, than with the emotional processing theory. At the same time, the evidence from human studies is very limited. However, the clearest test pitting the theories against each other, namely whether learning changes can be independent of changes to outcome processing has not been examined in any study yet. This question is addressed experimentally in chapter 3. There I find that prolonged increases in serotonin in humans can indeed enhance neural learning signals for both rewards and punishment in humans, in the absence of effects to neural outcome signals.

³ Differences between these studies may be due to the use of probabilistic (Rygula, 2014) or deterministic tasks, which differ in whether feedback is always veridical or can be misleading. Thus, in deterministic tasks, learning rates should always be maximal, while in probabilistic tasks, they should be smaller, i.e. more trials than just the last one should be remembered. Therefore, different cognitive learning processes may be involved in the tasks.

Overall, the studies reviewed here suggest that SSRIs have effects on both the processing of emotional social cues, as well as on plasticity and related learning. The answer which effect predominates long term also has practical/ clinical relevance. If one effect of SSRIs is to increase brain plasticity and make the brain more able to learn per se, then it is conceivable that it could make depression worse rather than better: if the environment that patients are in is negative, would they not learn even more about how negative it is? This view (Branchi, 2011) has been supported by a meta-analysis that found that the combinations of SSRIs with psychological therapy works better as treatment than SSRIs alone (Cuijpers, 2004) and by a study (Branchi et al., 2013) that found that SSRI treatment worsened symptoms of anhedonia in an animal model when the environment became more aversive (though this effect was not found in other studies (Bessa et al., 2009)). On the other hand, it is also possible that instead of just making learning faster or more flexible, plasticity itself could improve mood ('a plastic brain is a happy brain'). If this was the case, then other medications that directly target plasticity, but have none of the effects of SSRIs on social processing should also act as antidepressants. One way in which plasticity could be influenced is via the NMDA receptors that are involved with LTP and thus plasticity in all areas of the brain. In the next section, we will turn to review evidence for drugs other than SSRIs that modulate NMDA receptors and therefore potentially serotonin independent plasticity.

1.4 NMDA receptors and the plasticity hypothesis of depression

If plasticity is important in the treatment mechanism of serotonergic antidepressants, then other medications that also modulate plasticity, without having other effects on social or emotional processing, should also have antidepressant effects. One classical candidate for modulating plasticity is the NMDA receptor (Morris, 2013). Thirty years of research have identified the NMDA receptor as playing a crucial role in allowing plastic changes at synapses and related learning. NMDA receptors are a type of glutamatergic receptor that can induce learning when activated, using the computational principle of co-incidence detection, by allowing the influx of calcium ions that then lead to other processes such as synaptic remodelling (e.g. trafficking of AMPA receptors to the synapse or release of BDNF) which is involved with strengthening (LTP) or weakening (LTD) of synaptic connections. Thus, drugs that stimulate NMDA receptors could lead to increased plasticity. While it is not possible in living animals or humans to directly activate NMDA receptors with drugs because of problems with cell-toxicity, it is still possible to use partial agonists. One of the few licensed drugs with this property is d-cycloserine (DCS), a tuberculosis medication, which binds to the NMDA receptor's glycine binding site as a partial agonist. This means that at low concentrations, it makes it easier for the NMDA channel to be opened (i.e. become activated), while having the opposite effect at high concentrations (Lanthorn, 1994). In the following, I will firstly review evidence on whether DCS has an effect on depressive mood and secondly whether d-cycloserine can affect learning (as a proxy of plasticity).

1.4.1 D-cycloserine and depression

Not many studies have investigated whether d-cycloserine can affect depressed mood. The earliest evidence was found accidentally by Crane in the 1960 (described in Krystal et al. (2013)) who noted that tuberculosis patients treated with high doses of d-cycloserine (thus maybe

leading to NMDA blockade rather than activation) showed improvements in mood. Since then, only two trials have been performed with Heresco-Levy et al. (2006) finding that a dose of 250mg/day (in addition to patients' usual treatment) did not improve depression symptoms in treatment-resistant depression while the same group also found that a higher dose of 1g/day improved symptoms in treatment-resistant depression (Heresco-Levy et al., 2013).

On the other hand, the NMDA receptor antagonist ketamine has recently found in a meta-analysis to show promise as a treatment in depression with single dose regimes improving depressed mood for about one week (McGirr et al., 2015). While this could suggest that reducing – rather than increasing – plasticity is a useful treatment in depression, animal studies (reviewed in Duman et al. (2012)) have highlighted that a down-stream effect of ketamine is actually to increase markers of synaptic plasticity. Specifically, a single administration of ketamine has been found to increase synaptogenesis very rapidly (Li et al., 2010). This in turn was related to remedying the effects of chronic stress in a rodent model of depression (Li et al., 2011). This together with the fact that the half-life of ketamine is relatively short (~3h, Zarate et al. (2012)) thus suggests that ketamine's antidepressant effects are not mediated via blockade of NMDA receptors.

In conclusion, not enough data exists that could answer the question whether increasing plasticity in the form of stimulating NMDA receptors has antidepressant effects. One clear obstacle in these clinical trials has been that DCS only has effects at high enough doses but on the other hand can act as an antagonist at doses that are too high (Lanthorn, 1994). In other words, good titration is essential for its supposed treatment mechanism. Therefore, one strategy would be to use basic experiments in humans that capture processes related to learning in depression to see at what doses it improves learning and plasticity.

1.4.2 D-cycloserine and learning

Studies in animals have looked at diverse types of learning, finding that DCS can enhance reward learning (Golden and Houpt, 2007; Portero-Tresserra et al., 2013; Torregrossa et al., 2010) and fear extinction (Bouton et al., 2008; Walker et al., 2002). However, some studies also found no effects (Vurbic et al., 2011) or only weak effects (Torregrossa et al., 2010). On the other hand however, studies in humans (table 4) have not been as consistent, with most studies failing to find an effect of DCS on motor learning (Cherry et al., 2014; Feld et al., 2013; Gunthner et al., in submission; Kuo et al., 2008; Kuriyama et al., 2011a), fear learning (Guastella et al., 2007; Klumpers et al., 2012; Kuriyama et al., 2011b) or fear extinction consolidation (Kalisch et al., 2009), while only some found an improvement in motor consolidation (Kuriyama et al., 2011a) or increased fear learning consolidation (Kalisch et al., 2009). Additionally, while DCS has been proposed to enhance CBT for anxiety disorders via its effect on learning (Hofmann et al., 2013; Hofmann et al., 2015), a most recent meta-analysis failed to find such an effect (Ori et al., 2015). However, the power of the meta-analysis may have been reduced as studies were heterogeneous in terms of which diseases were treated, what concentration of DCS was used or at what time (with respect to the treatment session) DCS was administered.

In humans, one study has looked at the induction of plasticity by DCS specifically. Nitsche et al. (2004), using the same paradigm as discussed in the previous section, found that DCS could enhance the putatively NMDA-dependent after-effects of tDCS.

Study	Dose	Timing	Task	Results
Cherry (2014)	250mg	1h before task	Online motor balance learning	No effect in healthy controls or stroke patients on learning
			Online motor skill learning	No effect in healthy controls or stroke patients on learning
Feld (2013)	175mg	after learning	Motor sequence learning consolidation	No effect for consolidation during sleep or when awake
			Emotional picture memory consolidation	No effect for consolidation during sleep or when awake
			Declarative word memory consolidation	↑retrieval if DCS given before sleep retention, not if given before awake retention interval
Guastella (2007)	50mg 500mg	2-3h before extinction	Online, fear extinction	No effect on SCR or rating of extinguished stimulus No effect on SCR or rating of extinguished stimulus
Kalisch (2009)	500mg	learning and extinction	Fear learning consolidation, SCR	↑consolidation of fear learning
			Fear extinction consolidation, SCR	↓fear extinction
		2h before extinction	Online, fear extinction, SCR and FPS	No effect on fear extinction
Klumpers (2012)	250mg	2h before task	Online motor sequence learning	No effect on learning
Kuo (2008)	100mg	1.5h before task	Declarative word memory learning and consolidation	No effect on learning or consolidation
Kuriyama (2011)	100mg	1.5h before task	Online motor skill learning and consolidation	No immediate learning effect, but ↑consolidation at test on next day
Kuriyama (2011)	100mg	1.5h before task	Online fear acquisition and extinction	No immediate effects on fear acquisition or extinction
			Fear consolidation	No effects on fear recall. Effects on re-exposure unclear
Onur (2010)	250mg	4h before task	Online item categorization learning	↑categorization learning task
			Online spatial learning	No effect on online learning.
Gunthner & Scholl (subm.)	250mg	3h before task	Online motor sequence learning	No effect on online learning.

Table 4. Studies looking at the effects of d-cycloserine on learning. Abbreviations: skin conductance response (SCR), d-cycloserine (DCS)

In conclusion, evidence from animal studies mostly supports the view that DCS can enhance different forms of learning. In contrast, however, evidence from human studies is less conclusive. One reason for this may be the strong dose-dependency of DCS (Walker et al., 2002). While animal studies found that the higher the dose, the more learning is potentiated, human studies have generally used much lower doses (< 500mg) than the most effective doses in animal studies, due to concerns that in humans larger doses might produce antagonist effects on NMDA receptors. However, in this context, it is worth considering that the effective dose for potentiating NMDA receptors may also vary between brain areas because of local differences in glycine levels (which DCS competes with). Therefore, to establish whether DCS can increase plasticity and/or learning in brain networks related to depression, experiments should use tasks investigating processes affected in depression, such as reward and punishment learning, rather than ones relying on different neural systems such as motor learning. This question is addressed experimentally in chapter 4.

1.5 Summary and experimental questions

The introduction focused on reviewing the role of reward- and punishment-guided processing in depression, with particular focus on learning. Several studies have suggested that depression may be related to deficits in learning behaviour. However, I also highlighted that value-guided learning can be complex, involving many different cognitive and neural components. These include learning about reward or punishment information, even when there may be distracting stimuli, integrating different pieces of information and selection of actions. Because of this complexity, it is still unclear which ones of these processes are specifically affected by depression. To address this question, in chapter 2, I designed a novel reward and effort based learning task that measured and dissociated many diverse neural and cognitive components of complex learning behaviour in a naturalistic environment. I ultimately used this task to measure effects of depression (chapter 5). Prior to this, in chapter 3, I looked at the question whether the most common antidepressants, namely SSRIs, might modulate appetitive and aversive learning. This was based on animal studies, showing that SSRIs increase brain plasticity and enhance non-emotional learning. In chapter 4, I looked at whether a putative novel antidepressant, a partial NMDA agonist, might also influence value learning.

2. Neural mechanisms of learning real and hypothetical reward and effort

Summary

Naturalistic environments are complex and optimal behaviour therefore requires new, model-based, mechanism that learning in simpler environments does not (see chapter 1.1.2). In this chapter, we focus on a specific kind of model-based learning behaviour, namely how to allow relevant information to influence behaviour, while at the same time reducing the influence of irrelevant information on behaviour. In natural environments, a single choice can lead to multiple outcomes. Agents should learn which outcomes are due to their choices and therefore relevant for future decisions and which are stochastic in ways common to all choices and therefore irrelevant for future decisions between options. We designed an experiment in which human participants learnt the varying reward and effort magnitudes of two options and repeatedly chose between them. The reward associated with a choice was randomly real or hypothetical, i.e. participants only sometimes received the reward magnitude associated with the chosen option. The real/hypothetical nature of the reward on any one trial was, however, irrelevant for learning the longer-term values of the choices and participants ought to have only focused on the informational content of the outcome and disregarded whether it was a real or hypothetical reward. However, we found that participants showed an irrational choice bias, preferring choices that had previously led, by chance, to a real reward in the last trial. Amygdala and ventromedial prefrontal activity was related to the way in which participants' choices were biased by real reward receipt. By contrast, activity in dorsal anterior cingulate cortex, frontal operculum/anterior insula, and especially lateral anterior prefrontal cortex was related to the degree to which participants resisted this bias and chose effectively in a manner guided by aspects of outcomes that had real and more sustained relationships with particular choices, suppressing irrelevant reward information for more optimal learning and decision making.

2.1 Introduction

The environments in which animals and humans live are complex. Thus, to make the best decisions, agents must learn which choices are associated with good or bad outcomes, such as monetary rewards, or effort. They must then integrate information about these qualitatively different types of outcomes to make choices (Croxson et al., 2009; Pessiglione et al., 2007; Prevost et al., 2010; Rudebeck et al., 2006; Walton et al., 2002; 2003a). Outcomes may, however, have informational content that is independent of whether or not they are immediately rewarding. Imagine an animal foraging for berries on the higher branches of a tree after a strenuous climb. The animal tries to learn how many berries the tree carries to know whether to come back again. On a specific day, the weather may be good and the animal is able to gather a lot of berries from the tree, while on another day the weather may be bad and it has to abandon its food gathering prematurely. Despite not collecting any rewards, when learning about how good the tree is, it should only take into account how much food it saw hanging on the tree (this is the informational content of the outcome), but not how much food it managed to gather (this is the rewarding aspect of the outcome) on that day as that could depend on other interfering factors (such as the weather) and not the tree's value itself. In other words, sometimes the rewarding content, i.e. whether the berry reward was really experienced or only hypothetical, is a bad guide for future choices, which should be based on more abstract informational content.

In short, we tried to tackle the fundamental problem of how contingencies are learned when an outcome has, as is commonly the case, multiple components only some of which should be learned about. How does information in the irrelevant dimension (in our experiment this was the outcome's reward content) interfere with the learning of a contingency in the relevant dimension (in our experiment this was the outcome's informational content)? Because many brain systems are extremely sensitive to obtaining a reward (Vickery et al., 2011) we examined here whether this prominence of reward signals in some areas might bias behaviour and hamper learning based

on the informational content of the outcome. More importantly, we wanted to also test whether neural systems exist to counteract such biases. One region potentially equipped to contextualize reward appropriately is anterior prefrontal cortex (aPFC), because it has been implicated in complex behaviours such as pursuing alternative, hypothetical, and potentially novel future courses of action (Boorman et al., 2009; Donoso et al., 2014; Kolling et al., 2014).

We recorded brain activity using functional magnetic resonance imaging (fMRI) while participants learned the reward and effort magnitudes of two options and chose between them. Crucially, they only sometimes received the rewards associated with the option chosen. They were, however, always shown the associated reward magnitude regardless of whether the rewards were really received or hypothetical (i.e. the outcome's informational content was always provided even if the reward content was not). Thus, whether a reward was real or hypothetical was incidental and should not have affected participants' future choices, as the information needed for learning was the same in both conditions. Nonetheless, participants were biased towards repeating choices that led to real rewards. Ventromedial prefrontal cortex (vmPFC) and amygdala activity was related to this bias. APFC, frontal operculum/anterior insula (FO/AI), and dACC activity appeared to counteract such biases in several ways and mediated more flexible and optimal decision-making. Because only rewards, not effort costs, were real or hypothetical, we were furthermore able to test whether only reward signals or instead the relative value of repeating a choice was changed when rewards remained hypothetical (see methods). Not receiving a real reward had profound effects not only on the neural representation of rewards themselves, but also on representations of effort costs.

2.2 Methods

Participants

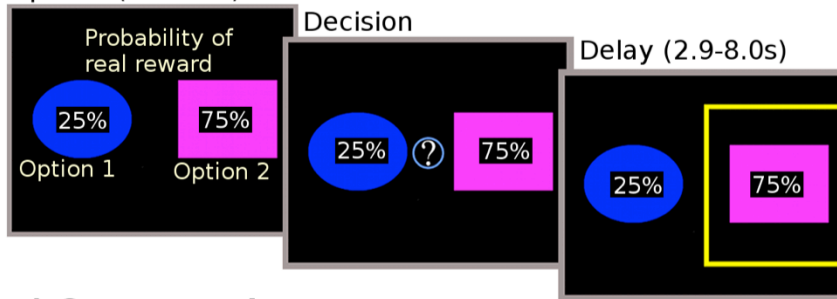
Twenty-one healthy participants (age 19-29 years, 10 female) took part in the study. The study was approved by the local ethics committee. One participant was excluded from data analysis because they repeatedly fell asleep during the MRI scan. Fourteen of the participants had taken daily placebo capsules for 2 weeks as part of another study.

Task

We designed a task to assess whether reward experience – more specifically whether a reward was really received or only hypothetical – changed decision-making even when the real/hypothetical nature of the outcome was an incidental feature of all choice outcomes. For this purpose, participants performed a learning task. They made repeated choices between two options with the aim of maximizing their monetary pay-off and minimizing the effort they needed to exert in an interleaved ‘effort phase’ (figure 1). On each trial there were three phases: first participants chose between two options (‘choice phase’), then they were shown the outcome of their choice (‘outcome phase’), then they had to exert the effort associated with the option they had chosen (‘effort phase’).

A) Decision phase

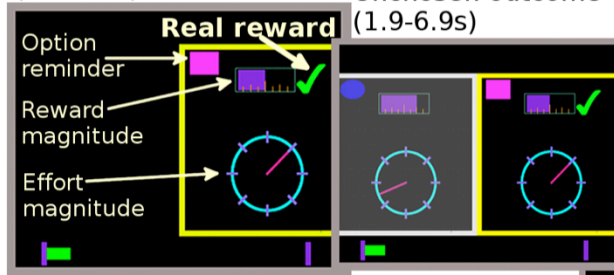
Options (1.4-4.5s)



B) Outcome phase

Chosen outcome
(1.9-2.1s)

Unchosen outcome
(1.9-6.9s)



C) Effort phase



D) Task schedule for reward and effort

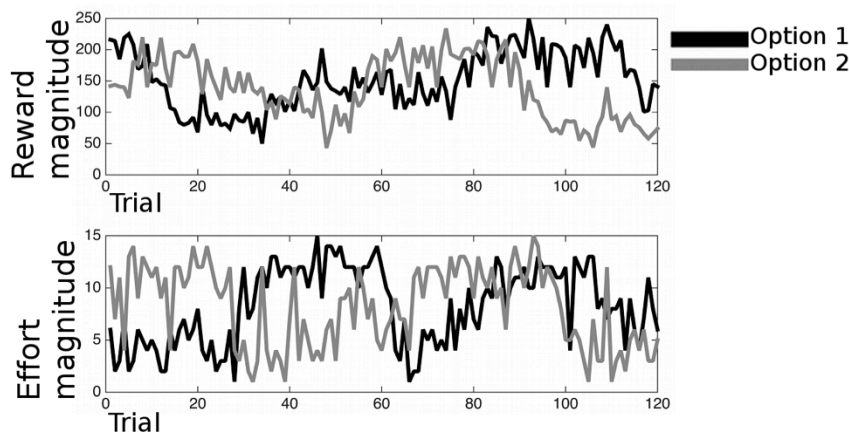


Figure 1. Task description: In the decision phase (A), participants were shown two options (i.e. choices), overlaid with the probability of receiving a reward for each choice. They could only decide after an initial monitoring phase (1.4-4.5s). The chosen option was then highlighted for 2.9-8s. In the following outcome phase (B), participants saw the outcome for the chosen option first (1.9-2.1s). The reward magnitude was shown as a purple bar (top of the screen), the effort magnitude was indicated through the position of a dial on a circle. Whether they received a reward was indicated by a tick mark (real reward, top display) or a red crossed out sign over the

reward magnitude (hypothetical reward, bottom display). If a reward was real, the reward was also added to a status bar at the bottom of the screen, which tracked rewards over the course of the experiment. A reminder of what option they had chosen was shown at the top of the screen. Then the reward and effort magnitudes were shown for the unchosen option (1.9-6.9s). Finally, participants performed the effort phase (C) where the number of targets was equivalent to the chosen effort outcome. Importantly, participants had to perform the effort phase on every trial, independent of whether the reward was real or hypothetical. An example schedule is shown in (D), with both the reward and effort magnitude values of the two choices.

In the decision phase (figure 1A), participants chose between two options using buttons on a trackball mouse. Each option had three independent attributes: a reward magnitude (reward points, later translated into monetary pay-off), an effort magnitude (amount of effort required in the effort phase), and a probability of being rewarded. The probability of each option was shown explicitly on the screen at the time of choice. In contrast, the reward and effort magnitudes of the options were not explicitly instructed and instead participants had to learn and track these slowly changing features of the two choices across trials. These magnitudes were drawn from normal distributions of which the means fluctuated pseudo-randomly and independently over the course of the experiment between three levels (low, mid, high), figure 1D. Participants were instructed to learn and keep track of the changing mean value of each magnitude across the experiment. Thus, the reward magnitudes the choices yielded constituted the only relevant reward information that could be tracked and learned, whereas the actual reward receipt and the probability thereof (whether the reward was real or hypothetical), should not have had any bearing on future decisions. Each of the four magnitudes was at each mean level equally often and only one of the reward or effort magnitude means was drifting at any one time. This was done to reduce unnecessary task complexity; no participant reported noticing this feature.

After the participants had selected an option, it was highlighted until the ensuing outcome phase. In the outcome phase (figure 1B), participants were first shown the reward and effort magnitudes of the option they had chosen, as well as whether they received a reward or not (in other words whether the outcome was a real secondary reinforcer indicating a specific monetary payment or

hypothetical). If they received a reward, the current trial's chosen reward magnitude was added to their total reward accumulated so far (figure 1B, on the bottom of the screen). They were then shown the reward and effort magnitudes for the option they had not chosen. During the outcome phase, participants could thus use the displayed information to update their estimates of the reward and effort magnitudes associated with the choices. Importantly, while they were shown whether they received a real reward or not, this should be irrelevant for their future choices as the informational content was the same. This is because how likely any choice was to be rewarded only depended on the probabilities that were explicitly cued at the time of choice and these probabilities changed randomly from trial to trial. In other words, whether a choice was likely to lead to a real or a hypothetical reward was not a feature of the choice that was sustained across trials or which differentiated it from the other choice and participants were made aware of this.

Finally, independently of whether the reward was real or hypothetical, participants had to perform the effort phase of the trial (figure 1C). Rather than simply exerting a brief force as in some studies of effort, here, participants had to exert a sustained effort (Croxson et al., 2009; Kurniawan et al., 2013; Rudebeck et al., 2006; Salamone et al., 1994; 2003; Walton et al., 2002; 2003a; 2006) by selecting circles that appeared on the screen using the trackball mouse. The circles were added to random positions on the screen in threes every three seconds (up to a total equal to the chosen effort magnitude). To make the task more effortful a random jitter (5 pixels, total screen size was 1280x800 pixels) was added to the mouse movement and circles only had a 70% probability of disappearing when clicked on. Furthermore, we pre-screened participants and only invited participants for the fMRI session if they had perceived the effort as aversive and were willing to trade-off money to reduce the effort that they needed to exert. The effort phase took place after the outcome phase rather than before so that the time of presentation of effort and reward outcome signals was matched.

Participants had 25s to complete the clicking phase and otherwise lost money equivalent to the potential reward magnitude of the chosen option ($1\% \pm 0.3\%$ of trials inside the scanner and $0.1\% \pm 0.0\%$ outside the scanner). Thus, the amount that could be lost was also independent of whether the reward had been real or hypothetical.

On most trials (100 out of 120) participants had to choose between the two options with changing reward and effort magnitudes. The reward magnitudes were set between 0 and 20 pence and the effort magnitudes were set between 0 and 15 circles that needed to be clicked. On the remaining trials ('Special-option-trials', SOTs), participants had to choose between one of the changing options and one of two fixed options whose values participants learned in a training session outside the scanner. The value of both fixed options was 7.5 pence, but one had a fixed effort magnitude of 4 circles and the other had one of 12 circles. The SOTs were included to ensure participants learned the values of each choice, rather than just their preference for one option over the other (a relative preference for one option over the other would not enable participants to choose effectively on the SOTs).

To summarize, participants needed to learn both the reward and effort magnitudes of the two options. Independent of these learnt quantities, participants were only rewarded on some trials, meaning that the reward magnitude was sometimes real and sometimes hypothetical. In contrast, the effort needed to be exerted on every trial. The inclusion of the effort dimension was crucial for arbitrating between different interpretations of our neural results (see section 'APFC's and FO/AI's effort cost and reward outcome representations change when the reward is real as opposed to hypothetical'). In short, it allowed us to disentangle whether areas only signalled rewards when they were real or whether they actually carried a signal in favour of switching to the alternative option, which was enhanced when rewards were hypothetical. This signal was

important because it appeared to allow participants to overcome a bias to stay, i.e. making the same choices, when the outcome was a real reward as opposed to an equally informative, but only hypothetically rewarding, outcome.

Participants performed 120 trials of the learning task inside the scanner and an additional 120 trials afterwards on the next day outside the scanner to increase the number of trials for the behavioural data analysis. Each participant performed the same two schedules in randomized order. Participants were informed about the features of the task in two training sessions before the scan, including the fixed number of trials they would perform. This ensured that they did not perceive low effort options as having a potentially higher monetary value because taking them might allow participants to move on to the next trial more quickly and to perform more trials with more chances to win money. Further details of the training were as follows: In the first training session (45min), participants performed a version of the task without a learning component, i.e. not only the probability, but also reward and effort magnitudes were explicitly shown. This training ensured that participants were familiar with the features of the task, for example, that they understood what the probability information meant. We also used this session to exclude participants before the fMRI session that did not find the effort sufficiently aversive to produce robust effects on behaviour. In a second training session (1h), we instructed participants about the learning task that they later performed in the fMRI scanner. At the end of the training, participants were queried about how they made decisions (specifically, they were asked “What are you thinking about when you’re making your decision”). All participants reported trying to learn the reward and effort magnitudes and using the explicitly cued probabilities to make decisions. No participants reported using information about whether reward had been real or hypothetical on the last trial when making choices. This suggested that participants were well aware of how to do the task before the beginning of the scan. If anything these thorough instructions should bias participants to try to counteract any tendencies they

might have to repeat choices based on receiving real rewards.

Interspersed with the 120 learning trials, there were 20 trials on which participants just had to indicate which option had a higher mean effort magnitude. These trials were included to ensure participants paid attention to the effort dimension. They were not given feedback about their choice. These trials were not included in the data analysis.

Experiment timings: The options were displayed for 1.4 to 4.5s before participants could make a choice. After the choice was made, the chosen option was highlighted for 2.9 to 8.0s. Next, the outcome was first displayed for the chosen option (1.9-2.1s), then for the unchosen option (1.9-6.9s). Participants then performed the effort exertion task (0-25s). Finally, the trial ended with an ITI (2.3-7.5s). The delays used ensured that we could examine the brain activity in the different phases of the task separately. There were no confounds between the brain activations in decision, outcome or effort exertion phase. All durations were drawn from Poisson-like distributions to increase design efficiency by having some long delays.

Behavioural analysis

In the behavioural analyses, we tested what information, both relevant and irrelevant for choice valuation, had an impact on decisions. We were however, most interested in whether a reward being real or hypothetical, influenced subsequent decisions. More specifically, we tested first whether having received a real or hypothetical reward on the last trial, would affect decisions on the next trial. The last trial's reward type (real vs. hypothetical) might imbue the option it was associated with (i.e. the chosen option) with positive (or negative, when no reward was received) affect. This would then make participants more likely to select that option again or more frequently. Additionally, we tested whether the reward type (real vs. hypothetical) changed how participants exerted effort in the subsequent effort phase of each trial.

All analyses were performed in Matlab and SPSS. When assumptions about statistical tests were violated, we used the Greenhouse-Geisser correction (for sphericity violations of ANOVA) or non-parametric Wilcoxon signed rank test (for violation of normality of t-tests).

Decision-making effects of real versus hypothetical reward

First, we performed analyses to establish that participants learned the reward and effort magnitudes in the task. We ran a logistic regression analysis predicting whether participants stayed with the same choice as made on the previous trial (chose it again on the current trial) or switched to the other option, based on the options' current reward probabilities (as already noted, these were explicitly indicated on each trial and because they varied randomly from trial to trial they could not be learned) and the reward magnitude outcomes and effort magnitude outcomes associated with the previous three trials. We also included a regressor denoting the last trial's reward type (real vs. hypothetical). This last regressor was our main focus of interest; it allowed us to test whether an aspect of the outcome that should have been irrelevant for learning biased decisions. All regressors, except the reward type (real vs. hypothetical), were coded as relative value differences (reward or effort magnitudes or probabilities) between the stay and the switch choice.

The multiple logistic regression was run in Matlab using `glmfit`, with a logit link function as the choice (stay or switch) being predicted was categorical. All regressors were normalised (as in all subsequent behavioural and fMRI regression analyses). For each participant, we obtained one β regression weight for each regressor. These were then tested for statistical significance across all participants.

For the analysis of the decisions, we excluded choices on the special-option-trials (SOTs; see above), i.e. trials in which decisions were not between the usual two options making it impossible to classify those choices into stays and switches in the usual way. On the trials after the SOTs, stay or switch was coded with respect to the trials before the SOTs, in which the two usual options were present.

Computational modelling of the decision behaviour

To look at the reward-type induced decision-bias in more details and test between different potential underlying mechanisms, we fitted different learning models to the behavioural data. Each model consisted of three main components. First, each model had estimates about the mean reward and effort magnitudes of each option. These were updated on every trial using a reinforcement-learning algorithm. Second, the magnitude estimates were integrated with the explicitly cued probabilities to calculate how valuable each of the two options was (i.e., their utility). Third, these two utilities were compared to predict participants' choices.

Our main question of interest was how reward type (real vs. hypothetical) influenced the decision-making based on the learnt information. To assess this, we first determined the best simple model using model comparison (see below). Next we modified this simple model to test different hypotheses of how reward type could influence decisions. Comparing the model fits we tested which hypothesis about underlying cognitive mechanism was best supported by the data.

Best simple model

We first wanted to determine the best simple model to capture participants' behaviour before modifying this model to also capture the impact of reward type. For this, we compared several models that differed in how the reward and effort magnitudes were learnt and in how reward magnitude and probability were integrated.

For each model, learning was modelled using a standard Rescorla-Wagner reinforcement-learning rule. On each trial, the predicted reward or effort magnitude of an option was updated based on the trial's outcome, as a function of the prediction error (PE):

$$PE_t = Outcome_t - Prediction_t$$

$$Prediction_{t+1} = Prediction_t + \alpha * PE_t$$

where α was the learning rate. Thus, the learning rate was a measure of how much participants updated their magnitude predictions when the outcome differed from their prediction. We fitted models that differed in their number of learning rates: some models used a shared learning rate for reward and effort magnitudes while others used separate learning rates. A last type of model used different learning rates for learning about the reward and effort magnitudes of the option that was chosen and the unchosen option.

The reward and effort magnitude predictions we thus obtained were then integrated together with the probabilities, explicitly shown in the task display, to yield the utility of each option. Based on previous studies (Scholl et al. (2014), see also chapter 4), we tested different methods for integrating reward magnitude and probability. First, we tested whether they were integrated optimally, i.e. multiplicatively (reward magnitude x probability):

$$Utility = \frac{1}{k} * Magnitude_{Reward} * Probability_{Reward} - \frac{\gamma}{k} * Magnitude_{Effort}$$

where γ is the effort weighting factor and k is a normalization factor (with $k=1+\gamma$).

Second, we tested whether instead participants used a heuristic and integrated them as a weighted sum:

$$Utility = \frac{1}{k} * Magnitude_{Reward} + \frac{\delta}{k} * Probability_{Reward} - \frac{\gamma}{k} * Magnitude_{Effort}$$

where γ is the effort weighting factor, δ is the probability weighting factor and k is a normalization factor (with $k=1+\gamma+\delta$).

All combinations of the different learning and integration models were fitted (total of four models). All models had in common that the utilities of the two options were compared with each other to predict participants' choices using a standard soft-max decision rule.

$$P(\text{Option}_A) = \frac{e^{\beta * \text{Utility}_A}}{e^{\beta * \text{Utility}_A} + e^{\beta * \text{Utility}_B}}$$

where β is the inverse temperature, reflecting a participant's ability to pick the option with higher utility.

To fit the different models, we used hierarchical Bayesian modelling, as implemented in JAGS (Plummer, 2003), which has been widely used previously (Ahn et al., 2011; Nilsson et al., 2011; Shiffrin et al., 2008). The key aspect of this hierarchical modelling approach is that parameter estimates for individual participants are assumed to come from a group-level distribution of those parameters. In other words, the estimates of individual participant parameters are constrained by the parameter estimates at the group level. This approach to model fitting requires the setting of prior distributions of the group level parameters to be estimated (mean and standard deviation). These priors are then updated based on Bayes law using Markov Chain Monte Carlo (MCMC) algorithms to obtain posterior distributions of the parameters. For each model, we obtained three chains with 25,000 samples after an initial burn-in of 25,000 samples. We checked for convergence using the JAGS in-built Gelman-Rubin statistic.

We set the priors as follows: the learning rate (α) was constrained to lie between 0 and 1. It was first transformed to a probit scale to facilitate hierarchical modelling (Nilsson et al., 2011). The priors for the mean (before transformation) were drawn from a normal distribution with mean 0 and variance 1. The priors of the standard deviation of the group distribution (before

transformation) were drawn from a uniform distribution between 0 and 10. The mean of the inverse temperature (β) was constrained to be between 0 and 10, with the precision (1/variance) drawn from a gamma distribution (with parameter settings shape=0.001, scale=0.001), which effectively produces a flat prior (Lee and Wagenmakers, 2014). The prior distributions for probability (δ) and effort weighting (γ) parameters were drawn from uniform distributions between 0 and 20 and the precision was again drawn from gamma functions as before. We confirmed after model-fitting that none of the parameters that we found lay too close to the boundaries of the priors that we used, ensuring that we were not misestimating the parameters due to incorrect constraints.

To determine which of the models that we fit best explained our data, we used model comparison based on DIC differences (Deviance information criterion, (Spiegelhalter et al., 2002)). DIC provides a measure of model fit, taking into account both how well the model explains the data and how complex the model is. DIC is a hierarchical modelling generalization of the AIC (Akaike information criterion). The smaller the DIC value, the better a model explains the data.

Modelling the effect of reward type:

Based on model comparisons of the above named simple models, we determined the best fitting model, which we found to be one based on a common learning rate for reward and effort and a heuristic utility (i.e. a weighted sum of probability, reward and effort magnitudes), see results.

To assess the behavioural effect of reward type (real vs. hypothetical), we then modified this best fitting model in the following three ways. First, we considered that one explanation for a difference in choice after real or hypothetical reward could be that participants only learn from real reward, but not from hypothetical reward ('NoLearningHypothetical' model). We modelled

this by setting the learning rate to 0 for the chosen reward outcome on trials where the reward was hypothetical. In all other instances – i.e. the reward was chosen and real or the reward of the unchosen option or of the effort dimension – we used the same learning rate (α):

$$Prediction_{t+1} = \begin{cases} Prediction_t + \alpha * PE_t, & \text{if reward real} \\ Prediction_t, & \text{if reward hypothetical} \end{cases}$$

We also tested a version of this hypothesis, in which instead of no learning for hypothetical reward, learning had a different learning rate for hypothetical than real reward ('LessLearningHypothetical' model).

As a second hypothesis, the difference in choice after real or hypothetical reward could be due to participants perceiving only real reward as rewarding ('NoRewardHypothetical' model). To test this, we modelled the hypothetical reward as having a reward magnitude of zero. The rest of the model was as before:

$$PE_t = \begin{cases} Outcome_t - Prediction_t, & \text{if reward real} \\ 0 - Prediction_t, & \text{if reward hypothetical} \end{cases}$$

We also tested a version of this hypothesis, in which instead of a reward magnitude of zero when reward was hypothetical, the reward magnitude was scaled by an additional free parameter.

As a third hypothesis, we considered that it is maybe not the reward magnitudes that are learnt differently, but instead, participants might unwittingly be biased towards choices that had led to real reward. In other words, this hypothesizes that participants find receiving a real reward more rewarding than a hypothetical reward, and that they are not only influenced by the informational content provided by both real and hypothetical outcomes but that they were also influenced by real reward receipt. In other words, when multiple aspect outcomes are received participants not only learn choice-outcome contingencies that are relevant to task performance but irrelevant outcome dimensions also affect their learning. This was modeled as a bias term when calculating the utility:

$$Utility = \frac{1}{k} * Magnitude_{Reward} + \frac{\delta}{k} * Probability_{Reward} - \frac{\gamma}{k} * Magnitude_{Effort} + \frac{\varepsilon}{k} \\ * RewardType_{LastTrial}$$

where δ was the probability weighting factor, γ was the effort weighting factor, ε was the reward type (real vs. hypothetical) weighting factor and k was a normalization factor ($k=1+\gamma+\delta+\varepsilon$).

We set the prior for the group mean of the weight of reward type (ε) between -10 and +10. This range including positive and negative values was chosen so as not, *a priori*, to bias values to be only positive (as our hypothesis would suggest), but to allow a free fit. We again used a flat distribution for the group standard deviation prior.

To see which of these hypothesis best explained the effects of reward type on decision, we again used model comparison based on DIC.

Effort exertion phase

In the effort phase, participants used a trackball mouse to click on circles appearing on the screen to make them disappear. We recorded every mouse click independently of whether the click made the target disappear or not and whether it was on the circles to be clicked or somewhere else on the screen. Thus, we could measure the clicking rate on each trial, as one proxy for motivation.

Here, we tested, using a regression analysis, whether the reward type (real vs. hypothetical) had an impact on how motivated participants were when exerting effort. Thus, the clicking rate on each trial was the dependent variable and the regressor of interest was whether the reward was real or hypothetical on the current trial (recall that the effort phase of each trial followed the reward outcome phase of each trial). As regressors of no interest we also included the chosen option's effort magnitude (i.e. the number of targets participants had to click), the unchosen option's effort magnitude, as well as the chosen option's and the unchosen option's reward

magnitudes. The regressions were run in Matlab (glmfit), using a log link function as the clicking rate data were constrained to be greater than zero.

In these analyses, we only included data from the behavioural session outside the scanner as the EPI sequence sometimes affected the reliability of the trackball mouse during the effort exertion phase during the scan. Because of this interference, the clicking performance recorded inside the scanner may not always have been a reliable measure. This was of course no problem outside the scanner.

MRI data acquisition

Structural MRI and fMRI measurements were taken using a Siemens 3 Tesla MRI scanner. For the fMRI, we used a Deichmann echo-planar imaging (EPI) sequence (Deichmann et al., 2003) [time to repeat (TR): 3000ms; 3x3x3mm voxel size; echo time (TE): 30ms; flip angle: 87°; slice angle of 15° with local z-shimming] to minimize signal distortions in orbitofrontal brain areas. This entailed orienting the field-of-view at approximately 30° with respect to the AC-PC line. We acquired between 1100 and 1300 volumes (depending on the time needed to complete the task) of 45 slices per participant. Additionally for each participant, anatomical images were acquired with a T1- weighted MP-RAGE sequence, using a GRAPPA acceleration factor of 2 (TR: 2200ms; TE: 4.53ms; inversion time: 900ms; voxel size: 1x1x1 mm on a 176x192x192 grid) [same protocol as (Chau et al., 2014; 2014)].

We used FMRIB's Software Library (FSL) (Smith et al., 2004) for image pre-processing and analysis. Functional images acquired were first spatially smoothed (Gaussian kernel with 5mm full-width half-maximum) and temporally high-pass filtered (3 dB cut-off of 100s). Afterward, the functional data were manually denoised using probabilistic independent component analysis (Beckmann and Smith, 2004), identifying and regressing out obvious noise components (Kelly et

al., 2010); we considered only the first 40 components of each participant which had the greatest impact to interfere with task data (total up to 550). We used the Brain Extraction Tool (BET) from FSL (Smith, 2002) on the high-resolution structural MRI images to separate brain matter from non-brain matter. The resulting images guided registration of functional images in Montreal Neurological Institute (MNI)-space using non-linear registrations as implemented in FNIRT (Jenkinson et al., 2012). The data was pre-whitened before analysis to account for temporal autocorrelations (Woolrich et al., 2001). Statistical analysis was performed at two levels. At the first level, we used an event-related general linear model (GLM) approach for each participant. On the second level, we used FMRIB's Local Analysis of Mixed Effects (FLAME 1) (Beckmann et al., 2003) with outlier de-weighting and tested the single group average. The main effect images are all cluster-corrected results with the standard threshold of $z=2.3$. We also analysed data in regions of interest (ROIs), extracted from spheres with a 3 voxel radius, identified in MNI standard space on the basis of orthogonal whole group analyses or used time courses for illustration purposes only (for example in figure 9).

fMRI analysis

In the fMRI analysis, we investigated the neural underpinnings of the effect of real versus hypothetical rewards on behaviour. We looked at these effects at two time points. First, we examined activity at the time when participants made decisions to examine whether brain activity reflected whether the reward had been real or hypothetical on the last trial, as well as whether the real/hypothetical nature of the reward influenced switch-stay signals. The second time point we examined was at the time of the outcome after the decision had been made. At this time point we examined, firstly, which brain regions were differently active when rewards were real or hypothetical and, secondly, whether the reward type (real vs. hypothetical) changed the sensitivity to the effort and reward magnitude outcomes.

Identifying areas coding real and hypothetical reward and effort in the outcome and decision phases

In the first design (GLM1), we investigated the neural effect of the reward being real or hypothetical at the time of the decision and at the time of the outcome. Additionally, we looked at how neural activity reflected the other key features of the options, namely their reward and the effort magnitudes and their reward probabilities. The regressors used in this design were as follows (see figure 2A for correlations between the regressors): We used three boxcar regressors, indicating the onset and duration of the decision phase (duration from the beginning of the trial until participants made a choice), the onset and duration of the outcome phase (from the appearance of the chosen outcome until the chosen and the unchosen outcomes were removed again from the screen) and lastly the effort exertion phase (from the appearance of the first target that participants needed to remove until they had removed the last target). We furthermore included the following parametric regressors in the decision phase: whether the reward of the chosen option had been real or hypothetical on the last trial, reward and effort magnitude predictions (derived from a standard Bayesian learning model, similar to that described in Behrens et al. (2007), see below) and the reward probabilities that had been displayed on the screen. In each case separate regressors for the chosen and the unchosen option were used. None of these regressors were correlated with participants' reaction time and inclusion of reaction time did not affect the brain signals reported in the decision phase. In the outcome phase we included the following parametric regressors: the reward type (real vs. hypothetical) delivered for the chosen option, the probability of reward for the chosen option, the reward magnitude outcomes and the effort magnitude outcomes for the chosen and the unchosen option. The onset of the regressors for reward receipt, chosen option reward probability, chosen option reward magnitude, and chosen option effort magnitude, were time-locked to the onset of the outcome phase; the onset and duration of the regressors indexing the reward magnitude and effort

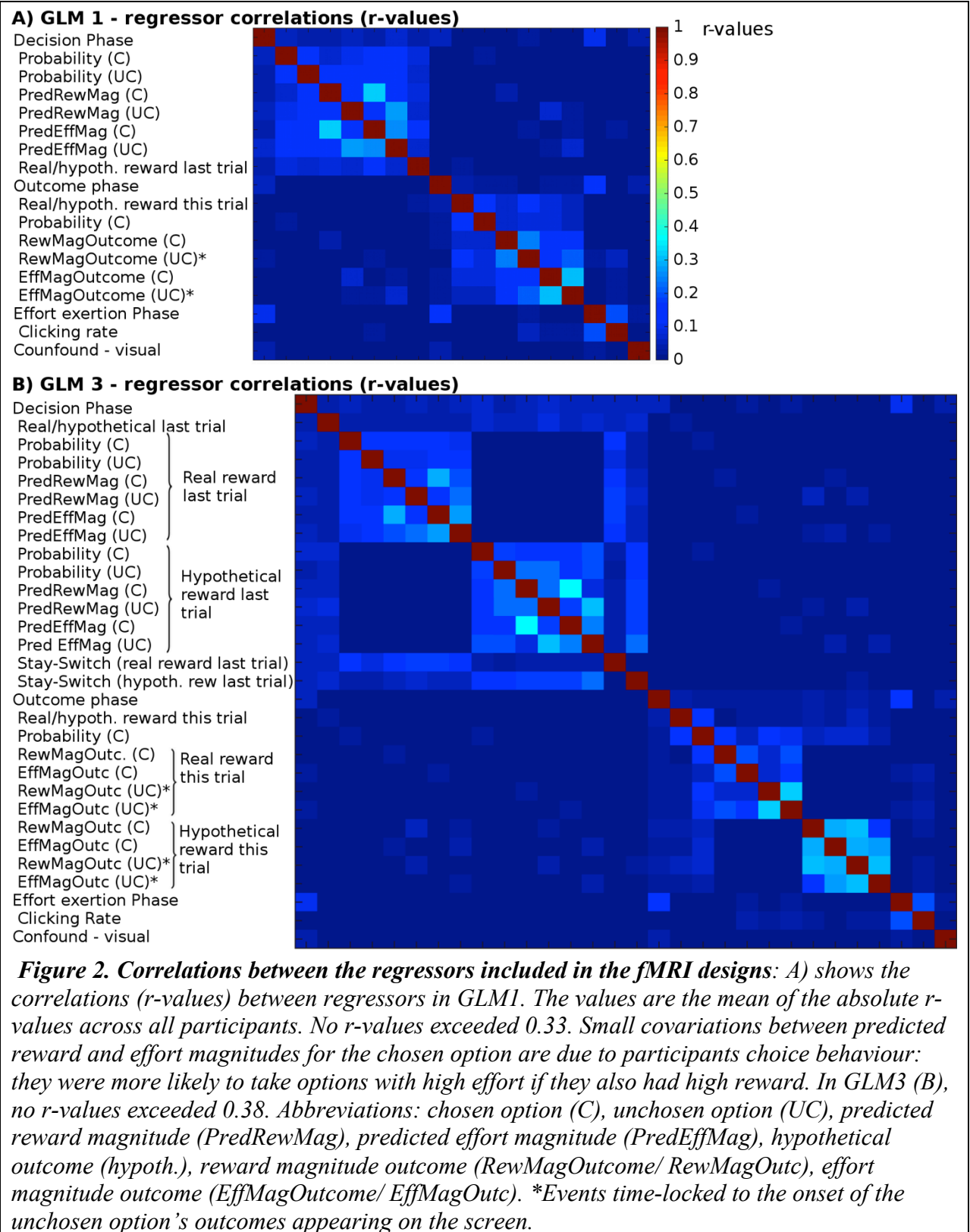
magnitude for the unchosen option were time-locked to their display (appearing about 2s after the outcomes of the chosen option). In the effort execution phase, we included the clicking rate as a parametric regressor. Finally, we included, as confound regressors, six movement regressors and a regressor indexing when additional visual stimuli were presented to warn participants that they had not clicked the targets on time and that the halfway point of the experiment had been reached.

We applied a Bayesian learner (similar to the one described in Behrens et al., 2007) to estimate the reward and effort magnitude predictions for each option for the fMRI analysis. We did this to obtain prediction regressors (of effort and reward magnitudes) for the fMRI analysis that could not possibly be biased by model specifications of a fitted model. However, it is of note that the regressors obtained in this way were highly correlated with those from the best fitting simple learning model ($r=0.95$). Therefore, we would not expect the fMRI results to differ were we to use those regressors instead. Furthermore, also note that most neural results are completely independent of the model used as we mainly focused on regressors that described model-free quantities (such as the shown probabilities at the time of the decision or the shown reward or effort magnitudes or reward type at the time of the outcome).

