Self- and other-regarding reinforcement learning: Disruptions in mental disorders and oxytocin's modulating role in healthy people

Shengchuang Feng

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Psychology

Brooks King-Casas, Chair
Pearl H. Chiu
Jian Li
Rachel A. Diana
Sheryl B. Ball

May 11, 2020

Blacksburg, VA

Keywords: depression, addiction, anxiety, PTSD, self-regarding learning, other-regarding learning, reinforcement learning, prediction error, oxytocin, dopamine

Self- and other-regarding reinforcement learning: Disruptions in mental disorders and oxytocin's modulating role in healthy people

Shengchuang Feng

ABSTRACT

It has been suggested that reward processing and related neural substrates are disrupted in some common mental disorders such as depression, addiction, and anxiety. An increasing number of psychiatric studies have been applying reinforcement learning (RL) models to examine these disruptions in self-regarding learning (learning about rewards delivered to the learners themselves). A review of RL alterations associated with mental disorders in extant studies will be beneficial for uncovering the mechanisms of these health problems. Although impaired social reward processing is common in some mental disorders [e.g., post-traumatic stress disorder (PTSD), social anxiety and autism], RL has not been widely used to detect the potentially disrupted social reward learning, especially for other-regarding learning (learning about rewards delivered to others). Meanwhile, it has not been clear whether some drugs, e.g., oxytocin (OT), can alter other-regarding learning, so they may serve as a therapeutic intervention when related deficits occur. In the present set of studies, we summarized common and distinct features in terms of self-regarding RL disturbances among depression, addiction and anxiety disorders based on previous findings (Paper I), tested whether behavioral and neural self- and other-regarding RL were impaired in PTSD with and without comorbid depression (Paper II), and investigated OT's behavioral and neural effects on self- and other-regarding RL in healthy males (Paper III). The results of our literature review showed that the commonalities in all three mental disorders were inflexibility and inconsistent choices, and the differences included decreased learning rates in depression, a higher weight to rewards versus punishments in addiction, and hypersensitivity to punishments in anxiety. The results of the PTSD study demonstrated impaired behavioral other-regarding learning in PTSD patients with and without depression, supposedly due to their hypervigilance to unexpected outcomes for others, as evidenced by the heightened responses in their inferior parietal lobule. The OT study detected OT's effects of attenuating behavioral other-regarding learning, as well as the neural coding of unexpected outcomes for others in the anterior cingulate cortex. These findings provide new evidence of self- and other-regarding RL alterations in mental disorders, reveal potential targets for their treatments, and bring caution for using OT as a therapeutic intervention.

Self- and other-regarding reinforcement learning: Disruptions in mental disorders and oxytocin's modulating role in healthy people

Shengchuang Feng

GENERAL AUDIENCE ABSTRACT

People learn to make choices to gain rewards and to avoid punishments delivered to themselves. As social animals, people also take account of outcomes delivered to others when learning. With the help of computational modeling, previous studies have found abnormal reward learning for oneself in people with mental health problems. To better understand mental illnesses, we summarized the similarities and differences of the learning abnormalities reported in previous studies about depression, addiction, and anxiety. We have found that people with these mental illnesses all tend to be inflexible and make more random choices when learning. As for the differences, people with depression tend to learn slower; people with addiction tend to see gaining rewards as more important than avoiding punishments; and people with anxiety tend to be oversensitive to punishments. Using computational modeling and imaging of brain function, we also tested whether learning for other was abnormal in post-traumatic stress disorder (PTSD), and found that, compared to healthy people, PTSD patients had slower learning for others' rewards, and the inferior parietal lobule, a brain region for processing social information, showed higher responses to unexpected outcomes for others. In another study, we examined whether oxytocin (OT), a neuropeptide that has been reported to change people's social functions, could influence reward learning for others in healthy males. The results showed that OT slowed down people's learning for others, and also decreased the neural learning signals in the anterior cingulate cortex, a region involved in processing other's outcomes. Our findings provide new information about how reward learning for oneself and others are changed in mental illnesses, reveal potential targets for their treatments, and bring caution for using OT as a therapeutic intervention.

ACKNOWLEDGMENTS

First of all, I would like to extend my sincere gratitude to my advisors Dr. Brooks King-Casas and Dr. Pearl Chiu for their support, guidance, and encouragement throughout my growth as a researcher during the past six years. Thank you for discussing my research projects with me, sharing your thoughts, and teaching me the ways of thinking in scientific research. I would also like to thank Dr. Jian Li and Dr. Rachel Diana for their insightful comments and suggestions in my preliminary exam, which directly helped me to revise and refine the review study in this dissertation. Dr. Li was also a committee member of my Master's thesis, and I want to thank him for his suggestions and encouragement in my writing of the thesis. I am also thankful for one class taught by Dr. Diana, in which I learned how to extract the most important information from a long book chapter. It continues to benefit my reading and writing. I took another class taught by Dr. Sheryl Ball, and I would like to thank her for extending my knowledge and interest from psychology to economics. Moreover, I appreciate the questions, comments, and suggestions from all the committee members during my dissertation defense.

Secondly, I thank Dr. George Christopoulos at Nanyang Technological University for his contribution to the oxytocin study and the PTSD study in this dissertation and Julia Julien for her assistance in data analysis for the PTSD study. I also appreciate the work of our coordinators who helped to collect data for these studies. I am equally grateful to previous and current members of the Chiu & King-Casas labs for their endless scientific and personal support. Meanwhile, I also want to thank my parents, who have been caring, supporting, and encouraging me since my birth (actually since ten months before my birth).

Last but not least, my appreciation and respect go to all medical workers around the globe who are fighting on the frontline to contain COVID-19 and other people who stay in their posts to help maintain an orderly society during this unprecedented pandemic. I believe as long as we human beings keep united and help each other, we will beat the virus and recover soon!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Since their birth, human beings have been learning to make decisions to gain rewards and to avoid punishments. Disrupted reward processing has been shown to be a transdiagnostic feature of multiple mental disorders (Zald & Treadway, 2017) and may account for their comorbidities. In the last decade, an increasing number of studies have been applying reinforcement learning (RL) models (Sutton & Barto, 1998) to understand details of these reward processing disruptions, but the common and distinct features in terms of RL across diagnostic boundaries remain vague. To address this issue, in Paper I, we made a systematic cross-disorder review for depression, addiction, and anxiety, which are the three most prevalent mental illnesses worldwide (Ormel et al., 1994; Steel et al., 2014).

Although these studies can provide rich information about reward deficits in psychopathy, most of them only focused on self-regarding reward learning. There has been a large body of evidence showing that people do not only attend to rewards received by themselves, but they also care about others' well-being (Christopoulos & King-Casas, 2015; Liu et al., 2019; McClintock, 1972; Messick & McClintock, 1968; Sul et al., 2015; Van Lange, 1999). Moreover, mental disorders are usually associated with deficits in other-regarding cognition (Palgi, Klein, & Shamay-Tsoory, 2017; Uekermann et al., 2008; Washburn, Wilson, Roes, Rnic, & Harkness, 2016; Wolkenstein, Schönenberg, Schirm, & Hautzinger, 2011). For example, the symptoms of post-traumatic stress disorder (PTSD) involve avoidance of people associated with the traumatic experience, blaming of self/others, and feeling distant from others (American Psychiatric Association, 2013; Stevens & Jovanovic, 2018). Disturbed other-perception (mentalizing) has also been reported in PTSD by many studies (Nietlisbach, Maercker, Rösler, & Haker, 2010; Plana, Lavoie, Battaglia, & Achim, 2014; Sharp, Fonagy, & Allen, 2012). These deficits possibly result in dysfunctional processing of rewards delivered to others. To test this hypothesis, in Paper II, we compared behavioral and neural learning of self- and other-regarding rewards in veterans with and without PTSD.

Given the disrupted social functioning (Palgi et al., 2017; Uekermann et al., 2008; Washburn et al., 2016; Wolkenstein et al., 2011) and reward processing (Zald & Treadway, 2017) in mental disorders and oxytocin (OT)'s influences on social cognition (Bartz, Zaki, Bolger, & Ochsner, 2011; Zik & Roberts, 2015), reward processing (Clark-Elford et al., 2014; Ide et al., 2018), and stress response (De Oliveira, Zuardi, Graeff, Queiroz, & Crippa, 2012; Heinrichs, Baumgartner, Kirschbaum, & Ehlert, 2003), OT has been suggested to be a potential therapeutic intervention for depression (Slattery & Neumann, 2010), addiction (McGregor & Bowen, 2012; Slattery & Neumann, 2010), and anxiety (Koch et al., 2014). Neurophysiological studies have demonstrated the close relationship between the oxytocinergic system and the dopaminergic reward system (Baribeau & Anagnostou, 2015; Grinevich, Knobloch-Bollmann, Eliava, Busnelli, & Chini, 2016; Hung et al., 2017) and OT's modulation of dopaminergic activity (Melis et al., 2007; Succu et al., 2008); therefore, it is possible that OT's therapeutic effects are mediated by its influences on the reward system. To test whether OT could causally affect reward processing, we implemented a double-blind, placebo (PL)-controlled design

in Paper III, and compared healthy males' behavioral and neural learning of self- and other-regarding rewards in the OT and PL conditions.

A taxonomy of reward processing

Previous literature has discussed the taxonomy of reward processing (McClure, York & Montague, 2004; Zald & Treadway, 2017), such as the classic division of reward attainment and reward anticipation (Berridge & Robinson 2003). In the present work, we apply a taxonomy that is compatible with previous ones and can also facilitate the discussion of related literature as well as studies in the present work. As a general term for any reward processes, reward processing can be divided into reward perception and reward learning. Reward perception reflects the static aspect of reward processing and can be defined as the subjective evaluation or experience of anticipating or receiving rewards. An individual can have a subjective evaluation of a received reward by simply "feeling" it. Reward learning, on the other hand, reflects the dynamic aspect of reward processing and can be defined as the process of acquiring the association between a stimulus (in classical conditioning) or an action (in instrumental conditioning) and rewards. It is dynamic in the sense that reward learning always involves prediction error (PE) —the discrepancy between the reward that a learner receives and the reward the learner predicts (Pearce, 2013; Rescorla &Wagner, 1972), and the learner acquires the associated value of the action or stimulus by minimizing the PEs. In a broader sense, the process of reward learning also includes reward perception because without the subjective evaluation or experience of rewards, no PEs can be calculated.

Reinforcement learning models and the neural prediction error signals

In RL studies, a typical experimental paradigm is the k-armed bandit task (usually $k \geqslant 2$; Sutton & Barto, 1998). It represents a simplified version of an environment, in which the learner acquires the contingencies between k options and their values through trial-and-error, with the goal of maximizing received rewards and/or minimizing received punishments. According to the basic RL model, reward learning in this task can be described as a process in which the expected value of selecting an option is updated by cumulating PEs weighted by a learning rate at each time step (Sutton & Barto, 1998). This updating process can be formally represented by the equation $Q_{(t)} = Q_{(t-1)} + \alpha*(V_{(t-1)} - Q_{(t-1)})$, in which $V_{(t-1)}$ denotes the received reward at time t–1 after the learner chooses an option, Q denotes the expected value of the option, and (V – Q) is PE (Rescorla & Wagner, 1972). α is the learning rate and describes how rapidly the learner uses PEs to update his/her expected value of that option. In the k-armed bandit task, the learner can choose among different options based on their expected values.

RL models are flexible and extensible in that we can incorporate additional parameters into the basic RL model to account for other components of the learning process. For example, the learner's sensitivity to rewards can be represented by a multiplier (ρ) on V, and the RL model becomes $Q_{(t)} = Q_{(t-1)} + \alpha*(\rho *V_{(t-1)} - Q_{(t-1)})$ (Huys, Pizzagalli, Bogdan, & Dayan, 2013). This new model helps to identify individual differences in reward sensitivity and can also serve as an example of our reward processing taxonomy, in which ρ reflects reward perception, while α reflects reward learning.

As for RL's neural substrates, a large number of functional magnetic resonance imaging (fMRI) studies have identified robust associations between PEs and the ventral striatum activity (Berns, McClure, Pagnoni, & Montague, 2001; Knutson, Fong, Adams, Varner, & Hommer, 2001; Li, McClure, King-Casas, & Montague, 2006). This is also consistent with the dopamine theory of learning based on animal studies, which show that PEs are encoded by phasic signaling of dopamine neurons in the ventral tegmental area projecting to the ventral striatum (Schultz, 2007; Schultz, Dayan, & Montague, 1997). In addition to the ventral striatum, other regions in a more extensive network are also involved in the encoding of PEs (Garrison, Erdeniz, & Done, 2013). This network includes the striatum, midbrain, insula, and some frontal and temporal areas, which are regions also most frequently found to be disrupted in several mental disorders (Fladung et al., 2009; Forbes et al., 2009; Plichta & Scheres, 2014).

When applying RL models in the study of mental disorders, participants' choice data in a learning task are fitted to an RL model, and the estimated model parameters can be compared between healthy controls and participants with mental disorders. In cases where participants' fMRI data are collected, trial-by-trial PE values can be correlated with the brain data, and the group differences in certain brain areas' encoding of PEs can be tested (Daw, 2011).

Other-regarding learning and social preferences

Apart from the above-mentioned learning in which the learner only receives rewards for him/herself, there are scenarios that involve rewards delivered to others. These other-regarding learning scenarios include at least two types: (i) the learner observes others making choices and receiving rewards (observational/vicarious learning; Burke, Tobler, Baddeley, & Schultz, 2010), and (ii) the learner makes choices impacting others' payoffs (Christopoulos & King-Casas, 2015). In our studies, we mainly focus on the latter type. This scenario can be presented as a combination of a two-armed bandit task and a reward allocation task, in which the learner chooses from options with different reward allocations between him/herself and others (Christopoulos & King-Casas, 2015).

To investigate reward learning in this scenario, a key step is to isolate other-regarding reward perception. A rich body of literature has shown that individuals vary a lot in social preferences or the extent to which they value rewards delivered to others (Fehr & Krajbich, 2014; McClintock, 1972; Van Lange, 1999). To capture such individual differences, various models were proposed (Battigalli & Dufwenberg, 2007; Charness & Rabin, 2002; Levine, 1998; Fehr & Schmidt, 1999; Van Lange, 1999). For example, the Fehr-Schmidt model (Fehr & Schmidt, 1999) takes account of the inequality of rewards between oneself and others. According to this model, individuals feel guilt when rewards for others are lower than rewards for themselves, and they feel envy when rewards for others are higher than rewards for themselves. The intensity of guilt and envy determines how they allocate resources. The Van Lange model (Van Lange, 1999) proposes that individuals do not only consider the inequality of payoffs but also weigh both their own and others' payoffs. Therefore, an individual may have a negative weight on rewards delivered to others, and see others' losses as his/her gains. These models were previously applied to non-learning reward allocation tasks and can benefit the study of other-regarding learning. As some of the first researchers examining other-regarding RL,

Christopoulos & King-Casas (2015) developed a new RL model with a social preference parameter similar to Van Lange (1999)'s other-regarding weight and separate learning rates for self-PEs and other-PEs, respectively. This model allows the dissociation between other-regarding reward perception and other-regarding reward learning, which lays a solid foundation for further research on this subject.

The neural substrates of other-regarding learning (in both scenarios mentioned above) have also been investigated in several studies and the frontal lobe is evidenced to process other-regarding learning signals (Apps, Rushworth, & Chang, 2016; Burke et al., 2010; Christopoulos & King-Casas, 2015). Both Burke et al. (2010) and Christopoulos & King-Casas (2015) detected other-regarding PE signals in the medial prefrontal cortex although the two studies applied different learning tasks. Another series of studies support the role of the anterior cingulate cortex (ACC) in encoding other-regarding information (Apps, Rushworth, & Chang, 2016), among which the ACC was found to process other's reward probabilities (Lockwood, Apps, Roiser, & Viding, 2015) and other's PEs (Apps, Lesage, & Ramnani, 2015) in humans, and reward allocations to others in monkeys (Chang, Gariépy, & Platt, 2013). Brain regions involved in theory of mind and other social processes, such as the temporoparietal junction (TPJ) and inferior parietal lobule (IPL), may also play a role in encoding other-regarding learning signals, as evidenced by decreased self-regarding PE signals of the TPJ for social outcomes (smiling and frowning faces) in PTSD (Cisler et al., 2015).

Oxytocin, social cognition, and reward learning

OT is a neuropeptide that has been shown to modify social cognitions, such as trust (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005) and cooperation (De Dreu et al., 2010; Israel, Weisel, Ebstein, & Bornstein, 2012) when administered extraneously. It has also been suggested to be a potential therapeutic intervention for several mental disorders (Slattery & Neumann, 2010; McGregor & Bowen, 2012; Slattery & Neumann, 2010; Koch et al., 2014), as it was found to increase dopaminergic activity in animals (Melis et al., 2007; Succu et al., 2008), reduce anxiety in stressful social contexts (De Oliveira et al., 2012; Heinrichs et al., 2003) and attenuate amygdala's responses to aversive facial stimuli in humans (Domes et al., 2007; Kirsch et al., 2005; Labuschagne et al., 2010).

Although many studies show OT's enhancement of prosocial cognition, 21% of the published studies of OT and prosociality reveal antisocial effects (Bartz et al., 2011). The directions of OT's effects on self-regarding learning are also mixed, with some showing enhanced learning (Hurlemann et al., 2010) and others showing detrimental effects (Clark-Elford et al., 2014; Ide et al., 2018). Bartz et al. (2011) suggested taking account of contexts and individual differences when examining the effects of OT. The social salience hypothesis also proposes that OT can increase the salience of social cues, magnifying prosociality when dealing with familiar or reliable others but diminishing prosociality when dealing with outgroup members (Shamay-Tsoory & Abu-Akel, 2016). Partially consistent with this hypothesis, a recent study also found OT's effects of increasing cooperation in competitors and individualists, and a nonsignificant trend of decreasing cooperation in prosocials when interacting with strangers (Liu et al., 2019). As suggested by these studies, a clear understanding of factors that may influence the effectiveness of OT is essential for using it as a therapeutic intervention.

The research of OT's effects on other-regarding learning has been lacking, but there is a noticeable pattern in self-regarding learning. It seems that incorporating social components (e.g., smiling faces) into the learning tasks is necessary for OT to have an effect (Clark-Elford et al., 2014; Evans, Shergill, & Averbeck, 2010; Hurlemann et al., 2010; Ide et al., 2018). Given that processing rewards delivered to others can be considered as a social component, it is likely that other-regarding learning will also be modulated by OT.

Present studies and hypotheses

The present work intended to fill the gaps in the existing literature with respect to (i) similarities and differences in self-regarding RL disturbances across depression, addiction, and anxiety, (ii) self- and other-regarding RL in PTSD and comorbid depression, and (iii) OT's effects on self- and other-regarding RL.

In Paper I, we summarized the behavioral and imaging findings from 54 studies using RL models to examine self-regarding learning in depression, addiction, or anxiety. We firstly summarized similar and different patterns in model agnostic performance, learning rates, and inverse temperatures (choice randomness), then we used additional model parameters to test some of the identified patterns. The neural PE signals were also compared between different mental disorders.

In Paper II, healthy veterans and veterans with PTSD (with and without comorbid depression) were recruited to perform a probabilistic social learning task (Christopoulos & King-Casas, 2015) with monetary payoffs for themselves and an anonymous partner while being scanned with fMRI. According to previous RL studies, anxiety is associated with increased learning rates (Aylward et al., 2019; Huang, Thompson, & Paulus, 2017; Vaghi et al., 2017), which might be due to their overestimation of the probability of bad outcomes (Aylward et al., 2019). Therefore, we hypothesized that PTSD patients would show increased self-regarding learning. Due to the high occurrence of social cognitive deficits in PTSD (Palgi et al., 2017; Washburn et al., 2016), other-regarding learning would be impaired in patients. The associated neural learning signals might show the same pattern.

In Paper III, we tested OT's effects on self- and other-regarding learning in healthy adult males. The participants completed the same fMRI task as in Paper II after being administered PL and OT. Learning parameters and related brain signals were compared between drug conditions. Previous studies suggest that including social components is necessary for OT to have an effect (Clark-Elford et al., 2014; Evans et al., 2010; Hurlemann et al., 2010; Ide et al., 2018), so we hypothesized that OT would not modify self-regarding learning. As for other-regarding learning, since OT can increase the salience of an anonymous partner (Shamay-Tsoory & Abu-Akel, 2016), the participants may care less about this partner's well-being after OT. Therefore, we hypothesized that participants in the OT condition might show decreased other-regarding learning.

Paper I: Reinforcement Learning Dysfunctions in Depression, Addiction, and Anxiety Disorders: A Systematic Review

Shengchuang Feng, Pearl H. Chiu, & Brooks King-Casas

ABSTRACT

Depression, addiction, and anxiety disorders are highly comorbid with each other, suggesting potential common cognitive and neural mechanisms. The last decade has witnessed a surge in the number of studies employing reinforcement learning (RL) models to examine the three types of disorders. In this paper, we summarize 54 of these studies to answer the question: What are the similarities and differences of the RL disturbances across these disorders, behaviorally and neurally? A systematic review reveals that learning inflexibility (possibly impaired model-based & model-free learning) and inconsistent choices seem to be the similarities across all these disorders. As for the differences, depression is associated with decreased learning rates, possibly in both reward and punishment domains. Addiction is characterized by a higher weight to rewards versus punishments. Abstinence may modulate addicts' sensitivity to rewards and punishments. Anxiety is featured by less optimism, hypersensitivity to punishments, and enhanced Pavlovian biases. Neurally, both depression and addiction are associated with impaired learning signals in key reward regions. Neural changes for anxiety include impaired learning signals in some reward regions as well as heightened neural responses to unexpected feedback. Limitations and future directions are also discussed at the end of the paper.

Keywords: depression, addiction, anxiety, reinforcement learning, learning rate, inverse temperature, prediction error, dopamine

INTRODUCTION

Depression, addiction (substance use) and anxiety disorders are among the most prevalent mental illnesses worldwide (Ormel et al., 1994; Steel et al., 2014). They are also highly comorbid with each other (Hirschfeld, 2001; Quello, Brady, & Sonne, 2005; Smith & Book, 2008), suggesting possible common cognitive and neural mechanisms. Specifically, there is increasing evidence that reward/punishment processing and related neural substrates are disrupted in these disorders (Aupperle & Martin, 2010; Chen, Takahashi, Nakagawa, Inoue, & Kusumi, 2015; Cremers, Veer, Spinhoven, Rombouts, & Roelofs, 2015; Luijten, Schellekens, Kühn, Machielse, & Sescousse, 2017; Vriends, Michael, Schindler, & Margraf, 2012). Benefiting from the development of neuroeconomics, researchers started to use reinforcement learning (RL) models to understand important aspects of several psychiatric and neurological disorders and hope to find common mechanisms across diagnostic boundaries (Maia & Frank, 2011; Montague, Dolan, Friston, & Dayan, 2012). As a result, the last decade has witnessed a surge in the number of RL studies of the three types of disorders. A systematic cross-disorder review is necessary to address the question: What are the similarities and differences of the RL disturbances across these three types of disorders, behaviorally and neurally? To answer this question, we will make a systematic summary of both behavioral and neuroimaging studies employing RL models in the research of depression, addiction, and anxiety disorders. In this review, we will first introduce RL and its neural substrates. Then we will summarize the characteristics and findings of these studies. Finally, we will make a general summary and discuss the limitations and future directions.

REINFORCEMENT LEARNING AND ITS NEURAL SUBSTRATES

In RL studies, a typical paradigm is the k-armed bandit task (usually $k \geqslant 2$; Sutton & Barto, 1998). It represents a simplified version of an environment, in which a learner learns the contingencies between options and outcomes despite uncertainties, with the goal of maximizing received rewards and/or minimizing received punishments. The basic form of this task is the two-armed bandit task, in which a learner chooses between two options. One is advantageous in that it is associated with a higher probability of a certain reward. The other, however, is associated with a lower probability of that reward. Without knowing these contingencies, the learner needs to learn them through trial-and-error. There are many variants of the task, with other components added to it. For example, the contingencies can be changing; more options/arms can be provided; and rewards and punishments can be delivered simultaneously by one option.

The learning process described above is called RL. It can be represented formally with the basic RL model (Sutton & Barto, 1998), in which the expected value (Q) of a certain option is updated according to a rule (Rescorla & Wagner, 1972):

$$Q_{(t)} = Q_{(t-1)} + \alpha*(V_{(t-1)} - Q_{(t-1)}). \qquad \text{(Equation 1.1)}$$

In this equation, $Q_{(t)}$ and $Q_{(t-1)}$ denote the expected value of an option at time t and t–1, respectively; $V_{(t-1)}$ denotes the actual value received by the learner at time t–1; $(V_{(t-1)} - Q_{(t-1)})$ is termed prediction error (PE), denoting the difference between the actual value and the expected value at time t–1. Positive PEs occur when received rewards are better

than expected. Negative PEs occur when received rewards are worse than expected. α is learning rate and describes how rapidly the learner uses PEs to update his/her expected value of an option. A higher learning rate indicates relying more on most recent information; a lower learning rate indicates relying more on past reward history. After the expected values of all options are updated, the learner's probability of choosing a certain option can be modeled with a standard softmax function (Luce, 1959):

$$P_{A(t)} = \frac{\exp(\beta * Q_{A(t)})}{\exp(\beta * Q_{A(t)}) + \exp(\beta * Q_{B(t)})} \qquad \text{(Equation 1.2)}$$

where $P_{A(t)}$ denotes the probability of choosing option A out of the two options (A and B) at time t; β is inverse temperature and can be used as a measure of the noisiness or randomness of the learner's choices.

There are at least three key advantages of RL models. Firstly, they allow us to make precise hypotheses about hidden variables in the learning process, which are very difficult to measure using conventional methods (Chen et al., 2015; Niv, 2009). Secondly, their flexibility and extensibility allow us to incorporate additional parameters into the basic RL model to account for other components of the task or other cognitive processes. Thirdly, when combined with neuroimaging techniques, they can provide information about how a learning component is implemented in a brain region rather than only reveal the location of that region (O'doherty, Hampton, & Kim, 2007).