In short, the Bayesian model estimated the current reward/effort magnitude predictions based on the previous trials for each of the four magnitudes (option A reward magnitude, option A effort magnitude, option B reward magnitude, option B effort magnitude) separately. It did this by taking into account the following properties of the experimental task: 1) The reward/effort magnitude outcomes were determined by separate underlying reward/effort distributions for each magnitude (normal distributions). 2) The mean of a magnitude distribution could sometimes change. 3) How quickly each mean changed could vary over the course of the experiment. Sometimes, the mean changed more quickly (high volatility) whereas at other times, it changes more slowly (low volatility). 4) The volatility of an attribute was not static but could also change over the course of the experiment. In other words, the parameters that the model estimated on

every trial for each attribute were the reward/effort magnitude prediction (mean and variance), the volatility, and the volatility change.

The model estimated these parameters for the current trial based on last trial's parameter likelihoods and the last trial's attribute outcomes. It did this using Bayes' rule, which is the most efficient way for updating beliefs given new evidence. In contrast to the fitted models, this model did not have any free parameters. For details on the specific mathematical implementation of the Bayesian learner, please refer to Behrens (2007). The only noteworthy change compared to Behrens et al. (2007) was that whereas they tried to predict binary outcomes using a beta-distribution we wanted to predict numerical outcomes using a normal distribution and thus added an additional parameter estimating the width of the normal distribution for the magnitudes.



Relating neural effects of real versus hypothetical reward to behaviour

We also looked for correlations, across participants, between brain representations of reward type (real vs. hypothetical) and the behavioural impact the reward type had on stay/switch

behaviour. For this we included, at the group level analysis stage of the fMRI data, a covariate denoting for each participant how much reward type on the last trial influenced subsequent stay/switch behaviour (GLM2). The covariate was thus the regression β weight from the behavioural analysis described above, for each individual participant, indicating how strongly the last trial's reward type influenced participants' stay/switch decisions. We looked for correlations between this covariate and the brain response to reward type at the time of the outcome phase. To assess whether the regions that we identified in this analysis made shared or independent contributions to the decision bias we also performed follow-up partial correlations. For this we first extracted each individual participant's fMRI contrast of parameter estimates (COPE) values for the contrast of reward type from ROIs placed inside the activation clusters using featquery (FSL). We then correlated the COPE from one region with the decision bias while controlling for the COPE from another region. Note that in this partial correlation analysis we are interested in whether the correlations in both areas explain at least partially independent variance or not. We tested this using partial correlations and mediator analyses (Sobel test, Preacher and Hayes (2004)).

Time course analysis – how reward and effort magnitude representations differ for real and hypothetical reward

Having identified which regions in the outcome phase coded effort and reward magnitudes, we then went on to test whether reward type (real vs. hypothetical) had an impact on the coding of reward magnitude and effort magnitude outcomes in these regions. This analysis is orthogonal to the analysis used in initial identification of the ROIs because the difference of the magnitudes in the real vs. hypothetical reward conditions is independent of the sum of the two conditions (i.e. the main effect of reward or effort outcomes) that was used to identify the regions. First, we extracted the blood oxygen level dependent (BOLD) signal from these regions. We then up-sampled the time-course tenfold, created epochs from the beginning of the outcome phase

onward and applied a GLM to every time point separately. By averaging the resulting β weights across participants we created the time-courses shown (the standard errors of the mean effect across participants are shown). We included as regressors: the relative reward magnitude outcome (chosen minus unchosen option) as separate regressors on trials when the reward was real or hypothetical, the relative effort magnitude outcome (chosen minus unchosen option), again as separate regressors for when the reward was real or hypothetical and finally we also included regressors for the main effect of reward type and to denote the expected probability of reward receipt. We statistically compared the reward outcome regressors in the real and hypothetical reward conditions, using two-tailed t-tests performed on the time course after haemodynamic convolution. For the haemodynamic convolution, we used a leave-one-out procedure to fit the haemodynamic response function. Specifically, for each participant, we used the data from all but that participant to determine the absolute peak (i.e. either the strongest peak or trough) of the time course in a window between 6s and 12s. We then aligned the peak of a canonical haemodynamic response function to this peak. The haemodynamic response function was made using `gammampdf` in matlab, with values $\alpha = 7^2/3^2$, $\beta = 7/3^2$. We then convolved the aligned haemodynamic response function with the omitted participant's time course to obtain one value per participant for each regressor of interest.

Stay-switch analysis

We performed a third whole-brain analysis (GLM3) to assess whether reward type (real vs. hypothetical) on the last trial also influenced brain signals directly related to the choice on the current trial, namely whether on the current trial the same choice was repeated as on the last trial (i.e. a stay decision) or whether the other option was chosen (switch). In other words, we wanted to identify regions that were not just differently affected by receipt of real and hypothetical reward outcomes but which were activated differentially as a result of an interaction between the outcome type and subsequent behaviour; we identified brain areas activated differently when

decisions were made in line with the bias suggested by last trial's reward type (i.e. stay decision after real reward and switch decision after hypothetical reward) or against it (stay decision after hypothetical reward and switch decision after real reward). In GLM3 (see figure 2B for a list of regressor correlations), we included again the same boxcar regressors for the onset and duration of the decision, outcome, and effort exertion phases. We included as our main regressors of interest in the decision phase whether the current trial was a stay or a switch choice but coded separately for instances on which there had been a real or a hypothetical reward on the last trial previous to it. We then looked whether stay/switch signals were affected by whether the reward was real or hypothetical on the last trial by looking at the difference between these two regressors (stay/switch after real reward minus stay/switch after hypothetical reward on the last trial). As confound regressors we included, first, a main effect regressor indicating the reward type on the last trial (regardless of whether the current trial was a stay or a switch decision). Additionally, we also included regressors indexing the predicted reward magnitudes, predicted effort magnitudes, and explicitly shown reward probabilities associated with both the chosen and unchosen options. Each of these quantities was included as separate regressors for when the reward had been real on the last trial and for when it had been hypothetical on the last trial. In the outcome phase, we included, as in GLM1, a regressor indicating current reward type (real vs. hypothetical) and what the probability of that reward type had been. Additionally, we included regressors, separately for trials with real or hypothetical reward, for the chosen and the unchosen reward magnitude outcomes and effort magnitude outcomes. The regressors in the effort exertion phase and the other motion confound regressors phase were as in GLM1.

Connectivity analysis

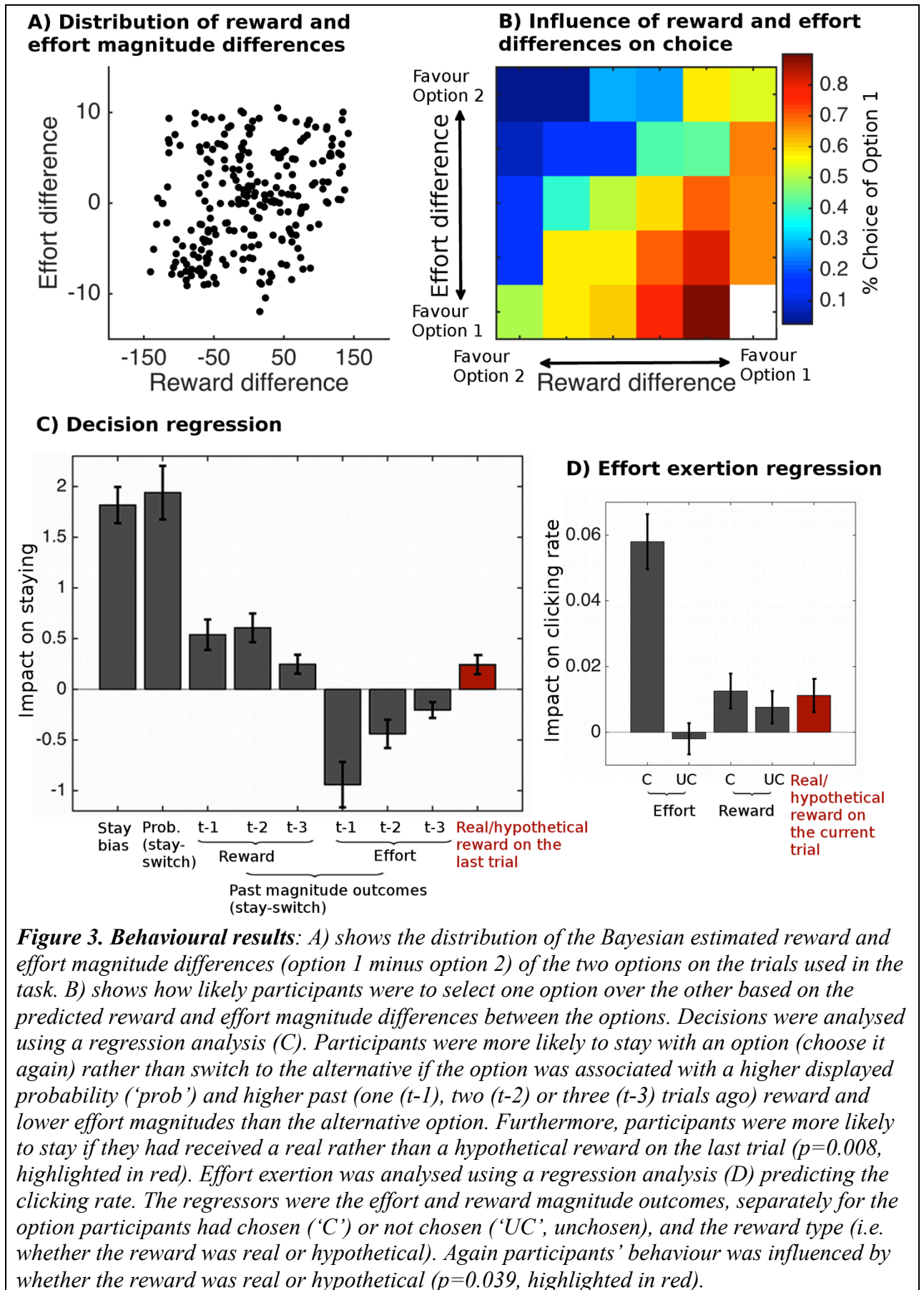
In the previous analysis (GLM3), we identified regions with activity reflecting the interaction between the decision to stay or switch and the reward type on the last trial. To further examine the relationship between one of those areas (for the sake of simplicity in the description of the

analysis procedure we note here that the region in question was the amygdala, which has previously been implicated in biased decision making (de Martino et al., 2006; Roiser et al., 2009)) and areas that we had identified as opposing biased decision making, we conducted an exploratory analysis of the interactions during task performance. More specifically, we performed a psychophysiological interaction analysis (PPI, O'Reilly et al. (2012)) to see whether connectivity between those types of region changed as a function of the decision x last reward type interaction effect. We hypothesized, based on previous research (Roiser et al., 2009), that if a region opposes the bias it could do so partly by suppressing activity in a region, such as the amygdala, that is associated with promoting the bias during the decision. If that were the case then a decrease in connectivity should be seen on trials when participants ended up making a potentially biased decision. In other words, choices in agreement with last trial's reward bias could be partly due to a failure of the bias-opposing region to suppress the bias-promoting region (the amygdala). To test this hypothesis we performed a PPI for every time point during the decision phase, equivalent to the other time course analyses. We included as main regressors of interest the time course extracted from the region identified by the bias consistent choice effect in GLM3 and the interaction between this timecourse and this bias effect. Specifically, the bias consistent choice effect regressor was binary, coding whether decisions were made in line with last trial's reward type (stay after real reward and switch after hypothetical reward) or against it. As usual for a PPI analysis, we included as confound regressors the main effect of the interaction between the current decision and the last trial's reward type, as well as the relative probabilities, reward and effort magnitudes (chosen minus unchosen), last trial's reward type as well as the global time course (Rogers et al. (2007), Friston et al. (1990)).

2.3 Results

Decision-making analysis

Participants performed a reward and effort learning task in which they tracked the continuously changing reward and effort magnitude values of two options. The range of reward and effort magnitude differences between the two options and the influence these differences exerted on participants' choices is shown in figure 3A and 3B. We found, as predicted, that participants were more likely to select an option associated with a higher expected reward magnitude and a lower expected effort magnitude. To test more formally whether participants learned the changing reward and effort magnitudes, we performed a regression analysis assessing the effect of reward magnitude and effort magnitude outcomes on recent trials and the reward probabilities (which were shown to participants at the time of the choice) (figure 3C). This analysis demonstrated that participants' decisions to repeat a choice (stay) or select the alternative option (switch) were influenced by the reward probabilities they were shown at the time they made a decision (these varied randomly from trial to trial) and that the history of reward magnitude and effort magnitude outcomes on recent trials (i.e. what should be learnt) also influenced whether participants stayed with their previous choice or switched to the alternative. Furthermore, the impact of more recent reward and effort magnitudes was larger than for reward and effort magnitudes received longer ago in the past, as is typically found in a learning task (ANOVA, main effect of time: $F(1.4,26.4)=4.536$, $p=0.031$).



Our main factor of interest in this decision analysis was whether participants' decisions were also influenced by whether the reward had been real or hypothetical on the last trial (reward type). To test this, we also included the last trial's reward type in the regression analysis described above. We found that participants were more likely to repeat a choice (stay) if they had received a real reward on the last trial (figure 3C, red bar, $z=2.65$, $p=0.008$, one-sample Wilcoxon signed rank test). To confirm this result in a simpler way, we divided all trials into two bins – those that occurred after real rewards and those that occurred after hypothetical rewards. We then calculated the percentage of trials for each participant on which they chose the same option again (stay) or selected the other option (switch) after the reward had been real or hypothetical on the last trial. We found, as before, that participants were more likely to stay when the reward had been real rather than hypothetical ($z=-3.10$, $p=0.002$, one-sample Wilcoxon signed rank test).

Computational modelling of the decision behaviour

To look at the reward-type induced decision-bias in more details, we fitted different learning models to the behavioural data. We first fitted four models to the data to determine which model to examine further to assess the effect of reward type (real vs. hypothetical). These models differed in the learning and the decision rules used. For each model, we computed the model fit using the DIC. DIC differences between models of 0-5 suggest no evidence in favour of a model, DIC differences of 5-10 suggests mild evidence in favour of a model, and DIC differences >10 suggest strong evidence in favour of a model (Spiegelhalter et al., 2002). The DIC scores are shown in table 1. A model with a shared learning rate for reward and effort and which integrated reward magnitude and probability in a heuristic manner by linear summation ('Addition1 α ' model) showed the best fit (DIC = 2828). This model explained 81% of participants' choices on average. We then analysed this model further with regard to the mechanism of how reward type could influence decisions. However we note that the second best model (which used separate

learning rates for reward and effort) showed a similar fit and so we also repeated all the analyses shown below for this alternative model. We do not report these analyses here because the conclusions drawn were identical to those drawn below. It is perhaps worth briefly pointing out that the model with different learning rates for the chosen and the unchosen option ('AdditionChosenUnchosen α ') also did not provide a better fit to the data than a model with just one learning rate ('Addition1 α '). Therefore, we cannot determine from our data whether the learning occurs separately for each option or for the relative value of the two options.

Simple models	DIC
Addition1 α	2828
Addition2 α	2833
AdditionChosenUnchosen α	2837
Multiplication1 α	4353
Multiplication2 α	4103
Reward type models (all based on Addition1α)	
DecisionBias	2780
NoLearningHypothetical	2837
LessLearningHypothetical	2822
NoRewardHypothetical	3388
LessRewardHypothetical	2830

Table 1. Results of the model fitting analysis. Deviance information criterion values (DIC) are reported; smaller values indicate a better fit.

Next, we fitted three model types to test different hypotheses of how reward type (real vs. hypothetical) could affect decisions (table 1). The first model type incorporated the hypothesis that reward type changed how participants learn about reward, more specifically whether – as could be expected from some basic reinforcement learning theories – hypothetical reward is not used for learning ('NoLearningHypothetical' model in table). We found that this model did not provide a better explanation for participants' behaviour (DIC = 2837). A variant of this model tested whether instead of no learning from hypothetical reward, learning from hypothetical reward might be reduced ('LessLearningHypothetical' model). This model only provided a marginally better fit to the data than the simplest 'Addition1 α ' model (DIC = 2822).

The second model tested whether hypothetical reward was perceived as non-rewarding ('NoRewardHypothetical' model in table). Again, this model did not provide a good explanation of participants' behaviour (DIC = 3388). A variant of this model tested whether instead of perceiving hypothetical reward as completely non-rewarding, it might be perceived as somewhat less rewarding ('LessRewardHypothetical' model). This model did not provide a better fit than the simple 'Addition1 α ' model (DIC=2830).

The third model tested whether after receiving a real reward when choosing an option, participants mistakenly perceive the option as having a higher utility and therefore are more likely to select it again ('DecisionBias' model in table). We found this model to provide the best fit to the data (DIC = 2780). While there have been criticisms of the usefulness of DIC for model comparison (Spiegelhalter et al., 2014), we note that the difference (on a logarithmic scale) that we found here is much larger (38 points) than the standard difference taken to suggest a large difference in model fit (10 points). Thus, even if the real difference is maybe somewhat smaller than the estimate, it would still be highly significant. Furthermore, we found that across participants, reward type had a consistent positive effect (mean= 0.19, 95% confidence interval: 0.01 to 0.39 table 2), replicating the result of the regression analysis that participants prefer options that had received a real rather than a hypothetical reward in the past. This suggests that participants can learn from both hypothetical and real reward, but that they irrationally behave as if choices that had been associated with a real reward were more valuable to them. The conclusions of this analysis are therefore the same as those drawn from the regression analysis.

	Mean	Standard error
Learning rate (α)	0.42	0.04
Inverse temperature (β)	3.85	0.2
Effort weight (γ)	1.54	0.3
Probability weight (δ)	1.34	0.17
Reward type weight (ϵ)	0.19	0.06

Table 2. Parameters of the best fitting model incorporating an effect of reward type ('DecisionBias' model). For each parameter, mean values and standard errors are shown.

It is of note that while this behavioural modelling is able to tell us that participants are biased towards an option that had been linked to a real reward, it does not distinguish between different possible neural mechanisms that could generate this behaviour. In other words, participants could simply have a weak bias or alternatively the bias is due to a competition between a brain mechanism that is biased by real reward (and thus drives the behaviour) and activity in another brain mechanism reflects the true, underlying task structure (namely that participants should ignore for future choice whether reward is real or hypothetical). To investigate this question further, we need to interrogate the neural signals.

Effort exertion phase

In the effort exertion phase, we tested for an effect of reward type (real vs. hypothetical) on the clicking rate (i.e. on how vigorously participants exerted an effort). A multiple regression analysis (figure 3D) revealed that if the reward was real as opposed to hypothetical then the clicking rate increased in the subsequent effort exertion phase ($t(19)=2.214$, $p=0.039$). This thus further supports the view that reward type was not ignored by participants, but instead affected their decision and effort exertion behaviour. The effect is consistent with what would be expected if participants perceived real rewards as more rewarding than hypothetical rewards.

FMRI results

The behavioural modelling was able to tell us that participants were biased towards staying with an option that had been linked to a real reward. However, the behavioural modeling could not distinguish between different possible neural mechanisms that could generate this behaviour. For example, the behavioural bias could simply be due to a brain mechanism that is weakly biased towards irrationally repeating a choice after real reward. Alternatively, the behavioural bias could be due to a competition between a brain mechanism that is biased by real reward (and thus drives the behaviour) and another brain mechanism with activity that reflects the true task

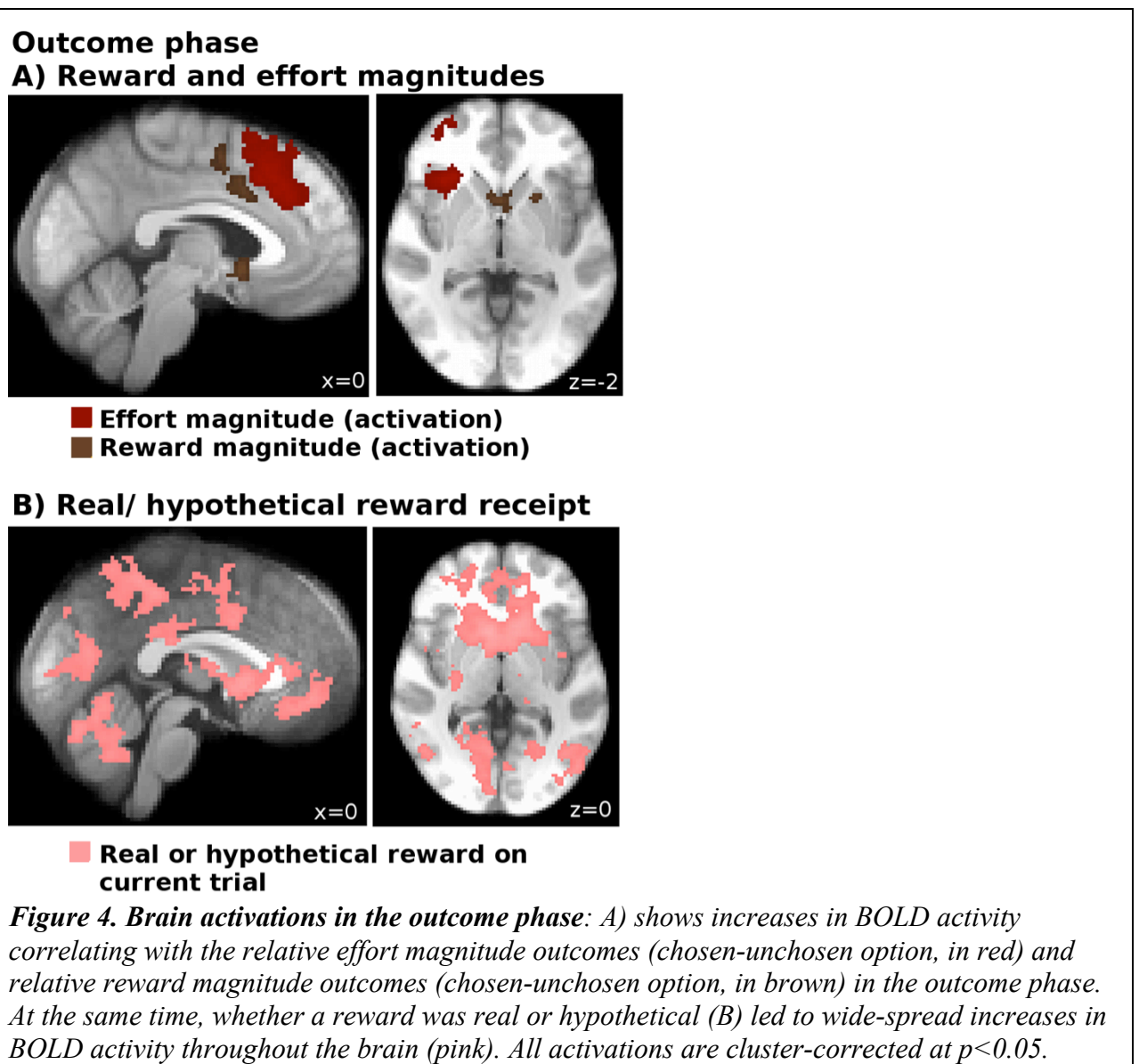
structure (namely that whether a reward is real or hypothetical should be irrelevant for future choice) and tries to oppose the bias. To investigate this question, we needed to interrogate the neural signals. We reasoned that the effect of reward type might be seen when the outcomes themselves were presented and participants needed to use the reward magnitude and effort magnitude information they received to update their expectations for their next choices. Alternatively, or additionally, an effect of reward type might be seen on the subsequent trial when participants next had to make a decision. Below we first consider reward type effects at the time of outcome processing and then in a later section we consider the effects of reward type on subsequent decision-related activity.

Reward magnitude and effort magnitude signals at outcome

When we examined activity during the outcome phase of trials (GLM1; see Methods), we found that the relative reward magnitude outcome (i.e. the chosen minus the unchosen reward magnitude) lead to an increase in BOLD activity in the ventral striatum and elsewhere (table 3, figure 4A, brown). There was an analogous effect of relative effort magnitude outcome (the chosen minus the unchosen effort magnitude) in frontal operculum/anterior insula (FO/AI), ACC and aPFC (figure 4A, red, see table 3 for a complete list of activations). We did not find any areas that showed a decrease in BOLD with either regressor. We note that we used relative outcome signals, rather than just the chosen outcome because in our task participants were shown information at outcome about both the chosen and the unchosen option. Therefore, for future decisions, the relevant quantity is how good an option is compared to the other option available. One could consider this quantity to be the relative evidence for switching/staying. However, in control analyses (not shown), we confirmed that the regions carried signals for both the chosen and the unchosen options separately (with reversed signs). Thus, our data does not speak to the issue of whether learning is about the relative evidence for the chosen versus the unchosen option or occurs separately in relation to the evidence for the chosen option and the

unchosen option. Both of these possibilities are compatible with both our fMRI and behaviour modelling findings, and further studies are needed to address this specific question.

Furthermore, we also note that as we used sustained effort, there is an inherent (and naturalistic) confound between amount of effort and duration of the effort (or delay). Comparison of our study results with other studies (Prevost et al., 2010) suggests that the effort related activations resembled those found in a pure effort task rather in a pure delay-discounting task.



Activations with Effort Outcome (C-UC)	x	y	z	z max	p	# Vox	Clus
FO/AI	30	24	-2	4.0	$4*10^{-4}$	605	4
ventral inferior frontal gyrus (44v ¹)	46	18	4	3.6			4
dACC/anterior rostral cingulate zone (RCZa)	4	36	32	3.7	$4*10^{-7}$		6
pre-SMA	2	18	54	3.2			6
inferior parietal lobule (PFm ²), right	52	-46	40	3.5	$5*10^{-6}$	954	5
inferior parietal lobule, (PFm ²), left	-28	-56	46	3.3	$4*10^{-3}$	448	2
area 46/ lateral frontal pole (FPI) ¹ , right	38	52	8	3.4	$6*10^{-4}$	581	3
dorsolateral prefrontal cortex (area 9/46v ³)	42	40	22	3.3			3
(additional separate cluster in area 9/46v)	46	18	38	3.3	0.01	393	1
Activations with Reward Outcome (C-UC)							
supplementary motor area/dACC/posterior rostral cingulate zone (RCZp ⁴)	2	0	52	3.5	0.01	431	1
dACC/ between posterior and anterior rostral cingulate zone (RCZp/RCZa ⁴)	8	10	38	3.3			1
inferior parietal lobule (PFop ²), left	-50	-42	28	3.8	$3*10^{-3}$	521	2
ventral striatum, left	-12	8	-10	4.1			3
ventral striatum, right	10	8	-10	3.7	$2*10^{-3}$	531	3
(cluster may extend to subcallosal cortex)	0	10	-6	3.3			3
Activations with reward type (real/hypothetical)							
putamen, right	14	12	-6	5.5	$2*10^{-87}$	37587	3
putamen, left	-18	14	-4	4.5			3
ventral striatum, left	-10	10	-8	5.1			3
ventral striatum, right	6	12	-2	5.0			3
ventromedial prefrontal cortex (vmPFC)	0	48	-4	3.4			3
perigenual ACC	-4	36	8	2.7			3
dorsal thalamus (extending to pulvinar), right	6	-18	14	3.1			3
dACC/posterior rostral cingulate zone (RCZp)	0	4	36	3.0			3
posterior cingulate (area 31 ⁴)	0	-48	50	2.8	$7*10^{-6}$	885	2
middle temporal gyrus, right	56	-48	-6	4.1			2
inferior parietal lobule (PF/PFt/hIP2 ²), right	44	-34	48	3.7			2
inferior parietal lobule (PF/PFt/hIP2 ²), left	-50	-40	50	3.5	0.03	309	1
occipital lobe	6	-76	8	3.7			1
cerebellum	4	-64	-16	4.1			1

Table 3. List of activations in the outcome phase (GLM1) for the difference in activity associated with the effort magnitude of the chosen option as opposed to the unchosen option (relative effort magnitude outcome), for the difference in activity associated with the reward magnitude of the chosen option as opposed to the unchosen option (relative reward magnitude outcome) and reward type (real vs. hypothetical). All results are cluster-corrected at $p < 0.05$. Shown are coordinates of local peaks (x, y, z), the max z score found at the peaks (z max), the p-value of each cluster, the number of voxels in each cluster (# Vox) and the index of the cluster the peak belongs to (Clus). Where indicated, anatomical labels were taken from ¹Neubert et al. (2014), ²Mars et al. (2011), ³Sallet et al. (2013), ⁴Beckman et al. (2009).

Representation of whether reward is real or hypothetical is widespread at the time of the outcome and relates to decision biases

We found widespread effects, in the outcome phase, of the reward type (real vs. hypothetical) in areas including the ventral striatum and the vmPFC (table 3, figure 4B, pink).

Notably, we also found activation with reward type in the aPFC, meaning that aPFC was more active (main effect) when outcomes were real rather than hypothetical (figure 4B, right hand panel, $t(19)=2.42$, $p=0.025$, statistical test in ROI, see the next section). At first glance, this might be surprising in as far as prior studies (e.g. Boorman et al., 2011) found that aPFC *deactivates* in proportion to the value of the option chosen. However, the result is consistent with the hypothesis that aPFC carries signals reporting the relative value of the choice not taken (the counterfactual choice), which would be the choice subjects would switch to if they could change their decision (and which they might switch to on the next trial) (Boorman et al., 2011). Thus, in our experiment, the activation found in relation with reward type is consistent with an enhanced switch signal that helps to overcome the stay bias introduced by real reward. If this interpretation is true, then individual differences in the strength of this activity should also relate to individual differences between participants in behavioural bias.

Therefore, we assessed next whether these BOLD increases related to the behavioural impact of reward type (real rewards biased participants towards staying with the same choice on the next trial, see figure 3C) using GLM2 (see Methods section *Relating neural effects of real versus hypothetical reward to behaviour*). Brain areas that are linked to the bias in behaviour induced by real rewards should be ones in which the difference in activity in response to a real versus a hypothetical reward is greatest in participants exhibiting the strongest behavioural bias. We found such an effect in vmPFC (figure 5, red). By contrast brain areas that are linked to resistance against the behavioural bias induced by real rewards should be ones in which the difference in activity in response to a real versus a hypothetical reward is greatest in participants

exhibiting the weakest behavioural bias. We found such an effect in aPFC and a dorsal part of ACC (figure 5, blue, table 4). Thus, this further supports the view that aPFC and dACC represent evidence in favor of switching to the alternative, and counteract signals advocating staying with the current choice in vmPFC when the reward was real as opposed to hypothetical.

Outcome phase - correlations between neural signals for real/ hypothetical reward and impact on decisions

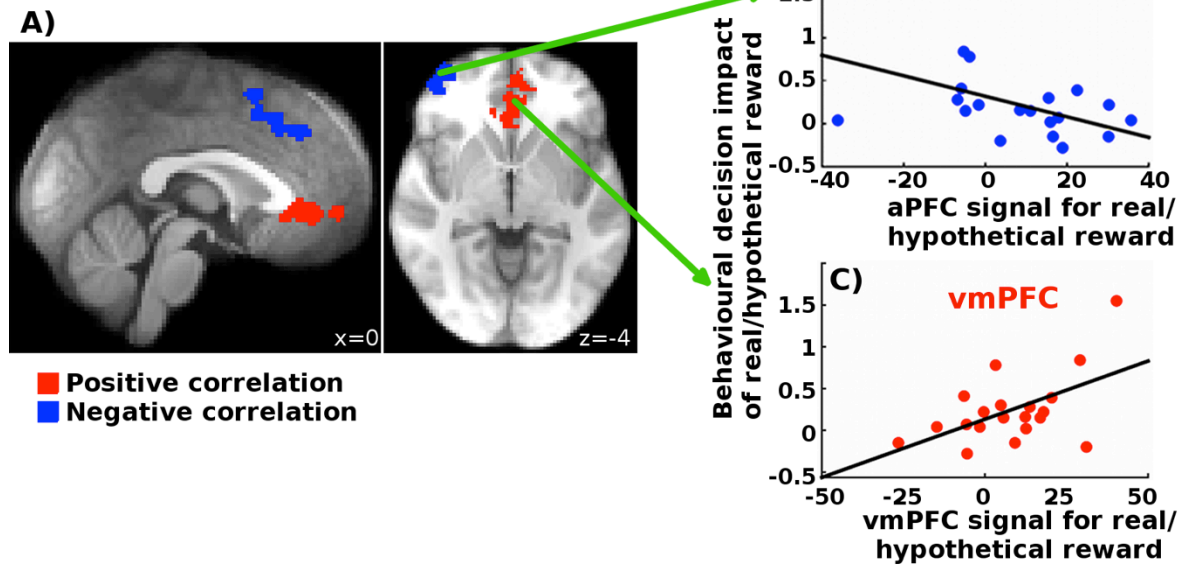


Figure 5. Correlations between the decision bias and the neural response to real versus hypothetical reward in the outcome phase: Participants who had a larger BOLD increase to real compared to hypothetical rewards in the vmPFC showed a larger behavioral bias (positive correlation in red). In contrast, participants who showed a larger activation to real compared to hypothetical rewards in the aPFC or the dACC showed a weaker behavioral bias (negative correlation in blue). All results are cluster corrected at $p < 0.05$. For illustration, we also show scatter plots of these correlations (B+C) using averages within spherical ROI's with a radius of 3 voxels in MNI space.

Positive correlations	x	y	z	z max	p	# Vox	Clus
vmPFC	-2	40	-2	3.2	$5 \cdot 10^{-3}$	417	1
Negative correlations							
Pre-SMA/dACC/anterior rostral cingulate zone (RCZa)	2	22	44	3.9	$6 \cdot 10^{-4}$	547	2
aPFC /lateral frontal pole (FPI) ¹	38	58	-2	3.5	$3 \cdot 10^{-3}$	438	1

Table 4. List of positive and negative correlations between individual differences in the neural signal for reward type (real vs. hypothetical) in the outcome phase and the behavioral decision bias (GLM2). Shown are coordinates of local peaks (x, y, z), the max z score found at the peaks (z max), the p-value of each cluster, the number of voxels in each cluster (# Vox) and the index of the cluster the peak belongs to (Clus) Where indicated, anatomical labels were taken from ¹Neubert et al. (2014).

To assess whether aPFC and vmPFC made independent contributions in relation to the decision bias, we performed partial correlations between the COPE values extracted from these regions (see table 4 for coordinates) and the behavioural impact of reward type (decision bias). We found that when controlling for aPFC correlations, the correlation between vmPFC and decision bias was still significant ($r=0.49$, $p=0.032$). Similarly the negative correlation between aPFC and the decision bias remained significant after controlling for vmPFC-behaviour correlations ($r=-0.498$, $p=0.03$). Similarly, we found that vmPFC is not a mediator for the effect of aPFC on behaviour ($z=-0.91$, $p=0.36$, Sobel test). Distinct neural processes linked to vmPFC and aPFC exert independent influences consistent with induction of the decision bias and with resistance to the decision bias. In contrast, when controlling for dACC, the correlation between aPFC and decision bias was no longer significant ($r=-0.34$, $p=0.16$) suggesting that activity in both of these areas reflects a common process related to resistance to the decision bias. Furthermore, while the activations of aPFC and dACC in response to real compared to hypothetical reward were strongly correlated ($p=0.642$, $p=0.002$), this was not true for aPFC and vmPFC ($r=-0.246$, $p=0.296$) or dACC and vmPFC ($r=-0.233$, $p=0.324$). This further supports the view that even though both aPFC and vmPFC were more active when a reward is real rather than hypothetical, activity in the two regions made independent and very different contributions to behaviour. In other words, aPFC did not exert its influence on behaviour by reducing the effect of reward on the vmPFC.

We note that as with any fMRI study, all results discussed here are correlational and further studies using techniques that can interfere with brain function will be needed to address questions about causality between brain and behaviour more directly.

APFC's and FO/AI's effort cost and reward outcome representations change when the reward is real as opposed to hypothetical

So far we have reported two main types of results. First, we have shown that reward magnitude and effort magnitude significantly affect activity in several frontal cortical regions in the outcome phase, when participants witness the consequence of their decisions. Second, we have shown that reward type (real vs. hypothetical) also affects activity in several frontal cortical regions in the outcome phases of trials and that this effect was related to the irrational behavioural bias to repeat choices associated with real as opposed to hypothetical rewards. Thus, the next obvious question was whether within areas coding reward magnitude and effort magnitude, there was any influence of reward type. This allowed us to arbitrate between two opposing hypotheses: First, a region could code real and hypothetical reward differently because it is sensitive to the rewarding aspect of an outcome. In this case, the effort magnitude should always be represented the same, independent of reward type, as the effort always needs to be executed regardless of the reward type. Second, as an alternative hypothesis, a region could represent reward and effort magnitudes differently when the reward is real if its activity is related to counteracting a decision bias – in this case we would expect it to represent reward and effort magnitudes more strongly when the reward is hypothetical in order to overcome the bias to stay with the rewarded option by enhancing the representation of the alternative option.

Using a timecourse analysis (figure 6), we found, to our surprise, that in the ventral striatum, there was no difference in the reward magnitude outcome effect for real vs hypothetical rewards ($t(19)=0.55$, $p=0.587$), even though there was a strong main effect i.e. BOLD increase when the reward was real vs. hypothetical ($t(19)=5.81$, $p<0.001$).

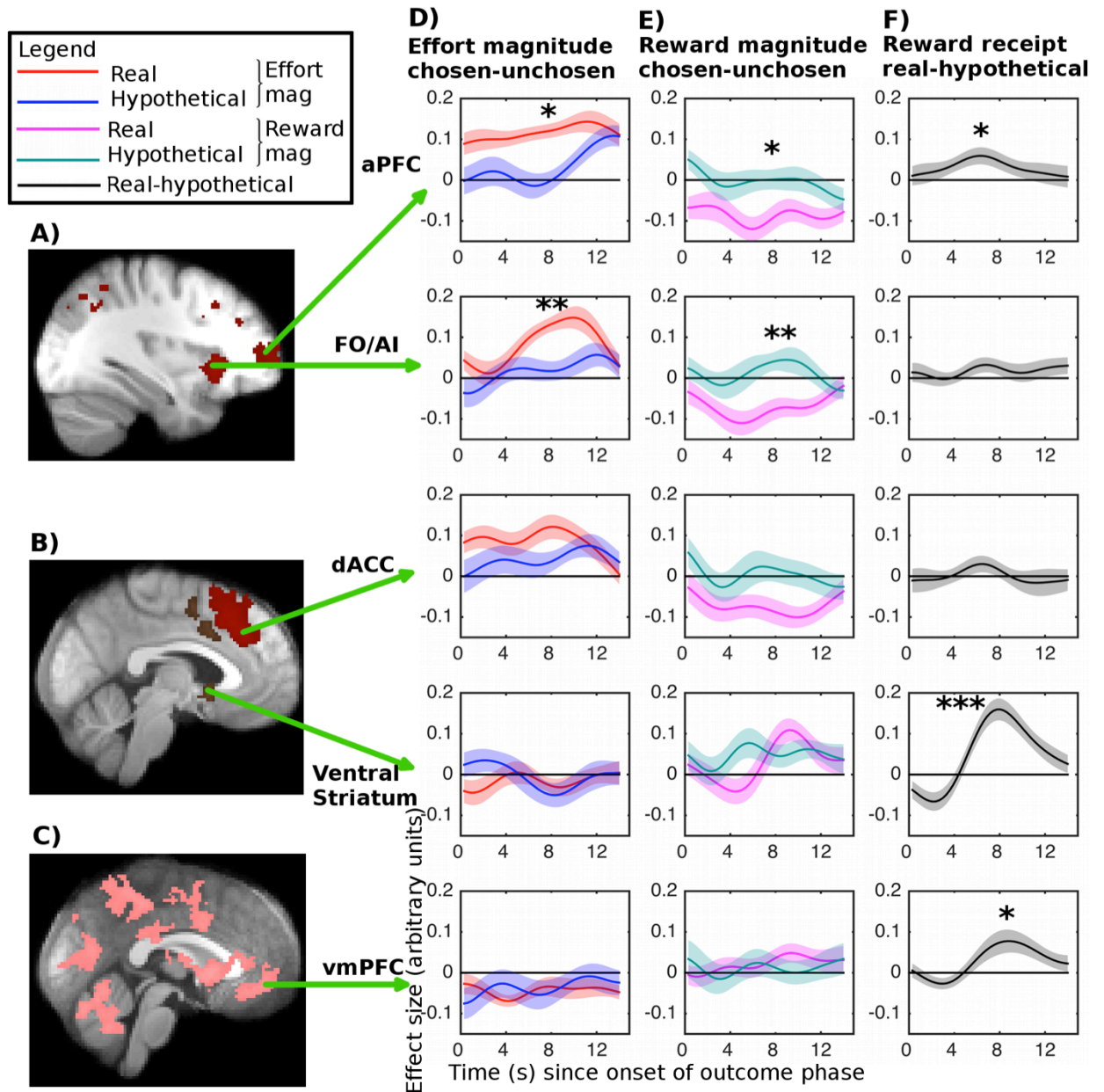


Figure 6. Timecourses from selected regions showing the main effect of real versus hypothetical reward and how the coding of the relative reward and effort magnitude outcomes is affected by the reward being real: A)-C) show the locations of the ROIs. Relative effort magnitude outcomes (chosen-unchosen option) (D) lead to a larger increase in BOLD when the reward was real rather than hypothetical in anterior prefrontal cortex (aPFC) and frontal operculum/ anterior insula (FO/AI), but not in ventral striatum. Similarly, relative reward magnitudes (chosen-unchosen option) (E) lead to a stronger decrease in BOLD when the reward was real rather than hypothetical in aPFC and FO/AI, but not in ventral striatum. Note (F) that whether the reward was real or hypothetical not only lead to an increase in BOLD in ventral striatum and ventromedial prefrontal cortex (vmPFC) but also in the aPFC. Significance [$*p < 0.05$, $**p < 0.01$, $***p < 0.001$] in D) and E) based on the result of paired two-tailed *t*-tests comparing the haemodynamically convolved timecourses from trials on which the reward was real or hypothetical. In F), significance tests were one-sample two-tailed *t*-tests. All ROIs were selected on the basis of an orthogonal contrast; aPFC, FO/AI and dACC ROIs were selected based on the whole-brain corrected contrast relative effort magnitude at the time of outcome (chosen option – unchosen option); the ventral striatum ROI was selected based on the whole-brain corrected contrast relative reward magnitude at time of outcome (chosen-unchosen)

option); the vmPFC ROI was selected based on the whole-brain corrected contrast real vs. hypothetical reward outcome. The ROIs were 3 voxels in radius in the case of all cortical regions (aPFC, FO/AI, dACC and vmPFC) and 2 voxels in radius in the case of the subcortical region (ventral striatum).

In contrast (figure 6), in aPFC and FO/AI, both reward magnitude and effort magnitude signals were significantly affected by reward type. In both regions, relative reward magnitude outcomes (chosen option reward magnitude outcome – unchosen option reward magnitude outcome) led to a BOLD decrease; related results have previously been reported (Boorman et al., 2009; Boorman et al., 2011; Boorman et al., 2013). But this effect was stronger when the rewards were real (aPFC: $t(19)=-2.314$, $p=0.032$; FO/AI: $t(19)=-3.671$, $p=0.002$). Similarly, relative effort magnitude outcomes (chosen option effort magnitude-unchosen option effort magnitude) led to a stronger increase in BOLD signal when the reward was real (aPFC: $t(19) = 2.27$, $p=0.035$; FO/AI: $t(19)=2.89$, $p=0.009$). This effect is striking because the effort that had to be exerted did not vary depending on whether reward was real or hypothetical. Instead the effect on effort and reward magnitude outcomes suggests that aPFC and FO/AI code the evidence in favor of the counterfactual option (the switch choice) more when there is more need to consider this because receiving a real reward biases behaviour towards staying with the current option. The pattern of results was qualitatively similar in dACC although it did not reach statistical significance (reward magnitude-related effects of reward type: $t(19)=-1.55$, $p=0.138$; effort magnitude-related effects of reward type: $t(19)=1.534$, $p=0.142$).

As already discussed in the last section, we also found an activation with real reward (main effect) in aPFC (figure 6, $t(19)=2.42$, $p=0.025$). This is not consistent with aPFC merely coding the value of the chosen option in an inverse or negative fashion. It is, however, like some of the other results found, consistent with the idea that aPFC codes the values of choices in a reference frame tied to the relative value of switching to the alternative choice as opposed to staying with the current choice.

APFC carries all information necessary for reigning in the impact of rewards on decisions and representing costs when necessary

From the series of analyses conducted so far aPFC repeatedly emerged as an area carrying various signals that are consistent with it playing a role in counteracting the bias to repeat choices (stay) introduced by real reward (figure 7). Firstly, aPFC activated when rewards were real rather than hypothetical, consistent with a switch signal to overcome the bias (GLM1, see also figures 4 and 6). Importantly, participants in whom this signal was stronger had a reduced behavioural bias or in other words, were better at overcoming the bias (GLM2, see also figure 5). Next, aPFC also showed an increase in the coding of the reward and effort magnitudes (GLM3, see also figure 6) when reward was real, again consistent with a role in overcoming the bias by enhancing the signal indexing the value of the alternative choice (switch) when real reward in other brain areas biased behaviour towards repeating the current choice (stay).

aPFC - overlap of neural effects

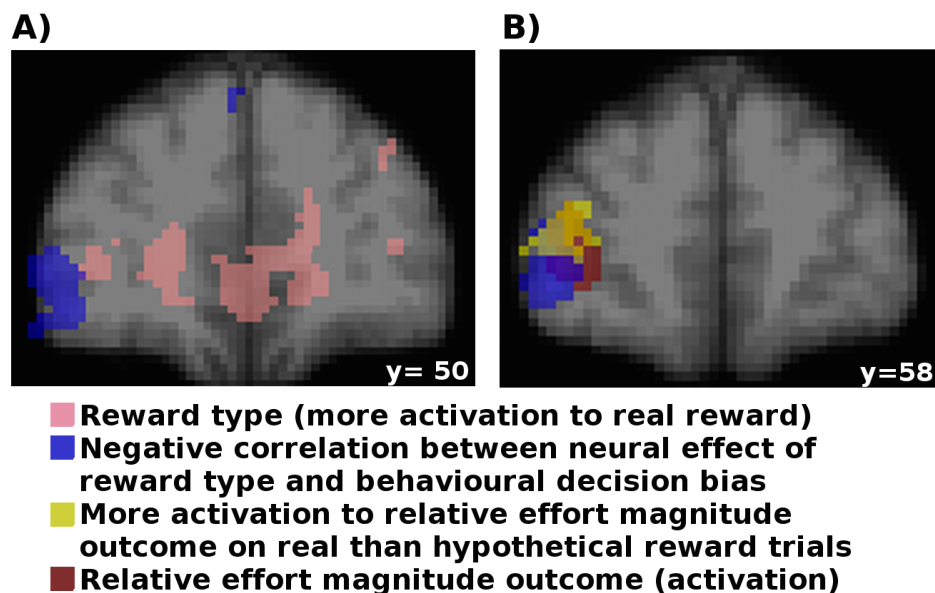


Figure 7. Various signals are present in the anterior prefrontal cortex (aPFC) during the outcome phase: Different signals are present in aPFC that are consistent with a role in overcoming a bias to stay with the current choice when there is real rather than hypothetical reward. Firstly (A), aPFC activity increases when rewards are real (pink) and participants with a stronger increase are better at overcoming the behavioral bias (blue). Secondly (B), the representation of effort magnitude outcomes increases in aPFC when reward is real (yellow). We also show the relative effort magnitude outcome contrast (red) used to identify our aPFC ROI (used for time courses in figure 6).

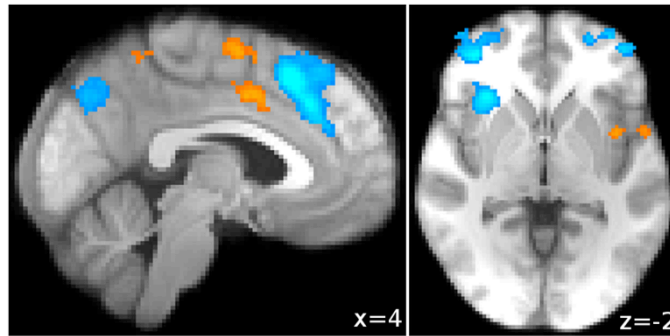
At the time of choice the vmPFC is sensitive to whether the reward was real or hypothetical on the last trial

To understand further what signals at the time of decision-making (as opposed to at the time of the previous trials' outcome phase) might mediate the real reward-induced bias to repeat choices we next investigated signals at the time of the decision phase. In addition to a regressor indexing the last trial's reward type (real vs. hypothetical), we also included regressors indexing all the rational decision variables that ought to have influenced decision making, i.e. the learnt reward and effort magnitudes and the explicitly shown probabilities (GLM 1, figure 2A). For simplicity of presentation, we combined these rational decision variables into a 'relative decision value' contrast. Specifically, the contrast for relative decision values included: chosen option value (reward magnitude + probability - effort magnitude) minus unchosen option value (reward magnitude + probability - effort magnitude). We found that relative decision value was associated with activity in several areas (figure 8A, table 5). For example relative decision value was associated with BOLD increases in a midcingulate region and decreases in dACC, FO/AI and aPFC. We also note that while the vmPFC did not carry a statistically significant signal of relative decision value, we found a sub-threshold activation.

The important result, however, is that whether the reward on the last trial had been real as opposed to hypothetical again affected vmPFC at the time of decision-making just as had been the case in the outcome phase (figure 8B, beige). This signal was significant even after whole brain cluster-based correction for multiple comparisons. The outcome-related and decision-related effects in vmPFC relating to whether rewards were real or hypothetical were distinct because of their large temporal separation. On average, 17 seconds elapsed between the outcome phase of one trial and the decision phase of the next trial (figure 1).

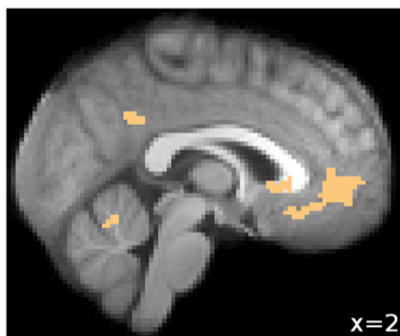
Decision Phase

A) Relative decision value



- Relative decision value activation
- Relative decision value deactivation

B) Real/ hypothetical reward receipt on the last trial



- Real or hypothetical reward on the last trial

Figure 8. Brain activations in the decision phase: Relative decision value (A), a linear contrast of the regressors for relative (chosen-unchosen option) predicted reward magnitude + shown reward probability – predicted effort magnitude led to increases (orange) and decreases (blue) in BOLD at the time of the decision. Regions activated in this contrast are regions in which activity is covarying with the value information that ought, rationally, to guide decision-making. Importantly, we also found increases in BOLD (B) in the vmPFC (beige) when a real rather than a hypothetical reward had been received on the last trial. VmPFC activity, therefore, covaries with an outcome feature that led to irrational biases in behavior. All results are cluster-corrected at $p < 0.05$.

Activations with relative decision value	x	y	z	z max	p	# Vox	Clus
inferior parietal lobule (PFop and PF/PFt/hlP2 ²), right	66	-30	32	4.4	1*10 ⁻⁷	1381	3
inferior parietal lobule (PFop and PF/PFt/hlP2 ²), left	-58	-38	34	3.9	7*10 ⁻⁷	1236	2
ventral premotor (area 6r ¹), left	-48	0	8	4.0	1*10 ⁻³	588	1
insula (middle), left	-38	2	4	3.3			1
posterior rostral cingulate zone (RCZp ⁴)	-2	2	48	3.8	1*10 ⁻⁷	1402	4
Deactivations with relative decision value							
preSMA	8	24	48	4.2	7*10 ⁻¹¹	2212	7
dACC/anterior rostral cingulate zone (RCZa ⁴)	-6	32	36	3.7			7
inferior frontal sulcus/ area 46 ¹ , right	40	48	-2	4.0	1*10 ⁻⁶	1186	5
dorsolateral prefrontal cortex (area 9/46V ³), left	-42	30	38	3.9	1*10 ⁻⁹	1868	6
aPFC/ lateral frontal pole(FPl), left	-24	60	4	3.6			6
inferior parietal lobule (PGa/PGp ²), right	48	-56	38	4.0	2*10 ⁻⁶	1119	4
inferior parietal lobule (PGa/PGp ²), left	-30	-64	34	3.3	2*10 ⁻⁴	727	3
FO/AI, right	32	26	-4	4.6	0.02	400	2
superior parietal lobe (area 7 ²)	6	-64	46	3.4	0.02	394	1
Activations with reward type (real/ hypothetical) on the last trial							
vmPFC	2	46	-2	3.3	4*10 ⁻¹²	2070	5
superior temporal gyrus	52	-8	-8	3.7	3*10 ⁻⁶	720	4
inferior parietal lobule (PGp ²), left	-36	-66	30	3.3	3*10 ⁻⁷	430	3
posterior cingulate (area 31 ⁴), right	16	-48	34	3.2	3*10 ⁻⁸	395	2
posterior cingulate (area 31 ⁴), left	-6	-54	22	3.0			2
cerebellum	-16	-48	-22	3.1	0.05	269	1

Table 5. List of activations and deactivations with relative decision value and of activation with last trial's reward type (real vs. hypothetical) during the decision phase (GLM1). Shown are coordinates of local peaks (x, y, z), the max z score found at the peaks (z max), the p-value of each cluster, the number of voxels in each cluster (# Vox) and the index of the cluster the peak belongs to (Clus). Where indicated, anatomical labels were taken from ¹Neubert et al. (2014), ²Mars et al. (2011), ³Sallet et al. (2013), ⁴Beckman et al. (2009).

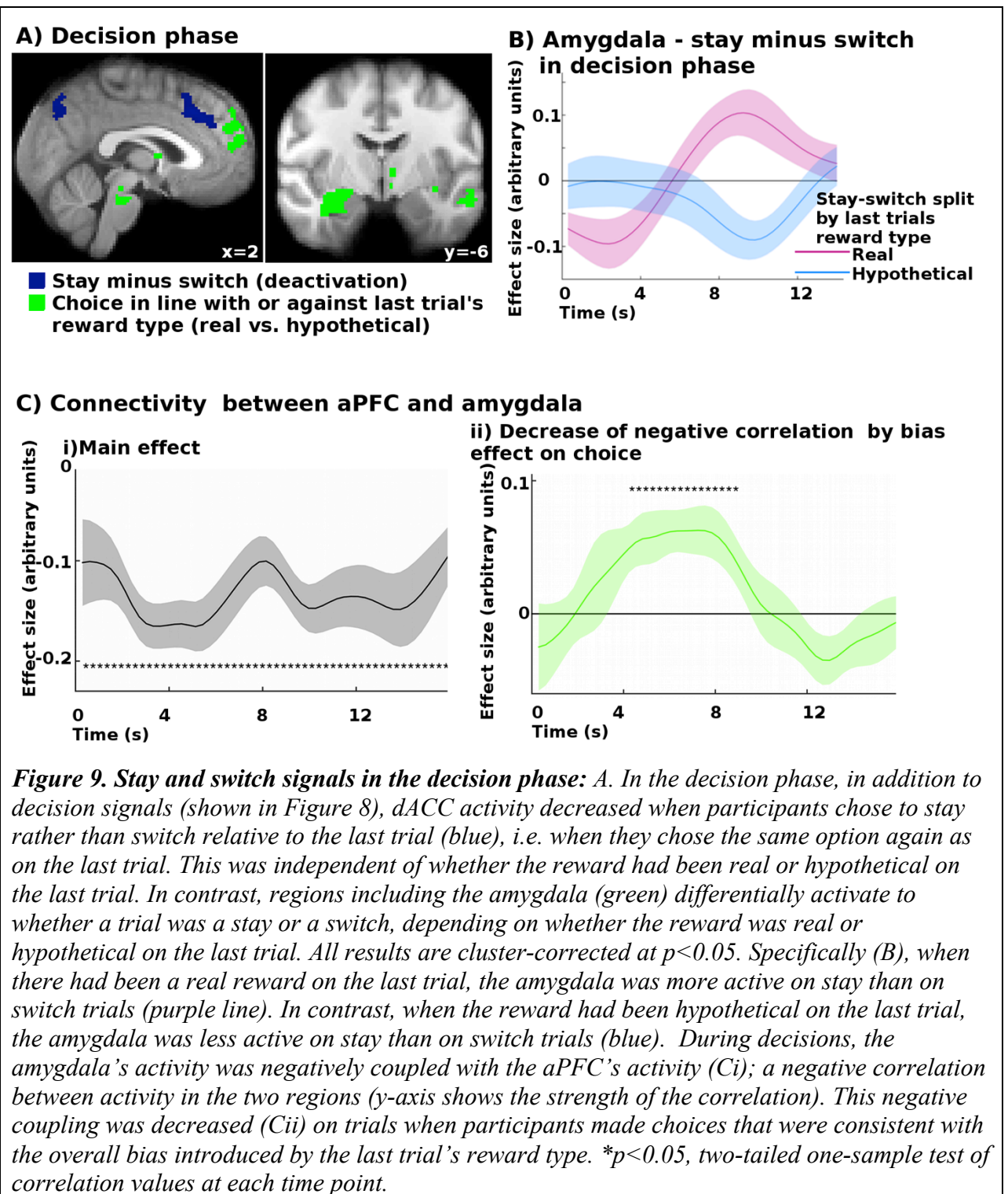
Amygdalar activity at the time of choice reflects the bias to repeat choices when reward was real as opposed to hypothetical on the last trial

So far, we have identified regions susceptible to the bias induced by real rewards because they activated differently when reward was real vs. hypothetical. Although such an activation pattern reflects the registration of the impact of the reward type, one could also imagine a region that then uses this information to bias behaviour, i.e. the decision to stay or switch. Previous studies

(de Martino et al. (2006), Roiser et al. (2009)) have found the amygdala to play such a role, more precisely they found the amygdala to activate more when decisions were in agreement with a frame (or bias) than when decision were made against it. In our case, this would translate to more activity when a decision is made to stay after a real reward or switch after a hypothetical reward; i.e. behaviour in both cases that is in line with the ‘frame’ introduced by the real/hypothetical reward on the last trial. And on the other hand, this would translate to reduced activity when a decision is made to switch after a real reward or stay after a hypothetical reward. We tested this hypothesis (GLM3) by including a stay-switch regressor and a regressor for last trial’s reward type (in addition to regressors controlling for the options’ values). This analysis then allowed examination of how such stay-switch signals interacted with the previous trials’ reward type: We compared activity associated with staying and switching on trials that followed either real or hypothetical rewards (i.e. a contrast of the regressors stay/switch on trials after a real reward minus stay/switch on trials after a hypothetical reward). The contrast revealed activations in the amygdala (figure 9A, green) and other regions (see table 6 for a full list of activations). For the next analyses we focused on the amygdala, because previous studies suggested this was a key area of interest (de Martino et al. (2006), Roiser et al. (2009)). However, further studies should also investigate the other areas found in the same contrast in more detail. To look at the effect in more detail, we extracted the BOLD timecourse from an ROI in the amygdala (figure 9B). We found that the amygdala was more active when participants decided to stay with the same option (repeat a choice) only after they had received a real reward on the last trial. In contrast, the amygdala was more active when participants decide to switch when the reward outcome on the last trial was only hypothetical. In other words, when the decision to stay or leave was made in line with the participants’ overall decision bias, the amygdala was more active.

As an exploratory analysis, we next hypothesized, by analogy with previous studies (Roiser et al., (2009)), that if the amygdala biases participants, its activity might be affected by an area that

opposes the bias. If this was the case, then decisions would be more likely to be made in line with the overall decision bias when connectivity between a frontal lobe area and amygdala is weaker (i.e. the frontal lobe area is less able to influence the amygdala). In our experiment, the frontal brain area most likely to have such a role was aPFC. To test this idea, we performed a PPI analysis in which we measured whether the functional coupling between amygdala and aPFC (measured as correlation in activity) was reduced when decisions were made in line with the overall decision bias. We found (figure 9Ci) that at the time of the decision aPFC and amygdala were, in general, negatively coupled ($p < 0.05$ at every time point, two-tailed one-sample t-test). This negative coupling was smaller (figure 9Cii) when decisions were made in line with the bias – shown as a reduction in the negative correlation between the two areas ($p < 0.05$ at 16 time points, two-tailed one-sample t-test). Thus in summary, we found that the functional connectivity or coupling between aPFC and amygdala may be linked to how well participants are able to prevent themselves from being biased. However, we also note that our design may not be the most ideal for examining such an interaction effect; in our task (in contrast to the task used by de Martino et al. 2006), decisions should be influenced by many factors and thus our categorization of whether a trial was in line with, or against, the bias was a simple one: for example on some trials, the rationally better option might also have happened to either be the option consistent or inconsistent with the bias. Thus further research is needed to probe the connectivity between these areas in more detail.



Deactivations with stay> switch	x	y	z	z max	p	# Vox	Clus
dACC/anterior rostral cingulate zone (RCZa)	-8	36	32	3.3	6*10 ⁻⁴	532	3
medial superior parietal cortex near IPS1/IPS2	4	-74	46	3.0	0.02	313	2
inferior parietal lobule (PGa ²), left	-40	-54	50	3.4	0.03	305	1
Activations for interaction (stay-switch) x (real/hypoth. reward on last trial)							
amygdala, right	30	-6	-20	3.1	7*10 ⁻⁶	910	5
putamen, right	28	14	-10	3.7			5
caudate, right	18	16	4	3.1			5
temporal fusiform cortex	-42	-32	-16	3.9	6*10 ⁻⁸	1379	6
putamen, left	-28	8	-10	3.6			6
hippocampus, left	-22	-24	-18	3.4			6
aPFC/medial frontal pole (FPm ³)	2	58	18	3.6	1*10 ⁻⁵	853	4
dACC/anterior rostral cingulate zone (RCZa ⁴), left	-14	10	40	3.2	2*10 ⁻⁴	669	3
brainstem	6	-24	-26	3.6	2*10 ⁻³	481	2
white matter in parietal cortex	22	-46	32	3.6	0.02	346	1

Table 6. List of activations for stay vs. switch and its interaction with last trial's reward type (real vs. hypothetical). Shown are coordinates of local peaks (x, y, z), the max z score found at the peaks (z max), the p-value of each cluster, the number of voxels in each cluster (# Vox) and the index of the cluster the peak belongs to (Clus). . Lists comprise main peaks and subpeaks (indented). Where indicated, anatomical labels were taken from ²Mars et al. (2011), ³Sallet et al. (2013), ⁴Beckman et al. (2009).

2.4 Discussion

In complex natural environments single choices can lead to multiple aspect outcomes. Humans should particularly learn about those aspects contingent upon choice – in our task such aspects were reward and effort magnitudes. Importantly, participants should not learn choice-outcome feature relationships when those outcome features have no contingent relationship with choice – in our task whether a reward was really received or only hypothetical. Participants were able to learn the real contingencies; they could use the informational content of reward, but they were also biased by its rewarding properties. At a neural level, we found areas possibly inducing (e.g. vmPFC and amygdala) and counteracting (aPFC, dACC, AI/FO) this bias.

Real rewards biased decisions

After obtaining real rather than hypothetical rewards participants were more likely to repeat choices. This irrational bias remained even after controlling for factors that should, rationally, have influenced decisions. This suggests participants perceived choices recently associated with real reward as preferable to those associated with hypothetical reward. Intriguingly, whether reward was real or hypothetical also affected how participants exerted effort. Using computational modelling, we found that this bias was not due to participants not learning about hypothetical rewards or treating hypothetical rewards as if no reward had occurred. Instead we found that the bias was best explained as participants simply preferring options that had led to a real reward independent of their other attributes that should rationally influence decisions (such as learnt reward or effort magnitudes or explicitly shown probabilities).

Brain networks driving decision bias

Two competing networks were related to the extent irrational biases manifested in behaviour. Activity patterns and between subject correlations suggested vmPFC and amygdala contributed to the bias while aPFC, together with FO/AI and dACC, counteracted it.

When participants saw choice outcomes vmPFC encoded whether reward had been real or hypothetical. We also found a positive correlation between this signal and participants' behavioural biases to repeat choices after real rewards. Participants whose vmPFC activity was more affected by whether reward was real were more biased.

VmPFC was not just active at choice outcome but also during decisions. Then vmPFC activity reflected whether reward had been real or hypothetical on the last trial but it did not significantly encode any quantity that should rationally have affected decisions such as reward and effort magnitudes or reward probabilities. These quantities were instead represented in aPFC, dACC, and FO/AI. This result recalls other studies that failed to find vmPFC activity when both rewards and effort costs had to be integrated before participants could choose (Burke et al., 2013; Croxson et al., 2009; Kurniawan et al., 2013; Prevost et al., 2010; but see also Skvortsova et al., 2014). It contrasts, however, with a large body of work confirming the importance of vmPFC in reward-guided decision-making (Rangel and Hare, 2010; Rushworth et al., 2012) and in subjective estimates of real reward contingencies (Tanaka et al., 2008).

VmPFC's role in inducing decision bias may be related to a more general role in choice repetition when this is advantageous. VmPFC activity at choice outcome predicts whether similar choices will be taken on the next opportunity (Boorman et al., 2011). Moreover, contrary to widespread beliefs about perseverative behaviour after frontal lesions, vmPFC lesions reduce rates at which advantageous choices are retaken (Noonan et al., 2012).

VmPFC has been linked to construction of task models to guide behaviour (Daw et al., 2011; Wunderlich et al., 2012a). In those studies, vmPFC contextualized real reward information relevant for decision-making in task models. In contrast, in our study, whether reward is real

should be completely ignored, rather than used within a model-framework. Thus our results suggest that while vmPFC may be good at contextualizing real reward within a task model, it may be less good at ignoring whether a reward is real or hypothetical.

Bias-related activity was also present in amygdala. Amygdala and vmPFC are monosynaptically interconnected and their activity is coupled (Carmichael and Price, 1995; Neubert et al., in press). Whereas vmPFC was always more active when the last trial had been rewarded, the amygdala was more active whenever choices accorded with the irrational behavioural bias that the vmPFC signal could induce. It may be helpful to consider this result in relation to the response potentiating effect that reward-related amygdala activity exerts in Pavlovian instrumental transfer (PIT) (Bray et al., 2008; Geurts et al., 2013a; Talmi et al., 2008) and to observations that amygdala choice coding changes when participants make decisions “framed” in terms of losses or gains and irrational changes in behaviour ensue (De Martino et al., 2006). It is also noteworthy that while we focused on the amygdala here, other areas were also more active when choices were in accordance with the irrational bias and further studies need to establish the specific roles they play.

Ventral striatum has also been implicated in the response-potentiating influences of reward in PIT (Bray et al., 2008; Geurts et al., 2013a; Prevost et al., 2012; Talmi et al., 2008) but these regions were less central to the behavioural biases identified in the present study. Ventral striatal activity increased identically when reward magnitude outcomes were larger regardless of whether outcomes were real or hypothetical although there was a main effect of reward type (real vs. hypothetical) that meant activity was greater when rewards were real as opposed to hypothetical. Thus, although ventral striatum was sensitive to reward type, this effect did not interact with reward magnitude. These results are similar to previous reports of no difference in coding of real or hypothetical rewards in ventral striatum (Bickel et al., 2009) and reports of both

fictive and real reward prediction errors in ventral striatum (Chiu et al., 2008; Lohrenz et al., 2007). These results are, however, surprising in the context of other brain regions that did code reward magnitudes differently when rewards were real rather than hypothetical.

A network including aPFC opposed the irrational bias in decision-making induced by real rewards

Having identified which regions bias behaviour when reward was real, we looked also for whether regions existed that opposed this bias. We found different activations in aPFC, FO/AI and dACC that contributed to overcoming the bias to repeat choices recently associated with real rewards. The results suggest that one reason that these regions code costs maybe not just because costs are aversive but also because they motivate change in behaviour when repetition is maladaptive.

The aPFC was particular interesting because it carried several important signals. In agreement with previous studies (Boorman et al., 2009; Boorman et al., 2011; Donoso et al., 2014), aPFC activation increased as the unchosen option's reward magnitude increased and deactivated with the chosen option's reward magnitude, suggesting aPFC represents the relative advantage of switching to the alternative option rather than staying with the current choice. We extend this finding by reporting a complementary pattern for the coding of learned effort cost magnitudes.

Coding of reward magnitude outcomes in aPFC and FO/AI was affected by whether reward outcome was real or hypothetical. The relative reward magnitude of the unchosen option was associated with increased activity when reward was real compared to hypothetical. This suggests it is not simply the hypothetical nature of choices that leads to them being represented in aPFC; if that had been the case then reward magnitude ought to have been represented more strongly when the outcome was hypothetical and effort representation should be unaffected by reward

type. Instead, these results suggest aPFC activity is driven by the importance reward and effort magnitudes have for switching to an alternative choice. Specifically, when a reward is real, brain regions such as vmPFC bias behaviour towards choice repetition. To overcome this bias it may be necessary to enhance the representation of the alternative choice in aPFC. Moreover, as functional coupling between amygdala and aPFC increased decisions were less likely to be made in a biased manner.

The importance of aPFC in overcoming the bias is further supported by the observation that reward type (real vs. hypothetical) led to an increase in activation in aPFC. This increase in activation would be surprising if aPFC only coded a negative value signal, as then real reward should have led to a deactivation compared to hypothetical reward. Instead, activation in aPFC is related to promoting the choice of the alternative option. In agreement with this interpretation, we also found that, participants with stronger activation to reward type in aPFC were better at counteracting the bias to repeat choices previously followed by real rewards. Thus, while individual variation in vmPFC activity was positively correlated with susceptibility to decision biases induced by real rewards the relationship was the opposite way round for aPFC. Further analysis demonstrated aPFC and vmPFC constitute two independent systems, one biasing behaviour towards staying with rewarded options and the other one counteracting this bias when it is important to do so. APFC, together with ACC, is not only involved in value-guided exploration (Daw et al., 2006; Kolling et al., 2012), but also in counteracting reward-based choice repetition when this should be avoided. In a similar vein Kolling et al., (2014) suggested activity in aPFC and dACC, as opposed to vmPFC, guides decision making in risky contexts to maximize longer term reward even when this did not maximize reward on the current trial.

Conclusion

In conclusion, these results highlight that human learning is subject to biases by the (emotional) experience of rewards. Neurally, this bias was linked to activity in vmPFC and amygdala. While on the other hand a network centred on aPFC showed activity patterns indicative of trying to overcome the bias. In the next chapters, we will examine whether these diverse processes of reward/ effort learning in the presence of emotional/experiential biases are affected by different pharmacological treatments of depression and symptoms of depression.

3. Serotonergic plasticity in concurrent reward and effort learning

Summary

Serotonin re-uptake inhibitors are the most common treatment for depression. While it is still unclear how they work, different theories have been proposed for their mechanisms of action (chapter 1.3). One prominent theory highlights the role of changes in valenced, or emotional, processing. On the other hand, another theory highlights the role of serotonin in neural plasticity, based on animal studies. In the following, we designed an experiment to test whether prolonged increases of serotonin could indeed improve plasticity related learning *in humans*. And furthermore whether these effects were secondary to changes in the processing of valence. For this, 29 healthy volunteers were randomly assigned to two weeks of the SSRI citalopram or placebo, in a double-blind between-subject design. We then measured their learning behaviour in the concurrent reward and effort learning task described in chapter 2, while monitoring their brain activity with fMRI. We found that prolonged increase in serotonin enhanced reward and effort learning (i.e. prediction error) signals in a widespread network of brain regions, including ventromedial, anterior cingulate, and posterior cingulate cortex. Importantly, enhanced reward- and effort-related learning signals manifested in the absence of changes to how serotonin affected the coding of the value of reward or effort receipts per se. This suggests that serotonin directly influences learning signals in humans beyond its putative role in the coding of appetitive or aversive signals. Behaviourally, we observed that these neural changes were accompanied by reward learning which was more robust to interference. Our results highlight the importance of studying the role of serotonin on longer timescales. Such effects may partly underlie the therapeutic impact of serotonergic drugs given to patients with psychological illnesses such as depression.

3.1 Introduction

To make good decisions in complex environments, humans and animals need to learn about and integrate different sources of information such as the good and bad aspects of outcomes. The neurotransmitter serotonin has been implicated in value-based choice and assumed to play a role in value learning, but even basic aspects of its operation and its function in such learning remain contested.

Serotonin has recently been implicated in simple aspects of reward-guided learning and decision making. For example, when serotonergic neuron activity is recorded on a millisecond timescale or serotonin levels are manipulated acutely, serotonin has been found to code information about aspects of good/appetitive (Fonseca et al., 2015; Liu et al., 2014; Miyazaki et al., 2014; Seymour et al., 2012) or bad/aversive (Amo et al., 2014; Cohen et al., 2015) outcomes or to relate to avoidance behaviors (Boureau and Dayan, 2011; Crockett et al., 2012a). On this basis, it has been suggested that it has a role in learning (Chamberlain et al., 2006a). However, other studies have found no effects (Guitart-Masip et al., 2014; Murphy et al., 2002; Schweighofer et al., 2008; Seymour et al., 2012; Worbe et al., 2015) or have not dissociated learning from altered responsiveness to the valence of the reinforcing events themselves. Thus, there is still no clear understanding of serotonin's role in value learning. Here we propose that this gap can be bridged by examining the effects of serotonin on value learning over a different timescale.

Based on animal studies, it is known that serotonin not only transfers information on millisecond timescales, but also acts over protracted time scales of days and weeks. In fact, at the neuronal level, prolonged increases in serotonin over such timescales lead to changes in plasticity (Bessa et al., 2009; Jitsuki et al., 2011). Increases in serotonin levels for several weeks, can re-introduce juvenile-like plasticity in visual cortex (Vetencourt et al., 2008) and the limbic

system (Karpova et al., 2011) in animals, in turn relating to behavioral improvements in learning about visual stimuli and fear extinction. Interestingly, the timeframe of plasticity observed in animals is very similar to the timeframe for antidepressants to take effect in patients so it may be particularly revealing to study serotonin's neural and behavioral effects during learning at this timescale.

We therefore tested here whether prolonged increases in serotonin affect appetitive and aversive learning independent of the effects serotonin may have on the coding of stimulus valence. We recruited 29 human participants, who received 20mg/day of the selective serotonin re-uptake inhibitor (SSRI) citalopram for two weeks or placebo. They subsequently performed a learning task while we measured their brain activity using functional magnetic resonance imaging (fMRI). In the task, participants concurrently learned about the value of two stimuli to choose between them. The value of each stimulus was determined by an appetitive (amount of monetary reward) as well as an aversive (amount of effort) dimension. This allowed us to tease apart the different neural effects serotonin might have on the representation of appetitive and aversive information. It also made it possible to dissociate the effect serotonin had on responses to receipt of positive and negative outcomes *per se* as opposed to the effect serotonin had on the impact positive and negative outcomes had *on learning*. We found that increased serotonin led to stronger reward and effort-related *learning* signals (i.e. reward and effort prediction errors, the difference between the received and the expected outcomes), in a widespread network of brain areas coding value information. At the same time, however, activity related simply to the receipt of reward/effort outcomes *per se* was unaffected. This suggests that serotonin directly influences learning signals in humans beyond its putative role in the coding of appetitive or aversive signals. Not only were neural reward-related learning signals more precise, but at a behavioral level we were able to show that reward-related learning behavior was also improved.

3.2 Methods

Participants

The study was approved by the local ethics committee. 46 healthy volunteers were recruited for the study screening session after providing written informed consent. They underwent psychiatric and medical screening before being included in the behavioural and MRI part of the study: They were included if they were right handed, they passed basic medical screening for current or past physical illness relevant to the study, were non- or light-smokers (<5 cigarettes a day), had taken no CNS-active medication during the last 8 weeks, were screened to be free of a past or present Axis 1 disorder on the Structured Clinical Interview for DSM-IV (First, 1996), had no first-degree relative with bipolar disorder, were fluent in English and had no counter-indications to fMRI. Female participants were neither pregnant nor breast-feeding. After the screening, 8 participants were excluded because of counter-indications to citalopram or the MRI scan, 6 participants were excluded because they were not sensitive to the effort (i.e. they did not find the effort task effortful or were willing to forfeit money to avoid the effort) and 2 participants dropped out before the MRI scan due to side-effects of citalopram. Of the participants taking part in the MRI session, data from one participant was excluded from analysis because he/she fell asleep during the MRI scan. There remained 15 participants in the citalopram group and 14 in the placebo group. The data of the participants in the placebo group is also included in chapter 2, focusing on the basic cognitive and neural effects. Based on previous studies on the effects of SSRIs on the perception of positive or negative stimuli (Harmer et al., 2004), if the expected effect of citalopram on learning is a 20% increase, a study with 15 participants per group has 90% power to detect this effect. However, it is of note that no human study has measured the effect of prolonged citalopram administration on learning; therefore, it is difficult to estimate an expected effect size.

Procedure

In a double-blind, between-subject design, participants were randomly assigned to two weeks of a clinical dose of citalopram (20mg/day, $n=15$) or placebo ($n=14$). They then performed a multi-dimensional magnitude learning task (see below) while we recorded their brain using fMRI. Additionally, they completed the following questionnaires during the initial screening visit and/or on the day of testing (i.e. after having taken citalopram or placebo for two weeks): Beck's Depression Inventory (BDI, Beck et al. (1996)), State-trait anxiety inventory (Spielberger and Gorsuch, 1983), Positive and negative affect schedule (Panas, Watson et al. (1988)), Cloninger scale (Cloninger et al., 1993). Blinding of participants to the drug condition was assessed by asking them informally which group they believed to be in.

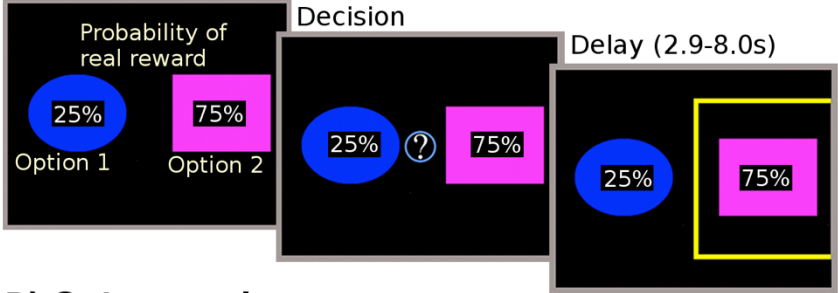
Magnitude learning task

The same learning task as detailed in chapter 2 was used. In short, in the task (Fig. 1), participants repeatedly chose between the same two options, aiming to choose the options maximizing their monetary gain and minimizing the effort they needed to exert to obtain the reward. When making their decision (Fig. 1A), they therefore had to take into account the independent reward and the effort magnitudes of the two options, which they had to learn and track from experience across trials as they slowly varied over the course of the experiment (Fig. 1D). At the time of the decision, they were additionally explicitly shown how likely each option was to lead to a reward outcome. If an option led to a real reward outcome, participants received the money as pay off, if not they were only shown how much money they could have won. After each choice, participants saw the outcome of their choice (Fig. 1B). At this time point, participants could learn about the two options. This task was designed to be challenging for participants: they had to simultaneously learn about all the different attributes, as well as learn about rewards in the same way when they were real or hypothetical. Thus, optimal learning

needed to represent all these components separately to *prevent them interfering with one another*.

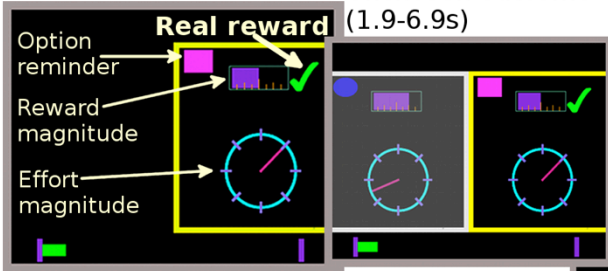
A) Decision phase

Options (1.4-4.5s)

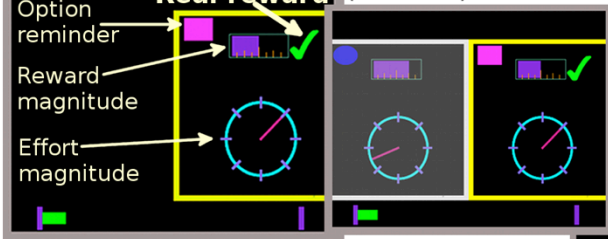


B) Outcome phase

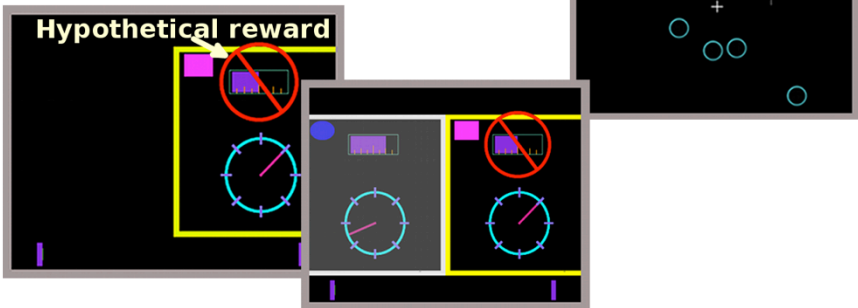
Chosen outcome (1.9-2.1s)



Unchosen outcome (1.9-6.9s)



C) Effort phase



D) Task schedule for reward and effort

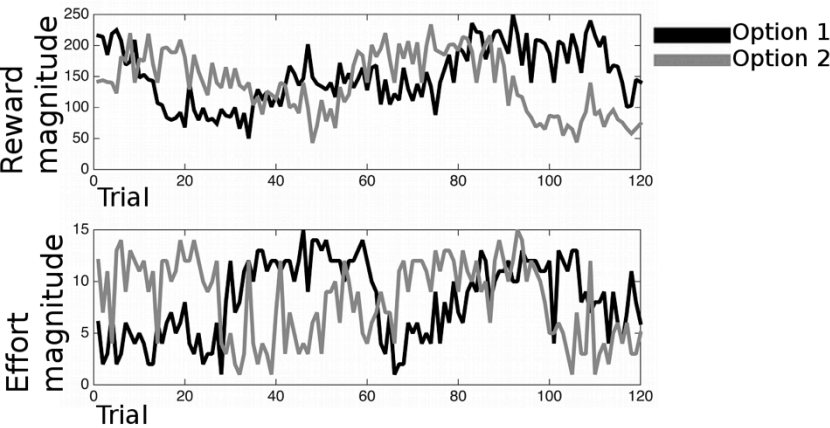


Figure 1. Task description: *In the decision phase (A), participants were shown two options (i.e. choices), overlaid with the probability of receiving a reward for each choice. They could only decide after an initial monitoring phase (1.4-4.5s). The chosen option was then highlighted for 2.9-8s. In the following outcome phase (B), participants saw the outcome for the chosen option first (1.9-2.1s). The reward magnitude was shown as a purple bar (top of the screen), the effort magnitude was indicated through the position of a dial on a circle. Whether they received a reward was indicated by a green tick mark (real reward, top display) or a red crossed out sign over the reward magnitude (hypothetical reward, bottom display). If a reward was real, the reward was also added to a status bar at the bottom of the screen, which tracked rewards over the course of the experiment. A reminder of what option they had chosen was shown at the top of the screen. Then the reward and effort magnitudes were shown for the unchosen option (1.9-6.9s). Finally, participants performed the effort phase (C) where the number of targets was equivalent to the chosen effort outcome. Participants had to perform the effort phase on every trial, independent of whether the reward was real or hypothetical. An example schedule is shown in (D), with both the reward and effort magnitude values of the two choices.*

Behavioural analyses

This study investigated how reward and effort learning was modulated by a prolonged increase in serotonin when both dimensions needed to be learned concurrently and could potentially interfere with one another. In this arguably more naturalistic situation, learning could be enhanced in several ways. For example, learning could be improved by changing the learning rate (i.e. making learning faster). Alternatively, learning could be made more robust in the presence of different stimuli that could otherwise interfere with learning. For example, learning about effort could interfere with learning about reward and making reward learning signals more robust could reduce this interference. The following analyses were performed to answer these questions. All analyses were performed in Matlab and SPSS. Greenhouse-Geisser corrections were used where appropriate.

Analysing the effects of citalopram on the speed of learning

In a first behavioural analysis (bGLM1), we tested whether learning speed was affected by citalopram. We ran a logistic regression analysis predicting whether participants ‘stayed’ with the same choice as made on the previous trial (chose it again on the current trial) or ‘switched’ to the other option, based on the options’ current explicitly cued reward probabilities and the reward and effort magnitude outcomes they had seen on the previous four trials. These

regressors were included as relative values, i.e. value in favor of the stay option minus in favor of the switch option to reduce the total number of regressors. We further included a regressor denoting the last trial's reward type (real vs. hypothetical). To investigate whether the groups differed in their learning speeds, the resulting regression weights for the past reward and effort magnitude outcomes were entered into an ANOVA with group (citalopram vs. placebo) as a between-subject factor and time (1, 2, 3, or 4 trials in the past) and valence (reward or effort) as within-subject factors.

As a complementary analysis, we also fitted a computational learning model. We followed similar procedures as described in Scholl et al. (2014), see chapter 4. In short, the model consisted of three main components: firstly the model had predictions of the mean reward/effort magnitudes underlying both outcomes. These were updated on every trial using a reinforcement-learning algorithm:

$$Prediction_t = Prediction_{t-1} + \alpha * PE_{t-1}$$

with

$$PE_{t-1} = Outcome_{t-1} - Prediction_{t-1}$$

where α was the learning rate. Thus, the learning rate was a measure of how much participants updated their reward/effort magnitude prediction when the reward/effort magnitude outcome differed from their expectation (i.e. depending on the prediction error). Secondly, the model then combined these reward/effort magnitude predictions together with the explicitly shown probabilities to calculate how valuable each option was (i.e. their utility):

$$Utility_{OptionA} = \frac{1}{k} * MagnitudePrediction_{Reward} + \frac{\gamma}{k} * Probability_{Reward} - \frac{\lambda}{k} * MagnitudePrediction_{Effort}$$

where k was a normalization constant with $k=1+\gamma+\lambda$, γ was the probability decision weight and λ was the effort magnitude decision weight. The Utility for option B was computed in the same

way. Thirdly, the model then compared the utility of the two options to predict how likely participants would be to choose either, using a standard soft-max decision rule:

$$P(\text{Option}_A) = \frac{e^{\beta * \text{Utility}_A}}{e^{\beta * \text{Utility}_A} + e^{\beta * \text{Utility}_B}}$$

where β reflected participants' ability to pick the option with the higher utility (i.e. the inverse temperature).

The data from each participant was fitted individually using Matlab's `fminsearch` routine, which adjusted the free parameters to minimize the difference between the predicted choice and the actual choice. Data from 'special option trials' was included for updating the reward/effort magnitude predictions, but was not used for fitting the model.

We fitted two separate models, one in which the same learning rate was used for learning about reward and effort magnitudes and one in which two separate learning rates were used, both yielding identical results in terms of group comparisons. We used t-tests to compare the learning rates between groups.

Reward learning interference

To assess whether citalopram protected reward learning from interference, we performed a regression analysis assessing how well information to be learnt from one trial (in the form of prediction errors, PEs see below) was used in the next trial when there was interference or when there was no interference (bGLM2, similar to (Chiu et al., 2008)). We first obtained reward PEs (RPEs) using the Bayesian learning model described in chapter 2. Including these RPEs in a regression analysis, it is possible to obtain a corresponding regression weight for each participant which measure how much they took the last trial's information to be learnt (i.e. the PEs) into account when making decisions on the next trial. Following this logic, we included in the regression analysis as our main effects of interest interactions between those relative RPEs (stay

minus switch option) and the factors that could cause interference: whether reward was real or hypothetical and how much the effort was worse than expected (i.e. relative effort prediction error, EPE). The regression weights of these interaction regressors then tell us how much participants' learning is subject to interference. As regressors of no interest we also included the relative explicitly shown probabilities, the relative reward and effort predictions (also obtained from the Bayesian model), the main effects of reward type (real vs. hypothetical), and the relative reward and effort prediction errors (see Fig. 2A for a list of regressors and correlations). To investigate whether there was interference, the thus obtained regression weights of the two interaction regressors (RPE x Last reward type and RPE x EPE) were entered into an ANOVA with group as a between subject factor. As the size of the regression weights indicated how much regressors influenced behavior, stronger interference was indicated by larger weights for the interaction regressors. To further probe the results of bGLM2, we compared to what extent participants could use RPEs on trials when reward was real or hypothetical using a regression analysis (bGLM3). We included as our two regressors of interest relative RPEs (stay minus switch option) separately on trials when reward was real or hypothetical. As regressors of no interest we also included the relative explicitly shown probabilities, the relative reward and effort predictions and the relative effort prediction errors. We compared the resulting regression weights for RPEs when reward was real or when reward was hypothetical using t-tests. One advantage of a regression analysis based on (Bayesian) value estimates and PEs, like in (Chiu et al., 2008; Rutledge et al., 2014), is that it makes behavioral and fMRI analysis more analogous.

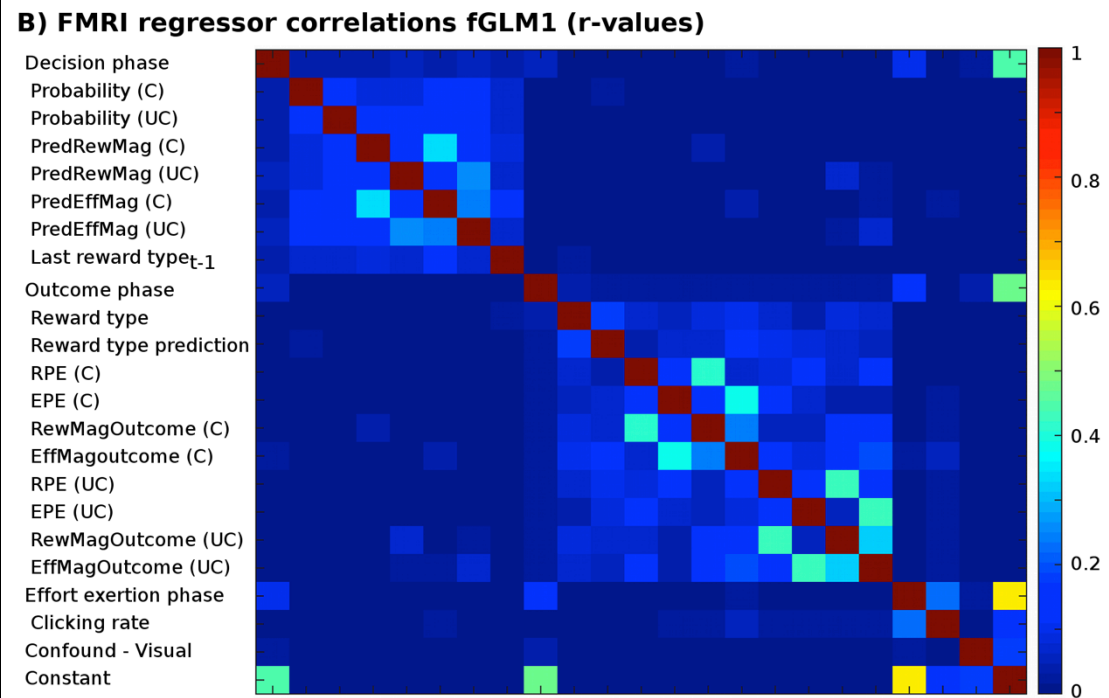
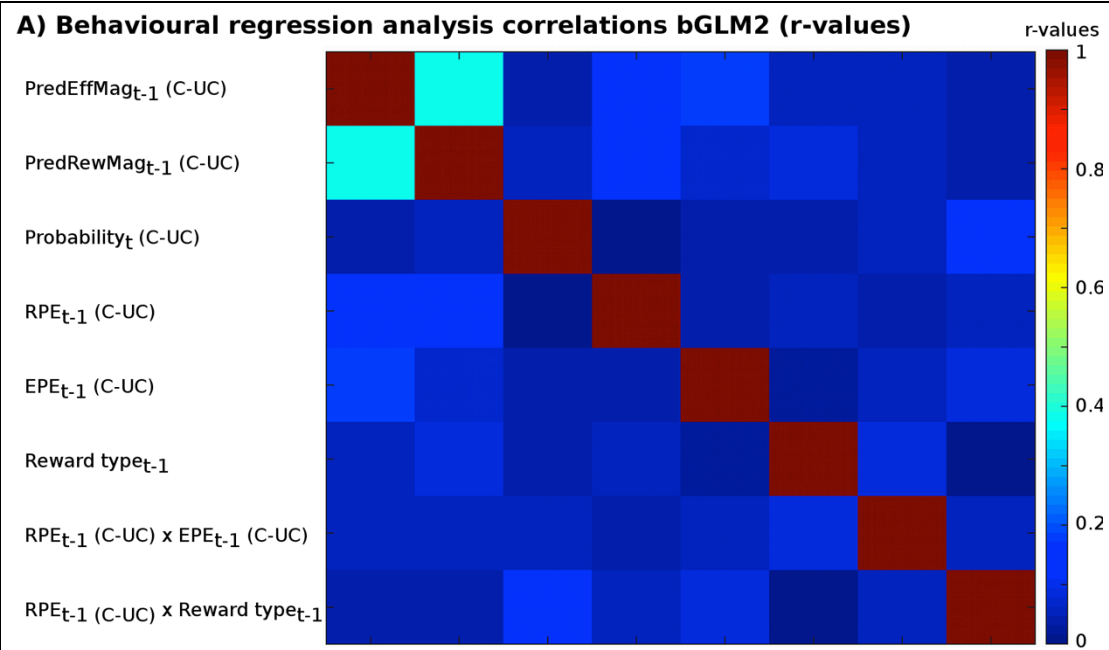


Figure 2. Correlations of regressors. Correlations of all the regressors used in the behavioural regression analysis (bGLM2) are shown in A). The values are the mean of the absolute correlation values (r-values) across all participants. No r-values exceeded 0.38. B) shows the correlations of the regressors used in the fMRI design fGLM1. No r-values exceeded 0.44. Abbreviations: A) Predicted effort/reward magnitude on trial $t-1$ (PredEffMag_{t-1}, PredRewMag_{t-1}), Effort/reward prediction error on trial $t-1$ (RPE_{t-1}, EPE_{t-1}), Interaction between reward prediction error and effort prediction error on trial $t-1$ (RPE_{t-1}(C-UC) x EPE_{t-1}(C-UC), Interaction between reward prediction error and reward type on trial $t-1$ (RPE_{t-1}(C-UC) x Reward type_{t-1}(C-UC), Chosen option (C), Unchosen option (UC); B) Predicted reward/effort magnitude (PredRewMag, PredEffMag), Reward type on the last trial before the current decision phase (Last reward type_{t-1}), Reward/effort prediction error (RPE, EPE), Reward/effort magnitude outcome (RewMagOutcome, EffMagOutcome), Chosen option (C), Unchosen option (UC)

MRI data acquisition and pre-processing

Structural MRI and fMRI measurements were taken using a Siemens 3 Tesla MRI scanner as detailed in chapter 2.

fMRI analysis

Our main goal for the fMRI analysis was to test whether citalopram changed learning and outcome signals for reward or effort. With fMRI, neural learning signals can be measured in the form of prediction errors (PEs), representing the difference between outcome and expectation, which should highly correlate with new learning. Thus, if citalopram increased synaptic plasticity and therefore induced more learning related changes in neural activity (Castren, 2013), this would be expected to manifest itself in increased PE signals. By contrast, other theories suggest serotonin is involved in the coding of the valence of outcomes (Boureau and Dayan, 2011; Daw et al., 2002). In this case, we would expect citalopram to affect the overall coding of reward or effort outcome signals in the brain, which could then in turn also lead to changed learning (and thus PEs). In our paradigm, PEs were sufficiently decorrelated from the outcomes themselves that the impact of both could be investigated in the same analysis (see below, also Fig. 2B). This was possible because participants needed to learn reward and effort magnitudes rather than probabilities.