One of the most prominent examples of these advantages is the neural substrates of PEs. In a typical RL study, the basic RL model or its variants are fitted to participants' trial-by-trial choice data to estimate the values of the free parameters (e.g., learning rates and inverse temperatures). Then, PE for each trial can be derived with the estimated parameters and correlated with functional magnetic resonance imaging (fMRI) signals. Multiple fMRI studies have consistently found robust associations between PEs and the ventral striatum activity (Berns, McClure, Pagnoni, & Montague, 2001; Knutson, Fong, Adams, Varner, & Hommer, 2001; Li, McClure, King-Casas, & Montague, 2006). This is also consistent with the dopamine theory of learning based on animal studies, which show that PEs are encoded by phasic signaling of dopamine neurons in the ventral tegmental area projecting to the ventral striatum (Schultz, 2007; Schultz, Dayan, & Montague, 1997). A more extensive network of brain regions associated with PEs in the reward condition (RPE) and the aversive condition (APE) are summarized in a meta-analysis (Garrison, Erdeniz, & Done, 2013). This network includes the striatum (ventral and dorsal), midbrain, insula, and some frontal and temporal areas, which are regions also most frequently found to be disrupted in several mental disorders (Fladung et al., 2009; Forbes et al., 2009; Plichta & Scheres, 2014). Due to their advantages and the robust PE-related activations in these brain regions, the RL models have become an ideal framework to examine the behavioral and neural alterations in the RL system across different disorders.

A SUMMARY OF RL STUDIES IN DEPRESSION, ADDICTION AND ANXIETY DISORDERS

Selection of studies and statistics of interest

When applying the RL framework to examine mental disorders, researchers often compare the free parameters of interest and the neural PE signals between a patient/subclinical group and a matched healthy control group. In this review, we mainly summarize human studies with these comparisons in depression, addiction, and anxiety disorders. We conducted searches for studies of these disorders in PubMed and Google Scholar databases [see **SUPPLEMENTARY INFORMATION (SI)**]. The final sample includes 21 (11 fMRI) depression studies, 20 (10 fMRI) addiction studies, and 17 (9 fMRI) anxiety studies. Two studies (Aylward et al., 2019; Mkrtchian, Aylward, Dayan, Roiser, & Robinson, 2017) had about 70% of participants in the disorder group with high comorbidity of depression and anxiety. One study examined depression and addiction as two separate factors (Feng, 2017), and another study had two separate groups with addiction and anxiety, respectively (Kanen, Ersche, Fineberg, Robbins, & Cardinal, 2019). Therefore, these four studies were included in two disorder categories (see **Table 1.1**). Although a majority of patients in another study had depression and anxiety (Huang, Thompson, & Paulus, 2017), it compared patients with high and low anxiety, this study was only included in the anxiety category. For all the other studies, only one type of primary disorder was observed in a majority of the disorder group despite comorbidities. In total, there are 54 studies included in this review.

For each study, we mainly focus on group differences in RL model parameters and neural correlates of PEs (if available) between participants with mental disorders and controls. In the following sections, we will describe the basic characteristics of these studies and discuss model-agnostic performance, model parameters, and neural PE signals in different mental disorders.

Basic characteristics of studies

In all 54 studies, most compared a mental disorder group or its subgroups with a control group. Five depression studies and two anxiety studies compared two nonclinical samples with high and low depression (Blanco, Otto, Maddox, Beevers, & Love, 2013; Frey, Frank, & McCabe, 2019; Frey & McCabe, 2020; Kunisato et al., 2012; Rouhani & Niv, 2019) or anxiety measures (Myers et al., 2013; Piray, Ly, Roelofs, Cools, & Toni, 2019). As mentioned above, one anxiety study compared clinical patients with high and low anxiety (Huang et al., 2017). One addiction study compared heavy and light drinkers (Gullo & Stieger, 2011). Moreover, the addicted participants in other four addiction studies may not meet clinical criteria for substance use disorder (Addicott, Pearson, Wilson, Platt, & McClernon, 2013; Feng, 2017; Lesage et al., 2017; Wei et al., 2018).

Apart from depressed nonclinical participants, depressed individuals in the depression studies were mostly with major depressive disorder (MDD). In the addiction studies, eight examined alcohol use disorder; four compared smokers and nonsmokers (Addicott et al., 2013; Feng, 2017; Lesage et al., 2017; Wei et al., 2018). The other studies were about cocaine, amphetamine, heroin, cannabis, and polydrug abusing. The addicted participants in 13 addiction studies were abstinent for different periods of time, ranging

from 12 hours to more than 3 months. In the 17 anxiety studies, all patients in three of the studies had a diagnosis of post-traumatic stress disorder (PTSD) and some of the patients had comorbid depression or other anxiety disorders (Brown et al., 2018; Cisler et al., 2015; Ross, Lenow, Kilts, & Cisler, 2018); six had patient with mainly obsessive-compulsive disorder (OCD; Carlisi et al., 2017; Hauser et al., 2017; Kanen et al., 2019; Murray et al., 2019; Norman et al., 2018; Vaghi et al., 2017); other studies with clinical groups all had mixed samples of anxiety disorders and MDD/depressive symptoms.

All studies employed a certain variant of the k-armed bandit task, some of which had changing contingencies between options and outcomes or reversals of advantageous and disadvantageous options. Rewards in these tasks included money, points, water, smiling/frowning faces, and symbolic feedback ("correct" or "wrong"). The characteristics of these studies are summarized in **Table 1.1**, **Supplementary Tables S1.1**, **S1.3**, and **S1.5**.

Model-agnostic performance

In general, participants with depression and addiction had worse performance than controls. Twenty depression studies tested the group differences in performance, and nine detected a significant decrease in depressives. Seventeen anxiety studies tested group differences in performance. Twelve found decreased and one found increased performance. In these studies, group differences in performance were more readily found in tasks with reward contingency reversals. More specifically, nine out of 10 reversal tasks in depression and addiction studies showed decreased performance, and one showed increased performance in the mental disorder group. This suggests cognitive inflexibility in depression and addiction (Beylergil et al., 2017; Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010). It is also easy to see that participants in the patient groups performed worse than controls in Iowa gambling tasks (IGT; Bechara, Damasio, Damasio, & Anderson, 1994). All seven IGTs with available group comparisons in addiction studies showed worse performance in addicted participants, suggesting possibly biased attention to rewards versus punishments in addiction (Kobayakawa, Tsuruya, & Kawamura, 2010; Yechiam, Busemeyer, Stout, & Bechara, 2005).

In anxiety studies, however, mixed results were seen. In 16 studies that tested performance, three found increased and five found decreased performance. In the three studies with increased performance, Myers et al. (2013) found enhanced accuracy in PTSD relative to controls. The authors provided a possible reason that, in their task, healthy controls tended to see an ambiguous no-feedback screen as rewarding (when it actually signaled wrong choices), resulting in less correct choices. Nonetheless, PTSD participants perceived that ambiguous feedback as more neutral (less rewarding), thus appeared to perform better. The other two studies (Kanen et al., 2019; Khdour et al., 2016) found higher accuracy or lose-switch in the punishment condition but not in the reward condition for patients relative to controls, this is probably due to anxious participants' hypersensitivity toward punishments, which may serve as a coping mechanism to reduce anxiety. In the studies showing decreased performance, two (Mkrtchian et al., 2017; Ousdal et al., 2018) applied the Pavlovian probabilistic go/no-go task (Guitart-Masip et al., 2012), in which the participants needed to act to avoid punishments and to inhibit their actions to get rewards. The anxious participants' worse

performance in this task suggests their heightened Pavlovian biases: excessive punishment avoidance and/or reward approach.

To summarize, model-agnostic results revealed learning inflexibility in depression and addiction disorders. Moreover, addicted participants might have more attentional bias to rewards versus punishments. Anxious participants tended to be less optimistic, hypersensitive to punishments, and have heightened Pavlovian biases.

Learning rate and inverse temperature

RL model parameters provide a tool to test the model-agnostic findings. As described above, the learning rate and inverse temperature are two basic parameters of RL models. In this section, we will discuss how the learning rate and inverse temperature can inform the mechanisms of mental disorders.

In **Table 1.1**, the pattern of learning rates shows that depression studies with significant group differences in this parameter mostly had decreased learning rates in depressed participants relative to controls (Chase et al., 2010; Dombrovski et al., 2010; Dombrovski, Szanto, Clark, Reynolds, & Siegle, 2013; Frey et al., 2019; Frey & McCabe, 2020), indicating impaired incorporation of reward information into the learning process in depression. Depression has been associated with altered responses to positive and negative feedback. Specifically, previous studies demonstrated hyposensitivity to rewards (Pizzagalli et al., 2009; Pizzagalli, Jahn, & O'Shea, 2005) but hypersensitivity to punishments (Beats, Sahakian, & Levy, 1996; Santesso et al., 2008) in depression. The depressive's medication states may also affect this pattern. It is suggested that antidepressants only have negligible effects on their hyposensitivity to rewards, but can attenuate their exaggerated punishment responses to the same level as their reward responses (Herzallah et al., 2013). Consequently, depressives on medication would have decreased learning in both reward and punishment domains. In the reviewed depression studies, five with clinical samples applied separate learning rates on positive and negative PEs (Aylward et al., 2019; Bakic et al., 2016; Chase et al., 2010; Dombrovski et al., 2010; Dombrovski et al., 2013; see note b of **Table S1.2** for differences between RPEs, APEs, positive PEs and negative PEs), and their results are generally consistent with these notions. Two of the studies found marginally decreased learning rates for positive PEs in depressives relative to controls (Chase et al., 2010) or in depressed suicide attempters relative to nonsuicidal participants (Dombrovski et al., 2013). Depressives in both studies were on medication. For learning rates for negative PEs, two studies found a marginal decrease (Chase et al., 2010; Dombrovski et al., 2010) of this parameter in medicated depressives. However, in an unmedicated anxious and depressed sample, increased values of this parameter were detected (Aylward et al., 2019). These results revealed a general pattern of decreased learning for both rewards and punishments in medicated depressives, and the impaired learning in the punishment domain may be due to medications.

In four addiction studies and four anxiety studies showing increased learning rates in the patient groups, seven had changing contingencies (including contingency reversals), suggesting that the altered learning rates may be related to task volatility. For a learner in a volatile environment, it is optimal to rely more on the most recent information and

employ a high learning rate. In support of this, Browning, Behrens, Jocham, O'reilly, & Bishop (2015) showed that participants' learning rates were higher in a volatile task than in a stable one. The altered learning rates in the patient groups imply that individuals with addiction and anxiety may over-adjust this parameter when faced with a fast-changing learning task, thus having impaired representation of the environment or model-based learning (Montague et al., 2012). As a side note, the increased learning rates for negative PEs in the anxious and depressed sample (Aylward et al., 2019) might not be solely due to their medication states. The changing contingencies in the learning task might also contribute to it.

Another important factor, abstinence, should also be considered when viewing the results from addiction studies in that the state of abstinence and satiety may influence reward processing differently (Addicott et al., 2012; Freeman, Morgan, Beesley, & Curran, 2012). It is posited that repeated exposure to drugs of abuse leads to a higher level of extracellular dopamine concentration (also termed tonic activity; Grace, 2000) and lower level of postsynaptic dopamine receptor availability in the ventral striatum as an effect of long-term adaptation to dependence (Nutt, Lingford-Hughes, Erritzoe, & Stokes, 2015; Volkow, Fowler, Wang, Swanson, & Telang, 2007). The high tonic activity can impose a sustained suppression on phasic dopamine signaling (Leknes & Tracey, 2008), through dopamine-releasing-inhibiting autoreceptors (Grace, 2000; Seeman & Madras, 1998). Meanwhile, lower striatal dopamine receptor availability may also diminish the phasic dopamine signaling. Consequently, sated participants with addiction disorder may have diminished reward learning, which would go back or be close to normal in abstinence. In support of this hypothesis, the sated smokers in Feng (2017) and Wei et al. (2018) showed decreased learning rates, and another study showed enhanced striatal PE signals in abstinent cocaine-dependent individuals than when they were sated, and this enhancement was associated with increased learning rates for positive PEs relative to learning rates for negative PEs (Wang et al., 2019). Lesage et al. (2017) also found comparable learning rates in abstinent smokers and nonsmokers. Inconsistent with this hypothesis, higher learning rates of sated smokers were found in Addicott et al. (2013). As discussed earlier, this pattern may be due to the volatile task (with six options and changing outcomes) in this study, in which participants might over-adjust their learning rates.

As a measure of choice randomness, the inverse temperature can also provide important information about mental disorders. In all studies with a significant group difference in this parameter, a consistent decreasing pattern can be found, suggesting that participants with these disorders were less likely to choose the option with the highest expected value. A lower inverse temperature can lead to lower accuracy and a higher number of trials to reach a learning criterion (Kunisato et al., 2012). Therefore, it is not surprising that, in all 10 studies with significant group differences in inverse temperature, five had a significant decrease in performance. This suggests a general inconsistency of choices across different disorders.

Additional parameters

The results of the model-agnostic performance and learning rates suggest learning inflexibility and a lack of representation of the environment across these three types of

disorders. Previous work has proposed three strategies in reward learning: model-based, model-free, and win-stay-lose-switch (WSLS; Daw, Niv, & Dayan, 2005; Worthy, Hawthorne, & Otto, 2013). Model-based learning involves acquiring the relationships between different states of the environment. Although very flexible, it requires high memory loads (Daw et al., 2005). Model-free learning does not involve a cognitive map of the environment. The learner only needs to track the reward history of different options and updates their expected values through trial-and-error (Sutton & Barto, 1998). WSLS strategy requires the least memory loads, in which the learner's choice in a trial is only based on the reward information from the previous trial (Worthy et al., 2013). Four of our reviewed studies specifically tested these strategies. Blanco et al. (2013) demonstrated that more participants with depression were better fitted by the basic RL model versus the model-based model than the nondepressed, suggesting less model-based learning in depressed participants, but Sebold et al. (2017) failed to find an altered balance between model-based and model-free strategies in detoxified alcohol abusers. Carlisi et al. (2017) and Norman et al. (2018) tested the balance between model-free and WSLS strategy and found that OCD teenagers relied more on WSLS compared to control participants (The results of additional parameters were summarized in **Tables S1.2**, **S1.4**, and **S1.6**). These results partially support our hypothesis and suggest impaired model-based learning in depression and impaired model-free RL in anxiety. This impairment also echoes previous studies showing memory deficits in mental disorders (Buckley, Blanchard, & Neill, 2000; Muller & Roberts, 2005; Rose & Ebmeier, 2006; Yan et al., 2014).

Another model-agnostic finding that can be tested is the attentional bias to rewards in addiction. This finding is strongly supported by studies applying the expectancy-valence learning model (Busemeyer & Stout, 2002), which has a parameter to measure the attentional weight allocated to rewards relative to punishments. Six addiction studies applied this parameter and four found significant group differences (Gullo & Stieger, 2011; Stout, Busemeyer, Lin, Grant, & Bonson, 2004; Tanabe et al., 2013; Yechiam et al., 2004), which consistently showed decreased attention to punishments (or more attention to rewards). This suggests that when facing both rewards and punishments, addicts tend to weigh rewards more than punishments, which is in agreement with their hypersensitivity to drugs despite potential harms (Robinson & Berridge, 2001).

In the previous section, we proposed a possible mechanism of higher tonic and lower phasic dopamine activity in the sated state but a reversed pattern in abstinence for addicts. Since tonic dopamine is related to motivation and the salience of environmental stimuli (Volkow, Fowler, Wang, & Swanson, 2004), subjective valuation of rewards may change in different conditions. This is partially supported by three studies incorporating reward and punishment sensitivity to their models (Ahn et al., 2014; Beylergil et al., 2017; Feng, 2017). In these models, the sensitivity parameter scales the received value. A higher sensitivity indicates a magnified subjective valuation of a stimulus. Feng (2017) found both increased reward and punishment sensitivity in sated smokers. Beylergil et al. (2017) and Ahn et al. (2014) found decreased punishment sensitivity in abstinent alcohol-dependent participants and abstinent heroin-dependent participants, respectively. An individual with unchanged reward sensitivity and decreased punishment sensitivity will pay more attention to rewards relative to punishments. It seems that this imbalance in

sensitivity is a likely explanation for the attentional bias to rewards in addiction, and it is possible that the attentional bias to non-drug rewards can also be modulated by abstinence.

As for model-agnostic findings in anxiety, there is also supporting evidence from model parameters. In Myers et al. (2013), the enhanced performance of PTSD participants seems puzzling, but parameter results can help to explain it. In the reward trials of their task, a smiling face with point gain signified correct choices; a no-feedback screen signified wrong choices. In the punishment trials, a frowning face with point loss signified wrong choices; a no-feedback screen signified correct choices. Since trial orders were randomized, the meaning of the no-feedback screen was ambiguous. The authors used a parameter to represent the value of the blank feedback and found it to be more neutral (closer to 0) among PTSD participants than controls. As a result, the control participants viewed the no-feedback screen as a positive value and made more wrong choices in the reward trials. It seems that PTSD participants made choices more optimally, but it may also suggest that they were less optimistic when interpreting ambiguous cues. In addition to the parameter suggesting less optimism, the hypersensitivity to punishments in PTSD was also evidenced by another parameter. Brown et al. (2018) employed a model with a dynamic learning rate modulated by an associability parameter, which was updated in each trial by a weight parameter and unsigned PE from the previous trial (Li, Delgado, & Phelps, 2011). This weight parameter during the loss condition was found to be higher in PTSD than control participants, suggesting a greater adjustment of loss learning after unexpected reward information (hypersensitivity to punishments).

Another feature of anxiety is facilitated Pavlovian fear conditioning and difficulty in extinguishing conditioned fear responses (Norrholm et al., 2011; Rabinak, Mori, Lyons, Milad, & Phan, 2017), which are associated with enhanced Pavlovian biases (e.g., excessive punishment avoidance). Two studies (Mkrtchian et al., 2017; Ousdal et al., 2018) used a Pavlovian probabilistic go/no-go task and employed a bias parameter to represent the Pavlovian bias. Ousdal et al. (2018) found enhanced bias among terror attack survivors, and the parameter was negatively associated with the number of days between times of testing and the traumatic event. Mkrtchian et al. (2017) found higher avoidance biases among depression and anxiety group when the participants were at the risk of electric shock but not in the safe condition. These results suggest that traumatic stress may induce long-lasting increases of Pavlovian biases, which may contribute to the avoidance behavior in anxiety disorders (Krypotos, Effting, Kindt, & Beckers, 2015) and a shift from goal-directed decision-making to more automatic forms of Pavlovian control (Schwabe & Wolf, 2013). Moreover, the results from these two studies also suggest that this shift may get alleviated as time lapses, but would be exacerbated when facing new stressors.

Imaging results of prediction error signals

The depression studies with imaging results showed a general pattern of decreased PE signals in reward regions of depressed participants (**Table S1.7**). Ten out the 11 fMRI studies compared PE signals between patients and controls. Five (Dombrovski et al., 2013; Frey & McCabe, 2020; Gradin et al., 2011; Kumar et al., 2018; Rothkirch, Tonn,

Köhler, & Sterzer, 2017) found only decreased, two (Geugies et al., 2019; Liu, Valton, Wang, Zhu, & Roiser, 2017) found only increased and one (Kumar et al., 2008) found both increased and decreased neural PE signals in depressives. As a core region of the dopaminergic reward system, the ventral striatum's coding of RPE signals was found to be attenuated for depressives in Kumar et al. (2008) and Gradin et al. (2011). In these two studies, water was used as the reinforcer and the participants were asked to refrain from drinking fluids from the night before the experiment. Therefore, they might be more engaged in the task and group differences could be more easily detected. As a complement to the decreased striatal RPE signals in Kumar et al. (2008) and Gradin et al. (2011), Liu et al. (2017) found increased APE activations in the ventral striatum of medication-free participants with depression, consistent with an earlier model-agnostic study (Ubl et al., 2015), in which medication-free depressives showed increased punishment-related PE signals in the right ventral striatum (Instead of being derived from RL models, the PEs in this study were the differences between actual rewards received by participants and potential rewards predicted by cues). These studies suggest that RPEs and APEs may be represented in an opposite manner in the striatum and this is in line with studies showing hyposensitivity to rewards and hypersensitivity to punishments in depression (Eshel & Roiser, 2010). They also suggest that stronger striatal activations might be more likely to be detected in medication-free depressives in that antidepressants might only dampen the exaggerated responses to punishments but keep hyposensitivity to rewards unchanged (Herzallah et al., 2013). Despite the inconsistent directions of the striatal coding of PEs, PE (including positive PE) signals in other reward regions still showed an attenuated pattern (**Table S1.7**), suggesting impairments in the dopaminergic reward system in depressives, which may be the neural substrates of their decreased model-agnostic performance and learning rates.

Similar to the depression studies, imaging results from addiction studies also showed a general pattern of decreased PE signals in regions implicated in reward learning (**Table S1.8**). Nine out the 10 fMRI studies compared PE signals between patients and controls. Five (Beylergil et al., 2017; Feng, 2017; Reiter et al., 2016; Tanabe et al., 2013; Wei et al., 2018) found only decreased and one (White et al., 2016) found both increased and decreased PE signals in addicts relative to controls. Three studies revealed PE differences in the region-of-interest (ROI) of the striatum (ventral striatum in Feng, 2017 and Tanabe et al., 2013; putamen in Wei et al., 2018), in which Feng (2017) and Wei et al. (2018) found decreased striatal PE signals in sated chronic smokers. As we proposed above, satiety may lead to impaired coding of PEs in the ventral striatum, which could be normalized in abstinence. In line with this hypothesis, several studies comparing abstinent addicts (mostly with alcohol dependence/abuse) and controls failed to find group differences in the ventral striatum PE signals even with lenient threshold or with ROI analysis (Beylergil et al., 2017; Deserno et al., 2015; Park et al., 2010; White et al., 2016). Inconsistent with this hypothesis, Tanabe et al. (2013) found diminished PE signals in the ventral striatum for abstinent polydrug users. The authors also mentioned that the substance use problems in their population were more severe than in previous studies and they used a more sensitive MRI pulse sequence (Tanabe et al., 2013). Therefore, it is possible that, the decreased PE signals can only be fully normalized by abstinence in less severe addictions, but in addicts with severe symptoms, abstinence may have limited effects.

The directions of PE signals in anxiety imaging studies were mixed (**Table S1.9**). Six out the nine fMRI studies compared PE signals between patients and controls. Two found weaker encoding of PEs in the salience network (Cisler et al., 2019), the ventral striatum/medial prefrontal cortex network, and the anterior insula network (Ross et al., 2018) in anxious patients, suggesting generally impaired learning. Three studies showed increased PE signals in the patients (Cisler et al., 2015; Hauser et al., 2017; Murray et al., 2019). In Cisler et al. (2015), increased PE signals in the temporoparietal junction (TPJ) were detected for PTSD patients when they made choices between two neutral face images with faces smiling or frowning as outcomes. Since the TPJ is implicated in theory of mind (Frith & Frith, 2006; Mar, 2011; Saxe & Kanwisher, 2003), this result reflects the heightened mentalizing of the patients when facing unexpected social rewards. In Hauser et al. (2017) and Murray et al. (2019), the increased PE signals in the anterior cingulate cortex and striatum among OCD participants may also suggest hypersensitivity to unexpected outcomes, which is consistent with the fMRI (Stern et al., 2011; Ursu, Stenger, Shear, Jones, & Carter, 2003) and electroencephalogram studies (Endrass & Ullsperger, 2014; Gehring, Himle, & Nisenson, 2000; Kathmann, Endrass, Klawohn, Grützmann, & Riesel, 2016) showing increased responses to error processing in OCD patients.

SUMMARY

In this work, we made a systematic review of RL studies in depression, addiction, and anxiety. Despite highly heterogeneous samples, experimental paradigms, RL models, and analysis methods, we still identified meaningful patterns. In general, participants with mental disorders performed worse and made more random choices (higher inverse temperatures). Poorer performance was more readily detected in more difficult tasks with contingency changes or reversals. In these tasks, participants tend to attend to the most recent information rather than past reinforcement history (higher learning rates), which may be optimal in a fast-changing environment. But participants with mental disorders may overreact and have overly high learning rates, suggesting learning inflexibility and possibly more reliance on model-free RL strategy or even simpler WSLS strategy. Since only two studies directly tested group differences in model-free and model-based learning, it is too early to draw firm conclusions.

RL alterations were also identified for each type of disorder. In depression, learning rates were found to be decreased in several studies and seemed to be attenuated in both reward and punishment domains. Consistent with decreased learning rates, impaired RPE signals and positive PE signals were also found in key reward regions. Meanwhile, increased APE signals in depressives may suggest enhanced learning for punishments. In addiction studies, the addicts had an increased attentional weight to rewards versus punishments. Moreover, they showed higher sensitivity to both rewards and punishments when sated but had decreased sensitivity to losses when in abstinence, which may contribute to the biased attentional weight to rewards. Similar to the neural pattern in the depression studies, addicted participants had weaker neural PE signals in key reward regions. In anxiety studies, anxious participants tended to view ambiguous stimuli less positively, be more concerned with punishments, and have stronger Pavlovian biases. Neurally, some of studies showed weaker encoding of the PE signals in anxious participants, but others demonstrated hypersensitivity to unexpected outcomes in PTSD and OCD patients.

This review mainly utilized common RL model parameters to make connections across different studies within each disorder as well as across three different types of disorders. The similarities and differences of RL disturbances were summarized, which may provide testable hypotheses for future research and may serve as a reference for other studies. Despite these meaningful findings, limitations should also be noted: (i) Due to the relatively small size of studies in each type of disorders and high heterogeneity in different studies, we could only conduct a qualitative review, which may limit its generalization; (ii) Many of the reviewed studies had participants with comorbid disorders and the results may reflect combined effects of more than one mental illness. Future empirical studies are recommended to use a factorial design, to investigate the effects of separate disorders as well as their combined effects on RL; (iii) we only summarized direct comparison results of overall performance, model parameters, and neural PE signals between a disorder group and a control group. Other important results (e.g., correlations, neural expected value signals) were not fully considered. Future reviews should pay more attention to these results, which may provide more new insights into the RL deficits in mental disorders; (vi) This review only focused on RL studies in depression, addiction, and anxiety disorders. Other types of computational models and other mental disorders were not included. Future studies are suggested to summarize studies with computational models applied on other economic games (e.g., trust game, prisoner's dilemma, etc.) as well as other mental disorders (schizophrenia, Parkinson's disease, etc.)

TABLES

Table 1.1. A summary of behavioral results in all included studies

Study	Participants in the disorder group	N	Mean age	Options	Change of contin-gency	Perfor-mance[a]	LR	IT
Depression								
Aylward et al., 2019	MDD/GAD/PAD[b]	44	29.0	4	C	ns	↑	NA[c]
Bakic et al., 2016	MDD	35	43.0	2	n	↓	ns	ns
Blanco et al., 2013	Nonclinical, depressed	38	NA	2	R	↓	NA	NA
Brown, 2018	MDD/dysthymia[b]	69	35.4	2	n	↓	NA	NA
Chase et al., 2010	MDD	23	46.2	2	n	ns	↓	ns
Dombrovski et al., 2010	MDD[d]	51	68.8	2	R	↓	↓	ns
Dombrovski et al., 2013	MDD[b, d]	33	66.3	2	R	↓	↓	↓
Feng, 2017[e]	MDD/dysthymia[b, d]	50	35.9	2	n	ns	ns	NA
Frey et al., 2019	Nonclinical, depressed	40	25.3	2	n	↓	↓	ns
Frey & McCabe, 2019	Nonclinical, depressed	21	23.2	6	n	↓	↓	NA
Geugies et al., 2019	MDD	36	47.0	2	n	ns	NA	NA
Gradin et al., 2011	MDD	15	45.3	2	n	NA	ns	ns
Kumar et al., 2018	MDD[b]	25	25.3	2	n	↓	ns	ns
Kumar et al., 2008	MDD	15	45.3	2	C	ns	NA	NA
Kunisato et al., 2012	Nonclinical, depressed	18	19.4	2	n	ns	ns	↓
Liu et al., 2017	MDD	21	30.7	2	n	ns	ns	ns
Mkrtchian et al., 2017	MDD/GAD/PAD[b]	43	28.8	4	n	↓	ns	NA
Moutoussis et al., 2018	MDD	39	34.1	2	n	ns	ns	NA
Rothkirch et al., 2017	MDD	28	36.3	2	n	ns	ns	ns
Rouhani et al., 2019	Nonclinical, depressed	101	NA	2	n	ns	ns	NA
Rupprechter et al., 2018	MDD	15	NA	2	n	NA	NA	↓
Addiction								
Addicott et al., 2013	Smokers (sated)	18	36.0	6	C	ns	↑	NA
Ahn et al., 2014	P. amphetamine/heroin dep.[d]	81	26.4	4 (IGT)	R	↓	↑	ns
Beylergil et al., 2017	Alcohol dependent	34	44.7	2	R	↓	ns	NA
Deserno et al., 2015	Detoxified alcohol abusers	13	45.1	2	R	↓	NA	NA
Feng, 2017[e]	Smokers (sated)[d]	45	36.3	2	n	ns	↓	NA
Gullo & Stieger, 2011	Heavy drinkers	22	21.0	4 (IGT)	n	↓	ns	ns
Kanen et al., 2019	Stimulant use disorder	17	34.3	2	R	↑	↓↑[f]	NA
Lesage et al., 2017	Smokers (abstinent)	24	35.8	2	R	↓	ns	↓
Lim et al., 2019	Cocaine addicts	68	38.0	2	n	NA	↓	ns
Mazas et al., 2000[g]	Alcohol abusers	27	21.6	4 (IGT)	n	↓	ns	ns
Myers et al., 2016	Heroin dependent	45	41.2	2	n	ns	ns	ns
Park et al., 2010	Alcohol abusers	20	42.5	2	R	↓	ns	ns
Reiter et al., 2016	Alcohol abusers	43	44.4	2	R	↓	ns	ns
Sebold et al., 2017	Detoxified alcohol abusers[d]	90	45.4	2	C	NA	ns	ns
Stout et al., 2004	Chronic cocaine abusers	12	36.9	4 (IGT)	n	↓	ns	↓
Tanabe et al., 2013	Polydrug abusers	16	15.6	4 (IGT)	n	↓	ns	↓
Wei et al., 2018	Smokers (sated)	38	24.5	4 (IGT)	n	↓	↓	NA
White et al., 2016	P. alcohol/cannabis abusers	32	36.6	4 (IGT)	n	↓	NA	NA
Yechiam et al., 2004[g]	Chronic cannabis abusers	25	NA	4 (IGT)	n	NA	↑	NA
Yechiam et al., 2005b[g]	Polydrug abusers	39	NA	4 (IGT)	n	NA	ns	ns

Study	Participants in the disorder group	N	Mean age	Options	Change of contin-gency	Perfor-mance [a]	LR	IT
Anxiety								
Aylward et al., 2019	GAD/MDD/PAD[b]	44	29.0	4	C	ns	↑	NA
Brown et al., 2018	PTSD[h]	39	32.3	2	n	ns	ns	ns
Carlisi et al., 2017	OCD	20	15.7	4 (IGT)	n	ns	ns	↓
Cisler et al., 2015	PTSD (GAD/MDD)[h]	25	34.7	2	C	ns	ns	ns
Cisler et al., 2019	Assault victims[b, d]	30	15.7	3	C	↓	ns	↓
Hauser et al., 2017	OCD	33	23.4	2	R	ns	ns	ns
Huang et al., 2017	Patients with high anxiety[b]	77	35.0[i]	3	C	ns	↑	ns
Kanen et al., 2019	OCD	18	35.4	2	R	↑	↑	NA
Khdour et al., 2016	GAD/SAD/PAD[d]	55	43.0	2	n	↑	ns	ns
Mkrtchian et al., 2017	MDD/GAD/PAD[b]	43	28.8	4	n	↓	ns	NA
Murray et al., 2019	OCD	18	32.1	2	n	↓	↓	ns
Myers et al., 2013	Nonclinical, Severe PTSD	48	54.0	2	n	↑	ns	ns
Norman et al., 2018	OCD	20	15.8	4 (IGT)	n	ns	ns	↓
Ousdal et al., 2017	Terror attack survivors[b]	25	19.6	4	n	↓	ns	NA
Piray et al., 2019	Nonclinical, high anxiety	21	NA	2	R	↓	ns	ns
Ross et al., 2018	PTSD (GAD/MDD)[h]	15	31.1	2	C	ns	ns	↓
Vaghi et al., 2017	OCD	24	41.3	NA	C	NA	↑	NA

Abbreviations: C, tasks with changing contingencies; dep., dependent; GAD, general anxiety disorder; IGT, Iowa gambling task; IT, inverse temperature; LR, learning rate; MDD, major depressive disorder; N, number of participants in disorder group; NA, not available; n, no contingency; ns, not significant; OCD, obsessive-compulsive disorder; P., past; PAD, panic anxiety disorder; PTSD, post-traumatic stress disorder; R, tasks with reversals of contingencies; SAD, social anxiety disorder; ↑, increased value in the disorder group relative to the control group; ↓, decreased value in the disorder group relative to the control group.

[a] Performance measures include mainly accuracy (see **Tables S1.2**, **S1.4**, and **S1.6** for details). Response time is not included as a measure of performance. [b] This group included both depressed and anxious participants. [c] NA is used in the table when the value of a variable was not provided, a model parameter was not used or estimated, or a between-group test was not conducted in the study. [d] This group consisted of two or more subgroups. ↓ and ↑ for these studies indicate any difference in a variable between a disorder subgroup and the control group (e.g., LR in Dombrovski et al., 2010) or subgroups with less severe symptoms (e.g., LR in Dombrovski et al., 2013). [e] The sample of this study included both depressed and addicted participants. When this study is discussed as a depression study, only group differences between depressed and nondepressed participants are considered. When this study is discussed as an addiction study, only group differences between addicted and nonaddicted participants are considered. [f] In this study, patients had increased LRs for positive outcomes/prediction errors (PEs) and decreased LRs for negative outcomes/PEs. [g] The data from this study were reanalyzed with RL models by Yechiam et al. (2005a). [h] All participants in this group had a diagnosis of PTSD and some of them had comorbid depression or other anxiety disorders. [i] This is the mean age for all participants in this study.

REFERENCES

Addicott, M. A., Baranger, D. A., Kozink, R. V., Smoski, M. J., Dichter, G. S., & McClernon, F. J. (2012). Smoking withdrawal is associated with increases in brain activation during decision making and reward anticipation: A preliminary study. *Psychopharmacology, 219*(2), 563-573.

Addicott, M. A., Pearson, J. M., Wilson, J., Platt, M. L., & McClernon, F. J. (2013). Smoking and the bandit: A preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and Clinical Psychopharmacology, 21*(1), 66-73.

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology, 5*, 1-15.

Aupperle, R., L, & Martin, P. P. (2010). Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues in Clinical Neuroscience, 12*(4), 517-531.

Aylward, J., Valton, V., Ahn, W.-Y., Bond, R. L., Dayan, P., Roiser, J. P., & Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nature Human Behaviour, 3*(10), 1116-1123.

Bakic, J., Pourtois, G., Jepma, M., Duprat, R., De Raedt, R., & Baeken, C. (2017). Spared internal but impaired external reward prediction error signals in major depressive disorder during reinforcement learning. *Depression and Anxiety, 34*(1), 89-96.

Beats, B., Sahakian, B. J., & Levy, R. (1996). Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. *Psychological Medicine, 26*(3), 591-603.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience, 21*(8), 2793-2798.

Beylergil, S. B., Beck, A., Deserno, L., Lorenz, R. C., Rapp, M. A., Schlagenhauf, F., . . . Obermayer, K. (2017). Dorsolateral prefrontal cortex contributes to the impaired behavioral adaptation in alcohol dependence. *NeuroImage: Clinical, 15*, 80-94.

Blanco, N. J., Otto, A. R., Maddox, W. T., Beevers, C. G., & Love, B. C. (2013). The influence of depression symptoms on exploratory decision-making. *Cognition, 129*(3), 563-568.

Brown, V. (2018). *Assessing and remediating altered reinforcement learning in depression.* (Doctor of Philosophy), Virginia Polytechnic Institute and State University, Blacksburg, VA. .

Brown, V. M., Zhu, L., Wang, J. M., Frueh, B. C., King-Casas, B., & Chiu, P. H. (2018). Associability-modulated loss learning is increased in posttraumatic stress disorder. *eLife, 7*, e30150.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience, 18*(4), 590-596.

Buckley, T. C., Blanchard, E. B., & Neill, W. T. (2000). Information processing and PTSD: A review of the empirical literature. *Clinical Psychology Review, 20*(8), 1041-1065.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253-262.

Carlisi, C. O., Norman, L., Murphy, C. M., Christakou, A., Chantiluke, K., Giampietro, V., . . . Mataix-Cols, D. (2017). Shared and disorder-specific neurocomputational mechanisms of decision-making in autism spectrum disorder and obsessive-compulsive disorder. *Cerebral Cortex, 27*(12), 5804-5816.

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychological Medicine, 40*(3), 433-440.

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neuroscience and Biobehavioral Reviews, 55*, 247-267.

Cisler, J. M., Bush, K., Steele, J. S., Lenow, J. K., Smitherman, S., & Kilts, C. D. (2015). Brain and behavioral evidence for altered social learning mechanisms among women with assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 63*, 75-83.

Cisler, J. M., Esbensen, K., Sellnow, K., Ross, M., Weaver, S., Sartin-Tarm, A., . . . Kilts, C. D. (2019). Differential roles of the salience network during prediction error encoding and facial emotion processing among female adolescent assault victims. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4*(4), 371-380.

Cremers, H. R., Veer, I. M., Spinhoven, P., Rombouts, S. A., & Roelofs, K. (2015). Neural sensitivity to social reward and punishment anticipation in social anxiety disorder. *Frontiers in Behavioral Neuroscience, 8*, 1-9.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704-1711.

Deserno, L., Beck, A., Huys, Q. J., Lorenz, R. C., Buchert, R., Buchholz, H. G., . . . Heinze, H. J. (2015). Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. *European Journal of Neuroscience, 41*(4), 477-486.

Dombrovski, A. Y., Clark, L., Siegle, G. J., Butters, M. A., Ichikawa, N., Sahakian, B. J., & Szanto, K. (2010). Reward/punishment reversal learning in older suicide attempters. *American Journal of Psychiatry, 167*(6), 699-707.

Dombrovski, A. Y., Szanto, K., Clark, L., Reynolds, C. F., & Siegle, G. J. (2013). Reward signals, attempted suicide, and impulsivity in late-life depression. *JAMA Psychiatry, 70*(10), 1020-1030.

Endrass, T., & Ullsperger, M. (2014). Specificity of performance monitoring changes in obsessive-compulsive disorder. *Neuroscience and Biobehavioral Reviews, 46*, 124-138.

Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry, 68*(2), 118-124.

Feng, S. (2017). *Association between reward sensitivity and smoking status in major depressive disorder*. (Master of Science), Virginia Polytechnic Institute and State University, Blacksburg, VA.

Fladung, A.-K., Grön, G., Grammer, K., Herrnberger, B., Schilly, E., Grasteit, S., . . . von Wietersheim, J. (2009). A neural signature of anorexia nervosa in the ventral striatal reward system. *American Journal of Psychiatry, 167*(2), 206-212.

Forbes, E. E., Hariri, A. R., Martin, S. L., Silk, J. S., Moyles, D. L., Fisher, P. M., . . . Axelson, D. A. (2009). Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *American Journal of Psychiatry, 166*(1), 64-73.

Freeman, T. P., Morgan, C. J. A., Beesley, T., & Curran, H. V. (2012). Drug cue induced overshadowing: selective disruption of natural reward processing by cigarette cues amongst abstinent but not satiated smokers. *Psychological Medicine, 42*(1), 161-171.

Frey, A.-L., Frank, M. J., & McCabe, C. (2019). Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. *Psychological Medicine*, 1-8.

Frey, A.-L., & McCabe, C. (2020). Impaired social learning predicts reduced real-life motivation in individuals with depression: A computational fMRI study. *Journal of Affective Disorders, 263*, 698-706.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

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

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science, 11*(1), 1-6.

Geugies, H., Mocking, R. J., Figueroa, C. A., Groot, P. F., Marsman, J.-B. C., Servaas, M. N., . . . Ruhé, H. G. (2019). Impaired reward-related learning signals in remitted unmedicated patients with recurrent depression. *Brain, 142*(8), 2510-2522.

Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction, 95*, S119–S128.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., . . . Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain, 134*(6), 1751-1764.

Guitart-Masip, M., Huys, Q. J., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage, 62*(1), 154-166.

Gullo, M. J., & Stieger, A. A. (2011). Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. *Drug and Alcohol Dependence, 117*(2), 204-210.

Hauser, T. U., Iannaccone, R., Dolan, R., Ball, J., Hättenschwiler, J., Drechsler, R., . . . Brem, S. (2017). Increased fronto-striatal reward prediction errors moderate decision making in obsessive–compulsive disorder. *Psychological Medicine, 47*(7), 1246-1258.

Herzallah, M. M., Moustafa, A. A., Natsheh, J. Y., Abdellatif, S. M., Taha, M. B., Tayem, Y. I., . . . Myers, C. E. (2013). Learning from negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. *Frontiers in Integrative Neuroscience, 7*, 1-9.

Hirschfeld, R. M. (2001). The comorbidity of major depression and anxiety disorders: Recognition and management in primary care. *Primary Care Companion to the Journal of Clinical Psychiatry, 3*(6), 244-254.

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational dysfunctions in anxiety: Failure to differentiate signal from noise. *Biological Psychiatry, 82*(6), 440-446.

Kanen, J. W., Ersche, K. D., Fineberg, N. A., Robbins, T. W., & Cardinal, R. N. (2019). Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. *Psychopharmacology, 236*(8), 2337-2358.

Kathmann, N., Endrass, T., Klawohn, J., Grützmann, R., & Riesel, A. (2016). Error-related brain activity as an endophenotype of obsessive-compulsive disorder. *European Neuropsychopharmacology, 26*(5), 894-895.

Khdour, H. Y., Abushalbaq, O. M., Mughrabi, I. T., Imam, A. F., Gluck, M. A., Herzallah, M. M., & Moustafa, A. A. (2016). Generalized anxiety disorder and social anxiety disorder, but not panic anxiety disorder, are associated with higher sensitivity to learning from negative feedback: Behavioral and computational investigation. *Frontiers in Integrative Neuroscience, 10*, 1-11.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport, 12*(17), 3683-3687.

Kobayakawa, M., Tsuruya, N., & Kawamura, M. (2010). Sensitivity to reward and punishment in Parkinson's disease: an analysis of behavioral patterns using a modified version of the Iowa gambling task. *Parkinsonism & Related Disorders, 16*(7), 453-457.

Krypotos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience, 9*, 1-16.

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., . . . Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology, 43*(7), 1581-1588.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain, 131*(8), 2084-2093.

Kunisato, Y., Okamoto, Y., Ueda, K., Onoda, K., Okada, G., Yoshimura, S., . . . 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*(4), 1088-1094.

Leknes, S., & Tracey, I. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience, 9*(4), 314-320.

Lesage, E., Aronson, S. E., Sutherland, M. T., Ross, T. J., Salmeron, B. J., & Stein, E. A. (2017). Neural signatures of cognitive flexibility and reward sensitivity following nicotinic receptor stimulation in dependent smokers: a randomized trial. *JAMA Psychiatry, 74*(6), 632-640.

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences, 108*(1), 55-60.

Li, J., McClure, S. M., King-Casas, B., & Montague, P. R. (2006). Policy adjustment in a dynamic economic game. *PloS One, 1*(1), 1-11.

Lim, T. V., Cardinal, R. N., Savulich, G., Jones, P. S., Moustafa, A. A., Robbins, T., & Ersche, K. D. (2019). Impairments in reinforcement learning do not explain enhanced habit formation in cocaine use disorder. *Psychopharmacology, 236*(8), 2359-2371.

Liu, W.-H., Valton, V., Wang, L.-Z., Zhu, Y.-H., & Roiser, J. P. (2017). Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. *Social Cognitive and Affective Neuroscience, 12*(9), 1520-1533.

Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.

Luijten, M., Schellekens, A. F., Kühn, S., Machielse, M. W., & Sescousse, G. (2017). Disruption of reward processing in addiction: An image-based meta-analysis of functional magnetic resonance imaging studies. *JAMA Psychiatry, 74*(4), 387-398.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience, 14*(2), 154-162.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103-134.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology, 626*(1), 83-86.

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-making biases, antisocial personality, and early-onset alcoholism. *Alcoholism: Clinical and Experimental Research, 24*(7), 1036-1040.

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., & Robinson, O. J. (2017). Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biological Psychiatry, 82*(7), 532-539.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16*(1), 72-80.

Muller, J., & Roberts, J. E. (2005). Memory and attention in obsessive–compulsive disorder: a review. *Journal of Anxiety Disorders, 19*(1), 1-28.

Murray, G. K., Knolle, F., Ersche, K. D., Craig, K. J., Abbott, S., Shabbir, S. S., . . . Bullmore, E. T. (2019). Dopaminergic drug treatment remediates exaggerated cingulate prediction error responses in obsessive-compulsive disorder. *Psychopharmacology, 236*(8), 2325-2336.

Myers, C. E., Moustafa, A. A., Sheynin, J., VanMeenen, K. M., Gilbertson, M. W., Orr, S. P., . . . Servatius, R. J. (2013). Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: Empirical data and computational model. *PloS One, 8*(8), 1-13.

Myers, C. E., Sheynin, J., Balsdon, T., Luzardo, A., Beck, K. D., Hogarth, L., . . . Moustafa, A. A. (2016). Probabilistic reward-and punishment-based learning in opioid addiction: Experimental and computational data. *Behavioural Brain Research, 296*, 240-248.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology, 53*(3), 139-154.

Norman, L. J., Carlisi, C. O., Christakou, A., Murphy, C. M., Chantiluke, K., Giampietro, V., . . . Rubia, K. (2018). Frontostriatal dysfunction during decision making in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(8), 694-703.

Norrholm, S. D., Jovanovic, T., Olin, I. W., Sands, L. A., Bradley, B., & Ressler, K. J. (2011). Fear extinction in traumatized civilians with posttraumatic stress disorder: Relation to symptom severity. *Biological Psychiatry, 69*(6), 556-563.

Nutt, D. J., Lingford-Hughes, A., Erritzoe, D., & Stokes, P. R. (2015). The dopamine theory of addiction: 40 years of highs and lows. *Nature Reviews Neuroscience, 16*(5), 305-312.

O'doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*(1), 35-53.

Ormel, J., VonKorff, M., Ustun, T. B., Pini, S., Korten, A., & Oldehinkel, T. (1994). Common mental disorders and disability across cultures: results from the who collaborative study on psychological problems in general health care. *JAMA, 272*(22), 1741-1748.

Ousdal, O., Huys, Q., Milde, A., Craven, A., Ersland, L., Endestad, T., . . . Dolan, R. (2018). The impact of traumatic stress on Pavlovian biases. *Psychological Medicine, 48*(2), 327-336.

Park, S. Q., Kahnt, T., Beck, A., Cohen, M. X., Dolan, R. J., Wrase, J., & Heinz, A. (2010). Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *Journal of Neuroscience, 30*(22), 7749-7753.

Piray, P., Ly, V., Roelofs, K., Cools, R., & Toni, I. (2019). Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals. *Journal of Neuroscience, 39*(8), 1445-1456.

Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., & Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *American Journal of Psychiatry, 166*(6), 702-710.

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological Psychiatry, 57*(4), 319-327.

Plichta, M. M., & Scheres, A. (2014). Ventral–striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: A meta-analytic review of the fMRI literature. *Neuroscience and Biobehavioral Reviews, 38*, 125-134.

Quello, S. B., Brady, K. T., & Sonne, S. C. (2005). Mood disorders and substance use disorder: A complex comorbidity. *Science & Practice Perspectives, 3*(1), 13-21.

Rabinak, C. A., Mori, S., Lyons, M., Milad, M. R., & Phan, K. L. (2017). Acquisition of CS-US contingencies during Pavlovian fear conditioning and extinction in social anxiety disorder and posttraumatic stress disorder. *Journal of Affective Disorders, 207*, 76-85.

Reiter, A. M. F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A., & Schlagenhauf, F. (2016). Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *Journal of Neuroscience, 36*(43), 10935-10948.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Robinson, T. E., & Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction, 96*(1), 103-114.

Rose, E. J., & Ebmeier, K. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders, 90*(2-3), 149-161.

Ross, M. C., Lenow, J. K., Kilts, C. D., & Cisler, J. M. (2018). Altered neural encoding of prediction errors in assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 103*, 83-90.

Rothkirch, M., Tonn, J., Köhler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain, 140*(4), 1147-1157.

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology, 236*(8), 2425-2435.

Rupprechter, S., Stankevicius, A., Huys, Q. J., Steele, J. D., & Seriès, P. (2018). Major depression impairs the use of reward values for decision-making. *Scientific Reports, 8*, 13798.

Santesso, D. L., Steele, K. T., Bogdan, R., Holmes, A. J., Deveney, C. M., Meites, T. M., & Pizzagalli, D. A. (2008). Enhanced negative feedback responses in remitted depression. *Neuroreport, 19*(10), 1045-1048.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences, 30*(5), 203-210.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends in Cognitive Sciences, 17*(2), 60-68.

Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D. J., Beck, A., . . . Neu, P. (2017). When habits are dangerous: Alcohol expectancies and habitual decision making predict relapse in alcohol dependence. *Biological Psychiatry, 82*(11), 847-856.

Seeman, P., & Madras, B. K. (1998). Anti-hyperactivity medication: Methylphenidate and amphetamine. *Molecular Psychiatry, 3*(5), 386-396.

Smith, J. P., & Book, S. W. (2008). Anxiety and substance use disorders: A review. *The Psychiatric Times, 25*(10), 19-23.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology, 43*(2), 476-493.

Stern, E. R., Welsh, R. C., Fitzgerald, K. D., Gehring, W. J., Lister, J. J., Himle, J. A., . . . Taylor, S. F. (2011). Hyperactive error responses and altered connectivity in ventromedial and frontoinsular cortices in obsessive-compulsive disorder. *Biological Psychiatry, 69*(6), 583-591.

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review, 11*(4), 742-747.

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

Tanabe, J., Reynolds, J., Krmpotich, T., Claus, E., Thompson, L. L., Du, Y. P., & Banich, M. T. (2013). Reduced neural tracking of prediction error in substance-dependent individuals. *American Journal of Psychiatry, 170*(11), 1356-1363.

Ubl, B., Kuehner, C., Kirsch, P., Ruttorf, M., Diener, C., & Flor, H. (2015). Altered neural reward and loss processing and prediction error signalling in depression. *Social Cognitive and Affective Neuroscience, 10*(8), 1102-1112.

Ursu, S., Stenger, V. A., Shear, M. K., Jones, M. R., & Carter, C. S. (2003). Overactive action monitoring in obsessive-compulsive disorder: evidence from functional magnetic resonance imaging. *Psychological Science*, 14(4), 347-353.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron, 96*, 348-354.

Volkow, N. D., Fowler, J. S., Wang, G.-J., & Swanson, J. M. (2004). Dopamine in drug abuse and addiction: Results from imaging studies and treatment implications. *Molecular Psychiatry, 9*(6), 557-569.

Volkow, N. D., Fowler, J. S., Wang, G.-J., Swanson, J. M., & Telang, F. (2007). Dopamine in drug abuse and addiction: Results of imaging studies and treatment implications. *Archives of Neurology, 64*(11), 1575-1579.

Vriends, N., Michael, T., Schindler, B., & Margraf, J. (2012). Associative learning in flying phobia. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(2), 838-843.

Wang, J. M., Zhu, L., Brown, V. M., De La Garza II, R., Newton, T., King-Casas, B., & Chiu, P. H. (2019). In cocaine dependence, neural prediction errors during loss avoidance are increased with cocaine deprivation and predict drug use. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4*(3), 291-299.

Wei, Z., Han, L., Zhong, X., Liu, Y., Zha, R., Wang, Y., . . . Wang, W. (2018). Chronic nicotine exposure impairs uncertainty modulation on reinforcement learning in anterior cingulate cortex and serotonin system. *Neuroimage, 169*, 323-333.

White, S. F., Tyler, P., Botkin, M. L., Erway, A. K., Thornton, L. C., Kolli, V., . . . Blair, R. J. (2016). Youth with substance abuse histories exhibit dysfunctional representation of expected value during a passive avoidance task. *Psychiatry Research: Neuroimaging, 257*, 17-24.

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review, 20*(2), 364-371.

Yan, W.-S., Li, Y.-H., Xiao, L., Zhu, N., Bechara, A., & Sui, N. (2014). Working memory and affective decision-making in addiction: a neurocognitive comparison between heroin addicts, pathological gamblers and healthy controls. *Drug and Alcohol Dependence, 134*, 194-200.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005a). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science, 16*(12), 973-978.