Whole-brain

In the first fMRI analysis, we investigated whether citalopram affected neural learning signals. For this, we looked at whether reward or effort prediction error signals differed between the groups, at the time of the outcome (fGLM1). The regressors used in this design were as follows (all correlations between regressors were $r < 0.44$, Fig. 2B): We used three boxcar regressors, indicating the onset and duration of the decision phase (from the beginning of the trial until participants made a choice), the onset and duration of the outcome phase (from the appearance of

the chosen outcome until the chosen and the unchosen outcomes disappeared from the screen) and lastly the effort exertion phase (from the appearance of the first effort target until participants had removed the last target). We included the following parametric regressors in the decision phase: whether the reward of the chosen option had been real or hypothetical on the last trial, reward and effort magnitude predictions (from the Bayesian learning model described above) and the reward probabilities that were displayed on the screen. In each case separate regressors for the chosen and the unchosen option were used (they were later combined to relative, i.e. chosen minus unchosen, values at the contrast level). In the outcome phase we included the following parametric regressors: the reward type (real vs. hypothetical) delivered for the chosen option, the reward probability for the chosen option, the reward and effort magnitude outcomes for the chosen and the unchosen option. As our main regressors of interest, we included the reward and effort prediction errors for the chosen and the unchosen option. The onset of all regressors for the chosen option were time-locked to the onset of the outcome phase; the onset and duration of the regressors for the unchosen option were time-locked to their display. In the effort execution phase, we included the clicking rate as a parametric regressor. Finally, we included, as confound regressors, six movement regressors and a regressor indexing when additional visual stimuli were presented to warn participants that they had not clicked the targets on time and that the halfway point of the experiment had been reached.

Time course analyses

We also used fGLM1 to identify regions for time course analyses. The aim of these analyses was to test whether citalopram could enhance reward or effort prediction error and/or outcome signals. All time courses were extracted, regressed and statistically tested as described in chapter 2. The leave-one out procedure for statistical testing of the regression weights (obtained by correlating regressors with the haemodynamic response functions) was adjusted for two groups in the following way: For each participant, we used the data from all but that participant of

his/her group (placebo or citalopram) to determine the absolute peak (i.e. either the strongest peak or trough) of the time course in a window between 6s and 12s. We then aligned the peak of a canonical haemodynamic response function to this peak. The haemodynamic response function was made using `gammampdf` in matlab, with values $\alpha = 7^2/3^2$, $\beta = 7/3^2$. We then convolved the aligned haemodynamic response function with the omitted participant's time course to obtain one value per participant for each regressor of interest.

In the outcome phase, the three sets of ROIs were regions coding reward type, relative (chosen minus unchosen option) reward outcomes and effort outcomes. In those areas we then tested whether the two groups differed in reward/effort PEs and in the coding of the reward/effort outcomes (fGLM2). Please note that these analyses were orthogonal to how the ROIs were identified.

In fGLM2, we included as regressors of interest the reward and effort prediction errors (separately for the chosen and the unchosen option), as well as the relative reward and effort magnitude outcomes. As regressors of no interest, we included the reward type (real vs. hypothetical) prediction error (i.e. the reward type of the trial minus the explicit probability of the chosen option). Prediction errors were included as separate regressors for chosen and unchosen options as previous studies (Boorman et al., 2011) suggested that different areas might carry signals for chosen and unchosen prediction errors. In contrast, chosen and unchosen outcome signals have been found to be present in the same areas with opposite signs (Boorman et al., 2009; Scholl et al., in press) and were therefore included as a relative value regressor. fGLM2 was time-locked to the onset of the outcome phase.

Control analyses

In the behavioural analysis (see results), we found that citalopram improved the learning from reward prediction errors in situations of interference. We therefore looked for related effects in the neural signal. In fGLM3 (analogous to the behavioural results shown in Fig. 3Bii, bGLM3), we included as regressors of interest the reward prediction error (of the chosen option) separately on trials when reward was real and on trials when reward was hypothetical. As regressors of no interest we included again the reward type prediction error (i.e. reward type of the trial minus explicit probability of the chosen option), the unchosen reward magnitude prediction error, the chosen and unchosen effort magnitude prediction errors and the relative (chosen minus unchosen) reward and effort magnitude outcomes. The ROIs in which we performed these analyses were again the five areas identified as either being sensitive to reward type or as coding a whole-brain corrected group difference in RPE. We performed a 2 (reward type, real vs. hypothetical reward) x 5 (area) x 2 (group) ANOVA on the reward prediction errors to test whether the reward prediction error signal in the citalopram group was selectively stronger compared to placebo in situation of interference by reward type; i.e. the effect of interest would be a reward type by group interaction. In fGLM4 (analogous to the behavioural analysis in Fig. 3Bi, bGLM2), we included as the main regressor of interest the interaction term between the reward prediction error and the relative (chosen minus unchosen) effort prediction error. As regressors of no interest we included the reward type prediction error, the reward prediction errors (chosen and unchosen option separately), the effort prediction errors (chosen and unchosen option separately), the interaction term between reward prediction error and reward type, as well as the relative (chosen minus unchosen option) reward and effort magnitude outcomes. We then performed a 5 (area) x 2 (group) ANOVA comparing the interaction term between reward prediction error and relative effort prediction errors across the groups.

3.3 Results

Participants

The two groups did not differ in any sociodemographic or personality parameters (Table 1A). Furthermore, participants did not differ in their BDI, trait, state anxiety, PANAS or persistence scores after taking citalopram or placebo (Table 1B, all $p > 0.4$). Neither was there a difference between the groups in the questionnaire scores after citalopram/placebo compared to baseline (Table 1C, all $p > 0.05$). It is however of note that there was a trend ($p = 0.07$ for value ‘blinding’, table 1B) for participants to correctly guess whether they had received citalopram or placebo.

Measure	Citalopram group			Placebo group			Group diff. (p-value)
	# Sub	Mean	Std. Err	# Sub	Mean	Std. Err	
A) Baseline							
Age	15	23.3	1.1	14	22.8	0.8	0.74
Gender (# female)	15	7.0		14	6.0		0.84
BDI	15	1.5	0.4	14	1.5	0.6	1.00
Trait anxiety	15	29.5	1.7	14	32.6	2.1	0.24
State Anxiety	14	32.1	2.6	13	32.9	2.3	0.84
PANAS (positive)	14	30.6	1.5	13	31.5	1.8	0.68
PANAS (negative)	14	14.3	1.1	13	13.9	1.2	0.82
B) Test							
BDI	15	3.4	0.7	13	2.9	0.7	0.60
Trait anxiety	15	30.4	2.0	12	33.2	2.9	0.43
State anxiety	14	28.6	1.3	13	30.8	1.7	0.32
PANAS (positive)	14	32.0	1.3	11	30.1	1.3	0.32
PANAS (negative)	14	12.4	1.0	11	12.7	1.1	0.84
Cloninger scale	15	13.4	0.8	13	12.9	1.1	0.68
Blinding	14	2.7	1.7	13	-1.9	1.7	0.07
C) Baseline-test							
BDI	15	-1.9	0.7	13	-1.3	0.5	0.45
Trait anxiety	15	-0.9	1.0	12	-1.8	1.9	0.66
State anxiety	13	3.9	2.3	12	2.8	2.1	0.73
PANAS (positive)	13	-1.8	1.3	10	2.1	1.6	0.07
PANAS (negative)	13	1.6	0.9	10	2.0	1.4	0.82

Table 1. Participant demographics and questionnaire scores. Participants did not differ in any baseline measurements (A), in any measurements at the time of the test, i.e. after two weeks of citalopram or placebo (B) or in any difference scores between the time of test and baseline (C). The number of subjects for who measures were available is given for each measurement (# Sub), together with the mean value of each group, the standard deviation and the results of a two-tailed t-test comparing the two groups. Abbreviations and questionnaires used: Beck's

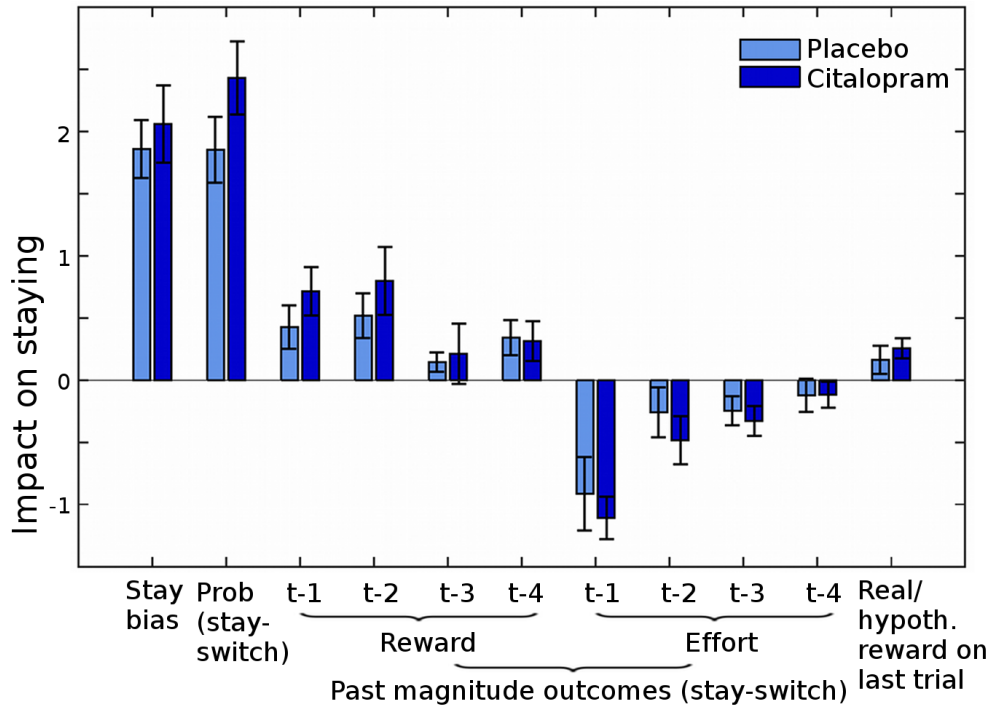
Depression Inventory (BDI, Beck et al. (1996)), State-trait anxiety inventory (Spielberger and Gorsuch, 1983), Positive and negative affect schedule (Panas, Watson et al. (1988)), Cloninger scale (Cloninger et al., 1993). Blinding of participants to the drug condition was assessed by asking them informally which group they believed to be in (positive values indicate citalopram, negative values placebo) and how certain they were (10=very certain, 0=very uncertain).

Citalopram makes reward learning more robust against negative interference

This study investigated how reward and effort learning was modulated by a prolonged increase in serotonin when both dimensions needed to be learned concurrently and could potentially interfere with one another. For this, we measured participants' brain activity while they performed a concurrent reward and effort learning task.

We first analyzed whether citalopram influenced the speed of learning about rewards or effort using a regression analysis (Fig. 3A, analysis bGLM1). While we found that both groups could learn about reward and effort, we did not find that citalopram affected the speed of learning (ANOVA, group x time: $F(3,81)=0.42$, $p=0.74$; group x valence (reward/effort) x time: $F(3,81)=0.02$, $p=0.995$). The advantage of such an ANOVA is that it tests for any difference in the use of past reward information over time, i.e. is not reliant on a particular change in shape (e.g. exponential decay). However, as a complementary analysis, we also measure speed of learning using a computational learning model. We found again that speed of learning (or learning rate) did not differ between the groups, neither when a common learning rate was fitted for learning about reward and effort magnitudes ($t(27)=0.6$, $p=0.53$) nor when two separate learning rates were fitted (reward: $t(27)=0.6$, $p=0.53$; effort: $t(27)=0.29$, $p=0.78$).

A) Decision regression



B) Learning interference

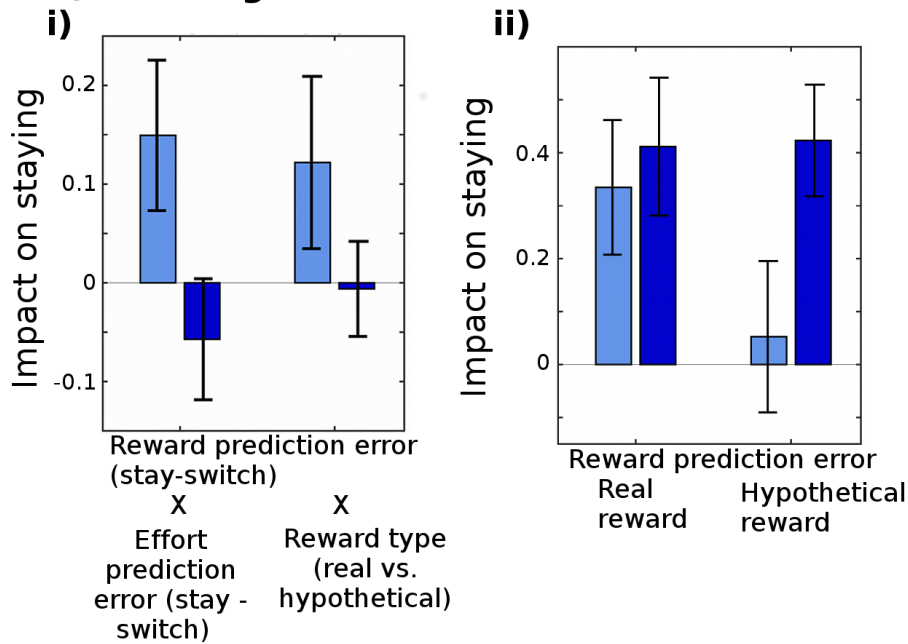


Figure 3. Behavioral results. A) A regression analysis (bGLM1) showed that participants were more likely stay with an option (i.e. chose it again) rather than switch to the alternative, as a function of the explicitly shown probabilities on that trial and higher reward and lower effort magnitude outcomes on the past one to four trials (t-1 to t-4). Participants were also more likely to stay with an option after they received real rather than hypothetical reward. The groups did not differ in their speed of learning (ANOVA, group $\times$ time: $F(3,81)=0.42$, $p=0.74$; group $\times$ valence (reward/effort) $\times$ time: $F(3,81)=0.02$, $p=0.995$), nor in the other parameters influencing switch/stay decision. B) To investigate how much learning about rewards was interfered with in the two groups we ran another regression analysis (Bi, bGLM2) predicting again staying or switching, but now including as regressors of interest the interactions between the reward prediction error and interfering factors, namely the reward type (real vs. hypothetical) and the

effort prediction error; other regressors of no interest (see methods), are not shown. Participants in the placebo group had larger interference effects in their reward learning than the citalopram group, i.e. the interaction terms had a larger impact on stay behavior (ANOVA, main effect of group $F(1,27)=6.14$, $p=0.02$). A closer look at the interaction (Bii, bGLM3) between the reward prediction error and reward type revealed that while both groups could use the reward prediction error to make choices when the reward was real (i.e. it impacted stay/switch behavior similarly), only the citalopram group could use it when the reward was hypothetical ($t(27)=-2.11$, $p=0.045$).

However, we did find that citalopram made learning more robust to interference: while placebo participants were not as good at using the reward learning signals across trials when there was interference, this effect was remedied by citalopram (bGLM2). More specifically, we performed a regression analysis examining how much participants could use reward learning signals (i.e. reward prediction errors) to make their choice and to what extent this was subject to interference. Reward prediction errors (RPEs) are a key component of classic learning theories. They drive learning or updating of predictions by indicating how unexpected the outcome is. Mathematically this is expressed as the difference between the actual outcome and the prior prediction/expectation. As our paradigm was complex, participants' learning about rewards could be subject to interference (e.g. through competition for limited processing resources). Interference with reward learning could occur if there are higher than expected effort outcomes (i.e. effort prediction errors, EPE). Secondly, the fact that rewards were only hypothetical rather than real on some trials could also affect people's ability to learn about reward magnitudes. We assessed whether SSRI administration affected such interference. We tested for the effects of interference in the regression analysis by including the interfering factors as interaction terms with the RPEs (bGLM2). A positive interaction term between RPE and reward type (real vs. hypothetical) or surprisingly high effort would then mean participants were not as efficient at using the RPE when reward was hypothetical compared to real, or effort was surprisingly high. Or in other words, that their learning about rewards was subject to interference. Indeed, the analysis showed that placebo participants' reward learning was more subject to interference compared to participants taking citalopram, i.e. the placebo participants had larger regression

weights for the interaction terms (Fig. 3Bi, ANOVA, main effect of group $F(1,27)=6.14$, $p=0.02$). Follow-up t-tests revealed that when the interference by effort effect was examined in isolation the difference was significant ($t(27)=2.13$, $p=0.04$). When the interference by hypothetical rewards was examined in isolation the effect was not significant ($t(27)=1.31$, $p=0.21$), but further analysis (bGLM3) revealed that it was only when rewards were real that both groups could use the reward prediction errors effectively on the next trial (Fig. 3Bii, $t(27)=-0.42$, $p=0.68$). In contrast, when rewards were hypothetical, the placebo group was no longer able to use reward prediction errors to guide effective decisions on the next trial, while this was remedied by citalopram ($t(27)=-2.11$, $p=0.045$).

In summary, we found that citalopram did not affect the overall speed of learning, but rather made reward learning more robust to interference.

FMRI

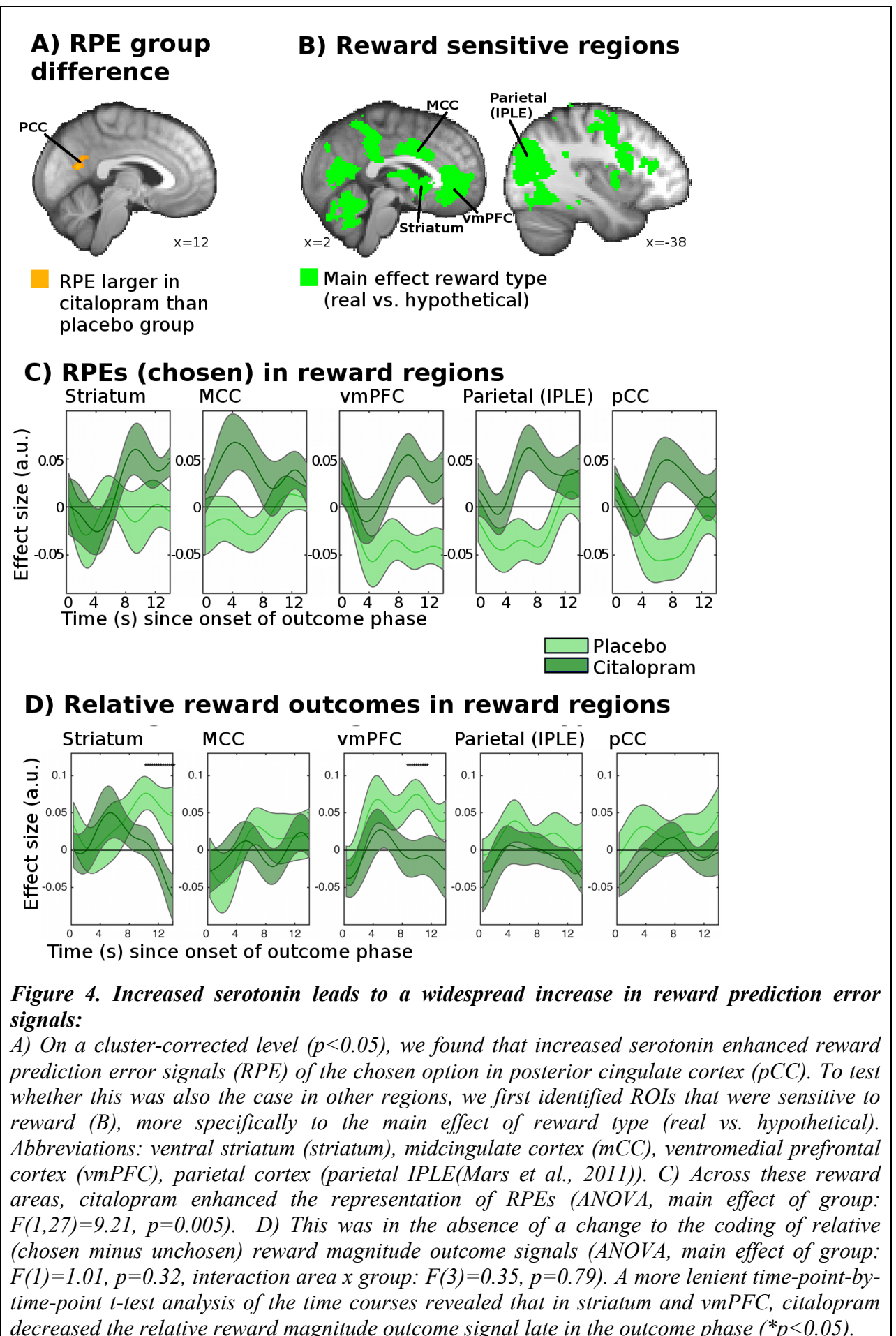
Our main goal for the FMRI analysis was to test whether citalopram changed learning and outcome signals for reward or effort. With FMRI, neural learning signals can be measured in the form of prediction errors (PEs), representing the difference between outcome and expectation, which should highly correlate with new learning. Thus, if citalopram increased synaptic plasticity and therefore induced more learning related changes in neural activity (Castren, 2013), this would be expected to manifest itself in increased PE signals. By contrast, other theories suggest serotonin is involved in the coding of the valence of outcomes (Boureau and Dayan, 2011; Daw et al., 2002). In this case, we would expect citalopram to affect the overall coding of reward or effort outcome signals in the brain, which could then in turn also lead to changed learning (and thus PEs).

Increased serotonin enhances reward prediction errors in reward sensitive areas

We measured neural signals related to learning at the outcome phase of the task. To test whether citalopram affected RPEs, we first selected regions of interest (ROIs) that activated to the receipt of reward (fGLM1, Fig. 4B, Table 2A, reward type regressor), and had previously been implicated in the processing of rewards or learning. These regions included striatum and ventromedial prefrontal cortex (vmPFC). We found (Fig. 4C, fGLM2), that in those ROIs, citalopram consistently enhanced the RPEs for the chosen option (ANOVA, main effect of group $F(1,27)=9.21$, $p=0.005$). When we repeated the same analysis on the whole brain, we additionally identified the posterior cingulate cortex as a region where citalopram led to stronger RPEs (Fig. 4A, fGLM1). We next tested whether this effect was specific to only some trial types, i.e. trials when there was interference on reward magnitude learning, as we had found in the behavioral analysis (see methods ‘Control analyses’, fGLM3 and fGLM4). We did not find that the groups differed in their RPEs more when there was interference because rewards were only hypothetical (interaction group x reward type condition: $F(1)=0.09$, $p=0.765$), or because the effort was surprisingly high (ANOVA, main effect of group: $F(1)=0.077$, $p=0.784$). In summary, citalopram had a general effect enhancing RPE signals in all contexts even though the enhanced effect of RPEs on behavior in the citalopram group was most apparent when there was a possibility of interference. This may be related to greater sensitivity of fMRI compared to behavioral analyses. Another set of ROIs that were sensitive to some aspect of reward were areas that deactivated with the relative (chosen minus unchosen) reward outcomes (Fig. 5, Table 2B). In these regions we did not find RPEs for the chosen option in either group, in agreement with previous studies (Boorman et al., 2011), and could therefore not test for group differences.

Next, we tested whether citalopram also affected the reward magnitude outcome signals (fGLM2). We found that citalopram did not increase reward magnitude outcome signals either in

the ROIs that activated to the receipt of reward (Fig. 4D) nor in the ROIs that deactivated with relative chosen reward magnitude outcomes (ANOVA, effect of group: $F(1)=0.03$, $p=0.86$). On the contrary, there was even a trend for reward outcome signals to be reduced in striatum and vmPFC by citalopram (Fig. 4D).



A) Reward type (real vs. hypothetical) - activations

	x	y	z	z max	p	# Vox	Clus
ventral striatum (bilateral)	-8	10	-6	4.12	$1*10^{-100}$	41620	3
ventromedial prefrontal cortex (vmPFC)	2	38	-8	3.57			3
midcingulate cortex	2	4	34	3.13			3
Posterior parietal cortex (IPLE/PGp)	-38	-74	30	5.31			3
Posterior cingulate cortex (cluster 9)	0	-34	36	3.32			3
lateral occipital, left	-42	-72	18	4.42			3
lateral occipital, right	40	-78	18	3.47			3
interior frontal gyrus	-40	8	20	3.99			3
middle temporal cortex	60	-46	-4	3.81	0.05	262	1
cerebellum	-18	-56	-30	3.35	$3*10^{-3}$	408	2

B) Relative effort outcome (C-UC) - activations

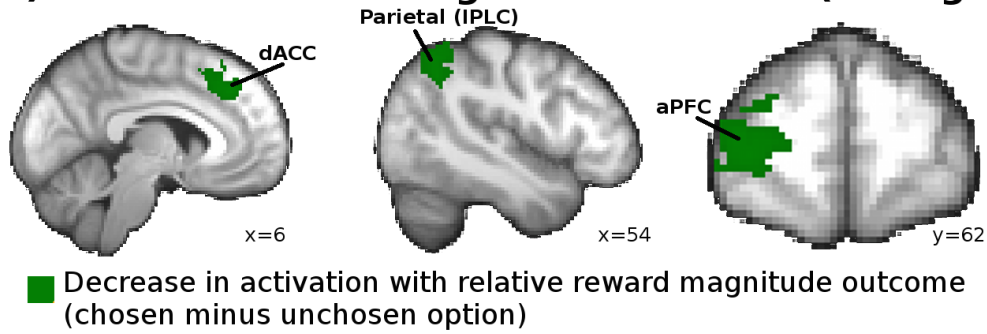
Posterior parietal, right (IPLC/PFm)	54	-46	42	4.18	$3*10^{-10}$	2003	6
Frontal operculum/ anterior insula, right	28	28	0	3.90	$2*10^{-8}$	1562	5
Cerebellum	-12	-80	-30	4.72	$1*10^{-6}$	1162	4
anterior prefrontal cortex (aPFC, areas 10 and 46), extending to dorsolateral prefrontal cortex	22	58	0	3.49	$3*10^{-4}$	668	3
Middle temporal gyrus	64	-38	-8	3.57	$1*10^{-3}$	567	2
dorsal anterior cingulate cortex (dACC)	4	36	30	2.81	$4*10^{-3}$	476	1

C) Relative reward outcome (C-UC) - deactivations

anterior prefrontal cortex (aPFC, areas 10 and 46 [3])	28	62	8	3.36	$1*10^{-4}$	809	4
posterior parietal , right (IPLC/PFm)	54	-46	42	3.54	$1*10^{-4}$	799	3
posterior parietal, left	-50	-58	42	3.64	$2*10^{-3}$	584	1
dorsal anterior cingulate cortex (dACC)	6	36	38	3.63	$1*10^{-3}$	638	2
dorsolateral prefrontal cortex (Area 9/ 46V)	36	24	38	3.41	$2*10^{-5}$	994	5

Table 2. Brain activity in the outcome phase. A) Areas activating more to real than hypothetical reward (main effect). B) Areas deactivating with the relative reward magnitude outcome (chose minus unchosen option). C) Areas activation with the relative effort magnitude outcome (chosen minus unchosen option). Shown are coordinates of local peaks (x, y, z), the max z score found at the peaks (z max), the p-value of each cluster, the number of voxels in each cluster (# Vox) and the index of the cluster the peak belongs to (Clus). Where relevant, references for area labels are indicated.

A) Relative reward magnitude outcome (both groups)



B) RPEs (unchosen option)

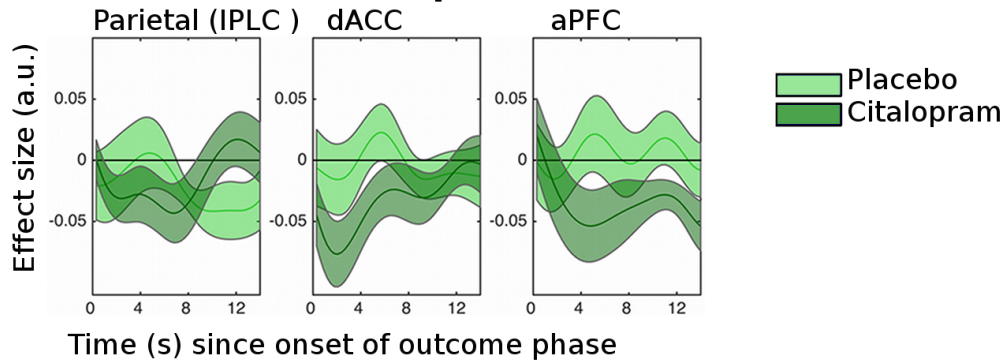


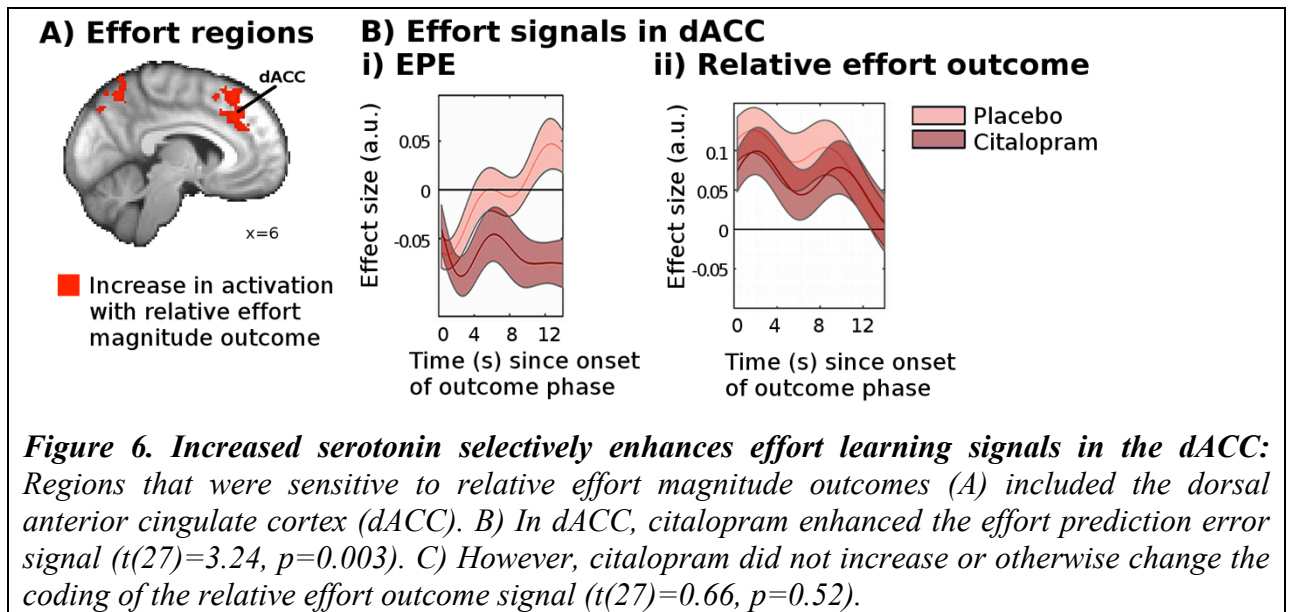
Figure 5. Reward prediction errors and outcome signals in regions that deactivated with relative reward magnitude outcomes. We identified areas in the outcome phase that decreased in activity with the relative (chosen minus unchosen option) reward magnitude outcomes (A) (see also Supplementary Table S2B). The relative reward magnitude outcome signals did not differ between the groups (ANOVA, main effect of group: $F(1)=0.03$, $p=0.86$). In those regions we did not find evidence for any reward prediction errors signals of the chosen option ($F(1)=1.10$, $p=0.31$) or any group difference (ANOVA, effect of group: $F(1)=0.25$, $p=0.62$). (B) We did however find evidence for weak RPE signals for the unchosen option (ANOVA, effect of intercept: $F(1)=4.85$, $p=0.04$). However, these did not differ between the groups (ANOVA, main effect of group: $F(1)=0.54$, $p=0.47$; group $\times$ area interaction: $F(2)=0.69$, $p=0.51$). All whole-brain results are cluster-corrected at $p<0.05$. Abbreviations: dorsal anterior cingulate (dACC), parietal (parietal – IPLC/ PFm (Mars et al., 2011)), aPFC (anterior prefrontal cortex).

Increased serotonin enhances effort prediction errors

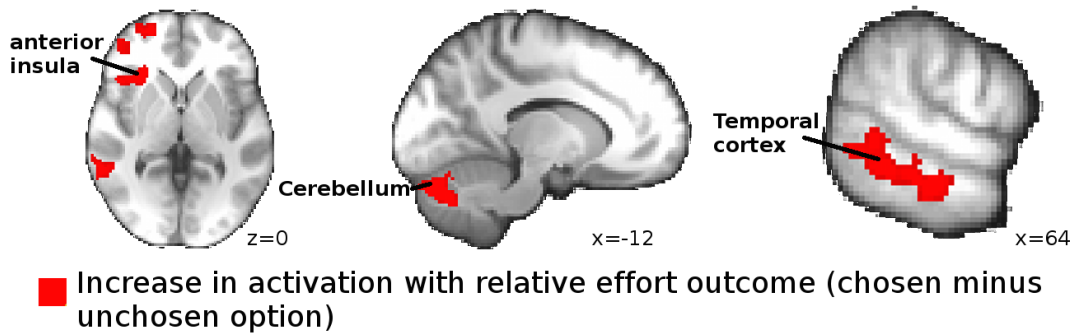
We performed similar analyses for effort prediction errors of the chosen option. In the brain regions activated by increased relative (chosen minus unchosen option) effort outcomes (Fig. 6A and 7A, Table 2C, fGLM1), we found only one of these areas to show stronger (i.e. more negative) effort prediction errors (EPEs) of the chosen option in the citalopram group (ANOVA: interaction effect group $\times$ area: $F(5,135)=2.62$, $p=0.027$; main effect of group: $F(1,27)=1.85$, $p=0.185$, fGLM2). More specifically, we found that in dACC, the citalopram group showed a stronger BOLD decrease with EPE (Fig. 6Bi, $t(27)=3.24$, $p=0.003$). No significant differences

were found in any of the other areas (Fig. 7B). Again, this was in the absence of changes to the coding of relative effort outcomes (Fig. 6Bii, fGLM2).

Together our fMRI results suggest that citalopram enhanced neural learning signals independently of changes to reward or effort sensitivity per se. This is in agreement with what would be predicted from an increase in synaptic plasticity in value-learning related brain regions by prolonged citalopram administration.



A) Additional areas for effort magnitude outcome



B) EPEs (chosen)

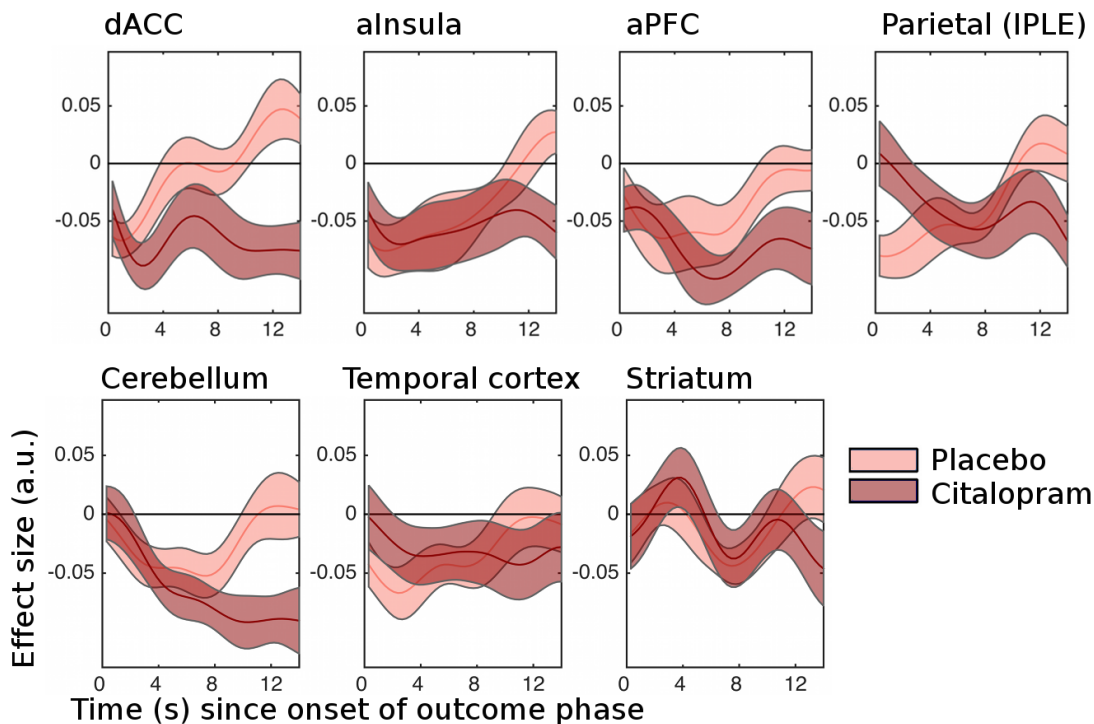


Figure 7. Effort prediction errors and outcome signals in regions that activated with relative effort magnitude outcomes. We identified areas in the outcome phase that increased in activity with the relative (chosen minus unchosen option) effort magnitude outcomes (A). We identified a total of 6 areas in this contrast (see also Supplementary Table S2C). The relative effort magnitude outcome signal did not differ between the two groups (ANOVA, main effect of group: $F(1)=0.28$, $p=0.61$). B) These areas also coded an effort prediction error (EPE) signal of the chosen option across the two groups (ANOVA, effect of intercept $F(1)=18.54$, $p<0.001$). While there was no group difference across all of these areas (ANOVA, main effect of group $F(1)=1.85$, $p=0.19$), there was a stronger EPE in some areas (ANOVA, interaction area $\times$ group $F(5)=2.62$, $p=0.03$). Follow-up t -tests revealed that citalopram enhanced EPEs selectively in dACC ($t(27)=3.2$, $p=0.003$). The ventral striatum was also tested a priori area of interest. There were again no group differences ($t(27)=-0.064$, $p=0.96$). All results in A) are cluster-corrected at $p<0.05$. Abbreviations: dorsal anterior cingulate cortex (dACC), anterior insula/ frontal operculum (aIns), anterior prefrontal cortex (aPFC), parietal cortex (Parietal, IPLE/PGp (Mars et al., 2011)), ventral striatum (striatum).

3.4 Discussion

Animal studies have highlighted links between increases in synaptic plasticity and learning with prolonged increases in serotonin levels. We investigated whether in humans prolonged administration of a selective serotonin re-uptake inhibitor (citalopram) at levels that are typically used in the clinic to treat psychological illness, which increases serotonin levels, could enhance neural and behavioural correlates of value learning. For this, participants performed a multi-dimensional learning task in which it was necessary to learn about both reward and effort. Neurally, we found that citalopram increased reward prediction errors (RPEs) in a widespread network of brain regions including ventromedial prefrontal cortex (vmPFC) and posterior cingulate cortex. At the same time, effort prediction errors (EPEs) were increased in dorsal anterior cingulate cortex (dACC). This change in learning signals occurred in the absence of changes to the overall outcome signals for reward or effort. Behaviourally, we found that citalopram made reward learning more robust or resilient to negative interference.

To understand what increased neural PEs after prolonged SSRI administration could be due to, it is important to first consider in what ways learning could be improved in our task. Firstly, in classic theories (Rescorla and Wagner, 1972), learning is about establishing expectation about *mean* magnitudes or probabilities. In this scenario, learning can be influenced by changing the speed of learning this expectation, i.e. by changing the learning rate. However, behaviorally, we found that citalopram did not affect the learning rate. Secondly, learning can also be improved by changing the *precision* of the representation of the learnt information, i.e. by reducing its noisiness. With a more precise estimate, the brain would be more capable of computing the prediction error, which would then lead to the stronger neural PE signals we found. On the neural level, a less noisy representation could be achieved by citalopram changing the properties of individual synapses or the number of synapses (Bessa et al., 2009) to allow better integration between the predictions and the outcomes (i.e. learning) or more generally to allow better

integration of neural information into a neural network itself. Again, such an improved integration does not necessarily mean faster learning, as it could equally mean using information more reliably and appropriately to inform future beliefs. Our behavioral data further support the claim that our neural signals reflect an increase in learning signals, as we found that behavioral measures of learning were improved. More specifically, learning about rewards was more robust to interference. This suggests that when the RPEs in the brain are stronger, there is a protective effect on learning in challenging situations, for example where negative interference draws resources away from the processing of key reward-related contingencies. Importantly, this result suggests that better learning does not necessarily mean faster learning, but rather being able to use the learnt information more consistently in situations in which interference may also occur.

One area that we identified as having increased RPE signals was the vmPFC. This area has repeatedly been identified as being involved with reversal learning (Boorman et al., 2009). Furthermore, it has been shown that depletion of serotonin levels in adjacent prefrontal areas impaired reversal learning in marmosets (Clarke et al., 2004; Clarke et al., 2005; Rygula et al., 2014). While our results further support the claim that serotonin affects the role of vmPFC in reversal learning, they also show that the effects of serotonin on learning were not caused by serotonin enhancing the coding of positive or negative outcomes per se. On the contrary, at a lower statistical threshold, we found that in vmPFC and striatum reward outcome signals were decreased, similar to what had been reported in other studies (Macoveanu et al., 2014; McCabe et al., 2010). Thus, increased serotonin levels can specifically enhance learning, potentially via changed synaptic plasticity. We also note that our results point to an effect of serotonin in learning that is not selective to vmPFC, as we found changes in the RPE signals in a wide range of areas. As we delivered SSRIs systemically, future studies need to establish the specific roles of the diverse regions we found to have increased RPEs. In contrast to the effects on RPEs, effort PEs were only affected in dACC. This could be the result of regional differences in the effects of

SSRIs (Baldinger et al., 2014). Alternatively, SSRI-induced changes in reward-related learning might have been more pronounced because related-related learning was more difficult in our paradigm while effort-related learning was less challenging.

Our results suggest that, at least when increased for a prolonged duration, the effect is not what would be expected if serotonin signaled only aversive events, i.e. acting as a simple opponent signal to dopamine. If increased serotonin levels had been signaling aversive information we would have expected the learning to only affect the aversive (here effort) domain. It is of note that most past studies emphasizing the relationship between serotonin, aversion, inhibition and learning (reviewed in(Faulkner and Deakin, 2014)) have looked at phasic information signaled by serotonin or short-term changes in serotonin, such as after depletion of the serotonin precursor tryptophan(Cools et al., 2008; Crockett et al., 2009; Crockett et al., 2012a; Evers et al., 2005; Geurts et al., 2013b) or a single dose of citalopram(Chamberlain et al., 2006b; Guitart-Masip et al., 2014). On the other hand, our results are in wider agreement with many recent recording studies from serotonergic neurons in the raphe nucleus, which have highlighted that even serotonergic neurons' phasic activity signals information about both aversive and *appetitive* stimuli on a short timescale (Cohen et al., 2015; Miyazaki et al., 2011). Recording studies have also suggested that some of the functions of serotonin may be mediated via other neuromodulators such as dopamine (Dremencov et al., 2009; Liu et al., 2014).

Serotonin is also known to lead to plasticity changes in the brain when its tonic levels are changed. Studies have found that serotonin levels naturally increase in situations where new learning occurs behaviorally (Jitsuki et al., 2011). Furthermore (Jitsuki et al., 2011; Karpova et al., 2011; Vetencourt et al., 2008), if serotonin levels are artificially increased by administering SSRIs for a prolonged time, similar changes occur to learning e.g. in form of visual adaptation or fear extinction. Molecularly, this has been linked to changes in synaptic plasticity (Jitsuki et al.,

2011), synapse remodeling (Bessa et al., 2009), neurogenesis in the hippocampus (Malberg et al., 2000) in animals. While both projections of serotonergic neurons and affected brain regions are very diverse, the type of neurons changed can be very selective (Jitsuki et al., 2011). Increased neural reward and effort learning signals are also in agreement with Palminteri et al. (2012) who found behaviorally enhancement of both reward and punishment learning after prolonged SSRI treatment of patients with obsessive compulsive disorder. Our data suggests a neural underpinning for this behavioral effect as well as extending the demonstration of behavioral effects of SSRIs to multi-dimensional learning in the context of interference.