Yechiam, E., Stout, J., Lamborn, C., Mussat-Whitlow, B., Liguori, A., & Porrino, L. (2004). *Differential influence of marijuana on decision processes in the Bechara gambling task.* Paper presented at the Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Yechiam, E., Stout, J. C., Busemeyer, J. R., Rock, S. L., & Finn, P. R. (2005b). Individual differences in the response to forgone payoffs: An examination of high functioning drug abusers. *Journal of Behavioral Decision Making, 18*(2), 97-110.

Reinforcement Learning Dysfunctions in Depression, Addiction, and Anxiety Disorders: A Systematic Review

SUPPLEMENTARY INFORMATION

Supplementary methods

The search used a combination of disorder-related keywords and reinforcement learning-related keywords in titles and abstracts. The depression-related keywords were set as ("depression" OR "depressive" OR "depressed" OR "anhedonia"). The addiction-related keywords were set as ("drug" OR "substance" OR "dependent" OR "abuse" OR "abuser" OR "addiction" OR "addictive" OR "addict" OR "addicted" OR "smoking" OR "smoker" OR "nicotine" OR "alcohol" OR "cocaine" OR "opioid" OR "cannabis" OR "amphetamine"). For anxiety studies, DSM-IV categorization of anxiety disorders was used, and the anxiety-related keywords were set as ["anxiety" OR "stress" OR "post-traumatic stress disorder (PTSD)" OR "obsessive-compulsive disorder (OCD)"]. The learning-related keywords were set as ["learning" AND ("reinforcement learning" OR "reward learning" OR "learning rate" OR "inverse temperature" OR "prediction error" OR "predictive error")]. The yielded articles for each search were checked and only human studies using reinforcement learning models were included in our pool. Both imaging and behavioral studies were included.

Supplementary tables

Table S1.1. Characteristics of participants in depression studies

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the depression group**	**Comorbidities of the depression group**	**Clinical measures of the depression group Mean (SD)**
Aylward et al., 2019	Participants with anxiety & mood disorders and healthy controls	D: 44 (28 f) C: 88 (50 f)	29 (8.7) 23 (5.1)	None	28 with GAD & MDD, 8 with GAD, 3 with PAD & MDD, 5 with MDD alone	20 (9.4) **BDI** (NA)[a] 47 (10.7) **STAI-S** (NA) 57 (8.2) **STAI-T** (NA)
Bakic et al., 2016	Participants with treatment resistant MDD and healthy controls	D: 35 (27 f) C: 44 (28 f)	43.00 (11.67) 37.89 (12.23)	Free from antidepressants, neuroleptics and mood stabilizers for at least 2 weeks	None	30.21 (10.27) **BDI-II** (Beck et al., 1996) 4.66 (2.25) **BDI-II** anhedonia (Becket al., 1996) 21.83 (5.63) **HDRS-17** (Hamilton, 1967) 58.97 (17.04) **TEPS** (Gard et al., 2006) 28.76 (9.02) **TEPS** consummatory (Gard et al., 2006) 30.21 (8.95) **TEPS** inhibitory (Gard et al., 2006) 7.31 (4.09) **SHAPS** (Snaith et al., 1995)
Blanco et al., 2013	Depressed and nondepressed individuals (A score of 16 on CESD as the cutoff)	D: 38 (NA) C: 95 (NA)	Young college students	NA	Anxiety symptoms	≥ 16 **CESD** (Radloff, 1977)
Brown, 2018	Participants with depression and healthy controls	D: 69 (49 f) C: 32 (20 f)	35.42 (11.46) 32.38 (10.75)	Some participants receiving medications	63 with MDD; 3 with dysthymia; 3 with MDD & dysthymia;	31.16 (8.00) **BDI-II** (Beck et al., 1996) 83.83 (9.63) **MASQ** anhedonia (Watson & Clark, 1991) 45.99 (9.05) **MASQ** general distress (Watson & Clark, 1991)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the depression group**	**Comorbidities of the depression group**	**Clinical measures of the depression group Mean (SD)**
					nicotine dependence; anxiety disorder	
Chase et al., 2010	Participants with MDD and healthy controls	D: 23 (10 f) C: 23 (11 f)	46.22 (2.25) 47.74 (2.14)	Receiving medications	Anxiety symptoms	26.78 (1.79) **BDI** (Beck et al., 1961) 32.39 (2.35) **BAI** (Beck et al., 1988) 26.17 (1.50) **MADRS** (NA) 40.27 (1.19) **SHAPS** (Snaith et al., 1995)
Dombrovski et al., 2010	Participants with MDD and healthy controls	DSA: 15 (9 f) DSI: 12 (5 f) DNS: 24 (13 f) C: 14 (2 f)	66.8 (7.8) 68.8 (5.6) 70.0 (7.4) 65.6 (4.9)	Receiving medications	Substance use	DSA: 19.2 (4.7), DSI: 20.2 (4.7), DNS: 18.5 (3.7) **HDRS** (Hamilton, 1960)
Dombrovski et al., 2013	Participants with MDD and healthy controls	DSA: 15 (7 f) DNS: 18 (12 f) C: 20 (12 f)	65.9 (6.3) 66.7 (5.7) 70.7 (8.7)	Receiving medications	Substance use; anxiety disorder	DSA: 12.9 (8.7), DNS: 11.1 (6.2) **HDRS** (Hamilton, 1960) DSA:7.9 (3.9), DNS: 6.7 (4.1) **EXIT** (Royall et al., 1992)
Feng, 2017	Smokers and nonsmokers with and without depression	DS: 25 (13 f) CS: 20 (3 f) DN: 25 (17 f) CN: 32 (21 f)	36.64 (1.91) 35.80 (2.25) 35.24 (2.38) 33.75 (2.08)	10 depressed participants taking antidepressants/ antianxiety medications	47 with MDD; 2 with dysthymia; 1with MDD & dysthymia; nicotine dependence; anxiety disorder	DS: 4.92 (3.00), CS: 4.86 (3.31) **FTND** (Heatherton et al., 1991) DS: 6.84 (4.15), DN: 6.40 (2.35) **SHAPS** (Snaith et al., 1995) DS: 35.56 (10.2), DN: 34.68 (5.8) **BDI-II** (Beck et al., 1996) DS: 90.04 (9.7), DN: 88.72 (6.3) **MASQ anhedonia** (Watson & Clark, 1991)
Frey et al., 2019	Participants with high (BDI>16) and low (BDI<8) depression	D: 40 (31 f) C: 52 (41 f)	25.33 (7.59) 24.02 (6.59)	No psychotropic medications in the past year	No recreational drugs in the past 3 months; smoking less than 5 cigarettes	30.73 (9.29) **BDI** (Beck et al., 1996) 17.39 (8.89) **RSAS** (Eckblad et al., 1982) 41.35 (8.05) **TEPS** anticipatory (Gard et al., 2006)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the depression group**	**Comorbidities of the depression group**	**Clinical measures of the depression group Mean (SD)**
					per week; anxiety symptoms	35.59 (6.64) **TEPS** consummatory (Gard et al., 2006) 120.63 (19.65) **SAQ** (Caballo et al., 2010)
Frey & McCabe, 2019	Participants with high (BDI>16) and low (BDI<8) depression	D: 21 (17 f) C: 22 (14 f)	23.20 (5.66) 22.45 (4.35)	No psychotropic medications in the past year	No recreational drugs in the past 3 months; smoking less than 5 cigarettes per week; anxiety symptoms	26.05 (9.63) **BDI** (Beck et al., 1996) 18.57 (6.43) **RSAS** (Eckblad et al., 1982) 57.75 (7.12) **STAI-T** (Spielberger et al., 1983) 24.38 (5.71) **PANAS** positive affect (Watson et al., 1988) 21.29 (7.27) **PANAS** negative affect (Watson et al., 1988)
Geugies et al., 2019	Participants with remitted recurrent MDD and healthy controls	D: 36 (26 f) C: 27 (19 f)	47 (range: 36-63) 41 (range: 36-65)	None	No current diagnosis of alcohol or drug dependence; no primary anxiety disorder	3.5 median of **HDRS-17** (Hamilton, 1967) 24.0 median of **SHAPS** (Snaith et al., 1995)
Gradin et al. 2011	Participants with MDD and healthy controls	D: 15 (9 f) C: 17 (10 f)	45.27 (12.32) 40.64 (11.87)	Receiving medications	Anxiety symptoms	22.93 (8.22) **BDI** (Beck et al., 1961) 6.27 (2.28) **BDI** anhedonia (Beck et al., 1961) 23.2 (4.3) **HDRS** (Hamilton, 1960) 54.60 (11.53) **SAS** (Spielberger, 1983)
Kumar et al., 2018	Participants with MDD and healthy controls	D: 25 (19 f) C: 26 (19 f)	25.25 (5.46) 26.31 (7.96)	No psychotropic medications in the past 2 weeks	Social phobia secondary to MDD	17.27 (3.99) **HDRS-17** (Hamilton, 1980) 26.26 (9.21) **BDI** (Beck et al., 1961) 33.40(4.22) **SHAPS** (Snaith et al., 1995)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the depression group**	**Comorbidities of the depression group**	**Clinical measures of the depression group Mean (SD)**
Kumar et al., 2008	Participants with treatment unresponsive MDD and healthy controls	D: 15 (9 f) UC:18 (11 f) MC:15 (9 f)	45.3 (12.3) 42.0 (12.8) 41.7 (12.0)	Stable medications for past 1 month	Anxiety symptoms	23.2(5.3) **HDRS-21** (Hamilton, 1960) 22.9 (8.2) **BDI** (Beck et al., 1961) 35.0(6.7) **SHAPS** (Snaith et al., 1995) 54.6(11.5) **SAS** (Spielberger, 1983)
Kunisato et al., 2012	Depressed and nondepressed individuals (A score of 16 on CESD as the cutoff)	D: 18 (13 f) C: 18 (13 f)	19.44 (0.98) 19.33 (1.14)	None	None	21.94 (6.17) **BDI-II** (Beck et al., 1996) ≥ 16 **CESD** (Radloff, 1977)
Liu et al., 2017	Participants with MDD and healthy controls	D: 21(12 f) C: 17 (10 f) 11 D and 13 C for modeling analysis	30.7 (8.9) 28.3 (5.2)	Medication-free for past 3 months	Anxiety symptoms	24.05(4.15) **HDRS-24** (Hamilton, 1967) 28.5(4.9) **SHAPS** (Snaith et al.,1995) NA **BDI-II** (Beck et al., 1997)
Mkrtchian et al., 2017	Participants with depression & anxiety and healthy controls	D: 43 (16 f) C: 58 (22 f)	28.8 (8.8) 26.7 (7.1)	No medications for the last 6 months	27 with GAD & MDD; 8 with GAD; 2 with MDD and PAD; 6 with MDD alone	20.05 (9.83) **BDI** (Beck & Steer, 1987) 56.65 (8.52) **STAI** (Spielberger et al., 1970)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the depression group**	**Comorbidities of the depression group**	**Clinical measures of the depression group Mean (SD)**
Moutoussis et al., 2018	Participants with MDD and healthy controls	D: 39 (22 f) C: 22 (11 f)	34.12 (9.47) 33.2 (8.25)	32 on medication; 28 receiving antidepressants	8 taking recreational drugs	16.9 (3.69) **HDRS-17** (Hamilton, 1960)
Rothkirch et al., 2017	Participants with MDD and healthy controls	D: 28 (15 f) C: 30 (22 f)	36.32 (11.88) 36.13 (11.96)	Medication-free for past 4 months	None	5.60 (3.60) **HDRS** (NA) 22.50 (3.71) **SHAPS-D** (Snaith et al., 1995; Franz et al., 1998) 33.00 (8.32) **BDI** (Beck et al., 1961)
Rouhani et al., 2019	Depressed (IDS score = 26-84) and nondepressed (IDS score = 0-13) individuals	D: 101 (NA) C: 184 (NA)	NA	NA	NA	26-84 **IDS** (Rush et al., 1996)
Rupprechter et al., 2018	Participants with MDD and healthy controls	D: 15 (12 f) C: 17 (13 f)	17-41 18-33	None	NA	24.7 (13.1) **BDI** (Beck et al., 1961) 12.3 (5.2) **HAD** anxiety (NA) 8.6 (4.8) **HAD** depression (NA) 17.3 (7.0) **HAMA** (NA) 17.7 (6.6) **MADRS** (NA) 37.8 (8.5) **SHAPS** (NA)

Abbreviations: BAI, Beck Anxiety Inventory; BDI, Beck Depression Index; C, control participants; CESD, Center for Epidemiological Studies Depression Scale; CN, control nonsmokers; CS, control smokers; D, Depressed participants; DN, depressed nonsmokers; DNS, nonsuicidal depressed participants; DS, depressed smokers; DSA, depressed suicide attempters; DSI, depressed suicide ideators; EXIT, Executive Interview; f, female; FTND, Fagerstrom Test for Nicotine Dependence; GAD, generalized anxiety disorder; HAD, Hospital Anxiety and Depression Scale; HAMA, Hamilton Anxiety Rating Scale; HDRS, Hamilton Depression Rating scale; IDS, Inventory of Depressive Symptomatology; MADRS, Montgomery-Åsberg Depression Rating Scale; MASQ, Mood and Anxiety Symptom Questionnaire; MC, medicated control; MDD, major depressive disorder; N, number of participants; NA, not available; PAD, panic disorder; PANAS, the Positive and Negative Affect Scale; RSAS,

Revised Social Anhedonia Scale; SAQ, Social Anxiety Questionnaire; SAS, Spielberger State Anxiety Scale; SD, standard deviation; SHAPS, Snaith-Hamilton Pleasure Scale; SHAPS-D, German version of the Snaith-Hamilton Pleasure Scale; STAI, State-Trait Anxiety Inventory; STAI-S, state anxiety subscale of State-Trait Anxiety Inventory; STAI-T, trait anxiety subscale of State-Trait Anxiety Inventory; TEPS, Temporal Experience of Pleasure Scale; UC, unmedicated control.

[a] NA is used in this table if the citation for a scale measure was not given in the reviewed studies.

Table S1.2. Behavioral tasks and results in depression studies

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature[a]	Other parameters
Aylward et al., 2019	Four-armed bandit task with and without threat of shock Real human faces as rewards (happy and fearful) Coded as 1 and –1 Changing probabilities	RL model with separate learning rates and sensitivities for positive and negative feedback, and lapse	ns for win-stay	**LR for +PE**[b] ns **LR for –PE** D > C	NA	**Reward sensitivity** ns **Punishment sensitivity** ns **Lapse** D > C
Bakic et al., 2016	Classification task (a variant of two-armed bandit task; symbolic reward) "correct" vs "incorrect" Coded as 1 vs 0 100% vs 0% 80% vs 20% 50% vs 50%	RL model with separate learning rates for positive and negative feedback and inverse temperature	ns for accuracy; D < C in lose-switch during the second half of the experimental session	**LR for +PE** ns **LR for –PE** ns	ns	NA
Blanco et al., 2013	Leapfrog variant of two-armed bandit task (gain of points) Initial: 10 points vs 20 points On any trial, the lower option could, with the probability of 7.5%, increase by 20 points.	RL model with learning rate fixed at 1 and inverse temperature; Ideal actor model with inverse temperature and volatility parameter	D < C[†] for accuracy	LR was fixed at 1.	**IT in RL model** NA **IT in ideal actor model** NA	**Volatility** NA

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature[a]**	**Other parameters**
Brown, 2018	Two-armed bandit task (gain & loss of money) ($0.70-0.80) vs ($0.20-0.30) (–$0.80- –0.70) vs (–$0.30- –0.20) 80% vs 20%	RL model with learning rate, sensitivity on more extreme outcomes, and outcome shift for gain and loss conditions respectively	D<C[†] for accuracy in reward trials; ns in punishment trials	NA	IT was fixed at 5.	NA
Chase et al., 2010	Two-armed bandit task with training and test phases (symbolic reward in the training phase) "CORRECT" vs "WRONG" Three pairs: 80% vs 20%; 70% vs 60%; and 60% vs 40%	RL model with separate learning rates for positive and negative feedback, and temperature	Training phase: ns for accuracy; D > C for RT Test phase: ns for accuracy; ns for RT	**LR for +PE** D < C[†] **LR for –PE** D < C[†]	**Temperature** ns	NA
Dombrovski et al., 2010	Two-armed bandit task with one reversal (symbolic reward) "CORRECT" vs "WRONG" 80% vs 20%	RL model with separate learning rates for positive and negative feedback, temperature and a memory parameter	ns in acquisition stage D > C for switch errors in reversal stage	**LR for +PE** ns **LR for –PE** DSI < C[†]	**Negative IT** ns	**Memory** DSA < C
Dombrovski et al., 2013	Two-armed bandit task with 12 reversals (symbolic reward) "CORRECT" vs "INCORRECT" 80% to 87% vs 13% to 20%	RL model with separate learning rates for positive and negative feedback, temperature and a memory parameter	DSA > C for probabilistic switch errors	**LR for +PE** DSA < (DNS+C) **LR for –PE** ns	**Negative IT** (DNS+DSA) > C	**Memory** ns

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature[a]	Other parameters
Feng et al., 2017	Two-armed bandit task (gain & loss of money) ($0.70-0.80) vs ($0.20-0.30) (–$0.80- –0.70) vs (–$0.30- –0.20) 80% vs 20%	RL model with learning rate and reward sensitivity	ns for accuracy	ns for (DS+DN) vs (CS+CN)	IT was fixed at 3.15.	**Reward sensitivity** ns for (DS+DN) vs (CS+CN)
Frey et al., 2019	Two-armed bandit task **Social condition:** "like" sign vs neutral "dislike" sign vs neutral 75 % vs 25 % **Nonsocial condition:** 5 pence vs nothing –5 pence vs nothing 75 % vs 25 %	RL model with learning rate, temperature, outcome valuation, and memory decay for social and nonsocial conditions	D < C[†] for accuracy of choosing the most rewarded and avoiding the most punished item	**Social LR** D < C **Nonsocial LR** ns	**Social temperature** ns **Nonsocial temperature** ns	**Social memory decay** ns **Nonsocial memory decay** ns **Social outcome valuation** ns **Nonsocial outcome valuation** ns
Frey & McCabe, 2019	Social learning task in which the participants learned the contingencies between six names and facial expressions Happy vs neutral Coded as 1 vs 0 Fearful vs neutral Coded as 1 vs 0 25 % vs 75 % 50 % vs 50 % 75 % vs 25 %	RL model with separate learning rates for reward and punishment conditions	D < C for accuracies of rating probability of seeing emotional faces in reward and aversion conditions	**Reward LR** D < C **Punishment LR** D < C	NA	NA

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature[a]**	**Other parameters**
Geugies et al., 2019	Pavlovian reward-learning task (juice as reward) Coded as 1 for juice delivery; 0 for no-juice Different probabilities of juice delivery (80%-20%) for conditioned stimuli	Temporal difference learning model	ns for wanting and liking ratings of juice	LR was fixed at 0.45.	NA	Discount factor was fixed at 1.0.
Gradin et al., 2011	Two-armed bandit task (water as reward) Coded as 1 vs 0 60% to 90% vs 0% to 20%	Temporal difference learning model	ns for pleasantness ratings of water; ns for RT	ns	ns	Discount factor fixed at 1.0.
Kumar et al., 2018	Two-armed bandit task (gain & loss of money) +$10 vs nothing Coded as 1 vs 0 –$10 vs nothing Coded as –1 vs 0 $0 vs nothing 80% vs 20%	RL model with separate learning rates and temperatures for reward and punishment conditions	D < C for accuracy for reward learning; ns for accuracy for punishment learning	**Reward LR** ns **Punishment LR** ns	**Reward temperature** ns **Punishment temperature** ns	NA
Kumar et al., 2008	Pavlovian reward-learning task (water as reward) Coded as 1 for water delivery; 0 for no-water; Probabilities of water delivery changed every 25 trials	Temporal difference learning model	ns for accuracy for reporting picture–water association; ns for pleasantness ratings of water	LR was fixed at 0.1.	NA	Discount factor was fixed at 1.0.

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature[a]**	**Other parameters**
Kunisato et al., 2012	Two-armed bandit task with training and test phases (monetary reward in training phase) +10 vs –10 yen Three pairs: 80% vs 20%; 70% vs 60%; and 60% vs 40%	RL model with separate learning rates for positive and negative feedback and temperature	D < C for choosing the most rewarding option in test phase	**LR for +PE** ns **LR for –PE** ns	**Temperature** D > C	NA
Liu et al., 2017	Two-armed bandit task (gain & loss of points) ±50 points vs 0 Coded as ±1 vs 0 80% vs 20%	Basic RL model	ns for accuracy; ns for RT	ns	ns	NA
Mkrtchian et al., 2017	Pavlovian probabilistic go/no-go task with and without threat of shock (four fractals representing four conditions: go/no-go to win 10 points and go/no-go to avoid losing 10 points) ±10 points vs 0 Coded as ±1 vs 0 80 % vs 20 %	RL model with separate learning rates and sensitivities for reward and punishment conditions, approach-avoidance bias, general action bias, and lapse	D < C for overall accuracy	**Reward LR** ns **Punishment LR** ns	NA	**Avoidance bias** D > C in threat condition ns in safe condition **Other parameters** ns
Moutoussis et al., 2018	Pavlovian probabilistic go/no-go task (four fractals representing four conditions: go/no-go to win and go/no-go to avoid to lose) upward arrow(win) vs horizontal line (null) downward arrow(lose) vs horizontal line (null) Coded as ±1 vs 0 80 % vs 20 %	RL model with learning rate, reward & punishment sensitivities, approach-avoidance bias, general action bias, and lapse	ns	ns	NA	ns

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature[a]	Other parameters
Rothkirch et al., 2017	Two-armed bandit task (gain & loss of money) ±50 cents vs 0 cents Coded as ±1 vs 0 80% vs 20%	RL model with separate learning rates and inverse temperatures for reward and punishment conditions	ns for accuracy	**Reward LR** ns **Punishment LR** ns	**Reward IT** ns **punishment IT** ns	NA
Rouhani et al., 2019	Learning the values of two scene categories in high and low risk conditions. In each trial, participants indicated the expected value of the scene category, and were shown the true value of that scene. They also finished a surprise memory test after the learning task.	RL model with learning rate which was calculated from participants' expected values and true values of the scenes	ns for performance (estimation of true values); D < C for remembering events with +PE; D > C for remembering events with –PE	ns	NA	NA
Rupprechter et al., 2018	Two-armed bandit task with passive viewing of fractural stimuli followed by reward (a £ symbol) or no feedback (coded as 1 or 0), and decisions between explicit probabilities and fractural stimuli	RL model without learning rate, but with a memory parameter of previously observed rewards and inverse temperature	NA for accuracy; ns for RT	NA	D < C	**Memory parameter** D < C

Abbreviations: C, control participants; CN, control nonsmokers; CS, control smokers; D, Depressed participants; DN, depressed nonsmokers; DNS, nonsuicidal depressed participants; DS, depressed smokers; DSA, depressed suicide attempters; DSI, depressed suicide ideators; IT, inverse temperature; LR, learning rate; NA, not available; ns, not significant; PE, prediction error; RL, reinforcement learning; RT, response time; †, marginally significant ($P < 0.1$).

[a] Some studies applied temperature, which is the reciprocal of inverse temperature. Some other studies applied the negative form of inverse temperature. [b] Different studies used inconsistent names for LR associated with different kinds of PEs. In the present study, LR associated with reward PEs (RPE) is named "reward LR"; LR associated with aversive PEs (APE) is named "punishment LR"; LR associated with positive PEs is

named “LR for +PE”; and LR associated with negative PEs is named “LR for –PE”. RPEs occur when outcomes only contain rewards of different amounts (e.g., $5 vs $1) or reward vs. omission (e.g., $5 vs $0). APEs occur when outcomes only contain punishments of different amounts (e.g., –$5 vs –$1) or reward vs. omission (e.g., –$5 vs $0). They are different from positive PEs and negative PEs, in which the PE values can only be positive or negative. RPEs and APEs can contain both positive and negative PE values.