The results of our study might shed light onto the mechanism of SSRIs as treatments for mental health disorders, such as depression. The earliest clinical effects of SSRI have been reported with the dosage and duration at which we administered citalopram to our participants (Castren, 2013). In fact, our results might suggest that SSRIs exert part of their clinical effect by enhancing how well patients can learn about positive relationships in complex environments even in the presence of negative interfering information which otherwise could prevent such learning. This effect may occur in addition to, or even underlie, previous reports that SSRIs shift the processing of social cues so that they are perceived as more positive (Harmer et al., 2009) In fact, these two effects, relating to learning and attentional biases, might interact with each other, resulting in patients perceiving the world as more positive and learning more reliably about this new more positive worldview. In this context, it is noteworthy that our behavioral results are in agreement with other studies that found that changes in serotonin levels affect how much negative stimuli can bias behavior (Evers et al., 2006a; Walker et al., 2009).

In conclusion, we found that prolonged increase in serotonin levels by an SSRI increased neural prediction errors signals during reward and effort learning. Concomitantly, reward learning in the face of negative interference was improved. Thus, prolonged increase in serotonin can

strengthen both appetitive and aversive neural learning signals, in a manner that is consistent with previous demonstrations of the impact of serotonin manipulation on neural plasticity. These results are also of clinical relevance, supporting theories that SSRIs' treatment effects on, for example, depression may be related to increases in neural plasticity and learning.

4. A role beyond learning for NMDA receptors in reward-based decision-making – a pharmacological study using d-cycloserine

Summary

In the last chapter we saw that serotonin can indeed modulate plasticity related learning in humans. If it is these plasticity effects that are responsible for improvements in mood, then other medications that modulate plasticity without having other effects on social or emotional processing should also modulate mood. One candidate mechanism that could be targeted is the N-methyl-D-aspartate (NMDA) receptor. These receptors have been shown to fulfil crucial functions in many forms of learning and plasticity. Animal studies have revealed that when their probability of activating is increased by administration of the partial NMDA agonist d-cycloserine, different forms of learning are enhanced. Whether this is also the case in humans has been less clear with experiments commonly failing to find improvements in learning such as motor learning or other non-emotional learning. However, what has not been examined yet in humans is whether d-cycloserine can enhance plasticity and learning in domains more related to depression such as learning about rewards and punishments. In this chapter therefore, we examined whether a partial NMDA agonist could modulate these forms of learning. Human participants were assigned to receive either 250mg of the partial NMDA agonist d-cycloserine (n=20) or matching placebo capsules (n=27). Reward-guided learning and decision-making were assessed using a task in which participants had to integrate learnt and explicitly shown value information to maximize their monetary wins and minimize their losses. We found no evidence for any changes in learning behaviour. This suggests that d-cycloserine, at least at the dose used in this study, is not well suited for studying how neural plasticity and learning are related to mood. However, our results also showed that while d-cycloserine had no effect on learning, it did improve decision-making. Decision-making was shifted towards a more optimal integration of the learnt and the explicitly shown information, in the absence of a direct learning effect. In conclusion, our results reveal a distinct role for NMDA receptors reward-guided decision-making. We discuss these findings in the context of NMDA's roles in neuronal super-additivity and as crucial for evidence integration for decisions

4.1 Introduction

Animals and humans live in ever changing and complex environments. They need to continuously track and learn about the changing properties of their environment and use them to behave adaptively. The neural mechanisms engaged in learning about the environment depend on the type of information being tracked. Nevertheless, the plasticity underpinning several of these various forms of learning has been argued to depend on the same molecular mechanism, i.e. N-Methyl-D-aspartate (NMDA) subtypes of glutamate receptors. Accordingly a number of studies have investigated whether NMDA receptor manipulations affect learning and memory (Bannerman et al., 2012; Bohn et al., 2003; Kuriyama et al., 2011a). Most of the studies looking at the role of NMDA receptors in reward learning have focused on simple tasks, such as a single association between a stimulus and an outcome or its reversal. However, in more ecological, and thus complex, scenarios, learning and decision-making often involve consideration of positive and negative aspects of potential outcomes as well as the integration of learnt information with information that is explicitly cued. Importantly, it has recently been shown that integration of information conveyed by different dimensions during decision making is an active process recruiting particular neural mechanisms (Burke et al., 2013; Stein and Stanford, 2008). Thus, to understand the full role of NMDA receptors in reward learning and decision-making, it might not be sufficient to only study a single simple component in isolation.

Interestingly, there is some, albeit indirect, evidence that NMDA receptors are not only important for learning but also for integration of information: NMDA receptor blockade has been reported to affect the integration of multisensory information in cat superior colliculus (Binns and Salt, 1996) or the integration of reward and delay in rats (Floresco et al., 2008). However, to our knowledge, it has not been investigated whether NMDA receptors play a role in human value-based decision-making and learning.

To examine the influence of changes in NMDA receptor activity on complex learning and decision-making behaviours in humans, we used the partial NMDA agonist d-cycloserine. D-cycloserine binds to the glycine site of the NMDA receptor. Glycine is a co-agonist of the NMDA receptor, meaning that NMDA receptors only open when both glycine and glutamate bind. D-cycloserine can thus increase the probability of glutamate release opening NMDA receptors, which in turn enhances NMDA receptor mediated activation. We designed a multi-attribute decision-making task in which participants learnt changing probabilities of gains and losses of two options. They made choices between options by integrating those learnt probability with explicitly cued information about gain and loss magnitudes. To assess potential effects of d-cycloserine on learning and decision-making, we used reinforcement learning models. Surprisingly, we did not find any evidence for a change in the rate at which participants learned about reward or punishment outcomes. However, we found that d-cycloserine improved decision-making. D-cycloserine led to a more optimal integration of the learnt probability information with the explicitly cued magnitude information.

4.2 Methods

Participants

The study was approved by the local ethics committee. 52 healthy volunteers (age 18 – 30) took part in the study. Participants were included in the study after undergoing a psychiatric and medical screening. Participants were included if they had a body mass index between 18-30 kg/m², passed basic medical screening for current or past physical illness relevant to the study, were non- or light-smokers (<5 cigarettes a day), had taken no CNS-active medication during the last 6 weeks, were not currently taking d-cycloserine, ethionamide or isoniazid, were screened to be free of a past or present Axis 1 disorder on the Structured Clinical Interview for DSM-IV (First et al., 2002), did not have a first-degree family member with a psychiatric history, and had fluent English skills. Female participants were neither pregnant nor breast-feeding.

Five participants (all in the placebo group) were excluded from the study at the point of data analysis because, when making decisions, they did not use the learnt probability information at all, but only the explicitly cued magnitude information to make their decisions, making it impossible to estimate their learning rate (corroborated by the logistic regression of past outcomes, see ‘Logistic Regression Analysis’ below). The exclusion assessment was made independently by two researchers (JS and NK) who were blind to their group assignments. There remained 20 participants in the d-cycloserine and 27 participants in the placebo group.

Performing a power calculation here is difficult as no study has looked at the effects of a similar dose of d-cycloserine on value learning in humans. A study on human motor learning (Kuriyama et al., 2011a) reported a doubling in learning speed (Cohen’s $d = 1.4$), based on this a sample size of 10 participants per group would be sufficient to detect an effect (at a significance level of 0.05, two-tailed) with a power of 0.8. However, we also note that many previous studies have failed to detect learning effects (see General Introduction 1.4). Therefore the actual effect size might be smaller than reported by Kuriyama et al. (2011a), which is why we included more participants than the minimum of 10 suggested by the power analysis. The groups were well-

matched on socio-demographic and personality parameters (Table 1). This experiment was collected as part of a larger study in which a third drug (hydrocortisone) was used (data not reported).

Demographics and questionnaire measurements							
	Pla(n=27)		DCS(n=20)		p		
Age	22.2±0.6		22.3±0.7		0.93		
Gender F:M	15:12		11:9		0.97		
BDI	2.0±0.5		0.9±0.4		0.14		
Education Years	16.9±0.3		15.9±0.5		0.10		
Trait anxiety	30.9±1.6		30.5±1.3		0.86		
BMI	22.0±0.4		22.0±0.5		0.99		
Neuroticism	5.2±0.9		5.6±0.9		0.33		
ACS Focusing	26.7±0.6		26.2±0.9		0.64		
ACS Shifting	34.7±0.8		34.5±1.0		0.31		
BIS	16.0±0.8		16.7±0.8		0.59		
BAS	24.4±5.8		23.6±4.8		0.59		
VAS items	Pla,before	Pla,after	DCS,before	DCS,after	p,before	p,after	p,diff score
Anxious	7.6±1.6	3.6±0.6	6.7±1.5	5.7±1.8	0.70	0.25	0.24
Sleepy	28.4±4.0	21.3±4.3	24.1±3.3	13.7±2.5	0.43	0.13	0.54
Flushed	9.1±2.0	3.2±0.8	7.1±1.9	3.7±0.7	0.50	0.62	0.31
Tearful	3.3±0.9	2.9±0.8	3.1±0.7	2.8±0.5	0.84	0.87	0.95
Nauseous	3.3±0.9	3.2±0.9	2.9±0.7	3.5±0.6	0.72	0.83	0.41
Hopeless	3.3±0.7	2.5±0.5	4.1±1.5	2.7±.5	0.61	0.85	0.65
Tremor	3.3±0.9	3.2±1.0	3.7±1.0	3.1±0.6	0.77	0.91	0.58
Sad	4.9±0.9	2.8±0.5	4.5±1.3	2.8±0.5	0.83	0.98	0.80
Dizzy	2.7±0.7	3.5±1.2	3.0±0.7	5.4±2.0	0.78	0.41	0.38
Depressed	2.9±0.5	2.6±0.5	3.1±0.8	2.5±0.5	0.85	0.90	0.65
Tachycardia	4.5±1.4	3.5±1.2	5.7±1.5	3.5±0.7	0.57	0.98	0.37
Alert	51.3±5.0	48.4±5.1	57.8±5.0	52.3±5.4	0.37	0.61	0.51

Table 1. Sociodemographic and questionnaire measurements. The following measurements were obtained before drug administration: Age, gender, Beck's Depression Inventory (BDI, Beck et al. (1996)), years of education at time of study, trait anxiety (Spielberger and Gorsuch, 1983), body mass index (BMI, weight(kg)/height(cm)²), neuroticism (Eysenck and Eysenck, 1975), attention control scale (ACS, Derryberry and Reed (2002)), behavioural inhibition, behavioural activation (BIS/BAS Scale, Carver and White (1994)). The values reported are mean values with standard errors and the p-scores from between subject t-tests (apart from the value for gender ratio, where a chi-squared test was used). Visual analogue scales were given to participants before and after drug administration. For each of the listed items, they were asked to indicate how they were feeling by placing a tick mark on a 100mm line, which was labelled 'not at all' on the left-hand side and 'extremely' on the right hand side. The values reported are the mean and standard error for the tick mark positions (in mm). P-values were calculated for the group differences in the first (at baseline, 'before') and second measurement (after drug administration, 'after') and on the difference scores (baseline – second measurement, 'diff score').

Procedure

In a double-blind, placebo-controlled design, participants were randomly allocated to a single dose of d-cycloserine (250mg) or matching placebo capsule. They fasted two hours before the testing visit. 250mg was chosen in agreement with recent studies (Klumpers et al., 2012; Onur et al., 2010). Participants were tested three hours after drug administration. According to product information (King's Pharmaceutical Ltd.), plasma peak levels are reached within 3-4h; other studies (Patel et al., 2011; van Berckel et al., 1997; Van Berckel et al., 1998) found that peak levels are reached within ~1h. However, given d-cycloserine's half-life of 8-12h (product information) or 15h (Patel et al., 2011), plasma levels would have been close to peak levels during testing, given either time-to-peak information. To assess potential subjective changes following d-cycloserine, participants completed questionnaires (Table 1) before capsule intake and before testing.

Probabilistic Instrumental Learning Task

Participants performed a probabilistic learning task with monetary wins and losses (Figure 1a). Participants made repeated choices between two options with the aim to maximize their monetary pay-off.

On each trial, participants had a choice between two options. Each option had four independent attributes: a reward and a loss magnitude, a reward and a loss probability. The magnitude determined how many points could be won (and lost) on this trial, while the probabilities determined how probable winning and losing was respectively. After participants selected one of the options, they were shown the outcomes for both options. However, only the option they had chosen contributed to the participants' earnings. In trials where the chosen option incurred both, wins and losses, the participants' earnings in that trial were the sum of both. Therefore, to maximize the overall gains, participants should be choosing the option on each trial with the

highest reward utility (reward magnitude x reward probability) and with the lowest loss utility (loss magnitude x loss probability).

Reward and loss magnitudes were explicitly cued at the time of choice and were randomly drawn from a flat distribution between 1 and 100 points. In contrast, the probabilities were not explicitly shown and had to be learnt across trials by observing the outcomes. The outcomes for one option could either be a win and a loss, only a win, only a loss, or neither win nor loss. The independent reward and loss probabilities determined the probability of these outcomes. The probabilities varied over the course of the experiment between 20% and 80%, with only one of the four probabilities varying at any given time (Figure 1c).

Before the experiment, participants were instructed about the task. They were informed that the magnitudes were drawn randomly and independently on each trial. They were also informed that the four probabilities were independent from each other. However, they were only told that the probabilities would change over the course of the experiment at varying speeds, not that it was always only one probability that was changing at any given time or that the minimum/ maximum probability values were 20% and 80%.

Participants first performed 30 training trials, followed by 381 test trials, which were included in the analysis. Each participant was tested using the same task schedule to allow for better group comparisons. At the start of the task participants were given £4 (400 points) to ensure that they had sufficient funds to sustain losses incurred even at the beginning of the experiment.

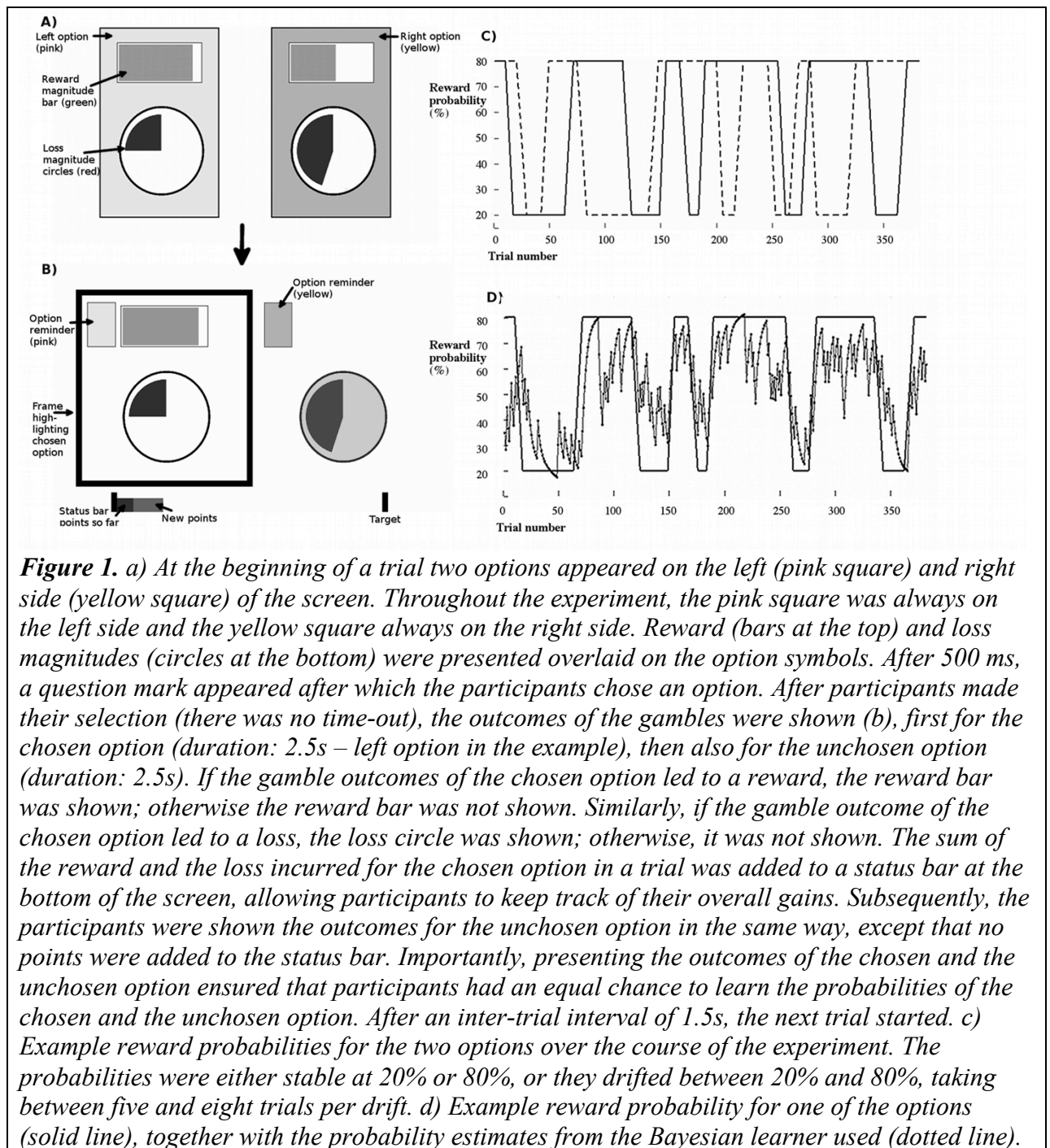


Figure 1. a) At the beginning of a trial two options appeared on the left (pink square) and right side (yellow square) of the screen. Throughout the experiment, the pink square was always on the left side and the yellow square always on the right side. Reward (bars at the top) and loss magnitudes (circles at the bottom) were presented overlaid on the option symbols. After 500 ms, a question mark appeared after which the participants chose an option. After participants made their selection (there was no time-out), the outcomes of the gambles were shown (b), first for the chosen option (duration: 2.5s – left option in the example), then also for the unchosen option (duration: 2.5s). If the gamble outcomes of the chosen option led to a reward, the reward bar was shown; otherwise the reward bar was not shown. Similarly, if the gamble outcome of the chosen option led to a loss, the loss circle was shown; otherwise, it was not shown. The sum of the reward and the loss incurred for the chosen option in a trial was added to a status bar at the bottom of the screen, allowing participants to keep track of their overall gains. Subsequently, the participants were shown the outcomes for the unchosen option in the same way, except that no points were added to the status bar. Importantly, presenting the outcomes of the chosen and the unchosen option ensured that participants had an equal chance to learn the probabilities of the chosen and the unchosen option. After an inter-trial interval of 1.5s, the next trial started. c) Example reward probabilities for the two options over the course of the experiment. The probabilities were either stable at 20% or 80%, or they drifted between 20% and 80%, taking between five and eight trials per drift. d) Example reward probability for one of the options (solid line), together with the probability estimates from the Bayesian learner used (dotted line).

Analysis

The behavioural analysis compared the effects of d-cycloserine versus placebo on a) learning of reward and loss probabilities, and b) integration of learnt (probabilities) and explicitly cued (magnitudes) information for guiding complex decisions. All analyses were performed in Matlab and SPSS.

Logistic regression analysis

To ensure that participants learnt the probabilities, we first assessed the impact of past outcomes (reward and loss) and the explicitly cued magnitudes (reward and loss) on choice, using a logistic regression analysis and normalised regressor estimates. We included regressors for the last five trial differences in reward and loss outcomes between the two options, as well as the differences in the explicitly cued magnitudes.

To investigate whether the groups differed in their learning speeds, the resulting regression weights for the past outcomes were entered into an ANOVA with group (d-cycloserine versus placebo) as a between subject factor and time (1,2,3,4 or 5 trials in the past) and valence (reward or loss) as within subject factors.

Modelling

To look at the learning effects more specifically and assess participants' strategies for the integration of information to make decisions, we used reinforcement-learning models to fit each participant's trial-by-trial behaviour.

Each model consisted of three main components. First, the model had estimates about the probabilities underlying the outcomes of both options. These were updated on every trial using a reinforcement-learning algorithm. Second, the probability estimates were integrated with the explicitly cued magnitudes to calculate how valuable each of the two options was (i.e. their utility). Third, these two utilities were compared to predict participants' choices. To determine the best parameter estimates for every participant, we used a standard log-likelihood maximization procedure.

When calculating how valuable each option is, participants might use different decision strategies for integrating learnt probabilities with explicit magnitudes. They could use a mathematically optimal strategy (utility as probability $\times$ magnitude). However, as this optimal strategy may be quite cognitively taxing, they could resort to a heuristic strategy (utility as a

weighted sum of probability and magnitude). To test which decision strategy participants used, we fitted different models to the data. To test for differences in strategy, we then compared how well each of these models explained the groups' behaviour. In addition, we fitted a third model, which directly estimated to what degree they used either decision strategy.

1. Optimal Model. This model assumed that participants integrated the learnt probabilities *optimally* with the explicit magnitudes (magnitude x probability). The learning of probabilities was modelled using a standard reinforcement-learning rule. On each trial, the estimated probability of an attribute was updated based on the trial's outcome, as a function of the prediction error:

$$Prediction_t = Prediction_{t-1} + \alpha * PE_{t-1}$$

with

$$PE_{t-1} = Outcome_{t-1} - Prediction_{t-1}$$

where α is the learning rate. Thus, the learning rate is a measure of how much participants updated their probability estimate when the outcome associated with an attribute differed from their expectation (e.g. the probability that the left option yields a reward). Separate learning rates were used for learning about wins and losses.

These predictions were combined optimally with the shown magnitudes:

$$Utility_{Reward} = Probability_{Reward} * Magnitude_{Reward}$$

The loss utility was computed in the same way and combined with the reward utility:

$$Utility_{total} = Utility_{Reward} - \lambda * Utility_{Loss}$$

where λ determines how much participants weighted the prospect of rewards versus losses.

A standard soft-max decision rule was used to predict the probability of choosing an option.

$$P(Option_A) = \frac{e^{Utility_A * \beta}}{e^{Utility_A * \beta} + e^{Utility_B * \beta}}$$

where β reflects a participant's ability to pick the option with higher utility.

To assess the effect of d-cycloserine on learning about wins and losses, we compared their respective learning rates between groups.

2. Heuristic Model. This model differed from the Optimal Model only in the decision strategy for integrating learnt probabilities with explicit magnitudes. Instead of being a product of probability and magnitude, utility was computed as a weighted sum.

$$Utility_{Reward} = \mu * Probability_{Reward} + Magnitude_{Reward}$$

where μ is the probability weighting factor, describing the relative importance of the learnt probability compared to the explicit. The loss utility was computed in the same way, sharing the same μ .

Again, we compared the groups' learning rates. Additionally, a change in learning could also manifest as a changed reliance on learnt compared to explicit information, we therefore also compared the probability weighting factors between the groups.

3. Hybrid Model. We used the Hybrid Model to examine whether d-cycloserine affected how participants integrated information, shifting them towards a more optimal, less heuristic decision strategy. The Hybrid Model computed utility as a weighted sum of the utilities from the Optimal and the Heuristic Model:

$$Utility = \omega * Utility_{Heuristic} + (1 - \omega) * Utility_{Optimal}$$

where ω is the heuristic weight factor, determining how much the overall utility is like the heuristic or the optimal utility. The higher ω (between 0 and 1), the more a participant relied on a heuristic decision rule.

Model comparison

If the groups differ in their decision strategy, this should also be reflected in how well the models incorporating the different strategies can explain behaviour. To assess this, we compared the

model fits using the Akaike information criterion (AIC). We calculated for each participant the AIC differences between the Optimal and the Heuristic model and also between the Heuristic and the Hybrid model. This allowed us to compare whether the d-cycloserine differed from the placebo group in how well their behaviour was explained by the Heuristic relative to the Optimal and by the Heuristic relative to the Hybrid model.

The data of each person was fit individually, using matlab's `fminsearch`. This was in contrast to analyses in chapter 2 which used a hierarchical fitting approach. The reason was that in the present study many more behavioural trials were available for each participant. Therefore, it was possible to obtain sensible fits for each model on the individual-subject level. When there is enough data available, this strategy has the advantage that it does not assume that the parameter values (e.g. learning rate) across participants are distributed normally. This is particularly relevant in a drug manipulation where participants might respond differently to the drug based on how well their bodies absorb the drug and their body mass etc.

To confirm the results from the modelling analysis using a different method, we performed an additional regression analysis with regressors analogous to the components of the Hybrid Model. We included regressors for the explicit magnitude differences, for the learnt probability estimate differences, as well as for the difference in optimal utilities (magnitude x probability). The probability estimates were obtained using a Bayesian Learner, like the one described in (Behrens et al., 2007), see below. These Bayesian probability estimates are the most accurate estimates a participant could have given the past outcomes (Figure 1d). If participants' decision strategy is more heuristic, the main effects should have a larger impact on behaviour. Conversely, if the decision strategy is more optimal, the interaction term should have a higher impact.

Bayesian Learner

In short, the model estimates the current reward/loss probabilities of an attribute (e.g. reward left option) based on the previous trials for each attribute separately. It does this by taking into account the following properties of the experimental task: 1) The reward/loss outcomes (either 1 or 0) are determined by underlying reward/ loss probabilities. 2) The probability of an attribute can sometimes change. 3) How quickly each probability changes can vary over the course of the experiment. Sometimes, the probability changes more quickly (the attribute's volatility is said to be high) whereas at other times, it changes more slowly (the volatility is said to be low). 4) The volatility of an attribute is not static but can also change over the course of the experiment.

In other words, the parameters that the model estimates on every trial for each attribute are the reward/loss probability, the volatility (i.e. the change in probability) and the volatility change. The model estimates these parameters for the current trial based on last trial's parameter likelihoods and the current attribute outcomes. It does this using Bayes' rule, which is the most efficient way for updating beliefs given new evidence. For details on the specific mathematical implementation of the Bayesian learner, please refer to Behrens (2007).

Learning about the unchosen option – control analysis

There is some evidence (Boorman et al., 2011) that different brain areas are used to learn about the chosen and the unchosen option, thus there is a possibility they could be affected differently by d-cycloserine.

To examine the effects of the chosen and alternative (unchosen) option we performed three regression analyses. In the first two analyses, we looked at the influence of the past chosen or unchosen outcomes on upcoming choices. In the third analysis, we looked at learning differences for the chosen and the unchosen option separately.

First, we ran a regression analyses, predicting switching or staying based on the probabilistic outcomes (1/0) for the chosen reward, the unchosen reward, the chosen loss and the unchosen loss. We also included regressors for the current reward and loss magnitudes. Second, we ran a regression analysis, again predicting switching or staying, based on the Bayesian probability estimates. The regressors were: Bayesian reward and loss probabilities for the past trial chosen and unchosen options and explicitly shown reward and loss magnitudes. Third, we also looked at whether learning from the chosen or the unchosen option differed. For this, we fitted a learning model. In the learning model, we fitted two separate learning rates for learning from the chosen and the unchosen option. To avoid over-parameterization, we did not split the learning rates further by learning rates for reward or loss. Apart from this, the model had the same parameters as the Heuristic Model described above.

4.3 Results

General performance

In the task, participants had to constantly track the independent win and loss probabilities of two options and integrate these with explicitly cued reward and loss magnitudes. The groups did not differ in the overall earnings or in the mean points won/lost (Table 2a).

a) Overall behaviour	Pla	DCS	p
Total money won	12.9±1.1	13.0±0.9	0.66
Mean points won	35.4±1.1	35.4±1.1	0.89
Mean points lost	21.8±1.4	21.6±0.3	0.67
Mean reward magnitude chosen	60.7±0.4	60.7±0.6	0.92
Mean loss magnitude chosen	48.0±0.5	48.0±0.5	0.96
Mean reward gamble outcome	0.6±0.0	0.6±0.0	0.9
Mean loss gamble outcome	0.4±0.0	0.4±0.0	0.86
b) Optimal decision model	Pla	DCS	p
LR reward	0.16±0.03	0.17±0.04	0.83
LR loss	0.27±0.05	0.22±0.04	0.52
Beta	0.06±0.006	0.07±0.007	0.40
Loss weight (λ)	0.90±0.07	0.90±0.07	0.94
fval	165±5	154±6	0.15
c) Heuristic decision model	Pla	DCS	p
LR reward	0.44±0.05	0.35±0.04	0.20
LR loss	0.43±0.05	0.38±0.06	0.52
Beta	0.04±0.00	0.04±0.00	0.81
Loss weight (λ)	0.88±0.05	0.87±0.05	0.86
Prob weight (μ)	71.7±9.6	77.8±15.1	0.72
fval	146±5	142±6	0.62
d) Hybrid decision model	Pla	DCS	p
LR reward	0.46±0.05	0.32±0.04	0.07
LR loss	0.42±0.05	0.36±0.05	0.47
Beta	0.05±0.00	0.06±0.00	0.08
Loss weight (λ)	0.89±0.05	0.90±0.04	0.93
Prob weight (μ)	69.70±10.8	119.8±63.8	0.38
Heuristic weight (ω)	0.84±0.03	0.59±0.23	4*10 ⁻⁵
fval	146±5	139±6	0.41

Table 2. a) Results of a general behavioural analysis. Parameter estimates (mean and standard error) for both groups and the results of significance tests for the Optimal decision model (b), the Heuristic decision model (c) and the Hybrid model (d).

Logistic Regression.

As a measure of learning, we assessed the impact of past reward/loss outcomes on choice, using a logistic regression analysis. We also included the currently displayed reward/loss magnitudes.

Participants were more likely to pick options with higher reward ($t(46)=18.6$, $p=5*10^{-23}$) and lower loss magnitudes ($t(46)=-14.9$, $p=3*10^{-19}$) (Figure 1a). They also chose options more frequently when they were associated with more past wins and fewer losses (Figure 2a), thus suggesting that they were able to learn from past outcomes.

To test whether learning differed between the groups, and maybe in dependence on reward and loss valence, we ran a 2 (group) x 5 (time point) x 2(valence) ANOVA on the regression weights of the past outcomes (Figure 2a). That participants learnt the reward/loss probabilities over time was evidenced by the fact that recent reward/loss outcomes influenced choices more than longer ago ones (main effect of time, $F(4,180) = 80.6$, $p<10^{-6}$). This effect of time was stronger for losses than rewards (interaction effect, time x valence, $F(4,180)=6.2$, $p=3.8*10^{-4}$), suggesting that loss probabilities were learnt more quickly.

Importantly, the groups neither differed in overall learning speed (time x group: $F(4,180)=0.9$, $p=0.42$), nor in their relative learning speeds for wins and losses (time x valence x group, $F(4,180)=0.3$, $p=0.89$).

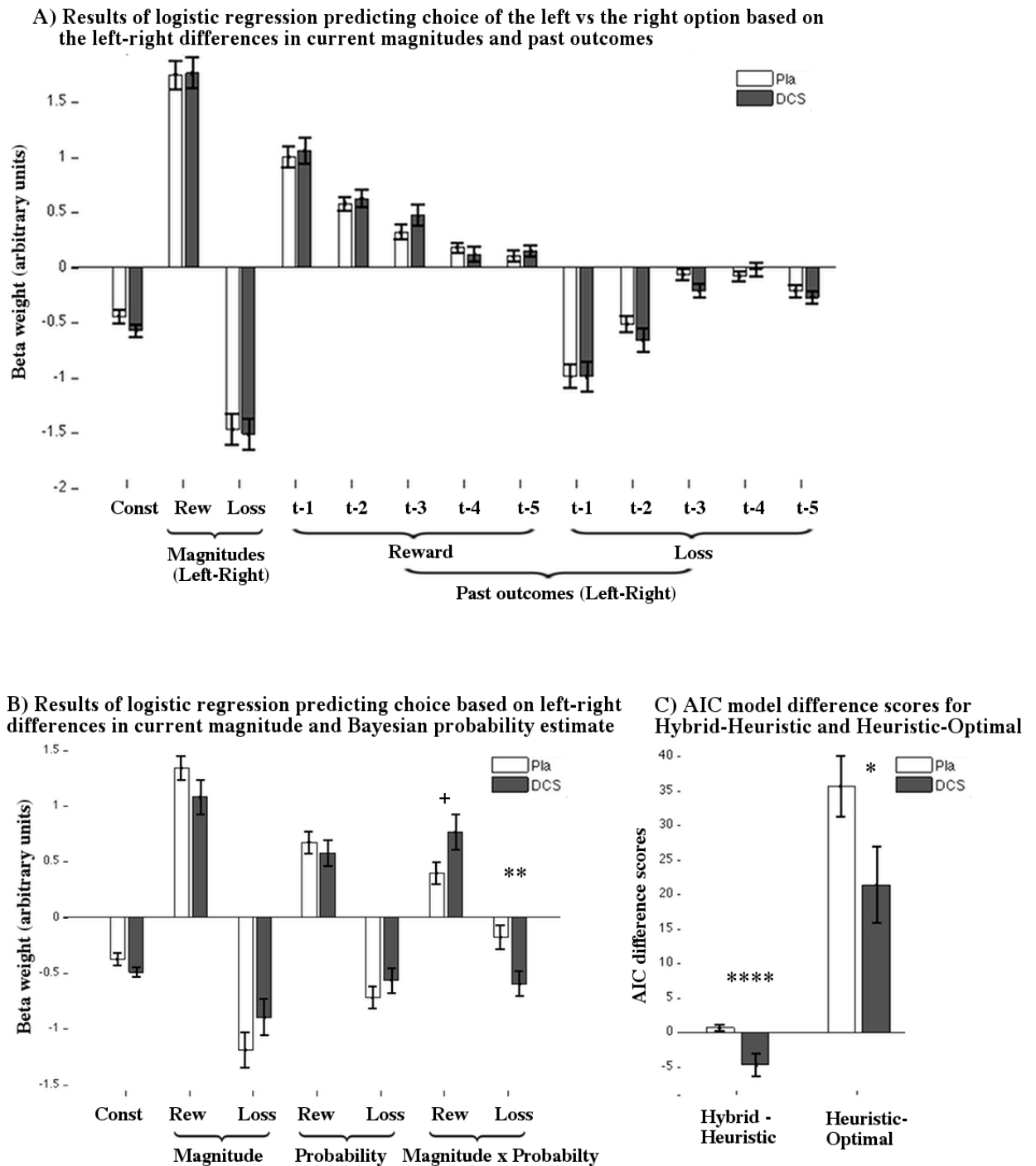


Figure 2. a) Decision weights (beta) for placebo (white) and d-cycloserine (grey), showing the decision impact of current magnitude differences (left-right) and past gamble outcome differences (left-right), for one to five trials in the past, on choice. b) Decision weights (beta) based on a regression using shown magnitude differences, probability prediction differences (estimated using a Bayesian model) and their interaction for both groups. c) AIC difference scores comparing the relative fit of the Hybrid to the Heuristic and of the Heuristic to the Optimal model in both groups. Error bars indicate SE. ⁺ $p=0.056$, * $p\leq 0.05$, ** $p<0.001$, **** $p<10^{-4}$

Modelling

The aims of the modelling analyses were to assess whether the groups differed in their learning rate (all three models), as well as whether they differed in their decision strategies (model three). Each model used a standard prediction-error learning rule for learning about reward/loss probabilities. The models only differed in how learnt probabilities were integrated with explicit reward/loss magnitudes.

1. Optimal Model. This model integrated magnitudes and probabilities optimally using a multiplicative rule. For assessing differences in learning, we compared the learning rate parameters between the groups. Surprisingly, the groups did not differ in their learning rates (reward: $t(45) = -0.2$, $p = 0.8$, loss: $t(45) = 0.7$, $p = 0.5$), nor did they differ in any other model parameter (Table 2b).

2. Heuristic Model. Participants might use a more heuristic and less optimal decision strategy to reduce task complexity. This was accounted for in the Heuristic model, in which the decision utility was calculated as a weighted sum of probability and magnitude. Here, we could detect a change in learning either as a changed learning rate or as a change in how much impact the learnt probabilities have on decisions relative to the explicit magnitudes. Again, we did not find any group differences for the two learning rates (reward: $t(45) = 1.3$, $p = 0.20$, loss: $t(45) = 0.7$, $p = 0.5$) or for the probability weighting factor ($t(45) = -0.36$, $p = 0.72$). The groups did not differ either in any other model parameter (Table 2c).

3. Hybrid Model. This model combined the Optimal and Heuristic Model's decision strategy using an additional weighting factor (ω , heuristic weight). ω indexed the degree to which participants used the heuristic relative to the optimal strategy. Again, the two groups did not differ in their learning rates (Table 2d). However, there was a strong difference in the heuristic

weight between the groups, with the d-cycloserine group relying more on an optimal rather than heuristic decision strategy compared to the placebo group ($t(45) = 4.6$, $p=3.6 \times 10^{-5}$).

Model Comparison. To investigate this shift in decision strategy further we compared how well each model explained the behaviour of each group, using the AIC (Figure 2c). If participants in the d-cycloserine group used a more optimal decision strategy, models incorporating the optimal decision rule should better explain their behaviour compared to the placebo group.

First, we compared the Optimal Model relative to the Heuristic model. In both groups, the Heuristic Model provided a better fit to the data than the Optimal Model. However, this effect was weaker in the d-cycloserine group than in the placebo group, as shown by a significant group difference ($t(45)=2.1$, $p=0.046$). This suggests that relative to the placebo, the d-cycloserine group used a more optimal strategy. We also compared the fits of the Heuristic to the Hybrid Model. Similarly, we found that the difference scores varied significantly between the groups ($t(45)=3.6$, $p=8.0 \times 10^{-4}$). For the d-cycloserine group, the Hybrid Model provided the best fit ($t(19)=3.9$, $p=0.01$; sum of AIC scores for difference between the Hybrid and the Heuristic Models: -94, with values <2 suggesting strong evidence in favour of the Hybrid Model). But for the placebo group, both models explained their behaviour equally well ($t(26)=1.6$, $p=0.11$; sum of AIC scores for the Hybrid minus the Heuristic Model: 18, with values >10 suggesting strong evidence in favour of the Heuristic model). These findings are in-line with those from the modelling analysis above and suggest that d-cycloserine shifted participants' behaviour to rely more on an optimal decision strategy.

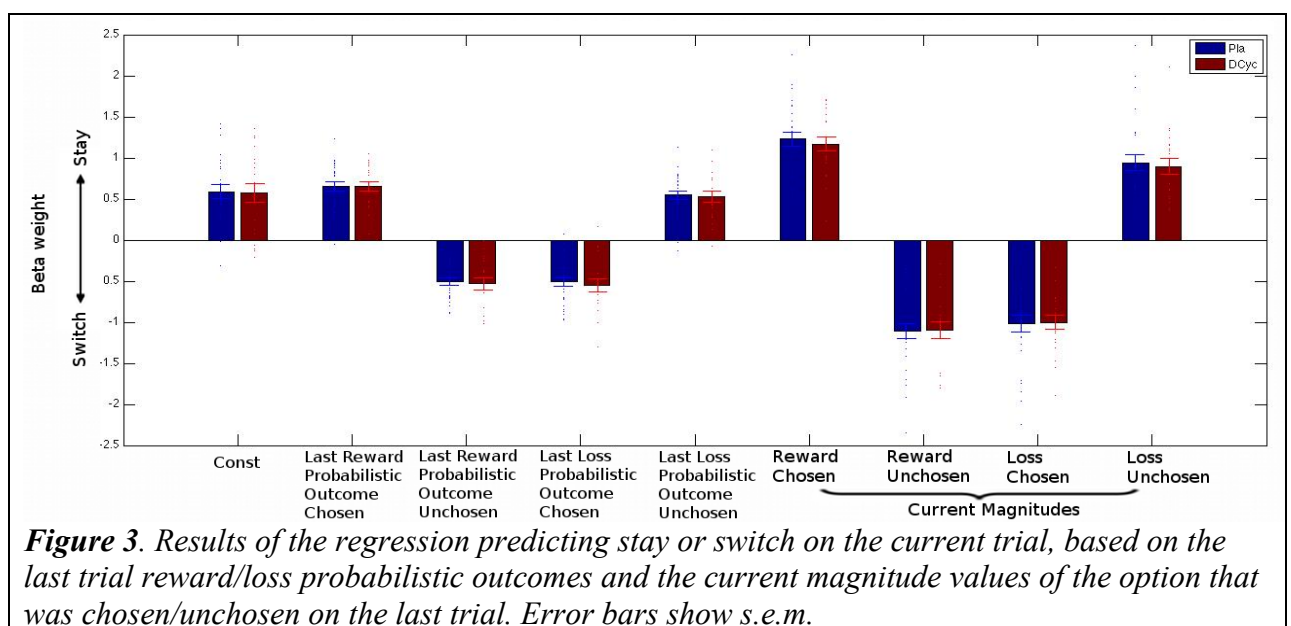
To investigate this further, we ran a logistic regression analysis, analogous to the Hybrid Model (Figure 2b). It included an interaction term for magnitude and probability (analogous to the Optimal Model), as well as each factor separately (analogous to the Heuristic Model). Similar to our Hybrid Model findings, the groups did not differ in the heuristic aspect: there was no difference between the main effects for magnitude and probability (magnitude: $F(1,45)=2.2$, $p=0.14$, probability: $F(1,45)=0.98$, $p=0.33$). In contrast, the groups differed on the optimal aspect

of the model: a 2-by-2 ANOVA analysis on the interaction regressors of magnitude and probability for reward and loss revealed that those regressors had a stronger impact on the decisions in the d-cycloserine group ($F_{(1,45)}=10.8$, $p=0.002$). Post-hoc tests revealed that this effect was significant for loss ($t_{(45)}=-2.7$, $p=0.01$) and was marginally significant for reward ($t_{(32.4)}=-2.0$, $p=0.056$).

Learning about the unchosen option – control analyses

To examine the effects of the chosen and alternative (unchosen) option we performed three analyses. In short, we found that d-cycloserine did not affect the usage of the unchosen option's outcomes for decision-making or learning.

In the first analysis, we predicted stay/switch behaviour based on the past (last trial) outcomes of the chosen and the unchosen option (figure 3, table 3A). In both groups, participants were influenced in their future choices by the past chosen and unchosen probabilistic outcomes. However, there was no difference between the groups in the impact of past chosen or unchosen reward or loss outcomes (reward chosen, $p=0.94$; reward unchosen, $p=0.71$; loss chosen, $p=0.67$, loss unchosen $p=0.78$).



Second, we ran a regression analysis, again predicting switching or staying, but this time based on the Bayesian probability estimates (figure 4, table 3B). Again, we found no group differences in the impact of the past chosen or unchosen options.

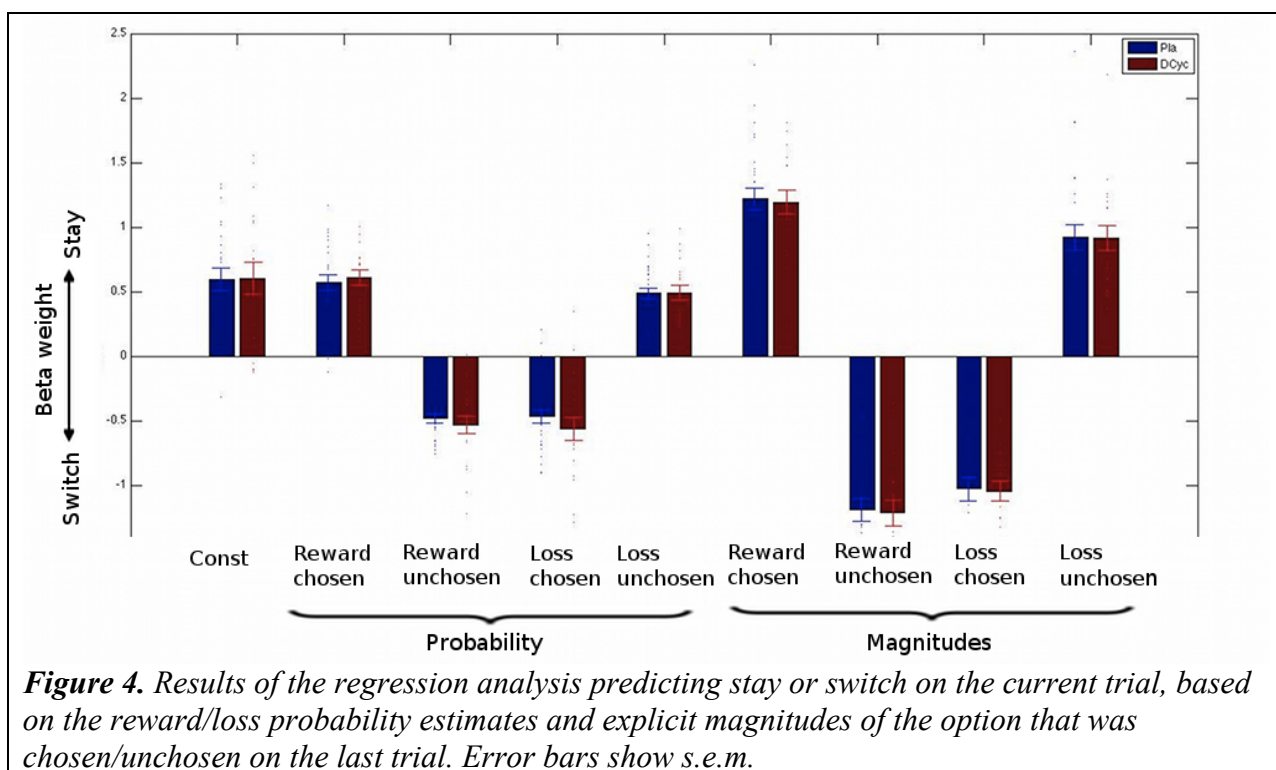


Figure 4. Results of the regression analysis predicting stay or switch on the current trial, based on the reward/loss probability estimates and explicit magnitudes of the option that was chosen/unchosen on the last trial. Error bars show s.e.m.

Third, we also looked at whether learning from the chosen or the unchosen option differed. For this, we fitted a learning model. In the learning model, we fitted two separate learning rates for learning from the chosen and the unchosen option. We found (table 3C) that the placebo and the d-cycloserine group did not differ in the learning rates for the chosen option ($p=0.52$) or for the unchosen option ($p=0.3$).

The results of these three analyses show that participants are influenced by the outcomes from the options they had chosen, as well as the outcomes of the alternative option. The d-cycloserine and the placebo group did not differ in how much their choices were influenced by the chosen or the unchosen options. Similarly, we did not find a difference between the groups in how much they learnt about the chosen or the unchosen options.

a) Last outcome GLM Switch-stay			
	Pla	DCS	p
Last reward outcome chosen	0.65±0.06	0.66±0.06	0.94
Last reward outcome unchosen	-0.50±0.04	-0.53±0.07	0.71
Last loss outcome chosen	-0.51±0.06	-0.55±0.08	0.67
Last loss outcome unchosen	0.55±0.05	0.52±0.07	0.78
b) Bayesian GLM Switch-Stay			
	Pla	DCS	p
Reward probability chosen	0.57±0.06	0.61±0.06	0.97
Reward probability unchosen	-0.48±0.04	-0.53±0.07	0.67
Loss probability chosen	-0.47±0.05	-0.56±0.09	0.49
Loss probability unchosen	0.49±0.04	0.49±0.06	0.34
c) Learning model, chosen and unchosen learning rates			
	Pla	DCS	p
Learning rate chosen option	0.48±0.05	0.43±0.04	0.52
Learning rate unchosen option	0.42±0.05	0.34±0.05	0.3
Beta	0.04±0.00	0.04±0.00	0.87
Loss weight	0.88±0.04	0.88±0.05	0.98
Prob weight	65.7±8.7	71.8±12.4	0.68
fval	147±5	143±5	0.57

Table 3. a) Results of the regression predicting stay or switch on the current trial, based on the last trial reward/loss probabilistic outcomes and the current magnitude values of the option that was chosen/unchosen on the last trial. Mean and standard error of the decision weights are shown for the two groups (Pla – placebo, DCS- d-cycloserine), as well as the results of between-subject t-tests comparing the placebo and the d-cycloserine groups. b) Results of the regression analysis predicting stay or switch on the current trial based on the reward/loss probability estimates and explicitly shown magnitudes of the option that was chosen/unchosen on the last trial. Decision weights are mean values with standard error; p-values are the results of between-subject t-tests. c) Parameter estimates of the learning model with separate learning rates for the chosen and the unchosen option (mean values with standard errors shown), p-values are the results of between-subject t-tests.

4.4 Discussion

Summary of the results

We studied the effect of changes in NMDA activity on human reward-guided learning and decision-making using d-cycloserine, a partial NMDA agonist. Participants had to learn changing probabilities of positive and negative outcomes over time and integrate them with explicitly cued reward and loss magnitudes in order to make decisions. While participants were able to track the changing probabilities, this learning was not modulated by d-cycloserine. If there had been an effect on learning, we would have expected either a change in learning rates or alternatively a changed reliance on the learnt compared to the explicit information, neither was the case.

However, d-cycloserine had a significant effect on decision-making. The drug shifted decision-making away from a heuristic approach towards a more optimal one. In the current context, more optimal behaviour meant that participants multiplied the learnt probability and the displayed magnitude information rather than adding them (heuristic behaviour). Note, however, that such multiplication might be more demanding especially since participants tracked several pieces of information at a time, making heuristic strategies an appealing way of reducing task complexity.

Absence of a learning effect

To our knowledge, there are no studies in humans that have looked at the effects of d-cycloserine on reward-guided learning and decision-making. However, animal studies suggest that d-cycloserine improves fear extinction (Walker et al., 2002), drug use extinction (Flavell et al., 2011) and appetitive conditioning (Golden and Houpt, 2007). However, our task is most similar to instrumental reward reversal learning task paradigms and it is still unknown whether this neurally distinct form of learning is also improved by d-cycloserine, with Golden and Houpt (2007) finding that at least simple reversal learning might not be.

Additionally, evidence for learning effects in human studies analogous to the animal studies above has been more inconsistent. While there is evidence from clinical trials that d-cycloserine may accelerate the therapy of different anxiety disorders (reviewed in Hofmann et al. (2013)) some experimental studies have not been able to replicate the basic fear extinction effects in humans (Guastella et al., 2007; Klumpers et al., 2012). Furthermore, clinical trials for alcohol or for cocaine use extinction have failed to find an enhancement by d-cycloserine (Kamboj et al., 2011; Price et al., 2013). This discrepancy between human and animal studies could be due to the strong dose dependency of d-cycloserine's effects (Walker et al., 2002): in general animal studies have used a higher dose (by about a factor of three) and injection rather than oral administration of the drug.

Changes in evidence integration for decision-making

We found that d-cycloserine led to a very selective shift in participants' decision strategy, making them integrate different sources of information in a more optimal manner, compared to the placebo group. More specifically, participants integrated the learnt probability and the explicitly cued magnitude more multiplicatively (i.e. non-linearly) and thus less additively (i.e. linearly).

Recent work has suggested functions for NMDA receptors beyond traditional long-term consolidation in learning. To our knowledge, nobody has yet tested whether changes in NMDA-mediated neural processing would affect evidence integration in a complex value-based decision-making task in humans. But there is some evidence that NMDA receptors can modulate higher cognitive functions, like working memory (Wang et al., 2013), attention (Herrero et al., 2013; Meuwese et al., 2013; Self et al., 2012) or delay discounting (Floresco et al., 2008). However, it is important to note that it is unlikely that the decision-making effects in our task are a working-memory effect in disguise. Collins and Frank (2012) have shown that in a standard human learning paradigm (similar to the one used here), learning rate captures a mixture of

cortico-striatal learning systems as well as prefrontal working memory. Therefore, if the observed effect were driven by a working-memory mechanism, we would have expected to see a change in at least one of the different measures we used to probe for learning effects. However, we did not see any such effects. Thus at least it appears likely that the maintenance of items in working memory was not affected; it would nevertheless be possible that another working memory ability, such as modulating items in working memory (here using them for decisions) might be affected. Future studies should address this.

There are at least two theoretical frameworks that are however in line with our results. First, NMDA receptors have been proposed as crucial for the integration of information over time during decision-making in a biophysical model (Wang, 2002). While this model was originally based on animal recordings during perceptual decision-making tasks (reviewed in Wang (2008)), recent human brain imaging studies have shown that it is also relevant for human reward-based decision-making (Chau et al., 2014; Hunt et al., 2012; Jocham et al., 2012). Simulation studies have suggested that modulation of the NMDA receptor activity should enhance or impair decision-making, depending on the baseline activation of NMDA receptors (Standage and Pare, 2011). Following these findings, the decision-making improvement in our study could be explained as shifting the NMDA receptor activation into a more favourable range for decision-making.

Second, there is evidence that NMDA receptors are important for neural computations when more than one input is integrated non-linearly. For example, Binns and Salt (1996) have shown that NMDA receptors are crucial for multisensory integration. More specifically, they found that the neuronal super-additivity between inputs of different modalities in the cat superior colliculus was strongly reduced by NMDA blockade. Furthermore, Self and colleagues (2012) have shown that NMDA receptors are crucial for the integration of top-down and bottom up information during figure-ground separation. In the context of our task the optimal way to integrate the probabilities and the magnitudes presented, is non-linearly (i.e. multiplicatively), rather than

linearly (additively). Therefore d-cycloserine may have led to more optimal integration by biasing processing towards more non-linear neural interactions.

One possible explanation for the occurrence of an effect on decision-making in the absence of a learning effect is that while both are supported by NMDA receptors, there is a different optimal range of NMDA activation supporting both functions (Standage and Pare, 2011). In addition, the strength of the effect of d-cycloserine might vary across brain areas, due to different NMDA-receptor subunit compositions or different local levels of glycine (see Yang and Svensson (2008) for a review). Therefore, the behavioural dissociation could be explained as due to different d-cycloserine sensitivities in reward decision- compared to reward learning-related brain areas.

Future directions

Future experiments will have to establish whether our results generalize to the integration of dimensions other than probability and magnitude, and to situations where all information to be integrated is explicitly shown rather than learnt.

It would also be interesting to manipulate the predicted choices for linear versus non-linear integration more widely to produce more distinct changes in choices and outcome between the groups, as well as testing whether other forms of non-linear integration or conditional reasoning (model-based) would also be affected.

Lastly, future decision-task designs could explicitly contrast which theoretical NMDA integration mechanism mentioned above drives the behavioural effect: while the biophysical model explanation would predict decision enhancements in situations where attributes are integrated linearly or non-linearly, the super-additivity mechanism would predict enhancements only where attributes need to be integrated non-linearly.

Conclusion:

We have found that enhancing the activation of NMDA receptors changed how humans

integrated evidence in a multi-attribute decision-making and learning task in the absence of more direct effects on learning. This suggests a role for NMDA receptors beyond their well-studied role in long-term consolidation and plasticity. Although we cannot clearly dissociate between different candidate mechanisms for our observed effect, either biophysical attractor networks or basic super-additivity models could plausibly explain our data.

5. Reward and effort learning in dysphoria

Summary

In chapters 3 and 4 we looked at whether and how different pharmacological manipulations could influence different aspects of reward-guided decision-making and learning. In the present chapter, we now turn to the question how these effects compare to those seen in depression, i.e. whether treatments reverse the effects seen in depression. If it turns out that the effects do not match, this could suggest cognitive mechanisms that future treatments should target. To answer this question, we measured the effect of dysphoria, on a concurrent reward and effort learning task, very similar to the task used in chapters 2 and 3. We collected data from 17 control participants and 17 dysphoric participants (included based on a Beck Depression Inventory Score above 15). Using a model-based analysis approaches, we could tease apart different components of reward- and effort-guided learning and decision-making. We then examined these components separately. We found that dysphoric participants took the information that they had learnt less into account when making decisions compared to controls. However, the two groups did not differ in their speed of learning. This result highlights the importance of using modelling approaches to tease apart the different components making up complex behaviours.

5.1 Introduction

Major depression is a heterogeneous disorder with varied emotional symptoms. One key symptom of depression is anhedonia - a reduced interest in or enjoyment of activities. This symptom may conceptually relate to the psychological construct of processing of valenced (rewarding or punishing) information. This is further supported by meta-analyses finding that processing of reward and punishment is biased, or ‘blunted’, in depression and that these biases affect diverse domains such as autonomic responses, memory and attention. This chapter focuses on one aspect of reward-guided behaviour, namely how humans learn how desirable and aversive stimuli in the environment are, how they update/revise this expectation over time and use it to make decisions. As discussed in the introduction (chapter 1.2.1), several studies have previously addressed the question whether depression affects learning but have given conflicting results. It has been proposed (Huys et al., 2013) that one reason for this heterogeneity of results is that a conceptual distinction needs to be made between learning as ‘improving performance with practice’ versus ‘the speed of learning associations’. While the speed of learning is what most studies were interested in measuring, what was actually often measured was improving performance with practice. However, performance itself is influenced by many other factors, such as sensitivity to rewards or punishments or being able to behave consistently, which are also known to be affected in depression (Blanco et al., 2013). Therefore, to be able to tease apart changes in performance into its components, including speed of learning, an appropriate experimental design and model-based analysis are necessary (Huys et al., 2013).

In addition, it is still unknown whether the presumed learning deficit for valenced stimuli is specific to only positive dimensions (reward) or also extends to negative ones, such as effort.

To measure whether valenced learning is affected by depression independent of other distinct aspects of valuation and learning, we tested 17 dysphoric and 17 control participants on a

learning task specifically designed to separate these different factors. We analysed the behaviour in a model-based fashion allowing the dissociation between distinct components necessary for task performance. The task was adapted from the concurrent reward and effort learning task described in chapters 2 and 3. In this task, participants again concurrently learnt the reward and effort magnitudes of two options. To select between both options, they had to use this learnt information. However, this time they integrated the learned information with explicitly shown probabilities of how likely each option was to require exertion of an effort, rather than probabilities of receiving reward. In other words, whereas the last chapters had real or hypothetical rewards, this study had real and hypothetical effort.

Effort was selected as the second dimension to be learnt because symptoms of depression, such as fatigue or loss of energy or reduced engagement in pleasurable activities, can be related to psychological concepts of motivation, i.e. the willingness to exert effort to obtain rewards. As reviewed in the introduction (chapter 1.2.3), several studies have suggested that effort-based decision-making is affected by depression with depressed participants being less willing to engage in effortful behaviour. However, in real life situations, it is not very common for choices to be presented together with explicit reward and effort; rather, humans need to learn from experience how rewarding and effortful different courses of actions are. It is therefore possible that this learning about effort requirements is specifically affected in depression, so that patients might not be willing to engage in effortful behaviour because, based on past experience, they have an exaggerated expectation of how effortful it will be.

Additionally, this design specifically allowed us to dissociate different potential effects of depression on overall learning performance, namely: 1) speed of learning (for both reward and effort), 2) reward vs. effort sensitivity, 3) use of the learnt information to make decisions and 4) effects of actual vs. hypothetical effort exertion. Based on the meta-analysis by Huys et al.

(2013) discussed in the introduction we hypothesise that dysphoric participants will differ in how well they use the learnt information to make decisions or in their sensitivity to reward or effort. While based on (Chase et al., 2010) we could expect an effect on the speed of learning.

5.2 Methods

Procedure

All participants were initially recruited through posters and an online questionnaire. Those who fulfilled the study criteria were then invited to a screening visit at least one week later followed by a testing visit involving performing a learning task while we recorded their brain activity using functional magnetic resonance imaging. This chapter focuses on the behavioural data only.

Initial recruitment

The study was approved by the local ethics council (CUREC Ref: MSD-IDREC-C2-2013-29). Dysphoric and control participants were recruited via poster and online advertising. Control participants were recruited with poster advertising ‘Healthy volunteers wanted’, while dysphoric participants were advertised with posters either headed ‘Are you feeling down?’ or ‘Are you feeling blue?’. Participants interested in taking part were directed to the study website which provided them with more information about the study (participant information sheet) and allowed them, after indicating that they consented, to fill in the BDI-II questionnaire (Beck Depression inventory, (Beck et al., 1996)). If they scored either no more than 5 (controls) or at least 15 points (classified as mild depression, dysphorics), they were then invited for the screening visit. As a study safety measure, if participants indicated on the BDI suicide item that they were considering suicide and provided contact information, this was passed on to the psychiatrist overseeing the study who then contacted them. About 250 participants completed the online questionnaire.

Screening visit

If they fulfilled criteria, participants were then invited for the screening visit, at least one week after filling in the initial questionnaires, to avoid including participants in the dysphoric group who only had a very short term negative mood fluctuation. In the screening visit, we ensured that

participants fulfilled the study inclusion criteria, after obtaining informed consent. The inclusion criteria for both groups were: no counter-indications to the MRI scan (2 controls and 1 dysphoric were excluded), no current psychoactive medication (i.e. all dysphoric participants were unmedicated for depression), no recreational drugs in the last month, no more than occasional tobacco consumption (<5 cigarettes per week), being right-handed, showing sensitivity to the effort manipulation used during the screening visit (see below for details; excluded 7 controls and 5 dysphorics), BDI in the range stated above (3 dysphorics were excluded). In the dysphoric group, participants were additionally excluded if they had current or past psychiatric comorbidities other than anxiety disorders, as assessed using the SCID (First, 1996), 3 participants were excluded. In the control group, participants were additionally excluded if they indicated any current or past psychiatric problems or any first degree relative with a psychiatric problem (4 participants were excluded). Furthermore, one participant in the dysphoric group withdrew from the study during the screening visit. Overall, 62 participants were screened and 36 were included in the main part of the study. Of these, however, data from 2 participants was excluded after it was collected because of reports of feeling unwell during the MRI scan or lack of sensitivity to the effort manipulation (which made analysis of the behavioural data impossible). This left a sample of 17 participants in each group. (Robinson et al., 2012a) reported that depressed patients showed a more than twice higher error rate in a learning task (Cohen's $d=1.14$). Based on this effect size, a study with 14 participants per group would be sufficient to detect an effect (at a significance level 0.05, two-tailed) with a power of 0.8. Robinson et al. (2012a)

As part of the screening visit, participants also completed several questionnaires and a computerized task measuring their effort sensitivity (see below) to ensure they were sensitive to the effort manipulation used in the study.

Testing visit

The testing visit took place one week after the screening visit for the dysphorics and at a variable time after the screening visit for the control group. The window of one week was chosen to reduce the likelihood of dysphorics showing spontaneous remission. As main part of the testing visit, participants performed a concurrent reward and effort learning task (see below) while we measured their brain activity using fMRI for about an hour. After a short break, they performed another schedule of the same task outside the scanner for an hour to increase the number of trials for data analysis. They also completed several questionnaires (see below).

Questionnaires

Depressive symptoms were measured using the Beck Depression Inventory (BDI-II, Beck et al. (1996)), the interviewer-led version of the Hamilton Rating Scale for Depression (HRSD, Hamilton (1960)) and the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I, First (1996)). BDI-II was scored as a total summed score across all items, as well as subscore for anhedonia (items 4, 12, 15, 21). For HRSD interview questions and scoring, guidelines from Williams (1988) were followed. Symptoms of anhedonia were further also assessed using the Snaith Hamilton Pleasure Scale (SHAPS, Snaith et al. (1995)). Responses were scored according to Franken et al. (2007) between 1 and 4, with higher values indicating higher symptoms of anhedonia, as well as by binarizing the data. Symptoms of apathy were assessed using the Starkstein scale (Starkstein et al., 2001), with higher values indicating greater apathy. Persistence was measured using the persistence subscale of the Temperament and Character Inventory (Cloninger et al., 1993), using a binary scaling with higher values indicating higher persistence. Persistence or willingness to exert effort was also measured using the ‘Work and Activity’ item of the HRSD (item 7). Anxiety was measured using the Spielberger State and Trait anxiety questionnaire (Spielberger and Gorsuch, 1983). Positive and negative affect was measured using

the Positive and Negative Affect Schedule (PANAS, Watson et al. (1988)), which includes separate subscores for subjective positive and negative affect.

During the screening visit, participants completed: BDI-II, HRSD, SCID, Trait/State anxiety, PANAS. During the test visit, they completed: BDI-II, State anxiety, PANAS, SHAPS, Persistence scale, Starkstein scale.

Learning task

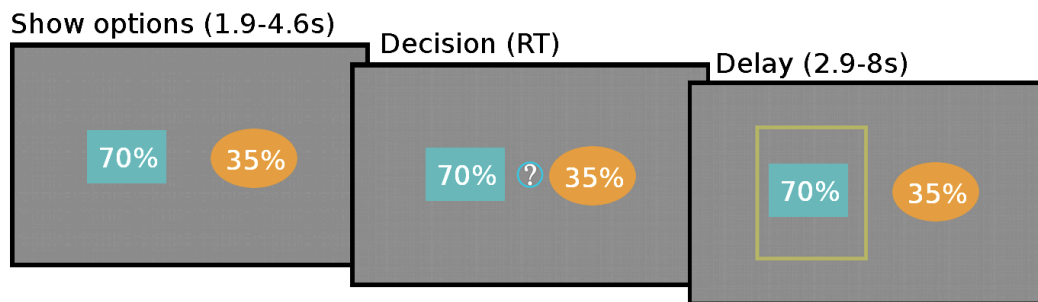
Participants performed a concurrent reward and effort learning task (fig 1), very similar to the one described in chapters 2 and 3, with some modifications (highlighted in detail below).

In short, the task was as follows: Participants repeatedly chose between the same two options, aiming to choose the options maximizing their monetary gain and minimizing the effort they needed to exert to obtain the reward. When making their decision (Fig. 1A), they therefore had to take into account the independent reward and the effort magnitudes of the two options, which they had to learn and track from experience across trials as they slowly varied over the course of the experiment. At the time of the decision, they were additionally explicitly shown how likely each option was to require exertion of an effort. If an option led to a ‘real effort’ outcome, participants needed to complete an effort exertion phase (see below) following the outcome phase; if not, the effort exertion phase was skipped. After each choice, participants saw the outcome of their choice. At this time point, participants could learn about the two options. This task was designed to be challenging for participants: they had to simultaneously learn about all the different attributes, as well as learn about efforts in the same way when they were real or hypothetical. Thus, optimal learning needed to represent all these components separately to prevent them interfering with one another. In the effort exertion phase, participants needed to perform a sustained effort involving clicking on targets as they appeared on the screen.

The specific modifications of this task compared to the one used in chapters 2 and 3 were:

Firstly, and most importantly, the meaning of the probability values shown at the time of the decision was changed to now relate to the effort rather than the reward magnitude. This meant that in the outcome phase, participants were shown for the chosen option whether they had to exert the effort associated with it in the following effort phase. If not, the effort phase was replaced by a short delay of 4s. In other words, in this task, the reward was always real, but the effort of the chosen option could be real or hypothetical. This change was made because many participants in the study in chapters 2 and 3 reported a strong aversion to the effort in the scanner and we considered it possible, that dysphoric participants were more sensitive to the effort. We therefore wanted to reduce the effort requirements of the task. However, we note that this also decreased easy comparison with the other chapters. Secondly, the effort phase was also slightly modified. The maximum allowed duration was increased from 25s to 30s and there was no longer a requirement to maintain central fixation. This was again done to reduce the effort requirement. Thirdly, if participants failed to complete the effort phase, they lost a fixed reward value equivalent to a full bar (100 points, i.e. 10 pence), rather than the reward value of the chosen option. Fourthly, there were no more effort query trials, where participants had to indicate which option had the higher reward, and the number of Special-option-trials (SOTs) was reduced to 16 out of 124. Sixthly, the stimuli and screen background were adapted to have approximately the same brightness (see Fig. 1). This was to reduce potential brain activations due to changes in brightness and furthermore to make the viewing of the stimuli more pleasant as some participants in the previous study had reported some discomfort at the brightness changes.

A Decision phase



B Outcomes phase

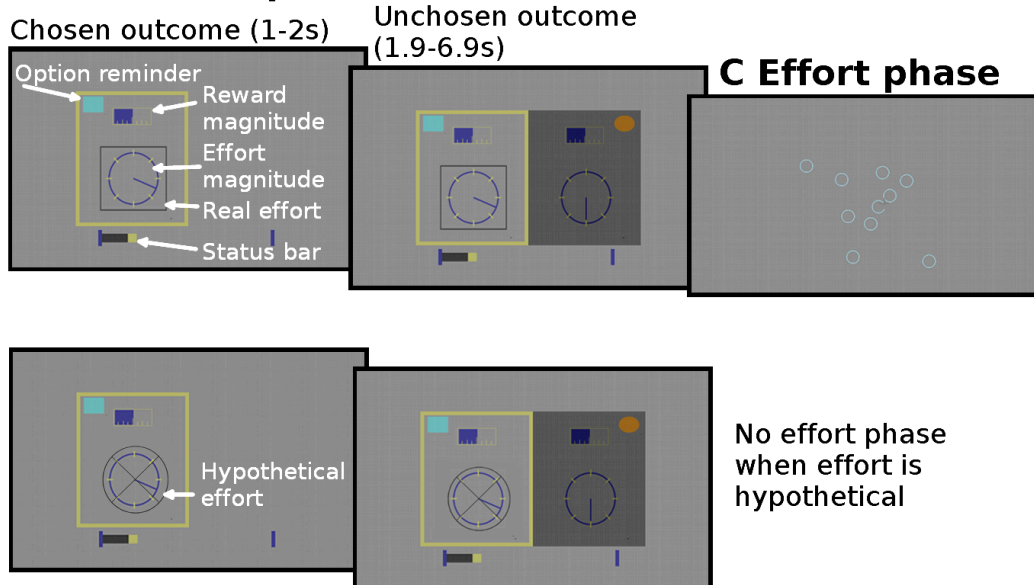


Figure 1. Task description. In the decision phase (A), participants were shown two options (i.e. choices), overlaid with the probability of receiving a reward for each choice. They could only decide after an initial monitoring phase (1.9-4.6s). The chosen option was then highlighted for 2.9-8s. In the following outcome phase (B), participants saw the outcome for the chosen option first (1.9-2.1s on all trials but the first three trials inside the MRI scanner, where the duration was 4s). The reward magnitude was shown as a purple bar (top of the screen), the effort magnitude was indicated through the position of a dial on a circle. The reward was added to a status bar at the bottom of the screen, which tracked rewards over the course of the experiment. Whether they had to exert an effort was indicated by either a box around the effort dial ('real effort', top display) or a crossed out circle on top of the effort dial ('hypothetical effort', bottom display). A reminder of what option they had chosen was shown at the top of the screen. Then the reward and effort magnitudes were shown for the unchosen option (1.9-6.9s). Finally, if effort was real, participants performed the effort phase (C) where the number of targets was equivalent to the chosen effort outcome.

Timings: Event timings were drawn from Poisson distributions to optimise separation of events:

Monitoring phase: mean 2.7s, min: 1.9s, max: 4.6s; RT: mean: 1.3s; Highlight chosen option

phase: mean: 4.4s, min: 2.9s, max: 8s; Outcome chosen option phase: mean: 2s, min: 1.9s, max:

2.1s (apart from first three trials: 4s); Outcome of both options phase: mean: 3.3s, min: 1.9s,

max: 6.9s; Effort exertion phase: during fMRI scan: mean:14.6s, outside scanner: mean 12.4s; ITI: when trial included the effort phase: mean 4.2s, min 2.3s, max 7.4s; when trial did not include effort phase: mean: 8.1s, min: 4.6s, max: 11.7s .

Data Analysis

All data analysis was performed in Matlab and SPSS. In ANOVA, when assumptions of sphericity were violated, Greenhouse-Geisser corrections were used. In t-tests, when the data was not normally distributed, non-parametric tests (one-sample Wilcoxon Signed rank test) were used.

Questionnaires

Different questionnaire measures were collected to further describe our dysphoric and control samples. These measurements also allowed putting different of self-reports of day-to-day function and feelings in relation to our behavioural task measurements. The data from each questionnaire was first recoded according to standard procedures as indicated in the questionnaire references provided above (e.g. reverse coding of some items). We then compared the data of each questionnaire between the two groups using two-tailed between-subject t-tests. When single items from one questionnaire were missing, we replaced them by the mean of the other items from that person. When whole questionnaires were missing from the testing day (BDI-II, PANAS, State anxiety, Cloninger, Starkstein: 1 participant; SHAPS: 10 participants), they were replaced, where possible (Cloninger and Starkstein data remained missing for one person), with data from the screening visit.

To assess whether the different questionnaires measured different cognitive phenomena, we correlated the summary scores from the different questionnaires (Pearson correlation

coefficient). This was done separately for each group so as to avoid spurious correlations due to main effect differences between the groups.

The different questionnaire scores were also correlated with relevant behavioural measurements. We correlated questionnaire scores across all participants ($n=34$) and also within just the dysphoric group. We also performed partial correlations across all participants, controlling for group assignment.

Decision and learning behaviour analysis

The aim of the decision analyses was to separate out different factors contributing to the reward- and effort- guided behaviour and see how they are affected by depression. Specifically, these components were: 1) learning speeds (i.e. learning rates) for effort and reward magnitude learning associated with the different options, 2) behavioural sensitivity to reward relative to effort, 3) behavioural sensitivity to learnt compared to explicitly shown value information, 4) behavioural stochasticity, or in other words ability to use the overall value difference (i.e. a combination of reward, effort and probability differences) of the two options to select between them consistently and accurately.

Distinguishing these different components was made possible by using computational analyses methods as well as the specific features of this concurrent learning task: 1) As two dimensions (reward and effort) needed to be learnt, we could dissociate whether only learning about reward or learning more generally, about both reward and effort was affected. 2) As participants needed to combine learnt (reward and effort magnitude) and explicitly shown (i.e. not learnt) dimension (effort probability), we could dissociate sensitivity to reward or effort from both sensitivity to learnt value information compared to value information in general and from choice stochasticity. 3) As a consequence of the effort probabilities, effort was sometimes real and sometimes

hypothetical, we could perform exploratory analyses similar to chapters 2 and 3, examining the effect of interferences of irrelevant interferences on model-based behaviour.

Regression analyses

Decision regression with learnt dimensions as past trial outcomes

In the first regression analysis (dGLM1), we predicted whether participants repeated the same choice as on the last trial ('stay') or selected the alternative choice ('switch') based on the information explicitly shown at the time of the choice and the information participants needed to learn. Specifically, the regressors were: effort probability (explicitly shown at the time of the choice), reward and effort magnitude outcomes from the past four trials, whether the effort had been real or hypothetical on the last trial. This is thus similar to a truncated reinforcement learning. In this and the following regression analyses of the decisions, all of the predictors (apart from whether effort had been real or hypothetical) were entered as difference between the option in favour of 'stay' minus in favour of 'switch'. This was done because it is the difference, not the absolute values that should influence decisions and to reduce the total number of regressors entered in the analysis. The regressions were performed using matlab's `glmfit` function with a 'logit' link function (binomial distribution). All regressors were normalised before being entered into the regression. As in chapters 2 and 3, we only included trials on which no 'special options' were present (216 out of 248 trials), as stay/switch was not defined on those trials.

This regression analysis was performed for each participant individually, based on the combined data from inside and outside the scanner. For each participant, we thus obtained one regression weight for each predictor. Significance tests were then performed on these regression weights. To test whether participants across both groups were sensitive to the different features they

should use to make their decisions (the learnt and explicitly shown information), we tested the combined data from both groups using ANOVAs or one-sample two-sided t-tests.

The following group differences were analysed: Firstly, to test whether the groups differed in their learning speed or the use of learnt information, the regression weights for the past reward and effort magnitudes (effort values were sign-reversed) were entered into a 2 (valence – reward/effort) x 4 (time steps in the past) x 2 (group) ANOVA. In this analysis, a difference in learning speed would be indicated by a group x time interaction, a different sensitivity to reward vs. effort would be reflected as a group x valence and interaction and finally, a different sensitivity to both types of learnt information would result in a main effect of group. Secondly, to test whether the groups differed in how they used the explicitly shown information, the regression weights for the shown probability were compared using t-tests. Thirdly, to test for differences in the tendency to repeat a previous choice ('perseveration'), we compared the constant, indicating a tendency to stay rather than switch beyond value differences of the option, of the regression analysis using a t-test.

Decision regression with learnt dimensions as Bayesian values

As explored in the introduction (chapter 1.1) and experimental chapters 2 and 3, a key component of learning models is the prediction error, i.e. the difference between the actual and the expected outcome. While the analysis above (dGLM1) could measure a proxy of learning (ANOVA: effect of time), it did not directly contain the prediction error itself, i.e. a key component of learning as discussed in the chapter 3, explicitly. To address this (dGLM2), we again followed the approach from chapter 3 (and also Chiu et al. (2008)): we used the same Bayesian model as described in chapters 2 and 3 to create trial-by-trial predicted reward and effort values for the current experiment. We then split these values into their components, namely the predicted value preceding the last outcome and the prediction error of the last

outcome. Including these regressors in the analysis from above we replaced the past trial outcomes for reward and effort magnitudes. We then tested whether the groups differed in their use of the Bayesian predictions and PEs using a 2 (Valence: reward/ effort) x 2 (type: last trial prediction / PE) x 2 (group) ANOVA. As follow-up tests, we performed two ANOVAs, separately for the last trial prediction weights and the last trial prediction error weights.

Decision regression – motivational fluctuations

As an exploratory analysis, we tested whether fluctuations in trial-by-trial motivation might have effects on choice. We hypothesized that having or not having to exert an effort on one trial might affect how much participants value the different dimensions at the time of choice on the next trial. For example, they might feel more motivated after not having had to exert an effort on one trial, leading them to take reward more into account on the current trial or conversely, they might feel less motivated after having had to exert an effort and might therefore try harder to avoid the effort on the current trial. To measure this effect, we performed a regression analysis (dGLM3), which included a measure of whether effort on the last trial affected how much the other decision relevant information was taken into account in the current trial. The three regressors of interest were interactions of effort type on the last trial (real vs. hypothetical) with: 1) relative (in favour of stay minus in favour of switch) probability, 2) relative reward and 3) relative effort magnitudes (from Bayesian model). As regressors of no interest, we also included the relative probability, learnt reward and effort Bayesian values and the effort type. We then entered the resulting regression weights for the interaction regressors in a 3 (type: probability, reward, effort) x 2 (group) ANOVA.

Effort exertion

Previous studies looking at the effects of depression on effort-based decision making have often examined mostly physical effort, such as squeezing of a hand dynamometer. Less is known

about how depression effects exertion of effort that also has cognitive components, such as movement planning. In the present study, the effort participants needed to exert was physical – moving a mouse cursor using a trackball mouse -, sustained – effort exertion phases lasted up to 30s – and also involved a cognitive planning element – participants needed to plan in which order to select the effort targets.

From this more complex effort behaviour, we could obtain several behavioural measurements. Firstly and most coarsely, we measured how long participants took on each trial to complete the effort and on what percentage of trials they managed to complete it successfully.

Secondly, we could also obtain two measure related to motivation. We measured the rate of clicking with the mouse during the effort exertion, which we found (chapter 2) to be sensitive to incidental reward levels. In the present experiment, the reward received (or of the unchosen option), as well as the probabilities associated with them should again be irrelevant for performing the effort: they have no influence on how the effort should be exerted or on the potential punishment for not completing the effort. To measure whether indeed participants' clicking rate showed this pattern in the current study we performed a regression analysis (eGLM1) predicting clicking rate, based on the probabilities and the effort and reward magnitude outcome of both the chosen and the unchosen options. We also included as confound regressor whether the data was collected in the session inside or outside the scanner. This was because while participants had to lie down inside the scanner, which made using a trackball mouse somewhat uncomfortable, they sat upright in front of a computer in the session outside the scanner. The regression was performed using matlab's glmfit with a 'log' link function (poisson distribution) because visual inspection of the clicking rate data suggested that it was not normally distributed (as also found in chapter 2).

We also measured the trajectories along which participants selected the effort targets as a measure of planning. Specifically, we recorded where participants clicked when they selected an effort target and at what time. From this we reconstructed the total trajectory of each trial and from this the relative trajectory length per target (as a measure that is more comparable across trials than the total trajectory length). We used this measure instead of the actual mouse cursor trajectory of mouse movement on each trial, as this data was not available. To test whether motivational factors affected this measure, we performed a regression analysis (eGLM2) with the same predictors as above, but this time using an ‘identity’ link function (normal distribution) as this data visually appeared normally distributed. The resulting regression weights for reward and probability were then entered into a 2 (valence: reward, probability) x 2 (choice: chosen, unchosen option) x 2 (group) ANOVA.

5.3 Results

Participant characteristics

The two groups were well matched in terms of age and years of education (all $p>0.8$). There were strong differences (all $p<10^{-4}$) between the groups in all questionnaires (table 1), i.e. they differed in symptoms of depression (BDI, HRSD), anxiety (Spielberger Trait/State anxiety), anhedonia (SHAPS, BDI anhedonia subscale), apathy (Starkstein scale), positive and negative affect (PANAS) and motivation (Persistence subscale of the Temperament and Character Inventory). Notably, while the BDI-II values in the dysphoric group (mean 32 ± 3.0) indicated severe depression, their HDSR values (mean 16.0 ± 1.5) only indicated mild depression. According to the SCID, 12 of the dysphoric participants fulfilled criteria for major depression while 5 did not. Incidentally, we also observed that several participants expressed confusion with respect to what the questions of the Starkstein Apathy questionnaire asked of them – e.g. whether ‘Are you concerned about your condition?’ meant whether they cared about how they felt or whether they thought they had health problems. This suggests, that the results from this questionnaire need to be interpreted cautiously.

	Control	Dysphoric	p-value
Aged	24.2±1.3	24.2±1.0	0.9
Education (yrs)	15.2±0.4	15.2±2.2	0.9
BDI - total	0.2±0.2	32.1±3.0	5*10 ⁻¹²
BDI - anhedonia	0.0±0.0	5.7±0.7	3*10 ⁻⁹
BDI - not anhed.	0.2±0.2	26.4±2.4	3*10 ⁻¹²
HRSD - total	0.5±0.2	16.0±1.5	10*10 ⁻¹²
HRSD - item 7	0.0±0.0	2.1±0.3	2*10 ⁻⁹
SHAPS - vas	17.9±1.1	33.7±1.5	1*10 ⁻⁹
SHAPS - binary	0.3±0.2	5.9±0.9	5*10 ⁻⁷
Persistence	15.5±0.8	9.4±1.1	2*10 ⁻⁴
Apathy	6.8±1.1	19.8±1.6	2*10 ⁻⁷
Trait anxiety	25.1±0.9	58.4±2.4	3*10 ⁻¹⁴
State anxiety	24.8±1.2	52.3±2.7	2*10 ⁻¹⁰
PANAS - positive	35.4±1.5	19.2±2.1	3*10 ⁻⁷
PANAS - negative	10.6±0.3	25.5±2.5	2*10 ⁻⁶

Table 1. Mean and standard errors of questionnaire scores for the control and the dysphoric group. P-values indicate the result of two-tailed between-subject t-tests. Abbreviations: BDI (Beck Depression Inventory) –total: summed score across all items, BDI – anhedonia: score of anhedonia items, BDI - not anhed.: score of BDI items excluding anhedonia items; HRSD (Hamilton Rating Scale for Depression) – total: summed score of all items, HRSD - item 7: score of the ‘work and activities item’; SHAPS (Snaith-Hamilton Pleasure Scale) – vas: sum of scores when coded ranging from 1 to 4, SHAPS – binary: sum of scores when coded as 0 or 1; Persistence – subscale from the Temperament and Character Inventory; Apathy - from Starkstein Apathy scale; Trait and state anxiety from Spielberger Trait/State anxiety scale; PANAS (Positive and Negative Affect Schedule): positive items and negative items subscores.

We calculated the correlations among questionnaire items (Fig. 2), separately for controls and dysphorics. Looking at these correlations is particularly relevant for the dysphoric group, as it can tell us whether – even though they differ from controls on all these measures – the questionnaires measure different symptoms of depression. However, it is important to keep in mind that in actually, this can only tell us whether we are measuring different dimensions of subjective reporting of symptoms, not whether we are measuring different behavioural or neural dimensions. For this it will be necessary to correlate these questionnaire measurements – if they cover different dimensions – with different behavioural and neural measures.

Overall, there were strong correlations among many of the different measures used (Fig. 2B). Particular exceptions to this pattern were persistence, apathy and positive PANAS scores. This suggests that while many of the questionnaires measured at least partially overlapping constructs, some measured separate constructs. Scores of depression measured with BDI or

HDSR were strongly correlated ($p < 0.01$). This suggests that they indeed both capture the same construct of 'depression'. Both questionnaires also strongly correlated with trait and state anxiety and with negative affect scores, which is unsurprising given the high rates of comorbidity and shared symptoms between depression and anxiety disorders (Mineka et al., 1998).

A questionnaire measure that is particularly relevant to reward processing is anhedonia. The two measures of anhedonia, BDI anhedonia items and SHAPS, correlated with each other ($p < 0.05$). SHAPS anhedonia did not correlate with the non-anhedonia items of the BDI, suggesting that it may measure a separate concept to the non-anhedonia items. However, it is also of note that the anhedonia items in the BDI strongly ($p < 0.001$) correlated with the non-anhedonia items of the BDI. Surprisingly, SHAPS anhedonia scores did not correlate with positive affect in the PANAS, suggesting reporting feeling pleasure in response to imagined positive events is not the same as reporting currently felt positive emotions. Furthermore, SHAPS anhedonia scores did not correlate either with persistence scores, suggesting that, at least to some extent, unwillingness to exert an effort in depression may be more than just the result of reduced enjoyment of the rewards that can be obtained through exerting an effort.

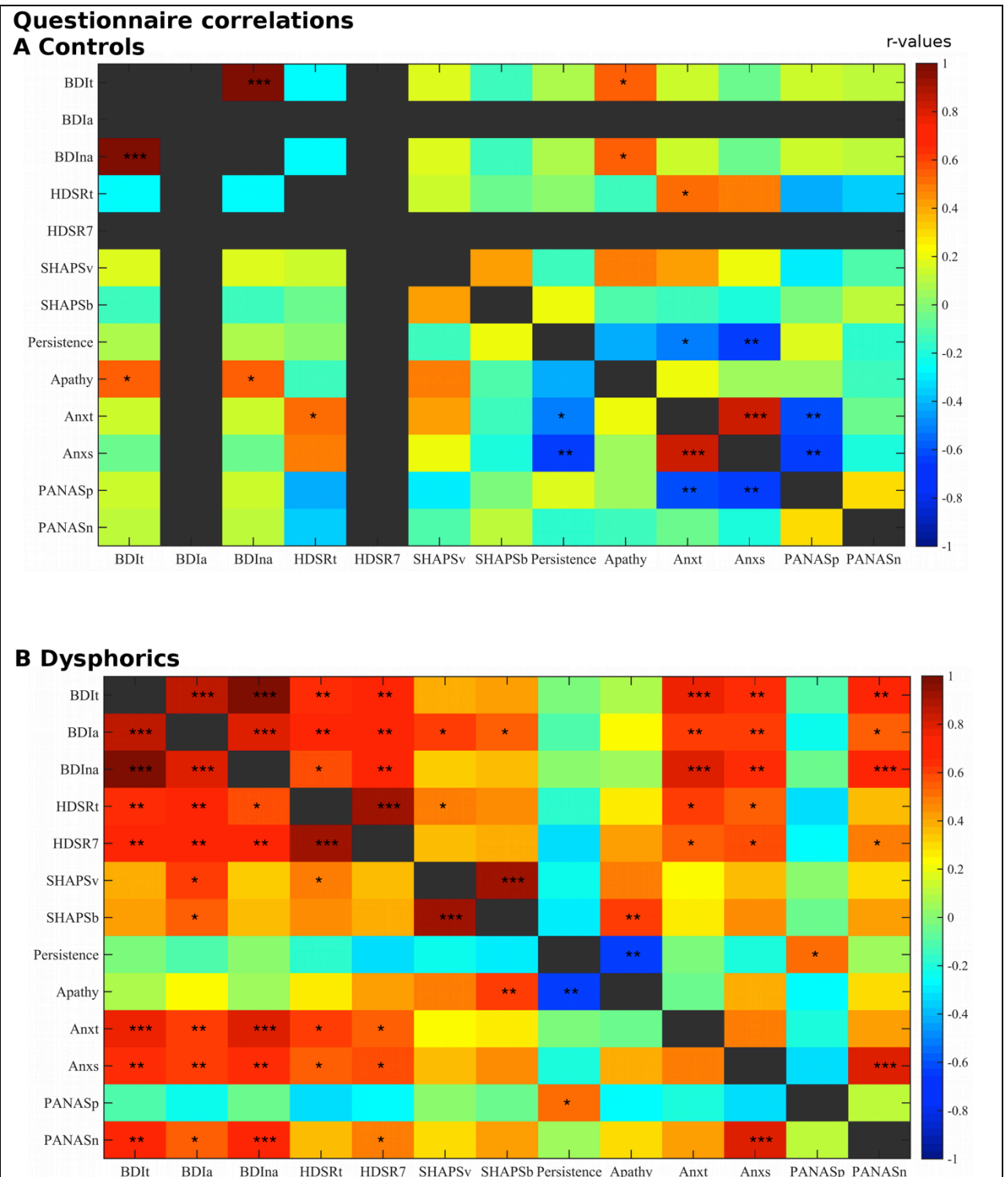


Figure 2. Correlations between questionnaire scores. A) Correlations within the control group ($n=17$), B) Correlations within the dysphoric group ($n=17$). Colours indicate the strength of the correlations (r -values), while significance of the correlations are indicated as: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Black squares indicate where no correlations could be computed for the control group because all participants indicated the same score.

Decision Regression analyses

Regression with learnt dimensions as past trial outcomes

We performed a regression analysis to assess whether dysphoria affected specific components of reward/effort guided decision making and learning (fig 3A, dGLM1). In the regression analysis, we predicted whether participants selected the same option again as on the last trial ('stay' with the last choice) or whether they selected the alternative option ('switch' to the alternative). We found that participants across the two groups used all the relevant information to make their choices: they were more likely to select an option if it had a lower probability of requiring effort, had been associated with higher reward magnitudes and lower effort magnitudes in the past (all $p < 0.001$). Furthermore, participants also evidenced learning, seen as a decrease in impact of outcome information that was more trials ago in the past (ANOVA, effect of time: $F(3, 96) = 63$, $p < 0.001$). Different from the results in chapters 2 and 3, in this specific analysis, there was no main effect of the irrelevant information – i.e. of whether effort had to be exerted on the last trial – affecting the current choice in either group ($t(33) = 0.09$, $p = 0.93$).

Next, we examined whether controls and dysphorics differed in their learning or how they used the learnt information to make choices. We found that they did not differ in their learning speed for reward or effort magnitudes (ANOVA, group x time interaction: $F(3, 96) = 0.08$, $p = 0.97$). However, dysphorics used the learnt information – across both effort and reward – less than controls when making decisions (ANOVA, main effect of group: $F(1, 32) = 4.79$, $p = 0.036$). This was in the absence of changes to the use of the explicitly shown probability information ($t(32) = -0.135$, $p = 0.89$). This pattern of results is in support of a reduced sensitivity to value information that was learnt. Additionally, while participants in general had a strong tendency to 'stay' with a previously chosen option, this again did not differ between the groups ($t(32) = 1.048$, $p = 0.30$).

Regression using Bayesian values of learnt reward and effort

In dGLM1, learning behaviour only be measured somewhat indirectly as slope of how past outcomes were weighted. To measure this learning again using a different approach, we performed a separate regression analysis, dGLM2 (Fig. 3Bi), which included prediction errors (PEs) for reward and effort as markers of learning. At the same time, this analysis also included the prediction prior to last trials outcome. This allowed us to find that the groups differed in how much their decisions were influenced by the last trial PEs compared to the previously learnt predicted values (ANOVA, type x group interaction $F(1, 32)=4.468$, $p=0.042$). Follow-up ANOVAs revealed that more specifically, dysphorics used the past predictions less than controls (ANOVA, main effect of group: $F(1, 32)=3.747$, $p=0.062$). This result was reminiscent of the analysis above (dGLM1) that also found that learnt information was used less by dysphorics. At the same time, the groups did not differ in how much their choices were influenced by the last PE ($F(1, 32)=0.646$, $p=0.427$). This result can again be related to dGLM1, which found no difference in the learning speed between the groups.

Relationship between decision making and questionnaire measures of depression and anhedonia

Lastly, we sought to see how measures of learning (i.e. regression weights for the prediction errors derived from the Bayesian model, Fig. 3Bi) correlated with questionnaire measures of either depression (BDI, HRSD) or anhedonia (SHAPS, BDIA) across both groups of participants combined, controlling for group (partial correlation). In other words, we tested whether these questionnaire measurements correlated with the behavioural measures beyond any main differences across the two groups. There were no correlations between any of the regression weights for either reward or effort reward PEs (i.e. measuring learning specific to one dimension) or their average (i.e. measuring learning across both dimensions) with any of the questionnaire measures (all $p>0.2$ for either parametric (Pearson) or non-parametric (Spearman))

correlations. Next, as a confirmatory analysis of the group difference trend seen in the above analysis, we analysed correlations between the questionnaire measures with how much learnt information was used (i.e. regression weights for the last trial predicted reward or effort value derived from the Bayesian model, Fig. 3Bi). How much participants used the learnt information correlated negatively (parameteric, Pearson correlation) with measures of depression (BDI: $p=0.019$, $r=-0.40$; HRSD: $p=0.044$, $r=-0.35$) and anhedonia (BDIa: $p=0.031$, $r=-0.37$; SHAPS: $p=0.010$, $r=-0.44$). This suggests that participants who were more depressed or more anhedonic relied less on the learnt information when making decisions. However, once group assignment was controlled for, these correlations were no longer significant (all $p>0.1$). Similarly, when we only tested these correlations within the dysphoric group, none reached significance, though there was a trend for the BDI scores towards significance ($p=0.07$, $r=-0.45$; BDIa: $p=0.19$, $r=-0.33$; BDIna: $p=0.06$, $r=-0.46$). This suggests that within the dysphoric group, people with higher symptoms of depression tend to also be more affected in their use of learnt information when making decisions.