Table S1.3. Characteristics of participants in addiction studies

Study	Participants	N (Number of females)	Age (SD)	Abstinence of the addiction group	Medications of the addiction group	Comorbidities of the addiction group	Clinical measures of the addiction group Mean (SD)
Addicott et al., 2013	Smokers and nonsmokers	A: 18 (10 f) C: 17 (10 f)	36 (8) 32 (12)	Sated	None	None	14 (6) carbon monoxide 16 (8) cigarettes/day 5.9 (2.3) **FTND** (Heatherton et al., 1991) 2 (4) alcohol drinks/week
Ahn et al., 2014	Past amphetamine-dependent participants, past heroin-dependent participants and healthy controls	AD: 38 (9 f) HD: 43 (8 f) C: 48 (10 f)	22.7 (3.7) 29.7 (5.0) 24.7 (4.9)	In protracted abstinence (> 3 months)	NA	None	AD: 3.3 (2.8), HD: 4.7 (2.7) **FTND** (Heatherton et al., 1991) AD: 6.62 (5.6), HD: 8.26 (6.4) **BDI-II** (Beck et al., 1996) AD: 66.13 (11.0), HD: 65.70 (9.9) **BIS-11** (Patton & Stanford, 1995) AD: 33.68 (7.7), HD: 36.12 (10.1) **STAI-S** (Spielberger & Gorsuch, 1983) AD: 38.58 (9.3), HD: 39.98 (10.1) **STAI-T** (Spielberger & Gorsuch, 1983)
Beylergil et al., 2017	Alcohol-dependent participants and healthy controls	A: 34 (0 f) C: 26 (0 f)	44.73 (8.27) 41.92 (9.59)	Abstinent for more than 1 week	No benzodiazepine or chlormethiazole medications for at least 1 week	Smoking in both groups	15.48 (7.73) **ADS** (Skinner & Horn, 1984) 17.48 (7.09) **OCDS** sum (Anton, 2000) 8.23 (10.06) **OCDS** craving (Anton, 2000) 5 (2.73) **FTND** (Heatherton et al., 1991)
Deserno et al., 2015	Detoxified alcohol abusers and healthy controls	A: 13 (0 f) C: 14 (0 f)	45.08 (5.97) 43.86 (9.23)	No current drug abuse other than nicotine	No medications for at least 4 plasma half-lives	Smoking in both groups	15.62 (7.91) **ADS** (Skinner & Sheu, 1982) 19.62 (8.19**) OCDS** sum (Anton, 2000) 39.39 (42.44) **OCDS** craving (Anton, 2000)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Abstinence of the addiction group**	**Medications of the addiction group**	**Comorbidities of the addiction group**	**Clinical measures of the addiction group Mean (SD)**
Feng et al., 2017	Smokers and nonsmokers with and without depression	DS: 25 (13 f) CS: 20 (3 f) DN: 25 (17 f) CN: 32 (21 f)	36.64 (9.55) 35.80 (10.06) 35.24 (11.9) 33.75 (11.77)	Sated	5 depressed participants were taking antidepressants/ antianxiety medications	Anxiety disorders for depressed participants	DS: 14.28 (1.88) CS: 12.20 (1.37) cigarettes/day DS: 4.92 (3.00) CS: 4.86 (3.31) **FTND** (Heatherton et al., 1991) DS: 6.84 (4.15) DN: 6.40 (2.35) **SHAPS** (Snaith et al., 1995) DS: 35.56 (10.2) DN: 34.68 (5.8) **BDI-II** (Beck et al., 1996) DS: 90.04 (9.7) DN: 88.72 (6.3) **MASQ** anhedonia (Watson & Clark, 1991)
Gullo & Stieger, 2011	Heavy and light drinkers of alcohol	HA: 22 (NA) LA: 22 (NA)	≈21 ≈23	NA	NA	None	**HA** ≈13 **AUDIT** (Saunders et al., 1993) ≈29 **PANAS** positive affect (Watson et al., 1988) ≈13 **PANAS** negative affect (Watson et al., 1988)
Kanen et al., 2019	Stimulant use (cocaine & amphetamine) disorder and healthy controls	A: 17 (3 f) C: 18 (3 f)	34.3 (7.4) 32.7 (6.9)	NA	No psychotropic medications	None	9.8 (11.2) **BDI-II** (Beck et al., 1996) 81.7 (9.7) **BIS-11** (Patton et al., 1995) 26.0 (7.8) **OCDUS** (Franken et al. 2002) 20.5 (5.4) age of onset of stimulant abuse 11.7 (7.4) years of stimulant abuse
Lesage et al., 2017	Smokers and nonsmokers	A: 24 (12 f) C: 19 (10 f)	35.8 (9.9) 30.4 (7.2)	Abstinent for 12 hours	No psychotropic medications	None	5.00 (1.9) **FTND** (Heatherton et al., 1991) 18.0 (10.6) years daily smoking 17.7 (7.9) cigarettes/ day

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Abstinence of the addiction group**	**Medications of the addiction group**	**Comorbidities of the addiction group**	**Clinical measures of the addiction group Mean (SD)**
Lim et al., 2019	Cocaine addicts and healthy controls	A: 68 (≈7 f) C: 55 (3 f)	≈38.0 (8.6) ≈41.3 (10.5)	Active users verified by urine screen	Methadone & buprenorphine	Opiate, cannabis and alcohol dependence; smoking in both groups	≈76.6 (9.6) **BIS-11** (Patton et al., 1995) ≈17.3 (13.4) **OCI-R** (Foa et al., 2002) ≈8.2 (1.9) **SRRS** (Hayaki et al., 2005) ≈4.2 (4.8) **AUDIT** (Allen et al., 1997) ≈0.4 (0.6) **DAST-20** (Skinner, 1982) ≈10.8 (6.1) cigarettes/day
Mazas et al., 2000	Alcohol abusers and healthy controls	A: 27 (12 f) C: 32 (18 f)	21.6 (2.4) 22.7 (2.1)	Abstinent	No psychotropic or antihistamine medications	Antisocial personality, cannabis	5.02 (2.87) drinking occasions/week 5.45 (2.82) drinks/occasion 16.5 (9.61) **DIS** anti-social symptoms **(**Robins et al., 1985)
Myers et al., 2016	Heroin-dependent participants and healthy controls	A: 45 (24 f) C: 35 (10 f)	41.2 (10.3) 39.0 (11.6)	With daily use	Treated with opioid medications	Schizophrenia depression, bipolar disorder	NA
Park et al., 2010	Alcohol abusers and healthy controls	A: 20 (0 f) C: 16 (0 f)	42.45 (1.8) 37.8 (2.22)	Abstinent for more than 1 week	No psychotropic medications for at least 4 plasma half-lives	Smoking in both groups	NA
Reiter et al., 2016	Alcohol abusers and healthy controls	A: 43 (9 f) C: 35 (10 f)	44.42 (10.21) 42.00 (10.49)	Abstinent for more than 8 days	No medications for at least 4 plasma half-lives; 1 participant had doxepin	Smoking in both groups	25.55 (9.78) **OCDS** (Anton et al., 1995) 2.04 (0.88) **ACQ** (Tiffany et al., 2000) 26.24 (8.72) **AUDIT** (Allen et al., 1997) 65.81 (9.18) **BIS** (NA)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Abstinence of the addiction group**	**Medications of the addiction group**	**Comorbidities of the addiction group**	**Clinical measures of the addiction group Mean (SD)**
Sebold et al., 2017	Detoxified alcohol abusers (abstainers & relapsers) and healthy controls	AS: 37 (7 f) RS: 53 (6 f) C: 96 (16 f) AS & RS classification based on regular assessments over 1 year after the task	45.7 (12.0) 45.2 (9.9) 43.6 (10.9)	Abstinent for an average of 21.4 (SD: 11.6) days for AS and 22.3 (SD: 12.4) days for RS	No psychotropic medications for at least 4 plasma half-lives	Depressive symptoms, smoking in all 3 groups	AS: 31.7 (4.4), RS: 32.8 (3.9) **AEQ** (Demmel & Hagen, 2004) AS: 3.9 (3.9), RS: 4.2 (3.7) **HADS** depressive symptoms (Zigmond & Snaith, 1983) AS: 10.3 (8.2), DS: 12.9 (8.4) **OCDS** craving (Mann & Ackermann, 2000)
Stout et al., 2004	Chronic cocaine abusers and healthy controls	A: 12 (3 f) C: 14 (0 f)	36.9 (10.3) 30.0 (6.1)	Abstinent	NA	Borderline or antisocial personality disorder, history of other drugs of abuse	NA
Tanabe et al., 2013	Polydrug abusers and healthy controls	A: 32 (13 f) C: 30 (15 f)	36.6 (9.0) 34.5 (7.5)	Abstinent for more than 0.1 years except for nicotine	NA	Antisocial personality disorder, smoking in both groups	NA
Wei et al., 2018	Smokers and nonsmokers	A: 38 (0 f) C: 37 (0 f)	24.5 (2.3) 23.7 (2.3)	Sated	NA	None	> 10 cigarettes/day NA **FTND** (Heatherton et al., 1991)
White et al., 2016	Past alcohol and cannabis abusers and healthy controls	A: 16 (7 f) C: 29 (8 f) A included 11 cannabis abusers, 3 alcohol	15.58 (1.55) 14.88 (1.25)	Abstinent for more than 2 weeks	No psychotropic medications	Use of other substances was likely.	75.50 **CBCL** conduct disorder (Achenbach, 2009)

Study	Participants	N (Number of females)	Age (SD)	Abstinence of the addiction group	Medications of the addiction group	Comorbidities of the addiction group	Clinical measures of the addiction group Mean (SD)
		abusers, and 2 abusers of both substances.					
Yechiam et al., 2004	Chronic cannabis abusers and healthy controls	A: 25 (NA) C: 16 (NA)	NA	Abstinent	NA	NA	NA
Yechiam et al., 2005	Polydrug abusers and healthy controls	A: 39 (NA) C: 37 (NA)	Between 18 and 35	Abstinent for 12 hours or more	No psychotropic medications	NA	NA

Abbreviations: A, participants with addiction; ACQ, Alcohol Craving Questionnaire; AD, past amphetamine-dependent participants; ADS, Alcohol Dependence Scale; AEQ, Alcohol Expectancy Questionnaire; AS, abstainers after alcohol detoxification; AUDIT, Alcohol Use Disorder Identification Test; BDI, Beck Depression Index; BIS, Barratt Impulsiveness Scale; C, control participants; CBCL, Child Behavior Checklist; CN, control nonsmokers; CS, control smokers; DAST, Drug Abuse Screening Test; DIS, the NIMH Diagnostic Interview Schedule, Version III-A; DN, depressed nonsmokers; DS, depressed smokers; f, female; FTND, Fagerstrom Test for Nicotine Dependence; HA, heavy drinkers of alcohol; HADS, the Hospital Anxiety and Depression Scale; HD, past heroin-dependent participants; LA, light drinkers of alcohol; MASQ, Mood and Anxiety Symptom Questionnaire; N, number of participants; NA, not available; OCDS, Obsessive-Compulsive Drinking Scale; OCDUS, Obsessive-Compulsive Drug Use Scale; OCI, Obsessive-Compulsive Inventory; PANAS, Positive and Negative Affect Schedule; RS, relapsers after alcohol detoxification; SD, standard deviation; SHAPS, Snaith-Hamilton Pleasure Scale; SRRS, Social Readjustment Rating Scale; STAI-S, state anxiety subscale of State-Trait Anxiety Inventory; STAI-T, trait anxiety subscale of State-Trait Anxiety Inventory.

Table S1.4. Behavioral tasks and results in addiction studies

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
Addicott et al., 2013	Six-armed bandit task (gain of points) with points of each arm gradually changed from trial to trial according to a biased random walk	RL model with learning rate, inverse temperature, and Kalman filter for unchosen options	ns for the number of points earned	A > C	NA	NA
Ahn et al., 2014	IGT (gain & loss of money) with disadvantageous decks gradually became advantageous	RL model with learning rate, response consistency (c), reward sensitivity, loss aversion, impact of gain on perseverance, impact of loss on perseverance, decay rate of perseverance, and RL weight	AD < C† and HD < C for selection of advantageous cards	HD > C†; ns for AD vs C	**c in IT = $3^c - 1$** ns	**Reward sensitivity** ns **Loss aversion** HD < C **Other parameters** ns
Beylergil et al., 2017	Two-armed bandit task with reversals (a smiling face as positive feedback and a frowning face as negative feedback) Coded as 1 vs –1 80% vs 20% or 50% vs %50	RL model with separate sensitivities for positive and negative feedback and same learning rate for the chosen and the unchosen option	A > C† for number of trials to meet learning criteria; A < C for number of correct choices	ns	IT was fixed at 1.	**Reward sensitivity for gain** ns **Reward sensitivity for loss** A < C

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
Deserno et al., 2015	Two-armed bandit task with reversals (a smiling face as positive feedback and a frowning face as negative feedback) Coded as 1 vs –1 80% vs 20% or 50% vs %50	Basic RL model	A > C for number of trials to reach learning criteria	NA	NA	NA
Feng et al., 2017	Two-armed bandit task (gain & loss of money) ($0.70-0.80) vs ($0.20-0.30) (–$0.80- –0.70) vs (–$0.30- –0.20) 80% vs 20%	RL model with learning rate and reward sensitivity	ns for accuracy; ns for RT	(DS+CS) < (SN+CN)	IT was fixed at 3.15.	**Reward sensitivity** (DS+CS) > (SN+CN) **Reward sensitivity for gain** (DS+CS) > (SN+CN) **Reward sensitivity for loss** (DS+CS) > (SN+CN)
Gullo & Stieger, 2011	IGT (gain & loss money)	RL model with learning rate, response consistency (c), and attention to gains relative to losses	HA < LA for selecting advantageous cards	ns	**c in IT = (t/10)^c** ns	**Attention to gains** HA > LA
Kanen et al., 2019	Two-armed bandit task with 18 reversals (a smiling cartoon face as positive feedback and a frowning face as negative feedback) Coded as 1 vs 0 ≈85% vs ≈15%	RL model with separate learning rates for +PE and –PE, reinforcement sensitivity, stimulus stickiness, and location stickiness	A > C for lose-shift	**LR for +PE** A < C **LR for –PE** A > C	IT was fixed at 1.	**Reinforcement sensitivity** ns **Stimulus stickiness** A > C **Location stickiness** ns

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature**	**Other parameters**
Lesage et al., 2017	Two-armed bandit task with reversals (gain & loss of money) +$1 vs –$1 75% vs 25%	RL model with learning rate, inverse temperature, and a parameter of bias toward one of the two options; Hidden Markov model (HMM) with inverse temperature, staying bias, perceived reversal probability and delay weight	A < C† for overall score; A < C† for lose-shift; ns for win-stay; A > C† for the number of rials to reach learning criterion	RL model ns HMM model NA	RL model **Log (IT)** ns HMM model **Log (IT)** A < C	RL model **Option bias** ns HMM model **Staying bias** ns **Perceived reversal probability** ns **Delay weight** ns
Lim et al., 2019	Deterministic learning task with left and right buttons on the screen leading to gaining 0 or 5 points.	RL model with learning rate, inverse temperature and a perseveration parameter	NA	A < C	ns	**Perseverative responding** ns
Mazas et al., 2000	IGT (gain & loss of money)	RL model with learning rate, response consistency (c), and attention to losses relative to gains	A < C† for selecting advantageous cards	ns	**c in IT = (t/10)^c** t is trial number. ns	**Attention to losses** ns
Myers et al., 2016	Probabilistic classification task (a variant of two-armed bandit task; gain & loss of points) ±25 points vs No feedback Coded as ±1 vs 0 80% vs 20%	RL model with critic and actor's learning rates for +PE and –PE, inverse temperature and a parameter of the tendency to repeat previous responses	ns for total points	**Critic LR for +PE** ns **Critic LR for –PE** ns **Actor LR for +PE** ns **Actor LR for –PE** ns	ns	**Tendency to repeat** A < C

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature**	**Other parameters**
Park et al., 2010	Two-armed bandit task with reversals (a smiling face as positive feedback and a frowning face as negative feedback) 80% vs 20% or 50% vs %50	RL model with separate learning rates for positive and negative feedback and inverse temperature	A > C for number of trials to meet learning criteria	ns	ns	NA
Reiter et al., 2016	Two-armed bandit task with 4 reversals (gain & loss of money) +10 vs –10 Eurocent coin 80% vs 20%	RL model with separate learning rates for the chosen option, weighted learning rates for the unchosen option, and inverse temperatures for positive and negative feedback	A < C for accuracy	**LR for +PE** ns **LR for –PE** ns **Weighted LR for +PE** ns **Weighted LR for –PE** ns	**Reward IT** ns **Punishment IT** ns **Weighted LR for –PE * punishment IT** A < C	NA
Sebold et al., 2017	Two-armed bandit task with two steps (gain & loss of money) Step 1: 70% vs %30 Step 2: The probability of outcomes for each arm gradually changed from trial to trial according to a random walk. +10 vs −10 Eurocent coin	Hybrid model (a mixture of model-based RL and temporal difference learning) with a parameter determining the balance between mode-free and model-based control	ns for model-based control	ns	ns	**Balance parameter** ns
Stout et al., 2004	IGT (gain & loss of money)	RL model with learning rate, response consistency (c), and attention to losses relative to gains	A < C for selecting advantageous cards	ns	**c in IT = (t/10)^c** A < C	**Attention to losses** A < C

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature**	**Other parameters**
Tanabe et al., 2013	IGT with play and pass responses (gain & loss of money)	RL model with learning rate, response consistency (c), and attention to losses relative to gains	The number of choosing good decks relative to bad decks improved faster for C than A†.	ns	**c in IT = (t/10)^{c}** A < C	**Attention to losses** A < C†
Wei et al., 2018	IGT (gain & loss of money)	Basic RL model	A < C for total score	A < C	IT was fixed at 1.	NA
White et al., 2016	Probabilistic passive avoidance task (a variant of four-armed bandit task with gain & loss of money) Four outcomes: +$5, +$1, –$1, –$5 Four arms: +$18.57, +$7.14, –$18.57, –$7.14 over every 10 trials	Basic RL model	A < C† for accuracy	NA	NA	NA
Yechiam et al., 2004	IGT (gain & loss of money)	RL model with learning rate, response consistency (c), and attention to losses relative to gains	NA	A > C	**c in IT = (t/10)^{c}** NA	**Attention to losses** A < C
Yechiam et al., 2005	IGT (gain & loss of money)	RL model with learning rate, response consistency (c), and attention to losses relative to gain	NA	ns	**c in IT = (t/10)^{c}** ns	**Attention to losses** ns

Abbreviations: A, participants with addiction disorder; AD, past amphetamine-dependent participants; C, control participants; CN, control nonsmokers; CS, control smokers; DN, depressed nonsmokers; DS, depressed smokers; HA, heavy drinkers of alcohol; HD, past heroin-dependent participants; IGT, Iowa gambling task; IT, inverse temperature; LA, light drinkers of alcohol; LR, learning rate; NA, not available; ns, not significant; PE, prediction error; RL, reinforcement learning; RT, response time; †, marginally significant ($P < 0.1$).

Table S1.5. Characteristics of participants in anxiety studies

Study	Participants	N (Number of females)	Age (SD)	Medications of the anxiety group	Comorbidities of the anxiety group	Clinical measures of the anxiety group Mean (SD)
Aylward et al., 2019	Participants with anxiety & mood disorders and healthy controls	A: 44 (28 f) C: 88 (50 f)	29 (8.7) 23 (5.1)	None	28 with GAD & MDD, 8 with GAD, 3 with PAD & MDD, 5 with MDD	20 (9.4) **BDI** (NA) 47 (10.7) **STAI-S** (NA) 57 (8.2) **STAI-T** (NA)
Brown et al. 2018	Veterans with PTSD and healthy controls	A: 39 (5 f) C: 29 (3 f)	32.3 (1.21) 33.3 (1.77)	Psychotropic medications	Depressive symptoms; nicotine dependence	66.7 (26.67) **CAPS** (Blake et al., 1995) 24.6 (13.61) **BDI-II** (Steer et al., 1999) 21.9 (8.43) **CES** (Lund et al., 1984)
Carlisi et al., 2017	Boys with OCD and healthy controls	20 (0 f) 20 (0 f)	15.7 (1.4) 15.1 (2.0)	Antidepressants	None	22.3 (5.8) **CY-BOCS** total (Goodman et al. 1989) 10.8 (3.6) **CY-BOCS** obsessions 12.0 (3.1) **CY-BOCS** compulsions **3.0 (2.7) SDQ** hyperactivity/inattention (Goodman & Scott, 1999)
Cisler et al., 2015	PTSD and healthy controls	A: 25 (25 f) C: 15 (15 f) 15 A and 14 C available for fMRI	34.7 (8.3) 30.87 (7.1)	NA	44% MDD; 52 % GAD; 4 % marijuana dependence; 16 % alcohol dependence	22.8 **BDI-II** (Beck et al., 1996) 55.7 **PCL-C** (Blanchard et al., 1996)
Cisler et al., 2019	Adolescent assault victims and healthy controls	A1: 12 (12 f) A2: 18 (18 f) C: 30 (30 f)	15.6 (1.44) 15.8 (1.6) 14.9 (2.9)	SSRI, SNRI, NDRI, mood stabilizer, benzodiazepine, stimulants, and other medications	A1: 50% PTSD; 50% MDD/BD; 66.7% anxiety disorder A2:	A1: 39.42 (29.03), A2: 45.28 (30.86) **CAPS** (Blake et al., 1995) A1: 27.17 (19.43), A2: 30.72 (15.18) **UCLA PTSD RI** (NA) A1: 5.67 (6.08), A2: 10.17 (5.97) **CBCL** anxiety (NA)

Study	Participants	N (Number of females)	Age (SD)	Medications of the anxiety group	Comorbidities of the anxiety group	Clinical measures of the anxiety group Mean (SD)
					55% PTSD; 50% MDD/BD; 77.2% anxiety disorder	A1: 4.92 (3.90), A2: 5.33 (2.99) **CBCL** depression (NA)
Hauser et al., 2017	Teenagers and adults with OCD and healthy controls	33 (12 f) 34 (21 f)	23.4 (9.5) 24.5 (11.2)	SSRI, neuroleptics, SNRI, and other medications	Depression, other anxiety disorders, ADHD, and anorexia nervosa	15.47 (9.87) **Y-BOCS** total (Goodman et al. 1989)
Huang et al., 2017	Clinical patients with high and low anxiety (PTSD, GAD, PAD & social phobia)	A: 77 (NA) C: 45 (NA) OASIS $\geqslant$9 and $\leqslant$8 for the two groups	35.03 (11.08) for A&C	Antidepressants, anxiolytics, antipsychotics, and mood stabilizers	Depression and alcohol dependence	11.51 (0.29) **OASIS** (Campbell-Sills et al., 2009)
Kenan et al., 2019	Participants with OCD and healthy controls	A: 18 (11 f) C: 18 (3 f)	35.4 (9.8) 32.7 (6.9)	SSRI	None	18.5 (10.0) **BDI-II** (Beck et al., 1996) 66.9 (9.7) **BIS-11** (Patton et al., 1995) 24.11 (13.0) **Y-BOCS** total (Goodman et al., 1989) 17.1 (11.0) age of onset of OCD 18.3 (10.6) years of OCD
Khdour et al., 2016	Participants with GAD/SAD/PAD and healthy controls	GA: 18 (NA) SA: 20 (NA) PA: 17 (NA) C: 18 (NA) Groups matched on gender	42.11 (5.70) 44.95 (4.27) 41.52 (4.93) 43.50 (6.84)	None	None	GA: 24.50 (2.98) SA: 24.50 (4.72) PA: 22.59 (3.98) **HAMA** (Hamilton, 1959)

Study	**Participants**	**N (Number of females)**	**Age (SD)**	**Medications of the anxiety group**	**Comorbidities of the anxiety group**	**Clinical measures of the anxiety group Mean (SD)**
Mkrtchian et al., 2017	Participants with depression & anxiety and healthy controls	A: 43 (16 f) C: 58 (22 f)	28.8 (8.8) 26.7 (7.1)	No medications for the last 6 months	27 with GAD & MDD; 8 with GAD; 2 with MDD and PAD; 6 with MDD alone	56.65 (8.52) **STAI** (Spielberger et al., 1970) 20.05 (9.83) **BDI** (Beck & Steer, 1987)
Murray et al., 2019	Participants with OCD and healthy controls	A: 18 (3 f) C: 18 (7 f)	32.1 (6.5) 35.6 (10.1)	SSRI	None	24.1 (6.8) **Y-BOCS** total (Goodman et al., 1989)
Myers et al., 2013	Veterans with severe PTSD symptoms and controls	A: 48 (0 f) C: 39 (0 f) A score of 50 on PCL-M as the cutoff	54.0 (8.4) 52.2 (10.8)	Psychoactive medications	NA	10.4 (11.0) **CES** (Keane et al., 1989) 21.1 (5.3) **AMBI** (Gladstone & Parker, 2005) 13.9 (6.4) **RMBI** (Gladstone & Parker, 2005) ≥ 50 **PCL-M** (Blanchard et al., 1996)
Norman et al., 2018	Participants with OCD and healthy controls	A: 20 (0 f) C: 20 (0 f)	15.76 (1.43) 15.15 (1.99)	NA	None	22.32 (5.97) **CY-BOCS** total (Goodman et al. 1989) 4.4 (3.03) **SDQ** hyperactivity/inattention (Goodman & Scott, 1999)
Ousdal et al., 2017	Survivors of a terror attack and healthy controls	A: 25 (17 f) C: 23 (14 f)	19.64 (1.35) 20.26 (2.30)	Only 1 participant used low-dose benzodiazepine	8 with PD; 7 with PTSD; 2 with GAD; 4 with major depressive episodes	NA
Piray et al., 2019	Participants with high and low scores on the Liebowitz social anxiety scale	A: 21 (21 f) C: 23 (23 f)	A&C: 20.7	None	None	31.00 (6.28) **Liebowitz social anxiety scale** (Liebowitz, 1987)

Study	Participants	N (Number of females)	Age (SD)	Medications of the anxiety group	Comorbidities of the anxiety group	Clinical measures of the anxiety group Mean (SD)
Ross et al., 2018	Participants with PTSD and healthy controls	A: 15 (15 f) C: 14 (14 f)	31.13 (6.38) 31.21 (7.2)	NA	47% MDD; 60% GAD; 7% marijuana dependence; 13% alcohol dependence	24.4 (14.4) **BDI-II** (Beck et al., 1996) 66.9 (9.7) **PCL-C** (Blanchard et al., 1996)
Vaghi et al., 2017	Participants with OCD and healthy controls	A: 24 (11 f) C: 25 (12 f)	41.33 (12.32) 40.68 (10.19)	SSRI	None	29.25 (11.87) **OCI-R** (Foa et al., 2002) 43.50 (10.72) **STAI-S** (Spielberger, 1983) 56.58 (7.91) **STAI-T** (Spielberger, 1983) 22.75 (4.32) **Y-BOCS** total (Goodman et al., 1989) 10.79 (2.41) **Y-BOCS** obsessions 11.96 (2.23) **Y-BOCS** compulsions

Abbreviations: A, participants with anxiety disorder; A1, assault victims with 1-2 assaults; A2, assault victims with 3 or more assaults; ADHD, attention deficit hyperactivity disorder; AMBI, Adult Measure of Behavioral Inhibition; BD, bipolar disorder; BDI, Beck Depression Index; BIS, Barratt Impulsiveness Scale; C, control participants; CAPS, Clinician Administered PTSD Scale; CBCL, Child Behavior Checklist; CES, Combat Exposure Scale; CY-BOCS, Children's Yale-Brown Obsessive-Compulsive Scale; f, female; fMRI, functional magnetic resonance imaging; GA, participants with general anxiety disorder; GAD, general anxiety disorder; HAMA, Hamilton Anxiety Rating Scale; MDD, major depressive disorder; N, number of participants; NA, not available; NDRI, norepinephrine–dopamine reuptake inhibitor; OASIS, Overall Anxiety Severity and Impairment Scale; OCD, obsessive-compulsive disorder; OCI-R, Obsessive-Compulsive Inventory-Revised; PA, participants with panic anxiety disorder; PAD, panic anxiety disorder; PCL-C, PTSD Checklist-Civilian version; PCL-M, PTSD Checklist-Military version; PTSD, post-traumatic stress disorder; RMBI, Retrospective Measure of Behavioral Inhibition; SA, participants with social anxiety disorder; SAD, social anxiety disorder; SD, standard deviation; SDQ, Strengths and Difficulties Questionnaire; SNRI, Serotonin and norepinephrine reuptake inhibitors; SSRI, selective serotonin reuptake inhibitors; STAI, State-Trait Anxiety Inventory; STAI-S, state anxiety subscale of State-Trait Anxiety Inventory; STAI-T, trait anxiety subscale of State-Trait Anxiety Inventory; UCLA PTSD RI, University of California–Los Angeles PTSD Reaction Index; Y-BOCS, Yale-Brown Obsessive-Compulsive Scale.