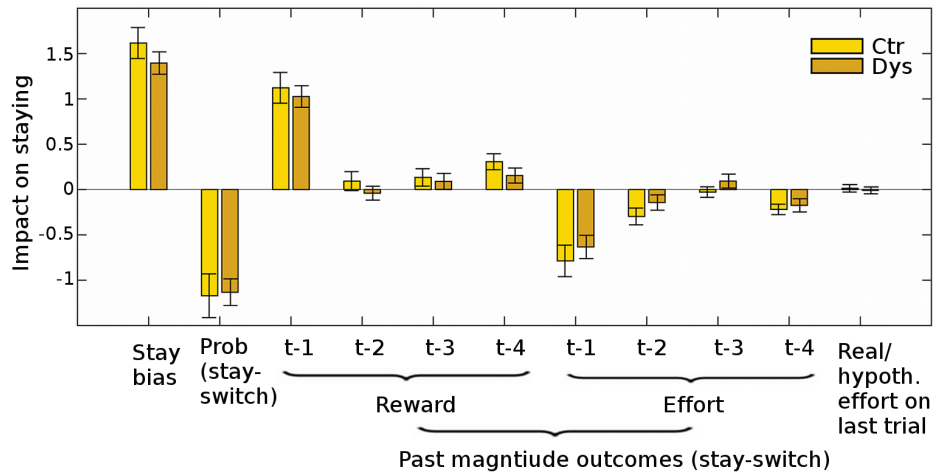
Decision making – motivational fluctuations

Lastly, we performed an exploratory analysis (dGLM3) to assess whether the effort being real or hypothetical – which might lead to motivational fluctuations - across trials influenced decisions. Specifically, we measured whether having to exert effort on one trial might influence participants in how much they take certain dimensions into account when deciding on the next trial. The analysis revealed that across both groups, participants took the decision relevant dimensions in account differently as a function of having had to do effort on the last trial (ANOVA, main effect of intercept: $F(1, 32)=4.49$, $p=0.042$). Follow-up t-tests/ non-parametric tests revealed that this was in particularly true for reward: participants took reward more into account after not having done effort (effort type interactions with reward: $z=-2.4$, $p=0.016$; with effort: $t(33)=0.249$, $p=0.81$; with probability: $z=1.359$, $p=0.174$). This however did not differ

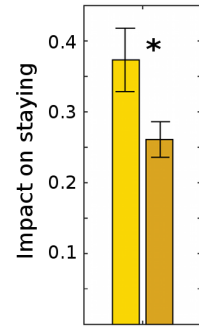
between the groups (all $p > 0.13$). As an aside, we also noted that, different from the results before, there was now a significant main effect of real/hypothetical ($t(33) = -2.713$, $p = 0.011$). Participants tended to chose an option more if it had been associated with hypothetical rather than real effort on the last trial. This effect was thus conceptually the same as the effect found in chapters 2+3 for real vs. hypothetical reward. This difference in results between this and the previous regression analyses could be due to the fact that, here, we controlled for other potential effects of effort type (having a main effect on the decision, as well as influencing other choice variables). In other words, these effects may only become clear when controlling for the other factors at the same time. Additionally or alternatively, it is possible that in this task the main effect of real/ hypothetical effort was weaker than the effect of real/ hypothetical reward in the previous chapters and therefore harder to detect. This could be so because effort type (real vs. hypothetical) might be processed in other brain areas than reward type and these areas may be less sensitive to biases.

A Decision regression with past trial outcomes

i All regressors

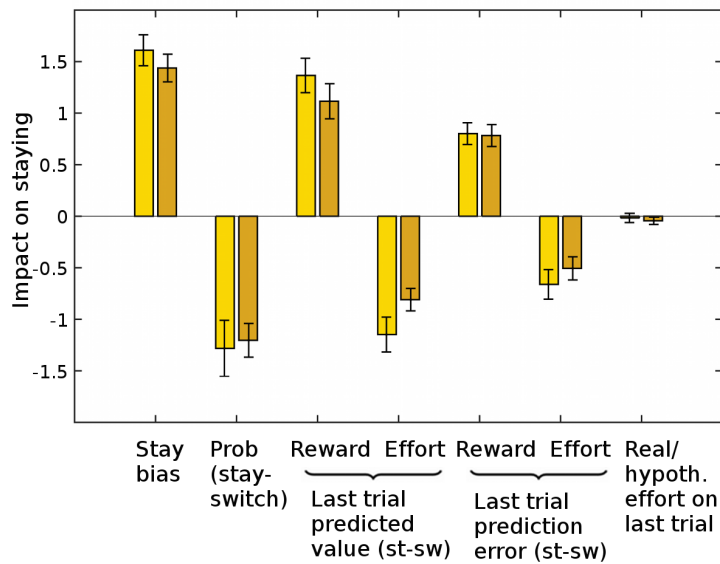


ii Average past outcomes



B Decision regression with predicted values

i Last trial predictions and PEs



ii Impact of real/hypothetical effort on next trial valuation

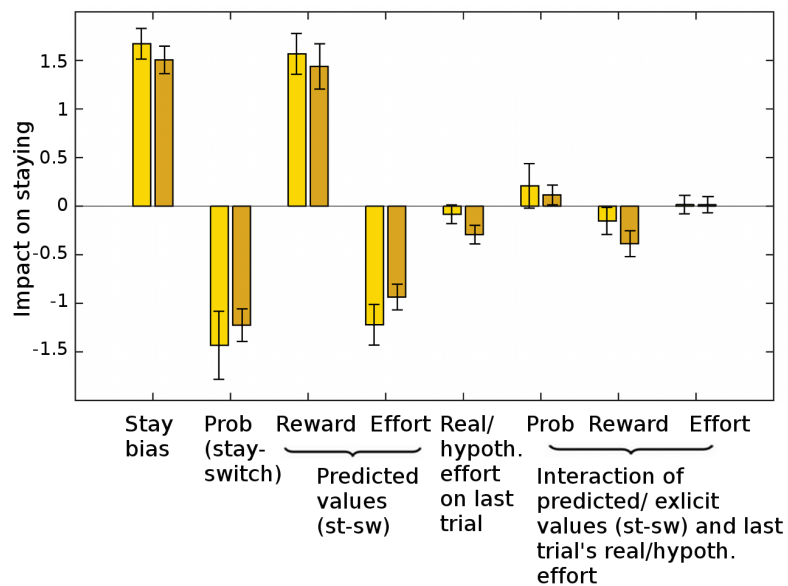


Figure 3. Decision data analysis. A) A regression analysis (dGLM1) showed that participants were more likely to stay with an option (i.e. chose it again) rather than switch to the alternative when relative (in favour of the 'stay' minus the 'switch' option) probability of effort shown at the time of choice was lower and when the reward/effort magnitude on the past one to four trials (t-1 to t-4) had been higher/lower (i). The dysphoric group used the learnt information less when deciding compared to the controls (ANOVA, main effect of group: $F(1)=4.79$, $p=0.036$), but did not differ in their speed of learning (ANOVA, group x time interaction: $F(3)=0.08$, $p=0.97$). This effect is shown again in ii) where the average regression weight of the learnt information (average of t-1 to t-4 for reward and (sign reversed) effort) is shown. B) A regression analysis (bGLM2) including the learnt dimensions as prediction errors (PEs) and predictions, based on a bayesian model, found that while dysphorics took the learnt predictions less into account when making decisions, but used the prediction error – i.e. learning signal – similarly. C) A regression analysis, again using predicted values from a Bayesian model, showed that across both groups, participants were more influenced by the learnt reward dimension when making their decision if they had not had to exert effort on the previous trial ($z=-2.4$, $p=0.016$).

Effort exertion behaviour

The effort exertion behaviour was analysed with the aim to understand whether dysphoria affects how effort is exerted and how this is modulated by motivational factors. Different from chapter 2, we could use data from both the session inside and outside the scanner for analysis of the execution phase as the trackball mouse measurements in the present study functioned with very high accuracy in both sessions. Participants needed on average to do the effort phase on 79.1 (inside the scanner) and on 79.8 (outside the scanner) trials out of 124. Out of these, they on average did not manage to complete the effort on 2.0% (inside the scanner) or 0.3% (outside the scanner) of trials. Dysphorics were more likely to not complete the effort successfully ($t(21.3)=2.42$, $p=0.024$; $99.4\pm0.2\%$ of trials completed successfully for controls and $98.3\pm0.4\%$ for dysphorics). The trials on which participants failed to complete the effort were excluded from further analyses. In the remaining trials, there was still a trend for the dysphorics to take longer to complete the effort phase ($t(32)=-1.864$, $p=0.071$; $12.9\pm0.3s$ for controls and $13.7\pm0.3s$ for dysphorics). Additionally, parametric correlations across all participants similarly showed that the time to complete the effort phase correlated positively with measures of depression (BDI: $p=0.042$, $r=0.35$; HRSD: $p=0.023$, $r=0.39$). Time to complete the effort phase also correlated with some measures of every day motivation, namely the HRSD item 7 ($p=0.0084$, $r=0.44$) and the Persistence score ($p=0.068$, $r=-0.32$). When either including group as a controlling factor in a

partial correlation, or performing the correlations only within the dysphoric group, the only correlation that remained significant was between time to complete the effort phase and HRSD item 7 ($p=0.047$, $r=0.49$; all others: $p>0.1$).

In the first analysis (eGLM1) we tested whether the reward or other valence dimensions affected motivational vigour, approximated by clicking rate (Fig. 4A). Different from the results in chapter 2, we failed to find such effects with either the reward of the chosen or the unchosen option affecting clicking rate (both $p>0.5$), or any group differences in this (both $p>0.3$). Similarly, we found no correlations between these measures of motivations with any questionnaire measurements (all $p>0.1$ for either parametric or non-parametric correlations).

In the second analysis (eGLM2), we tested whether the different valence aspects of the outcome, that should be irrelevant for executing the effort, influenced motivational trajectory planning (Fig. 4B), i.e. the average length of the path they took with the mouse between two targets. We found that across both groups together, this measure was modulated by the chosen and unchosen probabilities and reward magnitude outcomes (ANOVA, main effect of intercept: $F(1, 32)=18.363$, $p<0.001$): The higher the reward and the lower the probability of the chosen option (and vice versa for the unchosen option), the better participants were at planning efficient, i.e. shorter, distances between chosen targets. Qualitatively, this effect was present separately in both the data collected inside and outside the scanner. However, this effect differed between the two groups with the dysphorics showing this motivational modulation to a lesser extent (ANOVA, main effect of group: $F(1, 32)=4.897$, $p=0.034$).

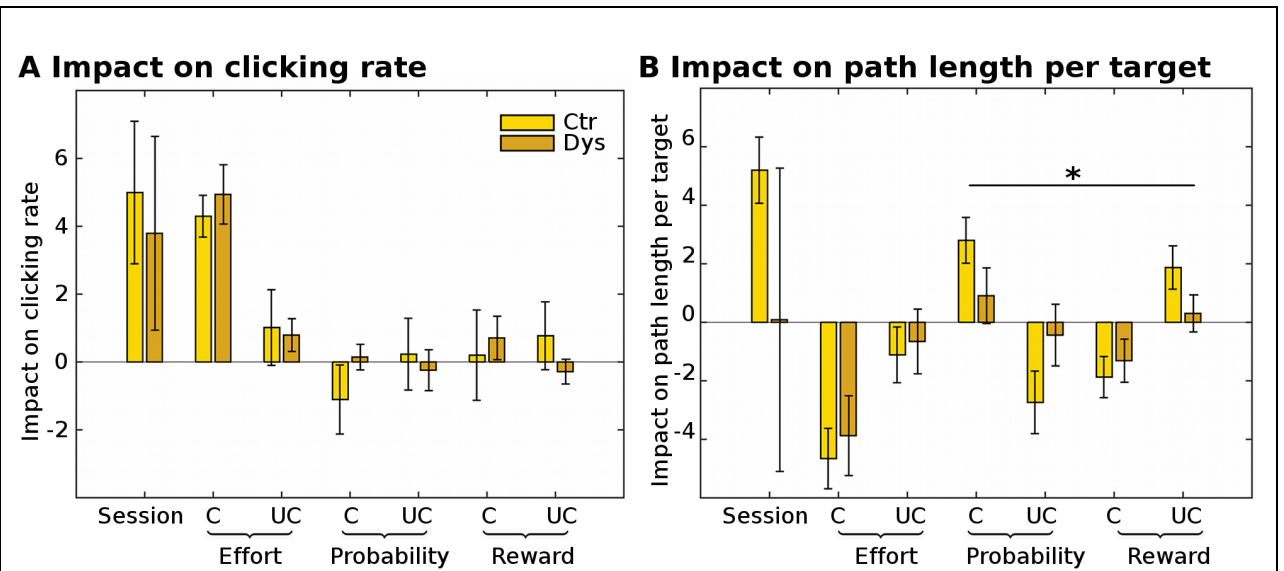


Figure 4. Effort exertion behaviour. *A)* Regression analysis of the effect of different motivational aspects on rate of mouse clicking. There was no motivational effect for reward outcomes of the chosen or the unchosen option. *B)* Regression analysis examining the effect of different motivational aspects on mouse movement trajectory planning (relative trajectory length per effort target). While controls were strongly affected by incidental motivational aspects of reward outcomes and effort probabilities of the chosen and unchosen options, this effect was reduced in the dysphoric group ($F(1)=4.897$, $p=0.034$). Note units of the y-axis are arbitrary

5.4 Discussion

For learnt information to guide behaviour adaptively it is important to be able to learn separately about different types of value information. Furthermore, when making choices the different dimensions of the learnt information need to be integrated. Depression has been associated specifically with impairments in reward learning (General introduction section 1.2.1). However, because value learning contains multiple aspects, it is still unclear whether the behavioural deficits seen in depression were due to impairment in the speed of learning, insensitivity to rewarding information (blunting) or impairment in using the learnt information to make decisions. To address this question, we tested dysphoric participants and healthy controls in a multi-dimensional learning task in which they learned concurrently about reward and effort magnitudes. Using this learned information as well as explicitly shown effort probability information, they repeatedly chose between two options.

Learning in depression

Using two separate model-based logistic regression analyses, we dissociated different components influencing decision making and learning. We found that while dysphoria did not affect the learning speed, it did reduce how much participants relied on the learnt information when making decisions.

This finding agrees with other studies reporting deficits in improving performance with practice across trials in depression (Cella et al., 2010; Herzallah et al., 2013; Mueller et al., 2015; Must et al., 2006; Pechtel et al., 2013; Remijnse et al., 2009; Taylor Tavares et al., 2008; Thoma et al., 2015; Vrieze et al., 2013). At the same time, our study is also in agreement with others that have not found *learning speed* – rather than other aspects of learning more generally – to be affected in depression (Gradin et al., 2011); Huys et al. (2013); Kumar et al. (2008); Kunisato et al. (2012). We however also note that on the other hand, some studies have found that depression

changes updating of predictions (Chase et al., 2010; Garrett et al., 2014; Korn et al., 2014). As discussed in detail in the General Introduction (section 1.2.1), these results may be specific to the updating of personal odds of negative life events in the light of information about population level risks specifically, rather than learning more generally.

Our results are particularly related to Huys et al. (2013) and Kunisato et al. (2012) who found that depression, or specifically the symptom anhedonia, was not related to a change in learning speed, but related to the sensitivity to reward or choice stochasticity. While sensitivity to rewarding (or punishing) information is conceptually different from overall choice stochasticity, previous studies have not dissociated between them. In our design however, we could differentiate between them because participants needed to learn about how rewarding and effortful two options were at the same time and integrate this with explicitly shown effort probabilities. We could therefore show that neither dysphoria, nor anhedonia more specifically, were associated with generally more stochastic decision making or behaviour, but instead dysphoria was specifically with a reduced sensitivity to learnt information that was either rewarding or punishing: while dysphoric participants used the learnt information less when making their decisions, they could use the explicitly shown information equally well.

There could be different reasons for the reduced use of learnt information. Firstly, it could be due to blunting, i.e. reduced sensitivity to different types, positive or negative, of valenced information that has been described in depression in a non-learning context before (Bylsma et al., 2008). However, if this was the case, it would not explain why dysphoric participants used the learnt information less to make choices, but not the, also valenced, explicitly shown information. Secondly, it could be that dysphoric participants were less motivated to do the task well and that this would particularly show in the most difficult component of the task, namely learning about the rewards and efforts, rather than just relying on the explicit information. If this

was true, then one could also have expected dysphoric participants to try avoiding other effortful (and thus requiring motivation) components of the task. However, dysphoric participants did not show a selectively increased sensitivity to avoiding the effort task.

Thirdly, it could be that dysphoric participants learn the rewarding and effortful information equally well, but thought that they did not, i.e. had reduced confidence in their learnt estimates (as suggested in other contexts (Fu et al., 2005; Spada et al., 2008; Yilmaz et al., 2011), but not universally (Sarisoy et al., 2014)). In this case, it would be reasonable to rely less on the learnt information compared to the explicit information, because they would perceive it as less reliable. To test more directly whether this was the case, future studies should measure confidence in learnt information more directly. Fourthly, it could be that, as suggested in other studies (Gradin et al., 2011; Kumar et al., 2008) our behavioural analyses were less sensitive to finding effects of learning speed than analyses of neural (fMRI) data could be. This could mean that there were actually learning effects that we could not detect with the behavioural analyses. To address this, fMRI data we acquired will be essential. It is noteworthy in this context that our design of learning reward/effort magnitudes, rather than probabilities, means that neural prediction error signals at the time of the outcome can be dissociated from outcome signals. This would allow answering the question more firmly which parts of the learning process are affected.

Additionally, we can use the fMRI data, to further understand difference between the groups in terms of effort and reward learning and choice, beyond what we can observe behaviourally.

Particularly, we can hopefully dissociate between neural activity which allowed the dysphoric group to 'function' in the paradigm compared to potentially distinct motivational activations in the controls.

Effort and motivation in depression

The effort that participants had to learn about was of mixed physical and cognitive nature. They had to move a mouse cursor (physical effort) to select targets on a computer screen. While doing so, they were free to plan and decide their trajectory and had to attend to different parts of space and the targets appearing i.e. (cognitive effort).

We found that dysphoric participants were more prone to failing to complete the effort task within the time limit and that symptoms of depression, and in particular every day motivation, were related to taking longer to complete the task. This is in agreement with a large body of previous research finding different kinds of motor retardation in depression (Buyukdura et al., 2011). It is theoretically possible that an increased risk of failing the task would make participants more sensitive to the effort dimension, despite such probabilities being generally very low. However, surprisingly, dysphoric participants were, if anything, less sensitive to the effort when making decisions. This result is furthermore in disagreement with some previous studies that found that depression or anhedonia was related to a decreased willingness to exert effort and a decreased willingness to exert more effort for more reward (which would correspond to an increase impact of the effort level on choice on our task) (Treadway et al., 2009; Treadway et al., 2012; Yang et al., 2014; Yang et al., 2015); though this result has not been found by all studies (Clery-Melin et al., 2011; Sherdell et al., 2012). This difference in results could be due to about 1/3 of our sample not reaching levels of clinical depression (though other studies that also failed to find effects of depression on effort decision behaviour (Clery-Melin et al., 2011; Sherdell et al., 2012) included MDD patients). Alternatively, differences could be due task differences (e.g. effort being embedded in a learning task, presenting options with varying levels of effort, rather than one default option) or our group sample consisting of large highly functioning university students.

This task also allowed us to measure both cognitive motivation, i.e. efficiency of participants' movements trajectories, and invigoration, i.e. speed of clicking. We found that the control group modulated their planning as a function of incidental motivational factors. When the relative reward and probability associated with their chosen option were more favourable, they optimized their movement trajectories to be shorter. This effect was absent in the dysphoric group. This suggests that in the domain of cognitive effort, depression is associated with a reduced sensitivity to incidental motivational stimuli. Again, we note that this was distinct from effects of the same dimensions on choice: the dimension of probability did not have less influence on the choices in the dysphoric group.

Surprisingly, while the questionnaire self-reports of the dysphoric participants about motivation in daily life suggested very marked motivational problems compared to controls, our measurements of effort-based decision-making and actual effort exertion behaviour only found relatively small differences. This could be because dysphoric participants' self-reports were exaggerated. Alternatively, the task used here did not tap into the same kind of motivation that is relevant for real-life behaviour, though we note that some features of the task, such as duration to complete the effort did correlate with self-report measures of real-life persistence. One striking difference from daily life is that in this study, participants *had to decide* on every trial between two concrete options. In contrast, in daily life, one may instead be faced with doing a default action – such as lying in bed – and then have first to motivate oneself that one should now decide whether to continue the default or engage in an effortful activity. Further studies, for example using designs inspired by the foraging literature (e.g. Constantino and Daw (2015), should look at this kind of situation in more detail.

Conclusion

Using a novel reward and effort learning task, we dissociated learning behaviour into different components. We found that while dysphoria did not affect the speed of learning, it reduced how much participants relied on the learnt information. This highlights the importance of dissociating different components of complex behaviours to understand which ones are affected by psychopathology.

6 General discussion

Depression is a multi-faceted illness with a diverse set of possible symptoms. Several of these symptoms are related to psychological aspects of reward or punishment processing, such as decision-making and learning. Using a computational approach, these concepts can be mathematically formalized, thus allowing measurements of specific cognitive and neural mechanisms. The experiments in this thesis then examined these processes in the context of depression and commonly used or putative treatments.

6.1 Neural and behavioural mechanisms of learning in complex, naturalistic environments

Experiments in chapters 2, 3 and 4 examined the basic cognitive and neural mechanisms of learning in complex, naturalistic environments. For this, I firstly designed novel tasks in which human participants needed to *concurrently* learn the reward and effort magnitude values of two options and chose between them based on this information. Learning in complex environments of this kind posed specific challenges for participants that went beyond those encountered in simpler environments in which for example only one kind of information needs to be tracked at any given time: Participants needed to simultaneously keep a large amount of information in mind and update or learn about it. Importantly, as part of this, they needed to learn about and remember the different kinds of information, i.e. about reward and effort, separately, rather than integrate them into one dimension such as ‘feeling of preference for one option’. Participants needed to then use this learnt information correctly, namely ensuring that they only based their decisions on relevant information, while at the same time ignoring irrelevant information that they might also be learning about. Furthermore, while choosing, it was also important to integrate information from the different dimensions using the appropriate strategy, e.g. representing value as a multiplicative interaction between reward magnitude and probability

rather than as a linear sum. In the experimental chapters 2 and 3, I showed that these different processes rely on different brain areas, as well being influenced by different neurotransmitter systems.

I found that participants were generally able to learn about this diverse information and use it for choices and that this learnt information was represented widely in neural networks that had previously been associated with learning and decision-making. However, I also found (chapter 2) that participants were biased by irrelevant reward information; they were biased towards choosing information that had in the past been paired with an irrelevant reward dimension. Neurally, I found that this bias was associated with processing in the ventral medial prefrontal cortex (vmPFC) and the amygdala. On the other hand, the ability to overcome this bias was associated with the lateral anterior prefrontal cortex (aPFC), another prefrontal area. Another way to look at the ‘ability to overcome a reward bias’ would be to call it ‘model-based’ behaviour: participants needed to use a model of the world – namely, only the relevant dimensions should be learnt while the irrelevant reward should not be learnt about and ignored for future decisions. It is of special note in this context that other studies in the past have identified the vmPFC with promoting model-based behaviour (see 1.2 General Introduction). The results of chapter 2 are therefore also important for highlighting that model-based behaviour is not a unitary process and instead, it might be more appropriate to consider each specific behaviour separately.

Another challenge of learning in naturalistic environment for participants was learning about each dimension of reward/punishment separately without being influenced by the other dimensions (chapter 3). I found that participants’ learning about reward information was interfered with by negative information present at the same time: the more negative the other information, the worse their learning about reward information. This effect could be due to

competition of the different dimensions to be learnt for limited processing resources. We found that this bias was ameliorated when the neural reward learning signals were enhanced by increasing serotonin levels (see below).

At the same time as the information needed to be kept separate during learning, it also needed to be integrated for making decisions. In other words, the good and bad of each option needed to be weighed up against each other to eventually produce one decision, namely choosing one option or the other. I found across all studies (chapters 2-5) that while participants took into account all information when making decisions, they did not integrate it optimally. Instead, they had a strong tendency to rely on a heuristic strategy, presumably for ease of implementation. Interestingly, I again found that this integration mechanism was specific and could be modulated selectively using a drug manipulation (NMDA study, chapter 4, see below).

In summary, chapters 2, 3 and 4 showed that successful learning behaviour in complex, naturalistic environments relies on a myriad of different cognitive and neural mechanisms that go beyond those observed in more simple environments. I also found that, because these mechanisms rely on different neural substrates, they could be modulated independently from each other using for example different pharmacological manipulations (discussed in more detail below). It will be important in the future to address whether these mechanisms are affected differently by diverse psychiatric disorders that involve deficits in reward guided learning or decision-making. If this were the case, it would suggest specific treatments tailored to these problems that could be most relevant for specific disorders.

One of the learning mechanisms that were not affected by the pharmacological manipulations used here was the ability to overcome the bias to repeat choices that had been associated with irrelevant rewards in the past, i.e. ignoring very salient information that is irrelevant for future

choices. Therefore, another future direction of research could be to examine whether this process can be modulated directly. As I found the lateral anterior prefrontal cortex to be particularly important in this process, a simple first step would be to modulate activity in this region. As this region lies on the surface of the brain, it is accessible to both electrical (transcranial direct current stimulation, tDCS) and magnetic (transcranial magnetic stimulation) brain stimulation. Therefore, an experiment could test firstly whether this region really plays a causal role. This would thus go beyond the correlational evidenced based on fMRI that was provided in this thesis. Secondly, an experiment could test whether it might even be possible to enhance this ability to overcome reward biases by enhancing activity in this area.

6.2 Serotonin and learning

The most common treatment for depression are selective serotonin re-uptake inhibitors (SSRIs). There are different theories of how they function that either highlight the role of serotonin in the coding of valence (how positive or negative something is) or in neural plasticity – the latter one based almost exclusively comes from animal studies (see 1.3 General Introduction). Experimental chapter 3 addressed the question whether firstly these presumed plasticity effects are found in humans when SSRIs are given at clinical relevant doses. Secondly, whether the plasticity effects lead to behavioural effects in behaviours relevant for depression, such as value based learning. And thirdly, whether the plasticity/ learning effects are simply secondary to changes in valuation of rewards or punishments as could be expect from a purely valenced-based view of SSRI action. For this, I measured the effects of *prolonged* (20mg citalopram/day for 2 weeks) manipulations of serotonin on behavioural and neural learning of both appetitive (monetary reward) and aversive (effort information) as a proxy for plasticity.

I found evidence for enhanced neural learning signals in both the appetitive and the aversive dimension. Concurrently, the behavioural results showed an improvement in reward learning

behaviour under interference. Thus this supports the view that prolonged increases in serotonin at clinically relevant doses can indeed improve neural and behavioural markers of learning. I also noted that this was in the absence to the neural coding of reward or effort receipts per se, thus suggesting that it is unlikely that the improvements in learning were secondary to changes in valence processing. These results should not be taken to mean that serotonin does not play a role in any valenced processing – there is also strong support for this view from other studies (see 1.3 General Introduction). Instead, these findings highlight that the function of a neurotransmitter such as serotonin can only be understood fully by studying it at different timescales as new functions that are not predicted from those on other timescales can emerge.

What is the clinical relevance of these findings? At first sight, the fact that neural learning signals about negative, and not just positive, information was enhanced might be a concern for the role of SSRIs in the treatment of depression. In this context, it has been proposed (see 1.3 General Introduction) that if SSRIs are given to treat depression they might have mostly beneficial effects on mood and depressive symptoms if the environment is positive – and new learning about how good the environment is can occur. But that at the same time, SSRIs might have mostly negative effects if the environment is negative – and thus new learning reinforces the negative views about the self and the environment further. However, it is also possible to look at this differently: As our experiment showed, better neural learning is not the same as faster learning. Instead, I found that better learning could mean stronger learning signals, i.e. learning signals that are more robust to interference. In other words, learning can be improved by allowing a clearer representation of the information to be learnt. In real life this could translate to having a clearer sense of the true contingencies in the environment. As an example, a person could in this way relearn correct predictions about negative outcomes such as that ‘feeling bad’ is not because of ‘being a guilty person’, but rather due to some very specific environmental event or internal thought. Whether our results really translate to this kind of real-life situation

needs to be borne out by further experiments. Another clinical aspect of our results is that learning was enhanced in general, which may be relevant for psychological therapies: it may be possible that the increased plasticity increases the speed of learning from therapy, which would suggest that giving SSRIs in combination with psychotherapy might also be relevant for other disorders that would not usually be treated with SSRIs.

As a future direction, it may be important to repeat the study with depressed, rather than just healthy control, participants. On one hand it seems likely that serotonin increase would lead to plasticity enhancements in this population, based on the robustness of the effect in healthy and depressed animals. On the other hand it is nevertheless possible that there may be differences, based on some studies suggesting that on a molecular level, some down-stream effects of serotonin may differ between depressed and healthy animals.

Lastly, in this study, I could dissociate learning from potential valence effects. While I did not find any valence shifts, there is also strong evidence (see 1.3 General Introduction) that prolonged administration of SSRIs shift the processing of social cues towards more positive, away from negative. As our study did not measure a social dimension, it cannot speak to how these effects relate to each other. It is well possible that both exist independently from each other and that in terms of clinical benefit, they interact with each other. It could be that social cues are perceived as more positive and that at the same time, relearning about these new more positive associations is speeded up. Future studies could investigate both phenomena in the same participants to determine how they relate.

While this study could answer the question that prolonged increases in serotonin can modulate plasticity related learning, it could not answer the question what about the plasticity is related to mood. This was for two reasons, firstly because the study used healthy volunteers, the SSRI did

not affect their mood. Secondly, because of the potential effects on social valenced processing, even if I had found an effect on mood, I would not have known whether they were really due to the plasticity. If it was the case that neural plasticity is directly related to mood (see 1.3.2 General Introduction for considerations why this might be so), then a medication that modulates plasticity without changing valenced processing should also lead to improvements in mood. I addressed this question in chapter 4 where I looked at another pharmacological manipulation of plasticity.

6.3 NMDA-receptors, learning and decision-making

In chapter 4, I sought to establish whether modulation of the activity of NMDA receptors could modulate learning and plasticity in humans. A large number of animal studies (see 1.4 General Introduction) have highlighted the importance of NMDA receptors in various kinds of learning. On the other hand, human studies have produced less consistent results, often failing to find an effect on learning. However, no human study had previously looked at its role in value-based learning, which might potentially be a more relevant kind of plasticity in depression, and thus for influencing mood, than other kinds of learning. Therefore, instead of measuring the effect of manipulating NMDA receptor activity on mood in depressed participants, I first thought to establish whether it could indeed modulate plasticity. For this I measured the effects of the partial NMDA agonist d-cycloserine (250mg, single dose) on concurrent reward (winning money) and punishment (losing money) learning. I found no evidence for any changes in learning for either reward or punishment. Similar, in the same participants, we also found no evidence for changes in motor sequence learning Günthner and Scholl et al. (in submission).

These results showed that d-cycloserine, at least at the dosage used here, is not well suited for studying whether a drug that modulates plasticity can modulate mood directly. We noted (see also 1.4 General Introduction) that one problem in using d-cycloserine in general is that while it acts as an agonist at low doses, which should enhance learning, its action reverses at higher

doses. It is unclear what the specific dose is for different brain regions at which this shift occurs. Therefore, future research might prefer to use another NMDA partial agonist, such as serine, that has more controlled properties.

The reason that I chose to use d-cycloserine however was that it has been found in several studies to increase the speed of different forms of cognitive behavioural therapy (CBT), which had been hypothesized to be due to its effects on learning and plasticity (Hofmann et al., 2015). Again, our results do not support this view. This however is not to say that d-cycloserine could not possibly speed up CBT (though there have been doubts about whether it in fact does speed up CBT). Instead, I suggest that it is possibly via a different mechanism. Based on theoretical considerations and some experimental data (see chapter 4.1 and 4.4), NMDA receptors have also been proposed to play a crucial role in other cognitive functions such as working memory or decision-making. Indeed, in chapter 4, I found that d-cycloserine improved how participants made decisions. Specifically, I found that participants were better at integrating the different dimensions that should influence decisions optimally, rather than relying on heuristic strategies. This was again further supported by related findings in a motor learning task (Günthner and Scholl et al. (in submission)), in which we found that d-cycloserine shifted how participants made decisions, making them decide more slowly, but also more accurately.

The importance of decision-making and the integration of evidence to make decision for psychotherapy have not been thoroughly investigated yet. It could be possible that DCS shifting a decision bias towards taking more time instead of ‘jumping to conclusions’ could be advantageous: DCS could help patients with anxiety disorders, who tend to decide too quickly that stimuli in the environment are threatening, to first reflect on the stimulus and then decide more accurately. As a consequence, it could be that patients can implement the instructions of psychotherapy more correctly and thus show enhancements by DCS. Future studies should

address this by measuring effects of DCS on decision-making in paradigms optimized for this purpose.

6.4 Depression and learning

In chapters 2-4, I identified different mechanisms of learning relying on anatomically and pharmacologically distinct processes. To know which – if any - of these is most relevant for depression, I lastly measured learning behaviour in dysphoric participants (chapter 5). I found that dysphoria did not affect how information about reward or effortful stimuli was learnt. Instead, depression affected how this information was used for learning: Dysphoric participants used the learnt information less compared to controls.

However, as a caveat to these results, it should be kept in mind, as also found in chapter 3 and previous studies (see 1.2 General Introduction), that neural learning signals can be more sensitive to subtle effects than behavioural measurements. This is potentially so because for any given behaviour, different brain areas are involved and code the same information redundantly (see chapters 2 and 3). Then even if learning processes might be affected in one brain area, another area may be able to compensate so that the overall behavioural output might not be affected. Therefore, it will be crucial in the future to analyse the neural data accompanying the behavioural data from chapter 5, to have more certainty that there really is no learning deficit in dysphoria.

This caveat aside, our results are not in support of a view that mild to moderate depression has a strong learning or plasticity component – at least in the domain of reward or effort guided learning measured here. Similarly, our results do not support the view that SSRIs help depression by restoring a plasticity deficit. This does not mean that increasing plasticity is not a clinically relevant function of SSRIs. Instead, it can be that increasing plasticity may be beneficial

because, as discussed above, it allows re-learning of faulty associations in depression or it could enhance the speed at which psychological therapy can lead to a re-learning of mental habits.

6.5 Technical implications

The findings of this thesis highlighted the advantage of using a computational approach to psychiatric questions. This approach allowed teasing apart different components of behaviour. Otherwise, this behaviour was just a series of binary choice on each trial, with many different cognitive and neural processes leading to a single output. However, the findings also showed difficulties with this approach: for complex designs, many trials, and therefore potentially long testing sessions are needed to acquire enough data. This could be challenging for more severely affected patient populations. In the future it may be worthwhile considering simulating data at the time of designing an experiment to find out the best ways to optimize paradigms to be at the same time sensitive to small differences in behaviour, but also efficient in the number of data required to measure these differences.

It is also of note that while there are advantages to adjusting experimental designs across specific studies, there can also be disadvantages. For example, using learning about monetary losses rather than about effort in chapter 4 allowed collecting more trials in a shorter time, which increased sensitivity of the behavioural analyses, mitigating the absence of neural data to further support conclusions. However, at the same time this meant that results were less comparable – it is unclear whether d-cycloserine might have had an effect on effort learning or whether citalopram might not have had an impact on neural learning signals for monetary losses. Similarly, while adjusting the learning paradigm in chapter 5 ensured that the task was not too effortful for the dysphoric participants to perform, it did not allow us anymore to look at the effects of real/ hypothetical reward on reward learning, which was the condition under which the behavioural effects of the serotonin manipulations were most obvious.

The present findings also underline the usefulness of neural data beyond its role in understanding what functions different brain areas have: Neural data can have a greater sensitivity than behavioural data, as detailed above. In chapter 3, I found that while a prolonged increase in serotonin levels only affected learning quite mildly, I could find larger neural differences in specific brain networks. Neural data can also answer questions about cognitive mechanisms that go beyond what can sometimes be revealed by modelling. In chapter 2, I found that using neural data I could determine the cognitive process of how reward biases are introduced and overcome more specifically.

6.6 Conclusions

In this thesis, I investigated neural and cognitive mechanisms of learning and plasticity in the context of depression and pharmacological treatments. In chapter 2, I found specific ways in which complex, naturalistic, learning recruits prefrontal brain circuits that are not engaged in the same way by more simple kinds of learning. Using the same paradigm, I also found (chapter 3) that a serotonergic antidepressant increased learning signals in the brain across both appetitive and aversive dimensions. On the other hand (chapter 4), d-cycloserine, a putative novel treatment for depression, did not lead to any improvement in reward or punishment learning. Finally (chapter 5), I compared these treatment effects to those found, in depression. Depression did not lead to a reduced speed of learning. However, depressed participants used the learnt information less than controls when making decisions. These results suggest that while depression and the most common treatment show some overlap in the processes they affect, this overlap is not perfect. Therefore, in the future, novel treatments should try to target those processes that were not affected by current treatments.

7. References

- Abler B, Seeringer A, Hartmann A, Gron G, Metzger C, Walter M, Stingl J (2011) Neural correlates of antidepressant-related sexual dysfunction: a placebo-controlled fMRI study on healthy males under subchronic paroxetine and bupropion. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 36:1837-1847.
- Abler B, Gron G, Hartmann A, Metzger C, Walter M (2012) Modulation of frontostriatal interaction aligns with reduced primary reward processing under serotonergic drugs. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32:1329-1335.
- Abramson LY, Seligman ME, Teasdale JD (1978) Learned helplessness in humans: critique and reformulation. *Journal of abnormal psychology* 87:49.
- Ahn WY, Krawitz A, Kim W, Busmeyer JR, Brown JW (2011) A Model-Based fMRI Analysis with Hierarchical Bayesian Parameter Estimation. *Journal of neuroscience, psychology, and economics* 4:95-110.
- Amo R, Fredes F, Kinoshita M, Aoki R, Aizawa H, Agetsuma M, Aoki T, Shiraki T, Kakinuma H, Matsuda M, Yamazaki M, Takahoko M, Tsuboi T, Higashijima S, Miyasaka N, Koide T, Yabuki Y, Yoshihara Y, Fukai T, Okamoto H (2014) The habenulo-raphe serotonergic circuit encodes an aversive expectation value essential for adaptive active avoidance of danger. *Neuron* 84:1034-1048.
- Anderson GM, Barr CS, Lindell S, Durham AC, Shifrovich I, Higley JD (2005) Time course of the effects of the serotonin-selective reuptake inhibitor sertraline on central and peripheral serotonin neurochemistry in the rhesus monkey. *Psychopharmacology* 178:339-346.
- Anderson IM, Richell RA, Bradshaw CM (2003) The effect of acute tryptophan depletion on probabilistic choice. *Journal of psychopharmacology* 17:3-7.
- Anderson IM, Del-Ben CM, McKie S, Richardson P, Williams SR, Elliott R, Deakin JF (2007) Citalopram modulation of neuronal responses to aversive face emotions: a functional MRI study. *Neuroreport* 18:1351-1355.
- Anderson IM, Juhasz G, Thomas E, Downey D, McKie S, Deakin JF, Elliott R (2011) The effect of acute citalopram on face emotion processing in remitted depression: a pharmacMRI study. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 21:140-148.
- Arce E, Simmons AN, Lovero KL, Stein MB, Paulus MP (2008) Escitalopram effects on insula and amygdala BOLD activation during emotional processing. *Psychopharmacology* 196:661-672.
- Artigas F, Romero L, de Montigny C, Blier P (1996) Acceleration of the effect of selected antidepressant drugs in major depression by 5-HT_{1A} antagonists. *Trends in neurosciences* 19:378-383.
- Bachatene L, Bharmuria V, Cattani S, Molotchnikoff S (2013) Fluoxetine and serotonin facilitate attractive-adaptation-induced orientation plasticity in adult cat visual cortex. *The European journal of neuroscience* 38:2065-2077.
- Baldinger P, Kranz GS, Haeusler D, Savli M, Spies M, Philippe C, Hahn A, Hoflich A, Wadsak W, Mitterhauser M, Lanzenberger R, Kasper S (2014) Regional differences in SERT occupancy after acute and prolonged SSRI intake investigated by brain PET. *NeuroImage* 88:252-262.
- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407-419.

- Bannerman DM, Bus T, Taylor A, Sanderson DJ, Schwarz I, Jensen V, Hvalby O, Rawlins JN, Seeburg PH, Sprengel R (2012) Dissecting spatial knowledge from spatial choice by hippocampal NMDA receptor deletion. *Nature neuroscience* 15:1153-1159.
- Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micah A, Dalley JW, Robbins TW (2010) Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 35:1290-1301.
- Baroncelli L, Sale A, Viegi A, Maya Vetencourt JF, De Pasquale R, Baldini S, Maffei L (2010) Experience-dependent reactivation of ocular dominance plasticity in the adult visual cortex. *Experimental neurology* 226:100-109.
- Batsikadze G, Paulus W, Kuo MF, Nitsche MA (2013) Effect of serotonin on paired associative stimulation-induced plasticity in the human motor cortex. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 38:2260-2267.
- Bavelier D, Levi DM, Li RW, Dan Y, Hensch TK (2010) Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:14964-14971.
- Beats BC, Sahakian BJ, Levy R (1996) Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. *Psychological medicine* 26:591-603.
- Beck AT, Steer RA, Brown GK (1996) Beck depression inventory-II. San Antonio.
- Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in FMRI. *NeuroImage* 20:1052-1063.
- Beckmann CF, Smith SM (2004) Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* 23:137-152.
- Beckmann M, Johansen-Berg H, Rushworth MF (2009) Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29:1175-1190.
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nature neuroscience* 10:1214-1221.
- Berger M, Gray JA, Roth BL (2009) The expanded biology of serotonin. *Annual review of medicine* 60:355-366.
- Berridge KC, Robinson TE, Aldridge JW (2009) Dissecting components of reward: 'liking', 'wanting', and learning. *Current opinion in pharmacology* 9:65-73.
- Bessa JM, Ferreira D, Melo I, Marques F, Cerqueira JJ, Palha JA, Almeida OF, Sousa N (2009) The mood-improving actions of antidepressants do not depend on neurogenesis but are associated with neuronal remodeling. *Molecular psychiatry* 14:764-773, 739.
- Beyer CE, Cremers TI (2008) Do selective serotonin reuptake inhibitors acutely increase frontal cortex levels of serotonin? *European journal of pharmacology* 580:350-354.
- Bhagwagar Z, Cowen PJ, Goodwin GM, Harmer CJ (2004) Normalization of enhanced fear recognition by acute SSRI treatment in subjects with a previous history of depression. *The American journal of psychiatry* 161:166.
- Binns KE, Salt TE (1996) Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus. *Journal of neurophysiology* 75:920-930.
- Blanco NJ, Otto AR, Maddox WT, Beevers CG, Love BC (2013) The influence of depression symptoms on exploratory decision-making. *Cognition* 129:563-568.
- Bohn I, Gierler C, Hauber W (2003) NMDA receptors in the rat orbital prefrontal cortex are involved in guidance of instrumental behaviour under reversal conditions. *Cerebral cortex* 13:968-976.
- Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (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.

- Boorman ED, Behrens TE, Rushworth MF (2011) Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS biology* 9:e1001093.
- Boorman ED, Rushworth MF, Behrens T (2013) Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multialternative choice. *Journal of neuroscience* 33:2242-2253.
- Boureau YL, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 36:74-97.
- Bouton ME, Vurbic D, Woods AM (2008) D-cycloserine facilitates context-specific fear extinction learning. *Neurobiology of learning and memory* 90:504-510.
- Branchi I (2011) The double edged sword of neural plasticity: increasing serotonin levels leads to both greater vulnerability to depression and improved capacity to recover. *Psychoneuroendocrinology* 36:339-351.
- Branchi I, Santarelli S, Capoccia S, Poggini S, D'Andrea I, Cirulli F, Alleva E (2013) Antidepressant treatment outcome depends on the quality of the living environment: a pre-clinical investigation in mice. *PloS one* 8:e62226.
- Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP (2008) The neural mechanisms underlying the influence of pavlovian cues on human decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:5861-5866.
- Brinkmann K, Gendolla GHE (2007) Dysphoria and Mobilization of Mental Effort: Effects on Cardiovascular Reactivity. *Motivation and Emotion* 31:71-82.
- Brinkmann K, Gendolla GH (2008) Does depression interfere with effort mobilization? Effects of dysphoria and task difficulty on cardiovascular response. *Journal of personality and social psychology* 94:146-157.
- Brinkmann K, Grept J, Gendolla GHE (2011) Dysphorics can control depressive mood's informational impact on effort mobilization. *Motivation and Emotion* 36:232-241.
- Browning M, Reid C, Cowen PJ, Goodwin GM, Harmer C (2007) A single dose of citalopram increases fear recognition in healthy subjects. *Journal of psychopharmacology*.
- Bruhl AB, Kaffenberger T, Herwig U (2010) Serotonergic and noradrenergic modulation of emotion processing by single dose antidepressants. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 35:521-533.
- Bruhl AB, Jancke L, Herwig U (2011) Differential modulation of emotion processing brain regions by noradrenergic and serotonergic antidepressants. *Psychopharmacology* 216:389-399.
- Burke CJ, Brunger C, Kahnt T, Park SQ, Tobler PN (2013) Neural integration of risk and effort costs by the frontal pole: only upon request. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:1706-1713a.
- Buyukdura JS, McClintock SM, Croarkin PE (2011) Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. *Progress in neuro-psychopharmacology & biological psychiatry* 35:395-409.
- Bylsma LM, Morris BH, Rottenberg J (2008) A meta-analysis of emotional reactivity in major depressive disorder. *Clinical psychology review* 28:676-691.
- Byrom NC, Msetfi RM, Murphy RA (2015) Two pathways to causal control: use and availability of information in the environment in people with and without signs of depression. *Acta psychologica* 157:1-12.
- Capitao LP, Murphy SE, Browning M, Cowen PJ, Harmer CJ (2015) Acute fluoxetine modulates emotional processing in young adult volunteers. *Psychological medicine* 45:2295-2308.
- Carmichael ST, Price JL (1995) Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *The Journal of comparative neurology* 363:615-641.
- Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology* 67:319.