Table S1.6. Behavioral tasks and results in anxiety studies

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
Aylward et al., 2019	Four-armed bandit task with and without threat of shock Real human faces as rewards (happy and fearful) Coded as 1 and –1 Changing probabilities	RL model with separate learning rates and sensitivities for positive and negative feedback, and lapse	ns for win-stay	**LR for +PE** ns **LR for –PE** A > C	NA	**Reward sensitivity** ns **Punishment sensitivity** ns **Lapse** A > C
Brown et al., 2018	Two-armed bandit task (gain & loss of money) ($0.70-0.80) vs ($0.20-0.30) (–$0.80- –0.70) vs (–$0.30- –0.20) 80% vs 20%	RL model with separate parameters for gain and loss trials: inverse temperature, reward sensitivity, decay parameters for the unchosen option and dynamic learning rate modulated by associability-weighted PE	ns for accuracy	**Unmodulated LR** ns	ns	**Associability weight in loss trials** A > C **Other parameters** ns
Carlisi et al., 2017	IGT (gain & loss money) with an anticipation period of 6s	RL model with learning rate, response consistency (c), reward sensitivity, loss aversion, impact of gain on perseverance, impact of loss on perseverance, decay rate of perseverance, and RL weight	ns for accuracy	ns	**c in IT = $3^c - 1$** A< C	**RL weight** A < C **Other parameters** ns

Study	**Task**	**Model**	**Model-agnostic performance**	**Learning rate**	**Inverse temperature**	**Other parameters**
Cisler et al., 2015	Social /nonsocial two-armed bandit task (social: faces as two arms; smiling vs frowning; Nonsocial: houses as arms; open vs locked) Probabilities changed every 25 trials	RL model with learning rate, inverse temperature and a volatility parameter	ns for accuracy	ns	ns	**Volatility** ns
Cisler et al., 2019	Social/nonsocial three-armed bandit task (social: neutral faces as arms; Nonsocial: houses as arms) $20 vs 0 Probabilities changed every 30 trials 80%, 50% and 20%	Anti-correlated Rescorla-Wagner model (the unchosen option was also updated using learning rate and PE) with separate learning rates for +PE and –PE and inverse temperature in social and non-social conditions	Mixed-effect models showed a negative relationship between correct responses and assault-exposure severity (P = 0.09).	ns	Mixed-effect models showed a negative relationship between IT and assault-exposure severity (P = 0.068).	NA
Hauser et al., 2017	Two-armed bandit task with reversals (gain or loss of 50 Swiss centimes) +50 vs –50 centimes 80% vs 20%	Anti-correlated Rescorla-Wagner model (the unchosen option was also updated using learning rate and PE) with a perseveration parameter	ns for accuracy; ns for number of reversals	ns	ns	**Perseveration** A < C
Huang et al., 2017	Changing point detection task (a variant of three-armed bandit task; gain of points as the weighted sum of reaction time, number of switches and accuracy) Three arms had changing contingencies: 9/13 vs 3/13 vs 1/13	RL model with inverse temperature and an adjustment parameter to the base learning rate when the arm with maximal value changes	ns for points earned	**Base LR** A > C	ns	ns

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
Kanen et al., 2019	Two-armed bandit task with 18 reversals (a smiling cartoon face as positive feedback and a frowning face as negative feedback) Coded as 1 vs 0 ≈85% vs ≈15%	RL model with separate learning rates for +PE and –PE, reinforcement sensitivity, stimulus stickiness (perseverance), and location stickiness	A > C† for lose-shift	**LR for +PE** ns **LR for –PE** A > C	IT was fixed at 1.	**Reinforcement sensitivity** ns **Stimulus stickiness** A < C **Location stickiness** ns
Khdour et al., 2016	Probabilistic classification task (A variant of two-armed bandit task; gain & loss of points) ±25 points vs No feedback 80% vs 20%	Basic RL model	ns for accuracy in reward condition; GA > C and SA > C in punishment condition	ns	ns	NA
Mkrtchian et al., 2017	Pavlovian probabilistic go/no-go task with and without the threat of shock (four fractals representing four conditions: go/no-go to win 10 points and go/no-go to avoid losing 10 points) ±10 points vs 0 Coded as ±1 vs 0 80 % vs 20 %	RL model with reward & punishment learning rates, reward & punishment sensitivities, approach-avoidance bias, general action bias, and lapse	A < C for accuracy	**Reward LR** ns **Punishment LR** ns	NA	**Avoidance bias** D threat > C threat; ns in safe condition **Other parameters** ns
Murray et al., 2019	Two-armed bandit task (gain or loss of 50 pence) +50 pence vs No feedback Coded as 1 vs 0 –50 pence vs No feedback Coded as –1 vs 0 0 vs No feedback 70% vs 30%	Basic RL model	A < C† for accuracy during reward trials; A < C for accuracy during punishment trials; ns for accuracy during neutral trials;	A < C†	ns	NA

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
			ns for RT in any condition			
Myers et al., 2013	Probabilistic classification task (a variant of two-armed bandit task; gain & loss of points) ±25 points vs No feedback (positive outcomes coded as 1; negative outcomes coded as –1) 80% vs 20%	RL model with separate learning rates for +PE and –PE, inverse temperature and the value of the no-feedback outcome	A > C for accuracy in reward condition	**LR for +PE** ns **LR for –PE** ns	**Temperature** ns	**No-feedback value** A < C
Norman et al., 2018	IGT (gain & loss money) with an anticipation period of 6s	RL model with learning rate, response consistency (c), reward sensitivity, loss aversion, impact of gain on perseverance, impact of loss on perseverance, decay rate of perseverance, and RL weight	ns for accuracy	ns	**c in IT = $3^c - 1$** A<C	**RL weight** A < C **Other parameters** ns
Ousdal et al., 2018	Pavlovian probabilistic go/no-go task (four fractals representing four conditions: go/no-go to win 1 Norwegian Krone and go/no-go to avoid losing 1 Krone) ±1 Krone vs 0 Coded as ±1 vs 0	RL model with reward & punishment learning rate, reward & punishment sensitivity, approach-avoidance bias, general action bias, and lapse	A < C† for overall accuracy; ns for RT	ns	NA	**Approach-avoidance bias** A > C
Piray et al., 2019	Two-armed bandit task (go & no-go as arms) with reversals	RL model with a dynamic learning rate modulated by associability-weighted PE	ns for accuracy after happy faces;	**Unmodulated LR** ns	ns	**Weight for dynamic LR after angry faces**

Study	Task	Model	Model-agnostic performance	Learning rate	Inverse temperature	Other parameters
	Four framings: emotion (angry vs happy faces) * valence (reward vs punishment) +10 cents vs 0 –10 cents vs 0 80% vs 20% 50 % vs 50%	squared, and separate angry and happy framing weight parameters quantifying the dynamic component of LR (the degree to which participants followed the Pearce–Hall associability rule)	A (stable–volatile trials) < C (stable–volatile trials) for accuracy after angry faces			A < C **Other parameters** ns
Ross et al., 2018	Two-armed bandit task (houses as arms) Locked houses cost $10 unlocked houses cost $0 Probabilities changed every 25 trials	Anti-correlated Rescorla-Wagner model (the unchosen option was also updated using learning rate and PE) with separate learning rates for +PE and –PE and inverse temperature	ns for accuracy	**LR for +PE** ns **LR for –PE** ns	A < C†	NA
Vaghi et al., 2017	Predictive inference task (particles landing in a location of a circle with small variations; locations changing on any trial with a probability of 0.125) Catching/missing resulted in +10/–10 points.	RL with learning rate for each trial (Since participants gave their estimated position in each trial, learning rate and spatial PE for each trial were directly calculated)	NA	A > C	NA	NA

Abbreviations: A, participants with anxiety disorder; C, control participants; GA, participants with general anxiety disorder; GAD, general anxiety disorder; IGT, Iowa gambling task; IT, inverse temperature; LR, learning rate; NA, not available; ns, not significant; PE, prediction error; RL, reinforcement learning; RT, response time; SA, participants with social anxiety disorder; †, marginally significant ($P < 0.1$).

Table S1.7. Group differences in neural prediction error signals in depression studies

Study	Scale of analysis	Depression< Control		Depression> Control	
		Region	MNI (x, y, z)	Region	MNI (x, y, z)
Brown, 2018	**(All PE)**[a] NA[b]	NA	NA	NA	NA
Dombrovski et al., 2013	**(Positive PE)** **Whole-brain**	R thalamus L/R STG L/R operculoinsular cortex L/R postcentral gyrus L/R SMA/cingulate	NA[c]	ns	NA
Feng, 2017	**(All PE)** **ROI** VS	ns	NA	ns	NA
Frey & McCabe, 2019	**(Reward PE)** **Whole-brain**	SPL/precuneus R insula R supramarginal gyrus R STL	–18 –58 68 48 –20 18 58 –32 24 68 –22 12	ns	NA
	ROI R VS L VS	ns ns	NA NA	ns ns	NA NA
	(Aversive PE) **Whole-brain**	ns	NA	ns	ns
	ROI R VS L VS	ns ns	NA NA	ns ns	NA NA
Geugies et al., 2019	**(Reward PE)** **Small volume** VS VTA	ns ns	NA NA	ns **sig**	NA 0 –21 –3
Gradin et al., 2011	**(Reward PE)** **Whole-brain**	L putamen L ventral striatum R ventral striatum L caudate R caudate/thalamus Midbrain R hippocampus	–28 4 0 –8 0 –6 8 2 –8 –6 2 16 8 –6 16 2 –24 –4 22 –22 –20	ns	NA
Kumar et al., 2018	**(Reward PE)** **Whole-brain**	ns	NA	ns	NA

	ROI				
	R VS	**sig**	NA	ns	NA
	L VS	ns	NA	ns	NA
	habenula	ns	NA	ns	NA
	R insula	ns	NA	ns	NA
	VTA	ns	NA	ns	NA
	(Aversive PE)				
	Whole-brain	ns	NA	ns	NA
	ROI				
	R VS	ns	NA	ns	NA
	L VS	ns	NA	ns	NA
	habenula	ns	NA	ns	NA
	R insula	ns	NA	ns	NA
	VTA	ns	NA	ns	NA
Kumar et al., 2008	**(Reward PE)**				
	Small volume	L ventral striatum	–24 6 –10	Rostral/sub-genual AC	2 54 6
		L dACC	–4 10 46	Retrosplenial cortex	–4 –60 26
				VTA	0 –21 –10
				Hippocampus	–17 –46 –10
Liu et al., 2017	**(Reward PE)**				
	Whole-brain	ns	NA	ns	NA
	(Aversive PE)				
	Whole-brain	ns	NA	R VS	6 9 –6
				L thalamus	–39 –6 9
				L substantia nigra	–9 –21 –9
Moutoussis et al., 2018	**(All PE)**				
	Whole-brain	ns	NA	ns	NA
	ROI				
	VS	ns	NA	ns	NA
Rothkirk et al., 2017	**(Reward PE)**				
	Small volume	L mOFC	–6 38 –11	ns	NA
	(Aversive PE)				
	Small volume	ns	NA	ns	NA

Abbreviations: AC, anterior cingulate; dACC, dorsal anterior cingulate cortex; L, left; MNI, Montreal Neurological Institute; mOFC, medial orbitofrontal cortex; NA, not available; ns, not significant; PE, prediction error; R, right; ROI, region-of-interest; sig, significant; SMA, supplementary motor area; STG, superior temporal gyrus; STL, superior temporal lobule; SPL, superior parietal lobule; VS, ventral striatum; VTA, ventral tegmental area.

[a] Reward PEs (RPEs) occur when outcomes only contain rewards of different amounts (e.g., \$5 vs \$1) or reward vs. omission (e.g., \$5 vs \$0). Aversive PEs (APEs) occur when outcomes only contain punishments of different amounts (e.g., –\$5 vs –\$1) or reward vs. omission (e.g., –\$5 vs \$0). They are different from positive PEs and negative PEs, in which the PE values can only be positive or negative. RPEs and APEs can contain both positive and negative PE values. "All PE" includes both RPEs and APEs. [b] Some studies did not analyze PEs or did not compare PE signals between patient and control groups. [c] The authors did not provide activation coordinates for this contrast.

Table S1.8. Group differences in neural prediction error signals in addiction studies

Study	Scale of analysis	Addiction < Control		Addiction > Control	
		Region	MNI (x, y, z)	Region	MNI (x, y, z)
Beylergil et al., 2017	**(All PE)** **Whole-brain**	 R superior PFC R middle PFC L middle PFC R angular gyrus	 25 8 63 40 33 43 27 23 45 –41 18 53 –41 11 53 –33 13 53 42 –62 43 27 –80 48 17 –72 50	ns	NA
Deserno et al., 2015	**(All PE)** **Small volume** VS	 ns	 NA	ns	NA
Feng, 2017	**(All PE)** **ROI** VS	 **sig**	 NA	ns	NA
Lesage et al., 2018	NA	NA	NA	NA	NA
Park et al., 2010	**(All PE)** **Whole-brain**	ns	NA	ns	NA
	ROI VS	ns	NA	ns	NA
Reiter et al., 2016	**(All PE)** **Whole-brain**	L mPFC L PCC	–8 62 12 –6 56 12 –2 –42 32	ns	NA
Sebold et al., 2017	**(All PE)** **Whole-brain**	ns	NA	ns	NA
Tanabe et al., 2013	**(All PE)** **Whole-brain**	L mOFC R mOFC R thalamus L insula L STG	–15 48 –18 15 33 –18 5 9 –29 0 –15 0 18 –18 12 15 –9 6 –39 30 24 –33 –30 24 –30 –18 12 –54 –3 –12 –51 –18 –5	ns	NA
	ROI DS VS mOFC lOFC	 **sig** **sig** **sig** **marginally sig**	 NA NA NA NA	 ns ns ns ns	 NA NA NA NA

Wei et al., 2018	**(All PE) Whole-brain**	ns	NA	ns	NA
	ROI Putamen	**sig**	NA	ns	NA
White et al., 2016	**(Reward PE) Whole-brain**	ns	NA	L PCC/precuneus /lingual gyrus L cuneus R cuneus	–17 –49 –5 –11 –48 19 13 –53 –7
	(Aversive PE) Whole-brain	L PCC/precuneus /lingual gyrus L cuneus R cuneus	–17 –49 –5 –11 –48 19 13 –53 –7	ns	NA

Abbreviations: DS, dorsal striatum; L, left; lOFC, lateral orbitofrontal cortex; MNI, Montreal Neurological Institute; mOFC, medial orbitofrontal cortex; mPFC, medial prefrontal cortex; NA, not available; ns, not significant; PCC, posterior cingulate cortex; PE, prediction error; PFC, prefrontal cortex; R, right; ROI, region-of-interest; sig, significant; STG, superior temporal gyrus; VS, ventral striatum.

Table S1.9. Group differences in neural prediction error signals in anxiety studies

Study	Scale of analysis	Anxiety < Control		Anxiety > Control	
		Region	**MNI (x, y, z)**	**Region**	**MNI (x, y, z)**
Brown et al., 2018	**(All PE)** **Whole-brain**	ns	NA	ns	NA
	ROI VS mPFC	ns ns	NA NA	ns ns	NA NA
Carlisi et al., 2017	NA	NA	NA	NA	NA
Cisler et al., 2015	**(All PE)** **Whole-brain**	ns	NA	L TPJ	–50 –27 16
Cisler et al., 2019	**(Negative PE)**[a] **ICA component** Salience network	**sig**	NA	ns	NA
Hauser et al., 2017	**(All PE)** **Whole-brain**	ns	NA	L ACC R putamen	–15 41 19 35 9 –2
Murray et al., 2019	**(Positive PE)** **ROI** VS	ns	NA	**sig**	NA
	(Negative PE) **ROI** ACC	ns	0 42 18	**sig**	0 42 18
Norman et al., 2018	NA	NA	NA	NA	NA
Piray et al., 2018	**(All PE)** **Whole-brain**	NA	NA	NA	NA
Ross et al., 2018	**(Positive PE)** [a] **ICA component** VS/mPFC Anterior insula L frontoparietal R frontoparietal Cingulate-preSMA	 **sig** **sig** ns ns ns	 NA NA NA NA NA	 ns ns ns ns ns	 NA NA NA NA NA

Abbreviations: ACC, anterior cingulate cortex; ICA, independent component analysis; L, left; MNI, Montreal Neurological Institute; mPFC, medial prefrontal cortex; NA, not available; ns, not significant; PE, prediction error; preSMA, pre-supplementary motor area; R, right; ROI, region-of-interest; sig, significant; TPJ, temporoparietal junction; VS, ventral striatum.

[a] In these studies, positive and negative PEs were not separately examined in the neuroimaging analysis. If the association between the timecourse of an ICA component and PEs was positive, the authors reported that this component encoded positive PEs. If this association was negative, they reported that this component encoded negative PEs. In Cisler et al. (2019), the salience network was negatively associated with PEs in anxious and control groups, but the association tended to be weaker in the anxious participants. Therefore, the authors reported that the "salience network encoding of negative PEs decreases with the severity of exposure to early-life assaultive violence." In Ross et al. (2018), the VS/mPFC and

anterior insula components were positively associated with PEs in controls, but these associations became weaker in PTSD patients; thus, the authors reported that they found "significantly less engagement of the VS/mPFC and anterior insula networks during positive PE encoding in the PTSD group."

Supplementary references

Addicott, M. A., Pearson, J. M., Wilson, J., Platt, M. L., & McClernon, F. J. (2013). Smoking and the bandit: A preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and Clinical Psychopharmacology, 21*(1), 66-73.

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology, 5*, 1-15.

Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research, 21*(4), 613-619.

Anton, R. F. (2000). Obsessive–compulsive aspects of craving: Development of the Obsessive Compulsive Drinking Scale. *Addiction, 95*, S211-217.

Anton, R. F., Moak, D. H., & Latham, P. (1995). The Obsessive Compulsive Drinking Scale: A self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. *Alcoholism: Clinical and Experimental Research, 19*(1), 92-99.

Aylward, J., Valton, V., Ahn, W.-Y., Bond, R. L., Dayan, P., Roiser, J. P., & Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nature Human Behaviour, 3*(10), 1116-1123.

Bakic, J., Pourtois, G., Jepma, M., Duprat, R., De Raedt, R., & Baeken, C. (2017). Spared internal but impaired external reward prediction error signals in major depressive disorder during reinforcement learning. *Depression and Anxiety, 34*(1), 89-96.

Beck, A., Ward, C., Mendelsohn, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893-897.

Beck, A. T., Guth, D., Steer, R. A., & Ball, R. (1997). Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behaviour Research and Therapy, 35*(8), 785-791.

Beck, A. T., & Steer, R. A. (1987). *BDI, Beck Depression Inventory: Manual*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Beylergil, S. B., Beck, A., Deserno, L., Lorenz, R. C., Rapp, M. A., Schlagenhauf, F., . . . Obermayer, K. (2017). Dorsolateral prefrontal cortex contributes to the impaired behavioral adaptation in alcohol dependence. *NeuroImage: Clinical, 15*, 80-94.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress, 8*(1), 75-90.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34*(8), 669-673.

Blanco, N. J., Otto, A. R., Maddox, W. T., Beevers, C. G., & Love, B. C. (2013). The influence of depression symptoms on exploratory decision-making. *Cognition, 129*(3), 563-568.

Brown, V. (2018). *Assessing and remediating altered reinforcement learning in depression.* (Doctor of Philosophy), Virginia Polytechnic Institute and State University, Blacksburg, VA. .

Brown, V. M., Zhu, L., Wang, J. M., Frueh, B. C., King-Casas, B., & Chiu, P. H. (2018). Associability-modulated loss learning is increased in posttraumatic stress disorder. *eLife, 7*, e30150.

Caballo, V. E., Salazar, I. C., Irurtia, M. J., Arias, B., & Hofmann, S. G. (2010). Measuring social anxiety in 11 countries. *European Journal of Psychological Assessment*.

Campbell-Sills, L., Norman, S. B., Craske, M. G., Sullivan, G., Lang, A. J., Chavira, D. A., . . . Stein, M. B. (2009). Validation of a brief measure of anxiety-related severity and impairment: The Overall Anxiety Severity and Impairment Scale (OASIS). *Journal of Affective Disorders, 112*(1-3), 92-101.

Carlisi, C. O., Norman, L., Murphy, C. M., Christakou, A., Chantiluke, K., Giampietro, V., . . . Mataix-Cols, D. (2017). Shared and disorder-specific neurocomputational mechanisms of decision-making in autism spectrum disorder and obsessive-compulsive disorder. *Cerebral Cortex, 27*(12), 5804-5816.

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychological Medicine, 40*(3), 433-440.

Cisler, J. M., Bush, K., Steele, J. S., Lenow, J. K., Smitherman, S., & Kilts, C. D. (2015). Brain and behavioral evidence for altered social learning mechanisms among women with assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 63*, 75-83.

Cisler, J. M., Esbensen, K., Sellnow, K., Ross, M., Weaver, S., Sartin-Tarm, A., . . . Kilts, C. D. (2019). Differential roles of the salience network during prediction error encoding and facial emotion processing among female adolescent assault victims. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4*(4), 371-380.

Demmel, R., & Hagen, J. (2004). The structure of positive alcohol expectancies in alcohol-dependent inpatients. *Addiction Research & Theory, 12*(2), 125-140.

Deserno, L., Beck, A., Huys, Q. J., Lorenz, R. C., Buchert, R., Buchholz, H. G., . . . Heinze, H. J. (2015). Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. *European Journal of Neuroscience, 41*(4), 477-486.

Dombrovski, A. Y., Clark, L., Siegle, G. J., Butters, M. A., Ichikawa, N., Sahakian, B. J., & Szanto, K. (2010). Reward/punishment reversal learning in older suicide attempters. *American Journal of Psychiatry, 167*(6), 699-707.

Dombrovski, A. Y., Szanto, K., Clark, L., Reynolds, C. F., & Siegle, G. J. (2013). Reward signals, attempted suicide, and impulsivity in late-life depression. *JAMA Psychiatry, 70*(10), 1020-1030.

Eckblad, M., Chapman, L., Chapman, J., & Mishlove, M. (1982). *The Revised Social Anhedonia Scale. Unpublished test*. T.R. Kwapil, Department of Psychology, University of North Carolina-Greensboro. 296 Eberhart Building, Greensboro, NC 27412-5001.

Feng, S. (2017). *Association between reward sensitivity and smoking status in major depressive disorder.* (Master of Science), Virginia Polytechnic Institute and State University, Blacksburg, VA.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological Assessment, 14*(4), 485-496.

Franken, I. H., Hendriks, V. M., & van den Brink, W. (2002). Initial validation of two opiate craving questionnaires: the Obsessive Compulsive Drug Use Scale and the Desires for Drug Questionnaire. *Addictive Behaviors, 27*(5), 675-685.

Franz, M., Meyer, T., Ehlers, F., Runzheirmer, P., & Gallhofer, B. (1998). German version of the Snaith-Hamilton-Pleasure Scale (SHAPS-D): Assessing anhedonia in schizophrenic patients. *European Psychiatry, 13*, 174s.

Frey, A.-L., Frank, M. J., & McCabe, C. (2019). Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. *Psychological Medicine*, 1-8.

Frey, A.-L., & McCabe, C. (2020). Impaired social learning predicts reduced real-life motivation in individuals with depression: A computational fMRI study. *Journal of Affective Disorders, 263*, 698-706.

Gard, D. E., Gard, M. G., Kring, A. M., & John, O. P. (2006). Anticipatory and consummatory components of the experience of pleasure: a scale development study. *Journal of Research in Personality, 40*(6), 1086-1102.

Geugies, H., Mocking, R. J., Figueroa, C. A., Groot, P. F., Marsman, J.-B. C., Servaas, M. N., . . . Ruhé, H. G. (2019). Impaired reward-related learning signals in remitted unmedicated patients with recurrent depression. *Brain, 142*(8), 2510-2522.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *Journal of Abnormal Child Psychology, 27*(1), 17-24.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., . . . Charney, D. S. (1989). Yale-brown obsessive compulsive scale (Y-BOCS). *Archives of General Psychiatry, 46*, 1006-1011.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., . . . Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain, 134*(6), 1751-1764.

Gullo, M. J., & Stieger, A. A. (2011). Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. *Drug and Alcohol Dependence, 117*(2), 204-210.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Social and Clinical Psychology, 32*(1), 50-55.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*(1), 56-62.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*(4), 278-296.

Hamilton, M. (1980). Rating depressive patients. *The Journal of clinical psychiatry, 41*, 21-24.

Hauser, T. U., Iannaccone, R., Dolan, R., Ball, J., Hättenschwiler, J., Drechsler, R., . . . Brem, S. (2017). Increased fronto-striatal reward prediction errors moderate decision making in obsessive–compulsive disorder. *Psychological Medicine, 47*(7), 1246-1258.

Hayaki, J., Stein, M. D., Lassor, J. A., Herman, D. S., & Anderson, B. J. (2005). Adversity among drug users: relationship to impulsivity. *Drug and Alcohol Dependence, 78*(1), 65-71.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerström test for nicotine dependence: A revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119-1127.

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational dysfunctions in anxiety: Failure to differentiate signal from noise. *Biological Psychiatry, 82*(6), 440-446.

Kanen, J. W., Ersche, K. D., Fineberg, N. A., Robbins, T. W., & Cardinal, R. N. (2019). Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. *Psychopharmacology, 236*(8), 2337-2358.

Khdour, H. Y., Abushalbaq, O. M., Mughrabi, I. T., Imam, A. F., Gluck, M. A., Herzallah, M. M., & Moustafa, A. A. (2016). Generalized anxiety disorder and social anxiety disorder, but not panic anxiety disorder, are associated with higher sensitivity to learning from negative feedback: Behavioral and computational investigation. *Frontiers in Integrative Neuroscience, 10*, 1-11.

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., . . . Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology, 43*(7), 1581-1588.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain, 131*(8), 2084-2093.

Kunisato, Y., Okamoto, Y., Ueda, K., Onoda, K., Okada, G., Yoshimura, S., . . . 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*(4), 1088-1094.

Lesage, E., Aronson, S. E., Sutherland, M. T., Ross, T. J., Salmeron, B. J., & Stein, E. A. (2017). Neural signatures of cognitive flexibility and reward sensitivity following

nicotinic receptor stimulation in dependent smokers: a randomized trial. *JAMA Psychiatry, 74*(6), 632-640.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22*, 141-173. doi:10.1159/000414022

Lim, T. V., Cardinal, R. N., Savulich, G., Jones, P. S., Moustafa, A. A., Robbins, T., & Ersche, K. D. (2019). Impairments in reinforcement learning do not explain enhanced habit formation in cocaine use disorder. *Psychopharmacology, 236*(8), 2359-2371.