- Castren E (2013) Neuronal network plasticity and recovery from depression. *JAMA psychiatry* 70:983-989.
- Cella M, Dymond S, Cooper A (2010) Impaired flexible decision-making in Major Depressive Disorder. *Journal of affective disorders* 124:207-210.
- Cerqueira JJ, Mailliet F, Almeida OF, Jay TM, Sousa N (2007) The prefrontal cortex as a key target of the maladaptive response to stress. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:2781-2787.
- Chamberlain SR, Muller U, Blackwell AD, Clark L, Robbins TW, Sahakian BJ (2006a) Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* 311:861-863.
- Chamberlain SR, Müller U, Blackwell AD, Clark L, Robbins TW, Sahakian BJ (2006b) Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* 311:861-863.
- Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW (2010) Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychological medicine* 40:433-440.
- Chase HW, Kumar P, Eickhoff SB, Dombrovski AY (2015) Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, affective & behavioral neuroscience* 15:435-459.
- Chau BK, Kolling N, Hunt LT, Walton ME, Rushworth MF (2014) A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature neuroscience* 17:463-470.
- Chau BK, Sallet J, Papageorgiou GK, Noonan MP, Bell AH, Walton ME, Rushworth MF (2015) Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. *Neuron* 87:1106-1118.
- Cherry KM, Lenze EJ, Lang CE (2014) Combining d-cycloserine with motor training does not result in improved general motor learning in neurologically intact people or in people with stroke. *Journal of neurophysiology* 111:2516-2524.
- Chiu PH, Lohrenz TM, Montague PR (2008) Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nature neuroscience* 11:514-520.
- Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC (2004) Cognitive inflexibility after prefrontal serotonin depletion. *Science* 304:878-880.
- Clarke HF, Walker SC, Crofts HS, Dalley JW, Robbins TW, Roberts AC (2005) Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25:532-538.
- Clarke HF, Walker SC, Dalley JW, Robbins TW, Roberts AC (2007) Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral cortex* 17:18-27.
- Clery-Melin ML, Schmidt L, Lafargue G, Baup N, Fossati P, Pessiglione M (2011) Why don't you try harder? An investigation of effort production in major depression. *PloS one* 6:e23178.
- Cloninger CR, Svrakic DM, Przybeck TR (1993) A psychobiological model of temperament and character. *Archives of general psychiatry* 50:975-990.
- Cohen JY, Amoroso MW, Uchida N (2015) Serotonergic neurons signal reward and punishment on multiple timescales. *eLife* 4.
- Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European journal of neuroscience* 35:1024-1035.
- Conrad CD (2010) A critical review of chronic stress effects on spatial learning and memory. *Progress in neuro-psychopharmacology & biological psychiatry* 34:742-755.
- Constantino SM, Daw ND (2015) Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*:1-17.

- Cools R, Blackwell A, Clark L, Menzies L, Cox S, Robbins TW (2005a) Tryptophan depletion disrupts the motivational guidance of goal-directed behavior as a function of trait impulsivity. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 30:1362-1373.
- Cools R, Calder AJ, Lawrence AD, Clark L, Bullmore E, Robbins TW (2005b) Individual differences in threat sensitivity predict serotonergic modulation of amygdala response to fearful faces. *Psychopharmacology* 180:670-679.
- Cools R, Robinson OJ, Sahakian B (2008) Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 33:2291-2299.
- Coull JT, Sahakian BJ, Middleton HC, Young AH, Park SB, McShane RH, Cowen PJ, Robbins TW (1995) Differential effects of clonidine, haloperidol, diazepam and tryptophan depletion on focused attention and attentional search. *Psychopharmacology* 121:222-230.
- Crockett MJ (2009) The neurochemistry of fairness: clarifying the link between serotonin and prosocial behavior. *Annals of the New York Academy of Sciences* 1167:76-86.
- Crockett MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29:11993-11999.
- Crockett MJ, Clark L, Apergis-Schoute AM, Morein-Zamir S, Robbins TW (2012a) Serotonin modulates the effects of Pavlovian aversive predictions on response vigor. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 37:2244-2252.
- Crockett MJ, Clark L, Roiser JP, Robinson OJ, Cools R, Chase HW, Ouden H, Apergis-Schoute A, Campbell-Meiklejohn D, Seymour B, Sahakian BJ, Rogers RD, Robbins TW (2012b) Converging evidence for central 5-HT effects in acute tryptophan depletion. *Molecular psychiatry* 17:121-123.
- Croxson PL, Walton ME, O'Reilly JX, Behrens TE, Rushworth MF (2009) Effort-based cost-benefit valuation and the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29:4531-4541.
- Cuijpers P (2004) Minor depression: risk profiles, functional disability, health care use and risk of developing major depression. *Journal of affective disorders* 79:71-79.
- David DJ, Samuels BA, Rainer Q, Wang JW, Marsteller D, Mendez I, Drew M, Craig DA, Guiard BP, Guilloux JP, Artymyshyn RP, Gardier AM, Gerald C, Antonijevic IA, Leonardo ED, Hen R (2009) Neurogenesis-dependent and -independent effects of fluoxetine in an animal model of anxiety/depression. *Neuron* 62:479-493.
- Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. *Neural Networks* 15:603-616.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876-879.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204-1215.
- Dayan P, Huys QJ (2009) Serotonin in affective control. *Annual review of neuroscience* 32:95-126.
- De Martino B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. *Science* 313:684-687.
- Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* 19:430-441.
- Del-Ben CM, Deakin JF, McKie S, Delvai NA, Williams SR, Elliott R, Dolan M, Anderson IM (2005) The effect of citalopram pretreatment on neuronal responses to neuropsychological tasks in normal volunteers: an FMRI study.

- Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 30:1724-1734.
- Denk F, Walton ME, Jennings KA, Sharp T, Rushworth MF, Bannerman DM (2005) Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology* 179:587-596.
- Derryberry D, Reed MA (2002) Anxiety-related attentional biases and their regulation by attentional control. *Journal of abnormal psychology* 111:225-236.
- Dickinson A (1985) Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences* 308:67-78.
- Diener C, Kuehner C, Brusniak W, Struve M, Flor H (2009) Effects of stressor controllability on psychophysiological, cognitive and behavioural responses in patients with major depression and dysthymia. *Psychological medicine* 39:77-86.
- Donoso M, Collins AG, Koechlin E (2014) Human cognition. *Foundations of human reasoning in the prefrontal cortex. Science* 344:1481-1486.
- Douglas KM, Porter RJ, Frampton CM, Gallagher P, Young AH (2009) Abnormal response to failure in unmedicated major depression. *Journal of affective disorders* 119:92-99.
- Dremencov E, El Mansari M, Blier P (2009) Effects of sustained serotonin reuptake inhibition on the firing of dopamine neurons in the rat ventral tegmental area. *Journal of psychiatry & neuroscience: JPN* 34:223.
- Duman RS, Li N, Liu RJ, Duric V, Aghajanian G (2012) Signaling pathways underlying the rapid antidepressant actions of ketamine. *Neuropharmacology* 62:35-41.
- Elliott R, Sahakian BJ, McKay AP, Herrod JJ, Robbins TW, Paykel ES (1996) Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychological medicine* 26:975-989.
- Eshel N, Roiser JP (2010) Reward and punishment processing in depression. *Biological psychiatry* 68:118-124.
- Evers EA, Cools R, Clark L, van der Veen FM, Jolles J, Sahakian BJ, Robbins TW (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 30:1138-1147.
- Evers EA, van der Veen FM, Jolles J, Deutz NE, Schmitt JA (2006a) Acute tryptophan depletion improves performance and modulates the BOLD response during a Stroop task in healthy females. *NeuroImage* 32:248-255.
- Evers EA, van der Veen FM, van Deursen JA, Schmitt JA, Deutz NE, Jolles J (2006b) The effect of acute tryptophan depletion on the BOLD response during performance monitoring and response inhibition in healthy male volunteers. *Psychopharmacology* 187:200-208.
- Eysenck HJ, Eysenck SBG (1975) Eysenck Personality Questionnaire (Junior & Adult): EdITS/Educational and Industrial Testing Service.
- Faulkner P, Deakin JF (2014) The role of serotonin in reward, punishment and behavioural inhibition in humans: insights from studies with acute tryptophan depletion. *Neuroscience and biobehavioral reviews* 46 Pt 3:365-378.
- Feld GB, Lange T, Gais S, Born J (2013) Sleep-dependent declarative memory consolidation--unaffected after blocking NMDA or AMPA receptors but enhanced by NMDA coagonist D-cycloserine. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 38:2688-2697.
- First MB (1996) User's guide for the structured clinical interview for DSM-IV axis I disorders: SCID-I: Biometrics Research Department, New York State Psychiatric Institute.
- Flavell CR, Barber DJ, Lee JL (2011) Behavioural memory reconsolidation of food and fear memories. *Nature communications* 2:504.
- Floresco SB, Tse MT, Ghods-Sharifi S (2008) Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 33:1966-1979.

- Fonseca MS, Murakami M, Mainen ZF (2015) Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Current biology* : CB 25:306-315.
- Franken IH, Rassin E, Muris P (2007) The assessment of anhedonia in clinical and non-clinical populations: further validation of the Snaith-Hamilton Pleasure Scale (SHAPS). *Journal of affective disorders* 99:83-89.
- Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RS (1990) The relationship between global and local changes in PET scans. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* 10:458-466.
- Fu T, Koutstaal W, Fu CHY, Poon L, Cleare AJ (2005) Depression, Confidence, and Decision: Evidence Against Depressive Realism. *Journal of Psychopathology and Behavioral Assessment* 27:243-252.
- Garrett N, Sharot T, Faulkner P, Korn CW, Roiser JP, Dolan RJ (2014) Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. *Frontiers in human neuroscience* 8:639.
- Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience and biobehavioral reviews* 37:1297-1310.
- Gerdelat-Mas A, Loubinoux I, Tombari D, Rascol O, Chollet F, Simonetta-Moreau M (2005) Chronic administration of selective serotonin reuptake inhibitor (SSRI) paroxetine modulates human motor cortex excitability in healthy subjects. *NeuroImage* 27:314-322.
- Geurts DE, Huys QJ, den Ouden HE, Cools R (2013a) Aversive Pavlovian control of instrumental behavior in humans. *Journal of cognitive neuroscience* 25:1428-1441.
- Geurts DE, Huys QJ, den Ouden HE, Cools R (2013b) Serotonin and aversive Pavlovian control of instrumental behavior in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:18932-18939.
- Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595.
- Godlewska BR, Near J, Cowen PJ (2015) Neurochemistry of major depression: a study using magnetic resonance spectroscopy. *Psychopharmacology* 232:501-507.
- Golden GJ, Houpt TA (2007) NMDA receptor in conditioned flavor-taste preference learning: blockade by MK-801 and enhancement by D-cycloserine. *Pharmacology, biochemistry, and behavior* 86:587-596.
- Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, Reid I, Hall J, Steele JD (2011) Expected value and prediction error abnormalities in depression and schizophrenia. *Brain : a journal of neurology* 134:1751-1764.
- Grillon C, Chavis C, Covington MF, Pine DS (2009) Two-week treatment with the selective serotonin reuptake inhibitor citalopram reduces contextual anxiety but not cued fear in healthy volunteers: a fear-potentiated startle study. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 34:964-971.
- Groenewold NA, Opmeer EM, de Jonge P, Aleman A, Costafreda SG (2013) Emotional valence modulates brain functional abnormalities in depression: evidence from a meta-analysis of fMRI studies. *Neuroscience and biobehavioral reviews* 37:152-163.
- Guastella AJ, Lovibond PF, Dadds MR, Mitchell P, Richardson R (2007) A randomized controlled trial of the effect of D-cycloserine on extinction and fear conditioning in humans. *Behaviour research and therapy* 45:663-672.
- Guitart-Masip M, Economides M, Huys QJ, Frank MJ, Chowdhury R, Duzel E, Dayan P, Dolan RJ (2014) Differential, but not opponent, effects of L -DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology* 231:955-966.
- Gunthner J, Scholl J, Favaron E, Harmer CJ, Johansen-Berg H, Reinecke A (in submission) The partial NMDA agonist d-cycloserine does not enhance motor learning but shifts the speed-accuracy trade-off.

- Hamilton JP, Etkin A, Furman DJ, Lemus MG, Johnson RF, Gotlib IH (2012) Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of baseline activation and neural response data. *The American journal of psychiatry* 169:693.
- Hamilton M (1960) A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry* 23:56.
- Hammen C (2005) Stress and depression. *Annual review of clinical psychology* 1:293-319.
- Harmer CJ, Bhagwagar Z, Perrett DI, Vollm BA, Cowen PJ, Goodwin GM (2003a) Acute SSRI administration affects the processing of social cues in healthy volunteers. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 28:148-152.
- Harmer CJ, Rogers RD, Tunbridge E, Cowen PJ, Goodwin GM (2003b) Tryptophan depletion decreases the recognition of fear in female volunteers. *Psychopharmacology* 167:411-417.
- Harmer CJ, Shelley NC, Cowen PJ, Goodwin GM (2004) Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *The American journal of psychiatry* 161:1256-1263.
- Harmer CJ, Mackay CE, Reid CB, Cowen PJ, Goodwin GM (2006) Antidepressant drug treatment modifies the neural processing of nonconscious threat cues. *Biological psychiatry* 59:816-820.
- Harmer CJ, Goodwin GM, Cowen PJ (2009) Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *The British journal of psychiatry : the journal of mental science* 195:102-108.
- Harmer CJ, Cowen PJ (2013) 'It's the way that you look at it'--a cognitive neuropsychological account of SSRI action in depression. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 368:20120407.
- Hayward G, Goodwin GM, Cowen PJ, Harmer CJ (2005) Low-dose tryptophan depletion in recovered depressed patients induces changes in cognitive processing without depressive symptoms. *Biological psychiatry* 57:517-524.
- Heresco-Levy U, Javitt DC, Gelfin Y, Gorelik E, Bar M, Blanzaru M, Kremer I (2006) Controlled trial of D-cycloserine adjuvant therapy for treatment-resistant major depressive disorder. *Journal of affective disorders* 93:239-243.
- Heresco-Levy U, Gelfin G, Bloch B, Levin R, Edelman S, Javitt DC, Kremer I (2013) A randomized add-on trial of high-dose D-cycloserine for treatment-resistant depression. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum* 16:501-506.
- Herrero JL, Gieselmann MA, Sanayei M, Thiele A (2013) Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* 78:729-739.
- Hershberger WA (1986) An approach through the looking-glass. *Animal Learning & Behavior* 14:443-451.
- Herzallah MM, Moustafa AA, Natsheh JY, Abdellatif SM, Taha MB, Tayem YI, Sehwal MA, Amleh I, Petrides G, Myers CE, Gluck MA (2013) Learning from negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. *Frontiers in integrative neuroscience* 7:67.
- Hofmann SG, Wu JQ, Boettcher H (2013) D-Cycloserine as an augmentation strategy for cognitive behavioral therapy of anxiety disorders. *Biol Mood Anxiety Disord* 3:1-10.
- Hofmann SG, Otto MW, Pollack MH, Smits JA (2015) D-cycloserine augmentation of cognitive behavioral therapy for anxiety disorders: an update. *Current psychiatry reports* 17:532.
- Holderbach R, Clark K, Moreau JL, Bischofberger J, Normann C (2007) Enhanced long-term synaptic depression in an animal model of depression. *Biological psychiatry* 62:92-100.

- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. *Nature neuroscience* 15:470-476, S471-473.
- Huys QJM, Pizzagalli DA, Bogdan R, Dayan P (2013) Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord* 3:12.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) Fsl. *NeuroImage* 62:782-790.
- Jitsuki S, Takemoto K, Kawasaki T, Tada H, Takahashi A, Becamel C, Sano A, Yuzaki M, Zukin RS, Ziff EB, Kessels HW, Takahashi T (2011) Serotonin mediates cross-modal reorganization of cortical circuits. *Neuron* 69:780-792.
- Jocham G, Hunt LT, Near J, Behrens TE (2012) A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature neuroscience* 15:960-961.
- Kalisch R, Holt B, Petrovic P, De Martino B, Kloppel S, Buchel C, Dolan RJ (2009) The NMDA agonist D-cycloserine facilitates fear memory consolidation in humans. *Cerebral cortex* 19:187-196.
- Kamboj SK, Massey-Chase R, Rodney L, Das R, Almahdi B, Curran HV, Morgan CJ (2011) Changes in cue reactivity and attentional bias following experimental cue exposure and response prevention: a laboratory study of the effects of D-cycloserine in heavy drinkers. *Psychopharmacology* 217:25-37.
- Kang HJ, Voleti B, Hajsan T, Rajkowska G, Stockmeier CA, Licznarski P, Lepack A, Majik MS, Jeong LS, Banasr M, Son H, Duman RS (2012) Decreased expression of synapse-related genes and loss of synapses in major depressive disorder. *Nature medicine* 18:1413-1417.
- Karpova NN, Pickenhagen A, Lindholm J, Tiraboschi E, Kuleskaya N, Ágústssdóttir A, Antila H, Popova D, Akamine Y, Sullivan R (2011) Fear erasure in mice requires synergy between antidepressant drugs and extinction training. *Science* 334:1731-1734.
- Kelly RE, Jr., Alexopoulos GS, Wang Z, Gunning FM, Murphy CF, Morimoto SS, Kanellopoulos D, Jia Z, Lim KO, Hoptman MJ (2010) Visual inspection of independent components: defining a procedure for artifact removal from fMRI data. *Journal of neuroscience methods* 189:233-245.
- Kemp AH, Gray MA, Silberstein RB, Armstrong SM, Nathan PJ (2004) Augmentation of serotonin enhances pleasant and suppresses unpleasant cortical electrophysiological responses to visual emotional stimuli in humans. *NeuroImage* 22:1084-1096.
- Klumpers F, Denys D, Kenemans JL, Grillon C, van der Aart J, Baas JM (2012) Testing the effects of Delta9-THC and D-cycloserine on extinction of conditioned fear in humans. *Journal of psychopharmacology* 26:471-478.
- Knutson B, Wolkowitz OM, Cole SW, Chan T, Moore EA, Johnson RC, Terpstra J, Turner RA, Reus VI (1998) Selective Alteration of Personality and Social Behavior by Serotonergic Intervention. *The American journal of psychiatry* 155:373-379.
- Kobayashi K, Ikeda Y, Sakai A, Yamasaki N, Haneda E, Miyakawa T, Suzuki H (2010) Reversal of hippocampal neuronal maturation by serotonergic antidepressants. *Proceedings of the National Academy of Sciences of the United States of America* 107:8434-8439.
- Kolling N, Behrens TE, Mars RB, Rushworth MF (2012) Neural mechanisms of foraging. *Science* 336:95-98.
- Kolling N, Wittmann M, Rushworth MF (2014) Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* 81:1190-1202.
- Korn CW, Sharot T, Walter H, Heekeren HR, Dolan RJ (2014) Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological medicine* 44:579-592.
- Kroenke K, Spitzer RL, Williams JBW (2001) The Phq - 9. *Journal of general internal medicine* 16:606-613.

- Krystal JH, Sanacora G, Duman RS (2013) Rapid-acting glutamatergic antidepressants: the path to ketamine and beyond. *Biological psychiatry* 73:1133-1141.
- Kumar P, Waiter G, Ahearn T, Milders M, Reid I, Steele JD (2008) Abnormal temporal difference reward-learning signals in major depression. *Brain : a journal of neurology* 131:2084-2093.
- Kunisato Y, Okamoto Y, Ueda K, Onoda K, Okada G, Yoshimura S, Suzuki S, Samejima K, Yamawaki S (2012) Effects of depression on reward-based decision making and variability of action in probabilistic learning. *Journal of behavior therapy and experimental psychiatry* 43:1088-1094.
- Kuo MF, Unger M, Liebetanz D, Lang N, Tergau F, Paulus W, Nitsche MA (2008) Limited impact of homeostatic plasticity on motor learning in humans. *Neuropsychologia* 46:2122-2128.
- Kuriyama K, Honma M, Koyama S, Kim Y (2011a) D-cycloserine facilitates procedural learning but not declarative learning in healthy humans: a randomized controlled trial of the effect of D-cycloserine and valproic acid on overnight properties in the performance of non-emotional memory tasks. *Neurobiology of learning and memory* 95:505-509.
- Kuriyama K, Honma M, Soshi T, Fujii T, Kim Y (2011b) Effect of D-cycloserine and valproic acid on the extinction of reinstated fear-conditioned responses and habituation of fear conditioning in healthy humans: a randomized controlled trial. *Psychopharmacology* 218:589-597.
- Kurniawan IT, Guitart-Masip M, Dayan P, Dolan RJ (2013) Effort and valuation in the brain: the effects of anticipation and execution. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:6160-6169.
- Lanthorn TH (1994) D-cycloserine: agonist turned antagonist. *Amino Acids* 6:247-260.
- Lee MD, Wagenmakers E-J (2014) *Bayesian cognitive modeling: A practical course*: Cambridge University Press.
- Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81:687-699.
- Li N, Lee B, Liu R-J, Banasr M, Dwyer JM, Iwata M, Li X-Y, Aghajanian G, Duman RS (2010) mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* 329:959-964.
- Li N, Liu RJ, Dwyer JM, Banasr M, Lee B, Son H, Li XY, Aghajanian G, Duman RS (2011) Glutamate N-methyl-D-aspartate receptor antagonists rapidly reverse behavioral and synaptic deficits caused by chronic stress exposure. *Biological psychiatry* 69:754-761.
- Lisman J, Grace AA, Duzel E (2011) A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in neurosciences* 34:536-547.
- Liu Z, Zhou J, Li Y, Hu F, Lu Y, Ma M, Feng Q, Zhang JE, Wang D, Zeng J, Bao J, Kim JY, Chen ZF, El Mestikawy S, Luo M (2014) Dorsal raphe neurons signal reward through 5-HT and glutamate. *Neuron* 81:1360-1374.
- Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America* 104:9493-9498.
- Macoveanu J, Rowe JB, Hornboll B, Elliott R, Paulson OB, Knudsen GM, Siebner HR (2013) Playing it safe but losing anyway--serotonergic signaling of negative outcomes in dorsomedial prefrontal cortex in the context of risk-aversion. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 23:919-930.
- Macoveanu J, Fisher PM, Haahr ME, Frokjaer VG, Knudsen GM, Siebner HR (2014) Effects of selective serotonin reuptake inhibition on neural activity related to risky decisions and monetary rewards in healthy males. *NeuroImage* 99:434-442.
- Malberg JE, Eisch AJ, Nestler EJ, Duman RS (2000) Chronic antidepressant treatment increases neurogenesis in adult rat hippocampus. *The Journal of Neuroscience* 20:9104-9110.

- Malenka RC, Bear MF (2004) LTP and LTD: an embarrassment of riches. *Neuron* 44:5-21.
- Mannie ZN, Williams C, Browning M, Cowen PJ (2014) Decision making in young people at familial risk of depression. *Psychological medicine* 45:375-380.
- Mars RB, Jbabdi S, Sallet J, O'Reilly JX, Croxson PL, Olivier E, Noonan MP, Bergmann C, Mitchell AS, Baxter MG, Behrens TE, Johansen-Berg H, Tomassini V, Miller KL, Rushworth MF (2011) Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:4087-4100.
- Marutani T, Yahata N, Ikeda Y, Ito T, Yamamoto M, Matsuura M, Matsushima E, Okubo Y, Suzuki H, Matsuda T (2011) Functional magnetic resonance imaging study on the effects of acute single administration of paroxetine on motivation-related brain activity. *Psychiatry and clinical neurosciences* 65:191-198.
- Matthews SC, Simmons AN, Strigo IA, Arce E, Stein MB, Paulus MP (2010) Escitalopram attenuates posterior cingulate activity during self-evaluation in healthy volunteers. *Psychiatry research* 182:81-87.
- McCabe C, Mishor Z, Cowen PJ, Harmer CJ (2010) Diminished neural processing of aversive and rewarding stimuli during selective serotonin reuptake inhibitor treatment. *Biological psychiatry* 67:439-445.
- McGirr A, Berlim MT, Bond DJ, Fleck MP, Yatham LN, Lam RW (2015) A systematic review and meta-analysis of randomized, double-blind, placebo-controlled trials of ketamine in the rapid treatment of major depressive episodes. *Psychological medicine* 45:693-704.
- Meuwese JD, van Loon AM, Scholte HS, Lirk PB, Vulink NC, Hollmann MW, Lamme VA (2013) NMDA receptor antagonist ketamine impairs feature integration in visual perception. *PloS one* 8:e79326.
- Mineka S, Watson D, Clark LA (1998) Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology* 49:377-412.
- Miyazaki K, Miyazaki KW, Doya K (2011) Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:469-479.
- Miyazaki KW, Miyazaki K, Tanaka KF, Yamanaka A, Takahashi A, Tabuchi S, Doya K (2014) Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Current biology : CB* 24:2033-2040.
- Morris RG (2013) NMDA receptors and memory encoding. *Neuropharmacology* 74:32-40.
- Mueller EM, Pechtel P, Cohen AL, Douglas SR, Pizzagalli DA (2015) Potentiated processing of negative feedback in depression is attenuated by anhedonia. *Depression and anxiety* 32:296-305.
- Munafò MR, Hayward G, Harmer C (2006) Selective processing of social threat cues following acute tryptophan depletion. *Journal of psychopharmacology* 20:33-39.
- Murphy FC, Rubinsztein JS, Michael A, Rogers RD, Robbins TW, Paykel ES, Sahakian BJ (2001) Decision-making cognition in mania and depression. *Psychological medicine* 31:679-693.
- Murphy FC, Smith KA, Cowen PJ, Robbins TW, Sahakian BJ (2002) The effects of tryptophan depletion on cognitive and affective processing in healthy volunteers. *Psychopharmacology* 163:42-53.
- Murphy SE, Norbury R, O'Sullivan U, Cowen PJ, Harmer CJ (2009) Effect of a single dose of citalopram on amygdala response to emotional faces. *The British journal of psychiatry : the journal of mental science* 194:535-540.
- Must A, Szabo Z, Bodi N, Szasz A, Janka Z, Keri S (2006) Sensitivity to reward and punishment and the prefrontal cortex in major depression. *Journal of affective disorders* 90:209-215.

- Neubert FX, Mars RB, Sallet J, Rushworth MFS (in press) Connectivity profiles reveal relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex.
- Nilsson H, Rieskamp J, Wagenmakers E-J (2011) Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology* 55:84-93.
- Nitsche MA, Jaussi W, Liebetanz D, Lang N, Tergau F, Paulus W (2004) Consolidation of human motor cortical neuroplasticity by D-cycloserine. *Neuropsychopharmacology* : official publication of the American College of Neuropsychopharmacology 29:1573-1578.
- Noonan MP, Walton ME, Behrens TEJ, Sallet J, Buckley MJ, Rushworth MFS (2010) Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences* 107:20547-20552.
- Noonan MP, Kolling N, Walton ME, Rushworth MF (2012) Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *The European journal of neuroscience* 35:997-1010.
- Norbury R, Taylor MJ, Selvaraj S, Murphy SE, Harmer CJ, Cowen PJ (2009) Short-term antidepressant treatment modulates amygdala response to happy faces. *Psychopharmacology* 206:197-204.
- Normann C, Schmitz D, Furmaier A, Doing C, Bach M (2007) Long-term plasticity of visually evoked potentials in humans is altered in major depression. *Biological psychiatry* 62:373-380.
- O'Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H (2012) Tools of the trade: psychophysiological interactions and functional connectivity. *Social cognitive and affective neuroscience* 7:604-609.
- Onur OA, Schlaepfer TE, Kukulja J, Bauer A, Jeung H, Patin A, Otte DM, Shah NJ, Maier W, Kendrick KM, Fink GR, Hurlmann R (2010) The N-methyl-D-aspartate receptor co-agonist D-cycloserine facilitates declarative learning and hippocampal activity in humans. *Biological psychiatry* 67:1205-1211.
- Ori R, Amos T, Bergman H, Soares-Weiser K, Ipser JC, Stein DJ (2015) Augmentation of cognitive and behavioural therapies (CBT) with d-cycloserine for anxiety and related disorders. *The Cochrane database of systematic reviews* 5:CD007803-CD007803.
- Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND (2013) Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America* 110:20941-20946.
- Outhred T, Hawkshead BE, Wager TD, Das P, Malhi GS, Kemp AH (2013) Acute neural effects of selective serotonin reuptake inhibitors versus noradrenaline reuptake inhibitors on emotion processing: Implications for differential treatment efficacy. *Neuroscience and biobehavioral reviews* 37:1786-1800.
- Outhred T, Das P, Felmingham KL, Bryant RA, Nathan PJ, Malhi GS, Kemp AH (2014) Impact of acute administration of escitalopram on the processing of emotional and neutral images: a randomized crossover fMRI study of healthy women. *Journal of psychiatry & neuroscience* : JPN 39:267-275.
- Palminteri S, Clair AH, Mallet L, Pessiglione M (2012) Similar improvement of reward and punishment learning by serotonin reuptake inhibitors in obsessive-compulsive disorder. *Biological psychiatry* 72:244-250.
- Patel DS, Sharma N, Patel MC, Patel BN, Shrivastav PS, Sanyal M (2011) Development and validation of a selective and sensitive LC-MS/MS method for determination of cycloserine in human plasma: application to bioequivalence study. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 879:2265-2273.
- Pechtel P, Dutra SJ, Goetz EL, Pizzagalli DA (2013) Blunted reward responsiveness in remitted depression. *Journal of psychiatric research* 47:1864-1869.

- Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau H, Dolan RJ, Frith CD (2007) How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* 316:904-906.
- Pittenger C, Duman RS (2008) Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropsychopharmacology* : official publication of the American College of Neuropsychopharmacology 33:88-109.
- Pizzagalli DA, Jahn AL, O'Shea JP (2005) Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological psychiatry* 57:319-327.
- Plummer M (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*, p 125: Vienna.
- Portero-Tresserra M, Marti-Nicolovius M, Guillazo-Blanch G, Boadas-Vaello P, Vale-Martinez A (2013) D-cycloserine in the basolateral amygdala prevents extinction and enhances reconsolidation of odor-reward associative learning in rats. *Neurobiology of learning and memory* 100:1-11.
- Preacher KJ, Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers* 36:717-731.
- Prevost C, Pessiglione M, Metereau E, Clery-Melin ML, Dreher JC (2010) Separate valuation subsystems for delay and effort decision costs. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:14080-14090.
- Prevost C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural correlates of specific and general Pavlovian-to-Instrumental Transfer within human amygdalar subregions: a high-resolution fMRI study. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32:8383-8390.
- Price KL, Baker NL, McRae-Clark AL, Saladin ME, Desantis SM, Santa Ana EJ, Brady KT (2013) A randomized, placebo-controlled laboratory study of the effects of D-cycloserine on craving in cocaine-dependent individuals. *Psychopharmacology* 226:739-746.
- Pryce CR, Azzinnari D, Spinelli S, Seifritz E, Tegethoff M, Meinlschmidt G (2011) Helplessness: a systematic translational review of theory and evidence for its relevance to understanding and treating depression. *Pharmacology & therapeutics* 132:242-267.
- Quan M, Zheng C, Zhang N, Han D, Tian Y, Zhang T, Yang Z (2011) Impairments of behavior, information flow between thalamus and cortex, and prefrontal cortical synaptic plasticity in an animal model of depression. *Brain research bulletin* 85:109-116.
- Raleigh MJ, McGuire MT, Brammer GL, Pollack DB, Yuwiler A (1991) Serotonergic mechanisms promote dominance acquisition in adult male vervet monkeys. *Brain research* 559:181-190.
- Rangel A, Hare T (2010) Neural computations associated with goal-directed choice. *Current opinion in neurobiology* 20:262-270.
- Remijnse PL, Nielen MM, van Balkom AJ, Hendriks GJ, Hoogendijk WJ, Uylings HB, Veltman DJ (2009) Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychological medicine* 39:1503-1518.
- Rescorla RA, Wagner AR (1972) *Classical conditioning 2: current research and theory*. Appleton Century-Crofts, New York:64-69.
- Robinson OJ, Frank MJ, Sahakian BJ, Cools R (2010) Dissociable responses to punishment in distinct striatal regions during reversal learning. *NeuroImage* 51:1459-1467.
- Robinson OJ, Cools R, Carlisi CO, Sahakian BJ, Drevets WC (2012a) Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *American Journal of Psychiatry* 169:152-159.
- Robinson OJ, Cools R, Sahakian BJ (2012b) Tryptophan depletion disinhibits punishment but not reward prediction: implications for resilience. *Psychopharmacology* 219:599-605.

- Robinson OJ, Overstreet C, Allen PS, Letkiewicz A, Vytal K, Pine DS, Grillon C (2013) The role of serotonin in the neurocircuitry of negative affective bias: serotonergic modulation of the dorsal medial prefrontal-amygdala 'aversive amplification' circuit. *NeuroImage* 78:217-223.
- Rogers BP, Morgan VL, Newton AT, Gore JC (2007) Assessing functional connectivity in the human brain by fMRI. *Magnetic resonance imaging* 25:1347-1357.
- Rogers RD, Everitt BJ, Baldacchino A, Blackshaw AJ, Swainson R, Wynne K, Baker NB, Hunter J, Carthy T, Booker E (1999) Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 20:322-339.
- Roiser JP, de Martino B, Tan GC, Kumaran D, Seymour B, Wood NW, Dolan RJ (2009) A genetically mediated bias in decision making driven by failure of amygdala control. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29:5985-5991.
- Romer Thomsen K, Whybrow PC, Kringelbach ML (2015) Reconceptualizing anhedonia: novel perspectives on balancing the pleasure networks in the human brain. *Frontiers in behavioral neuroscience* 9:49.
- Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MF (2006) Separate neural pathways process different decision costs. *Nature neuroscience* 9:1161-1168.
- Rushworth MF, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Current opinion in neurobiology*.
- Rutledge RB, Dean M, Caplin A, Glimcher PW (2010) Testing the reward prediction error hypothesis with an axiomatic model. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:13525-13536.
- Rutledge RB, Skandali N, Dayan P, Dolan RJ (2014) A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America* 111:12252-12257.
- Rygula R, Clarke HF, Cardinal RN, Cockcroft GJ, Xia J, Dalley JW, Robbins TW, Roberts AC (2014) Role of Central Serotonin in Anticipation of Rewarding and Punishing Outcomes: Effects of Selective Amygdala or Orbitofrontal 5-HT Depletion. *Cerebral cortex*.
- Salamone JD, Cousins MS, McCullough LD, Carriero DL, Berkowitz RJ (1994) Nucleus accumbens dopamine release increases during instrumental lever pressing for food but not free food consumption. *Pharmacology Biochemistry and Behavior* 49:25-31.
- Salamone JD, Correa M, Mingote S, Weber SM (2003) Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior: implications for studies of natural motivation, psychiatry, and drug abuse. *The Journal of pharmacology and experimental therapeutics* 305:1-8.
- Sallet J, Mars RB, Noonan MP, Neubert FX, Jbabdi S, O'Reilly JX, Filippini N, Thomas AG, Rushworth MF (2013) The organization of dorsal frontal cortex in humans and macaques. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:12255-12274.
- Sanacora G, Mason GF, Rothman DL, Krystal JH (2002) Increased occipital cortex GABA concentrations in depressed patients after therapy with selective serotonin reuptake inhibitors. *American Journal of Psychiatry* 159:663-665.
- Sarisoy G, Pazvantoglu O, Ozturan DD, Ay ND, Yilman T, Mor S, Korkmaz IZ, Kacar OF, Gumus K (2014) Metacognitive beliefs in unipolar and bipolar depression: a comparative study. *Nordic journal of psychiatry* 68:275-281.
- Scholl J, Gunthner J, Kolling N, Favaron E, Rushworth MF, Harmer CJ, Reinecke A (2014) A role beyond learning for NMDA receptors in reward-based decision-making-a

- pharmacological study using d-cycloserine. *Neuropsychopharmacology* : official publication of the American College of Neuropsychopharmacology 39:2900-2909.
- Scholl J, Kolling N, Nelissen N, Wittmann MK, Harmer CJ, Rushworth MF (2015) The Good, the Bad, and the Irrelevant: Neural Mechanisms of Learning Real and Hypothetical Rewards and Effort. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35:11233-11251.
- Scholl J, Kolling N, Nelissen N, Wittmann M, Harmer CJ, Rushworth MF (in press) The good, the bad and the irrelevant - neural mechanisms of learning real and hypothetical rewards and effort. *Journal of Neuroscience*.
- Scholl J, Kolling N, Nelissen N, Browning M, Rushworth MF, Harmer CJ (under review) Beyond negative valence - Serotonin enhances both reward and effort learning.
- Scholz J, Klein MC, Behrens TE, Johansen-Berg H (2009) Training induces changes in white-matter architecture. *Nature neuroscience* 12:1370-1371.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593-1599.
- Schweighofer N, Bertin M, Shishida K, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2008) Low-serotonin levels increase delayed reward discounting in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:4528-4532.
- Self MW, Kooijmans RN, Supèr H, Lamme VA, Roelfsema PR (2012) Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences* 109:11031-11036.
- Seligman MEP (1972) Learned helplessness. *Annual review of medicine* 23:407-412.
- Selvaraj S, Wylezinska M, Evans J, Jezzard P, Matthews PM, Cowen PJ (2006) Tryptophan depletion does not lower brain GABA levels in healthy volunteers. *Psychopharmacology* 187:131-132.
- Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32:5833-5842.
- Sharot T, Korn CW, Dolan RJ (2011) How unrealistic optimism is maintained in the face of reality. *Nature neuroscience* 14:1475-1479.
- Sherdell L, Waugh CE, Gotlib IH (2012) Anticipatory pleasure predicts motivation for reward in major depression. *Journal of abnormal psychology* 121:51-60.
- Shiffrin RM, Lee MD, Kim W, Wagenmakers EJ (2008) A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive science* 32:1248-1284.
- Simmons AN, Arce E, Lovero KL, Stein MB, Paulus MP (2009) Subchronic SSRI administration reduces insula response during affective anticipation in healthy volunteers. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum* 12:1009-1020.
- Skvortsova V, Palminteri S, Pessiglione M (2014) Learning to minimize efforts versus maximizing rewards: computational principles and neural correlates. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34:15621-15630.
- Smith SM (2002) Fast robust automated brain extraction. *Human brain mapping* 17:143-155.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 Suppl 1:S208-219.
- Smoski MJ, Felder J, Bizzell J, Green SR, Ernst M, Lynch TR, Dichter GS (2009) fMRI of alterations in reward selection, anticipation, and feedback in major depressive disorder. *Journal of affective disorders* 118:69-78.

- Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P (1995) A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry* 167:99-103.
- Spada MM, Nikčević AV, Moneta GB, Wells A (2008) Metacognition, perceived stress, and negative emotion. *Personality and individual differences* 44:1172-1181.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583-639.
- Spiegelhalter DJ, Best NG, Carlin BP, Linde A (2014) The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76:485-493.
- Spielberger CD, Gorsuch RL (1983) State-trait anxiety inventory for adults: Manual, instrument, and scoring guide: Mind Garden, Incorporated.
- Standage D, Pare M (2011) Persistent storage capability impairs decision making in a biophysical network model. *Neural networks : the official journal of the International Neural Network Society* 24:1062-1073.
- Starkstein SE, Petracca G, Chmerinski E, Kremer J (2001) Syndromic validity of apathy in Alzheimer's disease. *American Journal of Psychiatry* 158:872-877.
- Steele JD, Kumar P, Ebmeier KP (2007) Blunted response to feedback information in depressive illness. *Brain : a journal of neurology* 130:2367-2374.
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nature reviews Neuroscience* 9:255-266.
- Takahashi H, Yahata N, Koeda M, Takano A, Asai K, Suhara T, Okubo Y (2005) Effects of dopaminergic and serotonergic manipulation on emotional processing: a pharmacological fMRI study. *NeuroImage* 27:991-1001.
- Talbot PS, Watson DR, Barrett SL, Cooper SJ (2006) Rapid tryptophan depletion improves decision-making cognition in healthy humans without affecting reversal learning or set shifting. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 31:1519-1525.
- Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human pavlovian-instrumental transfer. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:360-368.
- Tanaka SC, Schweighofer N, Asahi S, Shishida K, Okamoto Y, Yamawaki S, Doya K (2007) Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PloS one* 2:e1333.
- Tanaka SC, Balleine BW, O'Doherty JP (2008) Calculating consequences: brain systems that encode the causal effects of actions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:6750-6755.
- Taylor M, Murphy SE, Selvaraj S, Wylezinska M, Jezzard P, Cowen PJ, Evans J (2008) Differential effects of citalopram and reboxetine on cortical Glx measured with proton MR spectroscopy. *Journal of psychopharmacology*.
- Taylor MJ, Freemantle N, Geddes JR, Bhagwagar Z (2006) Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Archives of general psychiatry* 63:1217-1223.
- Taylor MJ, Norbury R, Murphy S, Rudebeck S, Jezzard P, Cowen PJ (2009) Lack of effect of citalopram on magnetic resonance spectroscopy measures of glutamate and glutamine in frontal cortex of healthy volunteers. *Journal of psychopharmacology*.
- Taylor MJ, Godlewska BR, Norbury R, Selvaraj S, Near J, Cowen PJ (2012) Early increase in marker of neuronal integrity with antidepressant treatment of major depression: 1H-magnetic resonance spectroscopy of N-acetyl-aspartate. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum* 15:1541-1546.

- Taylor Tavares JV, Clark L, Furey ML, Williams GB, Sahakian BJ, Drevets WC (2008) Neural basis of abnormal response to negative feedback in unmedicated mood disorders. *NeuroImage* 42:1118-1126.
- Teyler TJ, Hamm JP, Clapp WC, Johnson BW, Corballis MC, Kirk IJ (2005) Long-term potentiation of human visual evoked responses. *The European journal of neuroscience* 21:2045-2050.
- Thoma P, Norra C, Juckel G, Suchan B, Bellebaum C (2015) Performance monitoring and empathy during active and observational learning in patients with major depression. *Biological psychology* 109:222-231.
- Torregrossa MM, Sanchez H, Taylor JR (2010) D-cycloserine reduces the context specificity of pavlovian extinction of cocaine cues through actions in the nucleus accumbens. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:10526-10533.
- Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, Zald DH (2009) Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PloS one* 4:e6598.
- Treadway MT, Zald DH (2011) Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience and biobehavioral reviews* 35:537-555.
- Treadway MT, Bossaller NA, Shelton RC, Zald DH (2012) Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. *Journal of abnormal psychology* 121:553.
- Treadway MT, Zald DH (2013) Parsing Anhedonia: Translational Models of Reward-Processing Deficits in Psychopathology. *Current directions in psychological science* 22:244-249.
- Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. *The European journal of neuroscience* 29:2225-2232.
- Tripp G, Alsop B (1999) Sensitivity to reward frequency in boys with attention deficit hyperactivity disorder. *Journal of Clinical Child Psychology* 28:366-375.
- Tse WS, Bond AJ (2002) Serotonergic intervention affects both social dominance and affiliative behaviour. *Psychopharmacology* 161:324-330.
- Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:4019-4026.
- van Berckel BNM, Lipsch C, Timp S, Wied CG, Wynne H, Van Ree JM, Kahn RS (1997) Behavioral and neuroendocrine effects of the partial NMDA agonist dycloserine in healthy subjects. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 16:317-324.
- Van Berckel BNM, Lipsch C, Gispen-de Wied C, Wynne HJ, Blankenstein MA, Van Ree JM, Kahn RS (1998) The partial NMDA agonist D-cycloserine stimulates LH secretion in healthy volunteers. *Psychopharmacology* 138:190-197.
- van der Veen FM, Evers EA, Deutz NE, Schmitt JA (2007) Effects of acute tryptophan depletion on mood and facial emotion perception related brain activation and performance in healthy women with and without a family history of depression. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 32:216-224.
- Vetencourt JFM, Sale A, Viegi A, Baroncelli L, De Pasquale R, O'Leary OF, Castrén E, Maffei L (2008) The antidepressant fluoxetine restores plasticity in the adult visual cortex. *Science* 320:385-388.
- Vickery TJ, Chun MM, Lee D (2011) Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72:166-177.
- Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M, Claes S (2013) Reduced reward learning predicts outcome in major depressive disorder. *Biological psychiatry* 73:639-645.

- Vurbic D, Gold B, Bouton ME (2011) Effects of D-cycloserine on the extinction of appetitive operant learning. *Behavioral neuroscience* 125:551-559.
- Walker DL, Ressler KJ, Lu K-T, Davis M (2002) Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *The Journal of Neuroscience* 22:2343-2351.
- Walker SC, Mikheenko YP, Argyle LD, Robbins TW, Roberts AC (2006) Selective prefrontal serotonin depletion impairs acquisition of a detour - reaching task. *European Journal of Neuroscience* 23:3119-3123.
- Walker SC, Robbins TW, Roberts AC (2009) Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. *Cerebral cortex* 19:889-898.
- Walton ME, Bannerman DM, Rushworth MFS (2002) The role of rat medial frontal cortex in effort-based decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22:10996-11003.
- Walton ME, Bannerman DM, Alterescu K, Rushworth MFS (2003a) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23:6475-6479.
- Walton ME, Bannerman DM, Alterescu K, Rushworth MFS (2003b) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *The Journal of Neuroscience* 23:6475-6479.
- Walton ME, Kennerley SW, Bannerman DM, Phillips PE, Rushworth MF (2006) Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural networks : the official journal of the International Neural Network Society* 19:1302-1314.
- Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, Rushworth MFS (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* 65:927-939.
- Wang M, Yang Y, Wang CJ, Gamo NJ, Jin LE, Mazer JA, Morrison JH, Wang XJ, Arnsten AF (2013) NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* 77:736-749.
- Wang X-J (2002) Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36:955-968.
- Wang XJ (2008) Decision making in recurrent neuronal circuits. *Neuron* 60:215-234.
- Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54:1063.
- Williams JBW (1988) A structured interview guide for the Hamilton Depression Rating Scale. *Archives of general psychiatry* 45:742-747.
- Willner P (2005) Chronic mild stress (CMS) revisited: consistency and behavioural-neurobiological concordance in the effects of CMS. *Neuropsychobiology* 52:90-110.
- Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* 14:1370-1386.
- Worbe Y, Palminteri S, Savulich G, Daw ND, Fernandez-Egea E, Robbins TW, Voon V (2015) Valence-dependent influence of serotonin depletion on model-based choice strategy. *Molecular psychiatry*.
- Wunderlich K, Dayan P, Dolan RJ (2012a) Mapping value based planning and extensively trained choice in the human brain. *Nature neuroscience* 15:786-791.
- Wunderlich K, Smittenaar P, Dolan RJ (2012b) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75:418-424.
- Yang CR, Svensson KA (2008) Allosteric modulation of NMDA receptor via elevation of brain glycine and D-serine: the therapeutic potentials for schizophrenia. *Pharmacology & therapeutics* 120:317-332.

- Yang XH, Huang J, Zhu CY, Wang YF, Cheung EF, Chan RC, Xie GR (2014) Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. *Psychiatry research* 220:874-882.
- Yang XH, Huang J, Lan Y, Zhu CY, Liu XQ, Wang YF, Cheung EF, Xie GR, Chan RC (2015) Diminished caudate and superior temporal gyrus responses to effort-based decision making in patients with first-episode major depressive disorder. *Progress in neuro-psychopharmacology & biological psychiatry* 64:52-59.
- Yilmaz AE, Gencoz T, Wells A (2011) The temporal precedence of metacognition in the development of anxiety and depression symptoms in the context of life-stress: a prospective study. *Journal of anxiety disorders* 25:389-396.
- Zarate CA, Jr., Brutsche N, Laje G, Luckenbaugh DA, Venkata SL, Ramamoorthy A, Moaddel R, Wainer IW (2012) Relationship of ketamine's plasma metabolites with response, diagnosis, and side effects in major depression. *Biological psychiatry* 72:331-338.
- Zhang WN, Chang SH, Guo LY, Zhang KL, Wang J (2013) The neural correlates of reward-related processing in major depressive disorder: a meta-analysis of functional magnetic resonance imaging studies. *Journal of affective disorders* 151:531-539.