Liu, W.-H., Valton, V., Wang, L.-Z., Zhu, Y.-H., & Roiser, J. P. (2017). Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. *Social Cognitive and Affective Neuroscience, 12*(9), 1520-1533.

Lund, M., Foy, D., Sipprelle, C., & Strachan, A. (1984). The Combat Exposure Scale: A systematic assessment of trauma in the Vietnam War. *Journal of Clinical Psychology, 40*(6), 1323-1328.

Mann, K., & Ackermann, K. (2000). Die OCDS-G: Psychometrische kennwerte der deutschen version der obsessive compulsive drinking scale. *SUCHT, 46*(2), 90-100.

Mazas, C. A., Finn, P. R., & Steinmetz, J. E. (2000). Decision-making biases, antisocial personality, and early-onset alcoholism. *Alcoholism: Clinical and Experimental Research, 24*(7), 1036-1040.

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., & Robinson, O. J. (2017). Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biological Psychiatry, 82*(7), 532-539.

Murray, G. K., Knolle, F., Ersche, K. D., Craig, K. J., Abbott, S., Shabbir, S. S., . . . Bullmore, E. T. (2019). Dopaminergic drug treatment remediates exaggerated cingulate prediction error responses in obsessive-compulsive disorder. *Psychopharmacology, 236*(8), 2325-2336.

Myers, C. E., Moustafa, A. A., Sheynin, J., VanMeenen, K. M., Gilbertson, M. W., Orr, S. P., . . . Servatius, R. J. (2013). Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: Empirical data and computational model. *PloS One, 8*(8), 1-13.

Myers, C. E., Sheynin, J., Balsdon, T., Luzardo, A., Beck, K. D., Hogarth, L., . . . Moustafa, A. A. (2016). Probabilistic reward-and punishment-based learning in opioid addiction: Experimental and computational data. *Behavioural Brain Research, 296*, 240-248.

Norman, L. J., Carlisi, C. O., Christakou, A., Murphy, C. M., Chantiluke, K., Giampietro, V., . . . Rubia, K. (2018). Frontostriatal dysfunction during decision making in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(8), 694-703.

Ousdal, O., Huys, Q., Milde, A., Craven, A., Ersland, L., Endestad, T., . . . Dolan, R. (2018). The impact of traumatic stress on Pavlovian biases. *Psychological Medicine, 48*(2), 327-336.

Park, S. Q., Kahnt, T., Beck, A., Cohen, M. X., Dolan, R. J., Wrase, J., & Heinz, A. (2010). Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *Journal of Neuroscience, 30*(22), 7749-7753.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768-774.

Piray, P., Ly, V., Roelofs, K., Cools, R., & Toni, I. (2019). Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals. *Journal of Neuroscience, 39*(8), 1445-1456.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385-401.

Reiter, A. M. F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A., & Schlagenhauf, F. (2016). Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *Journal of Neuroscience, 36*(43), 10935-10948.

Robins, L. N., Helzer, J. E., Spitzer, R., & Williams, J. (1985). *The NIMH Diagnostic Interview Schedule, Version III-A*. Washington, DC: Public Health Service.

Ross, M. C., Lenow, J. K., Kilts, C. D., & Cisler, J. M. (2018). Altered neural encoding of prediction errors in assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 103*, 83-90.

Rothkirch, M., Tonn, J., Köhler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain, 140*(4), 1147-1157.

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology, 236*(8), 2425-2435.

Royall, D. R., Mahurin, R. K., & Gray, K. F. (1992). Bedside assessment of executive cognitive impairment: The executive interview. *Journal of the American Geriatrics Society, 40*(12), 1221-1226.

Rupprechter, S., Stankevicius, A., Huys, Q. J., Steele, J. D., & Seriès, P. (2018). Major depression impairs the use of reward values for decision-making. *Scientific Reports, 8*, 13798.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The inventory of depressive symptomatology (IDS): psychometric properties. *Psychological Medicine, 26*(3), 477-486.

Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D. J., Beck, A., . . . Neu, P. (2017). When habits are dangerous: Alcohol expectancies and habitual decision making predict relapse in alcohol dependence. *Biological Psychiatry, 82*(11), 847-856.

Skinner, H. A. (1982). Drug Abuse Screening Test (DAST-20) 1982.

Skinner, H. A., & Horn, J. L. (1984). *Alcohol Dependence Scale (ADS) User's Guide*. Toronto: Addiction Research Foundation.

Skinner, H. A., & Sheu, W.-J. (1982). Reliability of alcohol use indices; The Lifetime Drinking History and the MAST. *Journal of Studies on Alcohol, 43*(11), 1157-1170.

Snaith, R. P., 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*(1), 99-103.

Spielberger, C. D. (1983). *State Trait Anxiety Inventory for Adults: Sampler Set: Manual, Test, Scoring Key*. Redwood City, CA: Mind Garden.

Spielberger, C. D., & and Gorsuch, R. L. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y) ("Self-Evaluation Questionnaire")*. Redwood City, CA: Mind Garden.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1970). *State-Trait Anxiety Inventory*. In. Palo Alto, CA: Consulting Psychologists Press.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology, 55*(1), 117-128.

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin & Review, 11*(4), 742-747.

Tanabe, J., Reynolds, J., Krmpotich, T., Claus, E., Thompson, L. L., Du, Y. P., & Banich, M. T. (2013). Reduced neural tracking of prediction error in substance-dependent individuals. *American Journal of Psychiatry, 170*(11), 1356-1363.

Tiffany, S. T., Carter, B. L., & Singleton, E. G. (2000). Challenges in the manipulation, assessment and interpretation of craving relevant variables. *Addiction, 95*, S177-187.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron, 96*, 348-354.

Watson, D., & Clark, L. A. (1991). *The Mood and Anxiety Symptom Questionnaire*. Department of Psychology, University of Iowa. Iowa City, IA.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Wei, Z., Han, L., Zhong, X., Liu, Y., Zha, R., Wang, Y., . . . Wang, W. (2018). Chronic nicotine exposure impairs uncertainty modulation on reinforcement learning in anterior cingulate cortex and serotonin system. *Neuroimage, 169*, 323-333.

White, S. F., Tyler, P., Botkin, M. L., Erway, A. K., Thornton, L. C., Kolli, V., . . . Blair, R. J. (2016). Youth with substance abuse histories exhibit dysfunctional representation of expected value during a passive avoidance task. *Psychiatry research: Neuroimaging, 257*, 17-24.

Yechiam, E., Stout, J., Lamborn, C., Mussat-Whitlow, B., Liguori, A., & Porrino, L. (2004). *Differential influence of marijuana on decision processes in the Bechara gambling task.* Paper presented at the Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Yechiam, E., Stout, J. C., Busemeyer, J. R., Rock, S. L., & Finn, P. R. (2005). Individual differences in the response to forgone payoffs: An examination of high functioning drug abusers. *Journal of Behavioral Decision Making, 18*(2), 97-110.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

Paper II: Self- and Other-Regarding Reinforcement Learning in Post-Traumatic Stress Disorder With and Without Comorbid Depression

Shengchuang Feng, George Christopoulos, Julia Julien, Pearl H. Chiu, & Brooks King-Casas

ABSTRACT

Impaired social cognition and reward processing are common in post-traumatic stress disorder (PTSD). The association between PTSD and deficits in social reward learning has also been observed. However, previous PTSD studies have been restricted to reinforcement learning (RL) about rewards delivered to oneself (self-regarding learning), ignoring the situations in which people perceive and learn about rewards delivered to others (other-regarding learning). In the current study, we used a probabilistic social learning task with different learning contexts, functional magnetic resonance imaging (fMRI) and neurocomputational analyses to test whether self- and other-regarding RL would be disrupted in combat-exposed veterans with PTSD (with and without comorbid depression) and whether individual differences in social preferences could modulate the processing of rewards delivered to others in these patients. We developed a computational model that could take account of context-dependent learning and individual differences in cooperativeness/competitiveness and found that PTSD patients, with or without depression, all showed decreased RL for other compared to healthy controls. Neurally, other-regarding surprise (magnitude of unexpected outcomes) signals in the right inferior parietal lobule (IPL) were higher in depressed PTSD participants than in controls, with nondepressed PTSD participants in between. The other-regarding surprise signals in the right IPL also had a positive correlation with a measure of avoidance & numbing across all participants. These results suggest generalized hypervigilance in PTSD about unexpected social rewards, regardless of the valence, which may result in impaired other-regarding learning. Self-regarding learning was not affected by PTSD, but comorbid depression was associated with increased learning for self, which could be tentatively accounted for by the reduced striatal gray matter volume in depressives. This study provides new evidence of impaired social RL in PTSD and may help to uncover behavioral and neural mechanisms of PTSD and reveal potential targets for its treatment.

Keywords: PTSD, depression, reinforcement learning, self-regarding learning, other-regarding learning, social reward, prediction error

INTRODUCTION

Post-traumatic stress disorder (PTSD) is a prevalent mental health problem among individuals who have experienced extremely stressful events (Nietlisbach & Maercker, 2009). The harm of the traumatic events can be directed to oneself or others, usually resulting in symptoms such as intrusive recollections or re-experiencing of the events, avoidance and numbing, hyperarousal, blaming of self/others, and feeling distant from others (American Psychiatric Association, 2000; Stevens & Jovanovic, 2019). Some of the symptoms reflect difficulties in social functioning (Frueh, Turner, Beidel, & Cahill, 2001). Moreover, disturbances in self-perception (Bluhm et al., 2012; Lanius, Frewen, Nazarov, & McKinnon, 2014) and other-perception (mentalizing; Nietlisbach, Maercker, Rösler, & Haker, 2010; Plana, Lavoie, Battaglia, & Achim, 2014; Sharp, Fonagy, & Allen, 2012) have also been reported in PTSD by many studies, implying that social cognitive deficits might be a key aspect of PTSD. Meanwhile, impaired reward perception (Elman et al., 2005; Hopper et al., 2008) and altered reward learning (Brown et al., 2018; Cisler et al., 2015; Myers et al., 2013; Ousdal et al., 2018; Ross, Lenow, Kilts, & Cisler, 2018) are present in PTSD as well, and they are more easily detected by tasks applying social stimuli (e.g., faces), suggesting that PTSD may be specifically related to deficits in social reward processing (Nawijn et al., 2015).

In previous PTSD studies, however, the reward tasks only involved the processing of rewards delivered to oneself, ignoring the situations in which people perceive and learn about rewards delivered to others (Christopoulos & King-Casas, 2015; Liu et al., 2019; Sul et al., 2015). As noted above, PTSD can develop following witnessing harms directed to others, and disrupted mentalizing is common in PTSD. Therefore, we ask if there is an association between PTSD and deficits in other-regarding reward processing. Given the evidence that the extent to which people value others' rewards varies at the individual level (McClintock, 1972; Messick & McClintock, 1968; Van Lange, 1999), we also ask whether individual differences in social preferences (other-regarding valuation) can modulate the processing of rewards delivered to others in PTSD. In the current study, veterans with combat-related PTSD and combat-exposed controls were tested on a probabilistic social learning task, in which they chose between two abstract patterns that had different reward contingencies. Their choices affected the monetary payoffs for both themselves and an unknown other (Christopoulos & King-Casas, 2015).

Reinforcement learning (RL) models have been used to describe the learning process in probabilistic learning tasks (Sutton & Barto, 1998), assuming that a learner assigns expected values to the abstract patterns and updates these values based on the prediction error (PE; the difference between the received value and expected value for each pattern). The PEs are positive when the learner obtains unexpected better values and are negative when the learner obtains unexpected worse values. As a teaching signal, the PEs drive the learner's estimation of a pattern's value to converge to its real value through trial-by-trial updating. In a social learning task with rewards for

oneself and others, an RL model can have separate self-regarding and other-regarding updating processes (Christopoulos & King-Casas, 2015). Based on this model, we tested whether veterans with PTSD and controls differed in these processes.

We also collected participants' neuroimaging data with functional magnetic resonance imaging (fMRI) and tested the group differences in neural representations of the self-regarding PEs in the ventral striatum and other-regarding PEs in the anterior cingulate cortex (ACC). The ventral striatum has been consistently found to encode self-regarding PEs (Berns, McClure, Pagnoni, & Montague, 2001; Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997); the ACC is suggested to encode other-regarding information (Apps, Rushworth, & Chang, 2016), including other's reward probabilities (Lockwood, Apps, Roiser, & Viding, 2015) and other's PEs (Apps, Lesage, & Ramnani, 2015). Moreover, some other brain regions, e.g., the amygdala, exhibit hyperactivation when PTSD individuals view positive and negative versus neutral social stimuli [e.g., happy (Killgore et al., 2014) and fearful faces (Armony, Corbo, Clément, & Brunet, 2005; Bryant et al., 2008; Rauch et al., 2000)], suggesting that these regions encode the magnitude rather than the valence of the stimuli. Based on this, we also examined the group differences in brain regions representing unsigned PE (surprise) signals (Fouragnan, Retzler, & Philiastides, 2018) for others.

Since depression and PTSD are highly comorbid (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995), and depression has been found to be associated with decreased reward sensitivity (Huys, Pizzagalli, Bogdan, & Dayan, 2013; Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008) and diminished striatal responses to reward (Pizzagalli et al., 2009), we also tested the effects of comorbid depression on self- and other-regarding behavioral learning and PE related neural activity.

METHODS

Participants

Ninety-eight combat-exposed US military veterans participated in this study. Written informed consent was acquired from all participants, and this study was approved by the Institutional Review Boards at Virginia Tech, Baylor College of Medicine, and the Salem Veterans Affairs Medical Center. All participants were English speaking, had verbal IQ greater than 80, and had normal or corrected to normal vision. PTSD and depression diagnoses were determined through the full Structured Clinical Interview for DSM-IV Axis I Disorders (First, Spitzer, Gibbon, & Williams, 2007). Exclusion criteria included: younger than 18 or older than 65; history of seizure disorder, stroke, hormone disorder; history of electroconvulsive therapy or chemotherapy for cancer; current pregnancy or menopause; fMRI contraindications; acute suicidal ideation; concurrent diagnosis of bipolar disorder, schizophrenia, schizoaffective disorder, delusional disorder, and organic psychosis.

After excluding six participants without current mental disorders but with traumatic

brain injury (TBI), nine participants with past or subthreshold PTSD, and another nine participants with missing behavioral data or abnormal performance [see **SUPPLEMENTARY INFORMATION** (**SI**)], we had 15 healthy controls (HC), 29 PTSD patients with no current depression (NP), and 30 PTSD patients with current depression (DP) in our data analysis (**Table 2.1**).

The three groups were not significantly different in sex ratio [$\chi^2(1)$'s ≤ 1.38, P's ≥ 0.24]. Compared to the controls, the PTSD groups were younger [one-way analysis of variance (ANOVA) for age: $F(2, 71) = 2.88$, $P = 0.063$; NP-HC: $t = -2.29$, $P = 0.025$; DP-HC: $t = -0.96$, $P = 0.34$; NP-DP: $t = -1.62$, $P = 0.11$], had fewer years of education [$F(2, 71) = 7.32$, $P = 0.001$; NP-HC: $t = -3.69$, $P < 0.001$; DP-HC: $t = -3.20$, $P = 0.002$; NP-DP: $t = -0.61$, $P = 0.54$] and had more combat exposure [$F(2, 66) = 7.64$, $P = 0.001$; NP-HC: $t = 11.37$, $P < 0.001$; DP-HC: $t = 8.52$, $P = 0.004$; NP-DP: $t = 2.85$, $P = 0.25$].

For imaging analysis, two participants with missing imaging data and 12 participants with excessive head motion in the scanners (cumulative translation > 5 mm and rotation > 5˚) were excluded, resulting in 60 participants. All tests on group differences and correlations across all three groups were conducted by controlling for sex, age, years of education, and combat exposure. Combat exposure had five missing participants for the remaining sample; therefore, there were 55 participants (12 HC, 19 NP, and 24 DP) in tests on group differences and correlations involving neuroimaging data.

Assessments

The following self-report measures were administered to all participants prior to the fMRI scanning session: demographic information, Beck Depression Index (BDI; Beck, Steer, & Brown, 1996), Clinician Administered PTSD Scale (CAPS; Blake et al., 1995), Combat Exposure Scale (CES, Lund, Foy, Sipprelle, & Strachan, 1984), PTSD Checklist-Military version (PCL-M; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996), Brief Traumatic Brain Injury Screen (BTBIS; Schwab et al., 2006), and Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). Besides, as an index of social preferences, the social value orientation (SVO; the preference for allocating rewards between oneself and others) was also assessed through a non-learning task with a sequential testing procedure (Christopoulos & King-Casas, 2015; Luce, 2014), in which participants serially made preference choices between two allocations of points for themselves and an anonymous partner (e.g., [Self: +50, Other: +85] vs. [Self: +85, Other: –50]; **Figure S2.1a**). Since each allocation can be represented by an arrow with a certain angle (**Figure S2.1b**), we can see choices as being made between two arrows. After each choice, the chosen arrow (allocation) was retained as one option in the next trial, and the unchosen arrow moved toward the chosen one with a certain step size. Based on this procedure, the two arrows gradually converged to each participant's preferred allocation. This assessment was repeated three times, with different initial allocations. The mean of the three measurements was used to

represent a participant's SVO (in degrees).

Experimental task

During the scanning session, participants performed a probabilistic social learning task with points for themselves and an anonymous partner, the same as in Christopoulos & King-Casas (2015). Prior to the scanning, participants were instructed to collect as many points as they could. They were also informed that: (i) they would never meet the other person or know each other's identity; (ii) the participant and the other person would be paid based on the outcomes of a random subset of the task trials; and (iii) the other person would not perform a task that could influence the participant's payoffs.

Given that the allocations between oneself and the other person can be represented as arrows with a certain angle, there are four kinds of arrows/allocations positioned in the four quadrants of a Cartesian coordinate system (x-axis for self and y-axis for other): Quadrant I [self-gain & other-gain], Quadrant II [self-loss & other-gain], Quadrant III [self-loss & other-loss], and Quadrant IV [self-gain & other-loss]. To examine the learning of different allocations, the task was designed to consist of six conditions or contexts, each including two of the four allocations (**Figure 2.1**). In each condition, participants could learn the reward contingencies of two abstract patterns by making choices between them. One pattern was associated with an 80% probability of a certain allocation and the other one was associated with an 80% probability of another allocation. The exact values for different allocations in each trial were randomly sampled from a uniform discrete distribution with a mean of +70 or –70 and a range of 20. There were 30 trials for each condition, which also formed one block. The order in which the six conditions/blocks were presented was pseudorandomized across participants. Between two blocks, there was an instruction screen to remind the participant that there would be "new sets of symbols".

The procedure of one trial was as follows: at the start, a fixation cross was shown at the center of the screen for 1s plus a value randomly selected from an exponential distribution with a mean of 1, truncated at 6. Then two abstract patterns were shown on the screen, allowing the participant to make a choice via a scanner-compatible button box within 3 s. The positions of the two patterns were randomized. After the participant's response, that chosen pattern was framed for 0.5 s plus a value also randomly selected from an exponential distribution with a mean of 1 and truncated at 6. Following this jittered confirmation, the outcomes for self and other were displayed for 2 s. The self- and other-outcomes were also randomly positioned to be above or under the center of the screen.

Behavioral analysis

Model-agnostic analysis

For the probabilistic social learning task, choices of the high value option for self and other and response time in all six conditions were calculated for each participant. The

group differences in these performance measures were tested using ANOVAs with diagnostic groups and task conditions as two factors and sex, age, and education as covariates. Trial-by-trial learning curves were also plotted for each condition. The learning curves depict the running average of the trial-by-trial proportion of participants that selected the high value option in each diagnostic group. To check if SVO would modulate learning in different conditions, separate learning curves for cooperative (SVOs > 5) and competitive (SVOs < –5) participants were plotted for each of conditions 1, 6, 3, and 4, in which they were supposed to only learn for themselves or others.

Computational modeling

To disentangle various cognitive components in the learning process, we fitted participants' choice data to different RL models. As mentioned in the introduction section, the basic form of an RL model (Sutton & Barto, 1998) describes an updating rule (Rescorla & Wagner, 1972) of the expected value for a certain option:

$$EV_t = EV_{t-1} + \alpha^* (V_{t-1} - EV_{t-1}). \qquad \text{(Equation 2.1)}$$

In this equation, EV_t and EV_{t-1} denote the expected value of an option at time/trial t and t–1, respectively; V_{t-1} denotes the actual value received by a learner after selecting that option at time t–1; $(V_{t-1} - EV_{t-1})$ is termed PE, denoting the difference between the received value and the expected value at time t–1. α is learning rate and can take values between 0 and 1. The higher it is, the faster the learner updates the expected value of an option and relies more on most recent information versus past reward history.

When multiple options are present, the learner can update the expected value for each option after it is chosen and delivers rewards. The learner's probability of choosing a certain option at each time point can be modeled with a standard softmax function (Luce, 1959):

$$P_{a,t} = \frac{\exp(\beta^* EV_{a,t})}{\exp(\beta^* EV_{a,t}) + \exp(\beta^* EV_{b,t})} \qquad \text{(Equation 2.2)}$$

where $P_{a,t}$ denotes the probability of choosing option *a* out of two options (*a* and *b*) at time t; β is inverse temperature and can be used as a measure of the noisiness or randomness of the learner's choices. It can take values equal to or larger than 0. The higher the inverse temperature is, the less random the choices are and the more likely the learner would choose the option with the highest expected value.

To account for learning for oneself and others, the gamma model (Equation 2.3) was proposed by Christopoulos & King-Casas (2015). In this model, the expected value for self (EV_S) and the expected value for other (EV_O) are updated through the learning rate (α_S) and PEs for self and the learning rate (α_O) and PEs for other, respectively. In

a learning task with each option associated with both self- and other-outcomes, the expected value of one option is calculated by summing up these two expected values. To represent the social preference of the learner, the outcomes for the other person are weighted by a gamma parameter, which is a discrete variable that can take the values –1, 0, and 1 in Christopoulos & King-Casas (2015)'s original paper. A competitive person would have the gamma as –1, meaning that he/she sees others' gains as his/her losses. A cooperative participant's gamma would be 1 in that he/she sees others' gains as his/her gains. A gamma of 0 suggests that the person is indifferent to others' outcomes. Given the fact that the SVO measure is a continuous variable and the SVOs of most participants in the present study were between –45° and +45° (The tangents of these two degrees are –1 and 1), γ was set as a continuous variable and bounded between –1 and 1.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * (V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * (\gamma * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \quad \text{(Equation 2.3)}$$

Based on the gamma model, we developed the angle distance model (Equation 2.4). This model takes account of the difference between the learner's preferred allocation and the received outcomes in each trial, both of which can be represented by arrows (**Figure 2.1c**). This difference is termed angle distance (A) and the units are radians. Instead of a single γ in the gamma model, the angle distance model uses the preferred allocation (η), the angle distance, and a weight on angle distance (κ) to dynamically transform outcomes received by the other person to subjective values for the learner. Here, η is the parameter representing the learner's social preference and was set as a continuous variable bounded between –1 and 1.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * (V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * ((\eta + \kappa * A_{t-1}) * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \quad \text{(Equation 2.4)}$$

It is possible that the learner's perception of outcomes for self is similarly influenced by the angle distance. Therefore, we also tested the following double angle distance model (Equation 2.5) with the outcome transformation mechanisms applied to both self- and other-outcomes.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * ((1 + \kappa_S * A_{t-1}) * V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * ((\eta + \kappa_O * A_{t-1}) * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \quad \text{(Equation 2.5)}$$

There are also models that take account of inequality of rewards between self and other. The Fehr-Schmidt model (Fehr & Schmidt, 1999) is a well-known model of this kind, and it can be written as follows for an RL task:

$$EV_t = EV_{t-1} + \alpha * (V_{S,t-1} - w_a * (V_{S,t-1} - V_{O,t-1}) - EV_{t-1}) \quad \text{if} \quad V_{S,t-1} - V_{O,t-1} > 0$$
$$EV_t = EV_{t-1} + \alpha * (V_{S,t-1} - w_b * (V_{O,t-1} - V_{S,t-1}) - EV_{t-1}) \quad \text{if} \quad V_{S,t-1} - V_{O,t-1} < 0$$
$$w_b \geq w_a \geq 0;\ w_a < 1 \quad \text{(Equation 2.6)}$$

where w_a measures participants' compassion or guilt when self is better off than other and w_b measures envy when self is worse off than other.

Another inequality model (Equation 2.7) proposed by Van Lange (1999) takes account of cooperation and egalitarianism in perceiving outcomes for self and other. It has three weight parameters (w_S, w_O, and $w_{|S-O|}$) respectively multiplied on self-outcomes, other-outcomes, and the differences between self- and other-outcomes.

$$EV_t = EV_{t-1} + \alpha * (w_S * V_{S,t-1} + w_O * V_{O,t-1} + w_{|S-O|} * |V_{S,t-1} - V_{O,t-1}| - EV_{t-1}) \quad \text{(Equation 2.7)}$$

The aforementioned basic RL model with no learning for other and the five other-regarding RL models were fitted using the participants' choice data. Hierarchical Bayesian analysis (performed with the Stan software package, version 2.16.0; Stan Development Team, 2017) was used to estimate the six models [see **SUPPLEMENTARY INFORMATION** (**SI**) for more details]. As part of the model estimation, PTSD and depression effects on the group-level mean (μ) for each learning parameter were tested using a regression equation:

$$\mu = \mu_{\_intercept} + PTSD * \mu_{\_slope1} + depression * \mu_{\_slope2} + covariate_i * \mu_{\_slope\_cov\_i} \quad \text{(Equation 2.8)}$$

in which *PTSD* and *depression* were dummy variables coded according to each participant's diagnostic status, and their effects on the group-level mean were represented by regression slopes. Since there were no participants with depression alone, no interaction term between PTSD and depression was included in the regression model. To control for sex, age, education, and combat exposure, these variables were entered as covariates (see **SI**). After model estimation, model-fit indices, including the integrated Bayesian information criterion (iBIC; Huys et al., 2012), widely applicable information criterion (WAIC; Watanabe, 2010), and leave-one-out cross-validation information criterion (LOOIC; Vehtari, Gelman, & Gabry, 2017), were calculated for all six models.

For the winning model, the 95% credible intervals or highest density intervals (HDI) of the regression slopes were used to determine the significance of PTSD and depression effects (Significant effects require that HDI does not overlap zero; Kruschke & Vanpaemel, 2015). We also calculated the proportion of samples in each slope's distribution that had the opposite sign to its mean. It can serve as a significance measure similar to frequentist *P* values (Mack, Preston, & Love, 2020), , although it was not used for the statistical inferences in the present study. As for the individual-level parameters, their estimates were extracted to generate PEs for imaging analysis and to test the correlations between learning parameters and other variables.

After the best-fitting model was selected, we conducted a parameter recovery analysis to verify that the parameters in the winning model could be reliably recovered and did not have the parameter identification problem (The parameter identification problem

occurs when choice data can be generated with more than one set of parameters; Greenberg & Webster, 1983). For each parameter, we randomly drew 50 values from a uniform distribution. These generative parameters were randomly assigned to 50 hypothetical participants to generate choice data using the winning model, and then new learning parameters were estimated using these choice data. High correlations between the generative and recovered parameters indicate that the estimation method can reliably capture the true parameter values.

Imaging analysis

The functional and anatomical imaging was conducted on 3 T Siemens Trio MR scanners (Siemens, Munich, Germany) at Baylor College of Medicine. SPM8 software package (Wellcome Trust Centre for Neuroimaging, London, UK) was used for imaging data preprocessing and analysis (see **SI** for details of image acquisition and preprocessing).

To examine the neural substrates of self-regarding PEs, other-regarding PEs, and other-regarding surprise values, three first-level general linear models (GLMs) were constructed for each participant using an event-related analysis procedure. In each GLM, the prepossessed functional imaging data were set as a dependent variable and the task events convolved with a hemodynamic response function (HRF) were set as independent variables. For all three GLMs, the events were the same, including the presentation of two options before choices, choices, confirmation of choices, and outcomes. Trial-by-trial self-PEs $[(1 + \kappa_S * A_t) * V_{S,t} - EV_{S,t}]$, other-PEs $[(\eta + \kappa_O * A_t) * V_{O,t} - EV_{O,t}]$ and surprise values (unsigned other-PEs) calculated from the best-fitting model were respectively entered into the three GLMs as a parametric modulator of the outcome event. Six head motion parameters were also included in the GLMs as regressors of no interest. Standard linear regression and parametric modulation analyses were performed to obtain beta maps for the three PE signals. Then contrast maps of surprise values were entered into the second-level random-effect whole-brain analysis using a one-sample t-test across all participants. A corrected threshold of $P < 0.05$ [false discovery rate (FDR) corrected at voxel-level and a minimum cluster size of 50 voxels] was used.

In the region-of-interest (ROI) analysis, the anatomical mask of the bilateral ventral striatum from the Oxford-GSK-Imanova structural and connectivity striatal atlases (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) was used as the self-PE ROI. Given the ACC's role in tracking other's rewards (Apps et al., 2016), the other-PE ROI was defined as a sphere with a radius of 6mm, centered at MNI (Montreal Neurological Institute template) coordinates [5 31 12], which are the middle point of the two peak voxels from Lockwood et al. (2015; MNI: [8,32,12], involved in processing other's reward probabilities) and Apps et al. (2015; MNI: [2, 30, 12], involved in processing other's PEs). The other-surprise ROI was defined as a sphere with a radius of 6mm, centered at the whole-brain peak coordinates from the second-level one-sample t-test. The three types of learning signals for each participant were extracted from the ROIs

in the corresponding beta maps and were compared between the three groups using one-way ANOVAs with sex, age, education, and combat exposure as covariates. The correlations between the learning signals and other variables were also tested.

RESULTS

Behavioral results

Model-agnostic performance

The three groups were not significantly different in choices of the high value option for self [in task conditions 1, 2, 5, and 6; group factor: $F(2, 68) = 0.86$, $P = 0.43$; task factor: $F(3, 213) = 1.07$, $P = 0.36$; interaction: $F(6, 213) = 0.68$, $P = 0.67$] or for other [in conditions 3 and 4; group factor: $F(2, 68) = 0.15$, $P = 0.86$; task factor: $F(1, 71) = 3.67$, $P = 0.060$; interaction: $F(2, 71) = 0.42$, $P = 0.66$], and were comparable in response time [group factor: $F(2, 68) = 0.42$, $P = 0.66$; task factor: $F(5, 355) = 2.33$, $P = 0.043$; interaction: $F(10, 355) = 1.43$, $P = 0.16$; **Figure S2.2**; see **SI** for more details].

Learning curves in conditions 1, 2, 5, and 6 showed increasing trends of choosing the high value option for self, indicating participants' acquisition of the reward contingencies for self. However, in conditions 3 and 4, where the participants were supposed to learn the rewards delivered to others, only rather flat learning curves were found (**Figure S2.3**).

In condition 3, when we only looked at the learning curves of the participants with the SVO measures larger than 5 (cooperative), the controls showed a trend of more choices of the high value option for other than the PTSD groups. For the participants with the SVO measures lower than –5 (competitive), the learning curves of the controls showed a trend of fewer choices of the high value option for other than the PTSD groups. These patterns suggest that, when the SVO was taken into account, the controls showed better learning than PTSD participants. No similar divergences of learning curves for other were present in condition 4 (**Figure 2.2b**). The learning curves for self did not show clear divergence for the three groups in condition 1 or 6 (**Figure 2.2a**).

Computational modeling results

The model comparison result indicated that the double angle distance model was the winning model, which had the smallest model-fit indices (**Table 2.2**). The highly significant correlations between generative and recovered parameters in parameter recovery suggested that the winning model and the estimation method could reliably capture the true values of the parameters (**Figure S2.4**).

Bayesian-estimated regressions from this model demonstrated no effects of PTSD on the group-level mean of self-regarding learning rates, but participants with depression had higher learning rates for self [$\mu_{slope2} = 0.81$, HDI: [0.29, 1.34], $P = 0.001$; **Figure 2.3**; **Table S2.2**]. Pairwise comparisons between groups showed that DP participants

had higher learning rates for self than NP, with controls lying in between (**Table S2.3**).

As for the other-regarding learning rate, a significant effect of PTSD was present [$\mu_{\_slope1} = –2.56$, HDI: [–4.01, –1.31], $P < 0.001$], with lower other-regarding learning rates in PTSD participants. Pairwise comparisons also showed higher learning rates for other in controls than in the two PTSD groups (**Figure 2.3**; **Tables S2.2** and **S2.3**). No effects of depression were found for this parameter.

For other model parameters, a marginal effect of PTSD was found for angle distance weights for self [$\mu_{\_slope1} = –0.33$, HDI: [–0.69, 0], $P = 0.025$], mainly driven by the less negative weights in HC than NP. Pairwise comparisons also revealed higher inverse temperatures in NP than HC (**Figure 2.3**; **Tables S2.2** and **S2.3**).

Imaging results

Prediction error signals

In the ventral striatum ROI, self-PE signals were significantly larger than 0 across all participants [$t(59) = 0.48$, $P < 0.001$], but no significant differences were found between the groups [one-way ANOVA: $F(2, 48) = 0.32$, $P = 0.73$; **Figure 2.4a**]. For the ACC ROI, other-PE signals were not different from 0 across all participants [$t(59) = –0.13$, $P = 0.70$], and no significant group differences were found [$F(2, 48) = 0.07$, $P = 0.93$; **Figure S2.4b**].

Other-regarding surprise signals

For the other-regarding surprise values (unsigned PEs), the whole-brain analysis of all participants revealed robust activations in the bilateral inferior parietal lobule (IPL), bilateral supplementary motor area and right middle frontal gyrus ($P < 0.05$, FDR corrected for multiple comparisons at voxel-level over the whole brain; **Figure 2.4c**; **Table S2.5**). In the IPL ROI centered at MNI coordinates [48, –51, 48] (the whole-brain peak coordinates), the other-regarding surprise signals showed significant group differences [$F(2, 48) = 3.98$, $p = 0.025$]. Post-hoc tests showed that DP had higher surprise signals than HC ($t = –2.60$, $p = 0.012$), with NP in between (**Figure 2.4d)**. After controlling for sex, age, education, and combat exposure, the IPL surprise signals for other were negatively correlated with the other-regarding learning rate ($r = –0.30$, $P = 0.026$); the IPL surprise signals for other were positively correlated with the PCL-M avoidance & numbing subscale across all participants ($r = 0.30$, $P = 0.025$; **Figure 2.4e**).

DISCUSSION

In the present study, we tested whether PTSD and depression were associated with deficits in other-regarding RL and whether individual differences in social preferences could modulate this association. We found that PTSD patients, with or without depression, all showed decreased other-regarding learning rates compared to healthy controls. Consistent with this, our model-agnostic regression analysis also indicated

worse predictions of choices based on other-outcomes in the PTSD groups (**Table S2.1**). These findings suggest that PTSD is associated with impaired other-regarding learning, providing support for previous findings showing deficits in social cognition (Plana et al., 2014; Sharp et al., 2012) and reward processing (Elman et al., 2005; Hopper et al., 2008; Ousdal et al., 2018) in PTSD.

It is worthwhile to mention that this PTSD effect could only be detected when individual differences in social preferences were taken into account. Cooperative individuals see rewards received by others similar to rewards received by themselves, so they tend to make choices that benefit others. On the contrary, competitive individuals see rewards for others as punishments for themselves, so they tend to make choices leading to others' losses (Christopoulos & King-Casas, 2015). In conditions 3 and 4 of our social learning task, the participants were assumed to mainly learn for others in that self is always winning (in condition 3) or always losing (in condition 4). As shown by the learning curves in condition 3, no sign of learning can be seen when participants with different social preferences were lumped together, but when cooperative or competitive participants were separated, controls demonstrated clear learning and they learned faster than PTSD patients. In condition 4, the participants were also assumed to learn for others, but no matter whether SVO is considered, no obvious learning in any group can be seen. This lack of learning might be attributed to the insecure context of condition 4, in which the participants only saw losing for themselves and would perceive others' outcomes as irrelevant or less salient. These results suggest that not only should social preferences be considered when examining social learning, but the learning contexts should also be taken account of.

The angle distance model and the double angle distance model were constructed to capture both the context-dependent learning as well as individual differences in social preferences. In the angle distance model, the other-outcomes are transformed as a function of both individual-level preferred allocations and trial-by-trial angle distances. This notion is supported by a recent study (Liu et al., 2019), in which participants were shown to use their individual preferred allocations as a reference point for rating potential self-other allocations in a reward evaluation task. The double angle distance model moved a step further by applying a similar dynamic transformation to self-outcomes, and it had the best model-fit among all candidate models. This model can help to explain some model-agnostic findings. If we derive the differences between expected values of the two options in the six conditions from the double angle distance model, we can see that the average difference in condition 4 (0.13) is much smaller than that in other conditions (all larger than 0.30), which might be the reason for the lack of learning in condition 4 (**Table S2.4**). This smaller difference in expected values also makes it more difficult to choose between the two options, which is reflected by the longer response time in condition 4 relative to condition 3 (**Figure S2.2b**).

Neuroimaging analysis failed to find group differences in PE signals for other in the

ACC (**Figure S2.5a**) but observed heightened IPL surprise signals in PTSD. Specifically, surprise signals in the right IPL were higher in PTSD participants with comorbid depression than in controls, and the strength of signals of nondepressed PTSD participants was intermediate between these two groups. Consistent with this pattern, the avoidance & numbing subscale of PCL-M was positively correlated with the surprise signals across all participants. The IPL is a key brain region in the mirror system (Van Overwalle & Baetens, 2009) and has been shown to be involved in agency (Chaminade & Decety, 2002), self-other discrimination (Uddin, Molnar-Szakacs, Zaidel, & Iacoboni, 2006), and processing moderately unfair monetary allocations in an ultimatum game (Polezzi et al., 2008). The avoidance & numbing subscale measures the extent to which the trauma victims avoid stimuli, including people, that remind them of traumatic events and feel estranged from others (Norris, Hamblen, Wilson, & Keane, 2004). The heightened responses of IPL in PTSD and its association with the avoidance & numbing subscale suggest that PTSD patients may have generalized hypervigilance about unexpected social rewards, and they may view the magnitude of the PEs as more relevant but ignore their valence. It is also postulated that PTSD patients interpret positive stimuli as threatening or aversive (Frewen, Dean, & Lanius, 2012), and thus both positive and negative PEs are seen as negatively-valenced stimuli. As a result, this hypervigilance and negative interpretation of positive stimuli may disrupt their attention to and incorporation of new reward information, further leading to impaired learning of other-regarding rewards. This explanation gains partial support from Cisler et al., (2015)'s study. They found that PTSD females showed greater responses to social PEs in the left temporoparietal junction, a mentalizing key region, in a probabilistic learning task with smiling faces as reinforcers, but in a trust game, the same participants showed decreased learning rates for reciprocated money from investees.

As for self-regarding learning, our Bayesian-estimated regressions demonstrated an association between depression and increased learning rates for self. The increased learning rate indicates more attention to the most recent reward information, but faster forgetting of past reward history (Busemeyer & Stout, 2002), suggesting impaired attentional or memory function related to self-regarding rewards in depression. In our study, although self-PE signals in the ventral striatum were similar across all groups, reduced gray matter volume of the ventral striatum was observed in the depressives and there was a robust negative correlation between the ventral striatum volume and learning rates for self across all participants (**Figure S2.5a**, **b**). Similarly, an earlier study found altered sensitivity to rewards in healthy males with reduced striatal volume (Barrós-Loscertales et al., 2006). These studies point to the possibility that the structural properties of the ventral striatum might be associated with some specific aspects of reward processing (e.g., learning rates for self) but other aspects may remain intact (PE signals for self).

The depressives showed another difference with the other two groups: in HC and NP, the SVOs measured in a non-learning task and the preferred allocations estimated

from the best-fitting model were positively correlated, suggesting that the two variables reflect the same construct. However, this positive correlation disappeared in the depressed group (**Figure S2.5c**). One possibility is that the winning model did not fit the DP's choice data very well, so the estimation of the preferred allocations was not reliable. It turned out that the winning model's predictive accuracy of the choices was comparable between the three groups (HC: 0.72; NP: 0.73; DP: 0.76); therefore, this possibility is excluded. Another explanation is that the PTSD patients with comorbid depression had unstable social preferences, which might be different depending on the structure of the tasks or their physiological state. A possible cause of this is medication. Previous studies have shown that antianxiety and antidepressant agents can modify some cognitive functions like attention and memory (Amado-Boccara, Gougoulis, Littre, Galinowski, & Loo, 1995; Hindmarch, 1999). Therefore, the depressed group with more participants on medication might perform very differently in the two tasks. Consistent with this, we observed larger discrepancy between the two social preference measures in participants who were on medication than those who were not. Besides, the size of the discrepancy was also positively correlated with the number of medications across all participants (**Figure S2.5d**).

To summarize, the present study, by applying a newly developed RL model, tested self- and other-regarding learning in PTSD with and without depression. Along with some other learning studies (e.g., Cisler et al., 2015), it expands the scope of PTSD research from social reward perception to social reward learning. The findings from this study also help to uncover the mechanisms of neural and behavioral social learning deficits in PTSD, not only lending new tests for hypotheses and theories from previous studies, but also providing potential targets for the treatment of PTSD. Moreover, the interpretation of some results in this study remains unclear, e.g., the unstable social preferences in depressed PTSD patients, which can serve as hypotheses to be tested by future studies. Despite its implications, some limitations of our study should also be noted. Firstly, since it is correlational rather than causal, we cannot exclude the possibility that the impaired social learning predisposes to PTSD initiation. To address this issue, future studies can apply prospective research to collect data before new army recruits enter their military service, and test what factors contribute to the development of PTSD (Hendler & Admon, 2016). Secondly, our participants did not include a depression only group, preventing us from testing PTSD and depression as two separate factors. Future PTSD studies are suggested to recruit participants that can facilitate the separation of effects from different comorbid disorders. Lastly, our participants were predominantly males and all were veterans; therefore, our findings are yet to be replicated in females or other populations in future research.

TABLES AND FIGURES

Table 2.1. Demographic information and scale measures

	Healthy Controls (HC; N = 15)	Nondepressed PTSD (NP; N = 29)	Depressed PTSD (DP; N = 30)
Sex	0 females	2 females	5 females
Ethnicity (proportion of white)	12/15	16/29	21/30
Age (years)	37.27 ± 2.63	30.90 ± 2.27	34.60 ± 1.76
Education (years)	16.00 ± 0.48	14.03 ± 0.29	14.30 ± 0.31
WTAR	114.64 ± 2.59	104.41 ± 2.19	103.90 ± 1.92
CES (combat exposure)[a]	13.07 ± 2.68	24.44 ± 1.92	21.59 ± 1.43
TBI	0/15	20/29	20/30
On medication	0/15	11/29	20/30
Antidepressants	0/15	9/28	18/30
Antianxiety agents	0/15	4/28	4/30
Stimulants	0/15	1/28	3/30
Antipsychotics	0/15	0/28	5/30
Mood stabilizers	0/15	0/28	8/30
Current smokers	1/14	5/19	7/26
Past smokers	3/14	6/19	8/26
Past MDD/dysthymia	3/15	21/29	0/30
Current MDD/dysthymia	0/15	0/29	30/30
Past alcohol dependence/abuse	9/15	19/29	18/30
Past drug dependence/abuse	2/15	4/29	6/30
PCL-M	27.73 ± 3.13	53.93 ± 2.00	61.43 ± 2.18
Cluster B (re-experiencing)	7.80 ± 0.96	16.10 ± 0.82	16.70 ± 0.86
Cluster C (avoidance & numbing)	10.73 ± 1.36	20.28 ± 0.94	24.67 ± 1.03
Cluster D (hyperarousal)	9.20 ± 0.94	17.55± 0.72	20.07 ± 0.66
CAPS (past month)	5.53 ± 1.70	57.86 ± 3.80	73.67 ± 3.24
CAPS (lifetime)	14.87 ± 4.08	92.55 ± 3.62	97.07 ± 3.09
BDI	5.60 ± 1.17	18.97 ± 1.56	30.87 ± 1.91
SVO (in degrees)	0.80 ± 7.46	4.10 ± 3.95	-6.28 ± 4.58

Abbreviations: BDI, Beck Depression Index; CAPS, Clinician Administered PTSD Scale; CES, Combat Exposure Scale; MDD, major depressive disorder; PCL-M, PTSD Checklist-Military version; TBI, traumatic brain injury measured with the Brief Traumatic Brain Injury Screen (BTBIS); SVO, social value orientation; WTAR, Wechsler Test of Adult Reading. Data are represented as mean ± standard error or as proportions of participants in a group.

[a] This measure is available for 25 NP and 29 DP participants.

Table 2.2. Model-fit indices of candidate models

Model	iBIC	LOOIC	WAIC
Basic RL model (Only self-outcomes included)	12070.1	12167.4	12166.3
Gamma model	11208.3	11354.8	11352.1
Angle distance model	10713.5	10965.8	10970.6
Double angle distance model	10693.2	10844.4	10839.9
Fehr-Schmidt model	11091.5	11277.1	11272.6
Van Lange model	11007.4	11167.8	11161.8

The best-fitting model is underlined.

Abbreviations: iBIC, integrated Bayesian information criterion; LOOIC, leave-one-out cross-validation information criterion; WAIC, widely applicable information criterion.

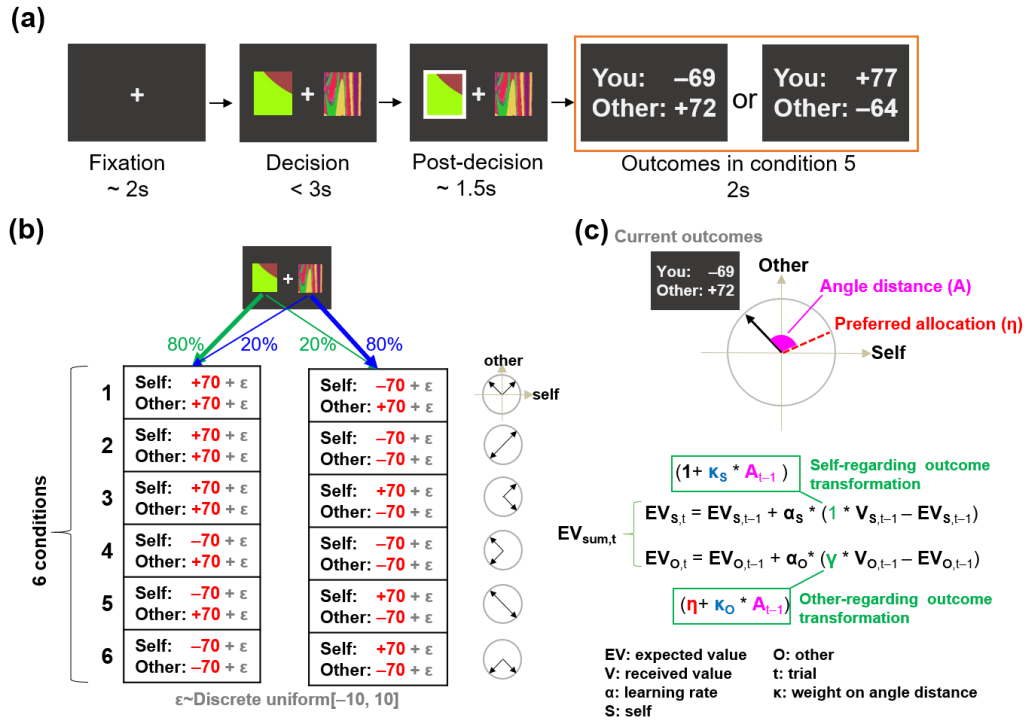

Figure 2.1. Probabilistic social learning task and the double angle distance model. (a) One example trial from condition 5 of all six conditions in the probabilistic social learning task. Participants chose between two abstract patterns, viewed the outcomes of the choice, and learned over time which was the better option. Their choices affected the payoffs for themselves and an anonymous partner. **(b)** Participants made 30 choices in each condition and learned the contingencies between the patterns and outcomes. The two possible outcomes in each condition can be represented by two vectors/arrows with certain angles in a Cartesian coordinate system with x-axis for self and y-axis for other. In conditions 1 and 6, other is always winning or losing, so it can be assumed that there is mainly learning for self. On the contrary, in conditions 3 and 4, self is always winning or losing, so presumably there is mainly learning for other. In conditions 2 and 5, we can assume that there is learning for both self and other. **(c)** The double angle distance model is adapted from the gamma model (Christopoulos & King-Casas, 2015), in which the expected values for self (EV_S) and other (EV_O) are independently updated as in the basic RL model. In the updating of EV_O, the reward (V_O) delivered to others is transformed by γ. Our new model takes account of the angle between a learner's preferred allocation and the outcomes in each trial, both of which can be represented by vectors. Hence, the other-regarding transformation parameter γ can be substituted by ($\eta + \kappa * A_{t-1}$), in which η is the preferred allocation and A_{t-1} is the angle distance between the two vectors representing η and outcomes at trial t–1. Parameter κ is the weight representing the extent to which the angle distance can influence the other-regarding preference transformation. The self-regarding transformation parameter is 1 in the gamma model, and can be modified similarly as γ.

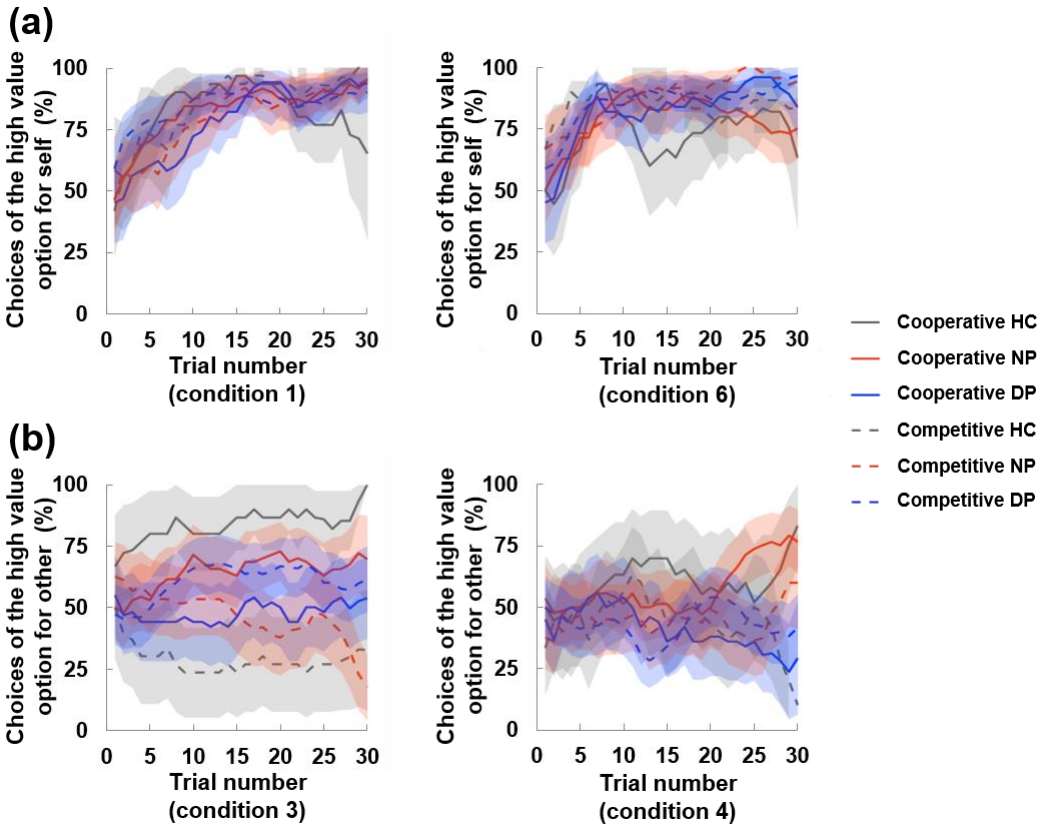

Figure 2.2. Learning curves for self and other in participants with cooperative and competitive social preferences. (a) In task conditions 1 and 6, cooperative and competitive participants in all three groups showed learning of the high value option for self. **(b)** In task condition 3, cooperative HC showed learning of the high value option for other and competitive HC showed learning of the low value option for other. The other two groups showed less learning. In task condition 4, the curves did not show clear learning or differences for the three groups. Cooperative participants included 6 HC, 14 NP, and 10 DP participants with social value orientation measures (SVOs) higher than 5. Competitive participants included 6 HC, 12 NP, and 17 DP participants with SVOs lower than –5. The learning curves depict the running average (window size = 5; mean ± standard error) of the trial-by-trial proportion of participants in each group that selected the high value option. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD.

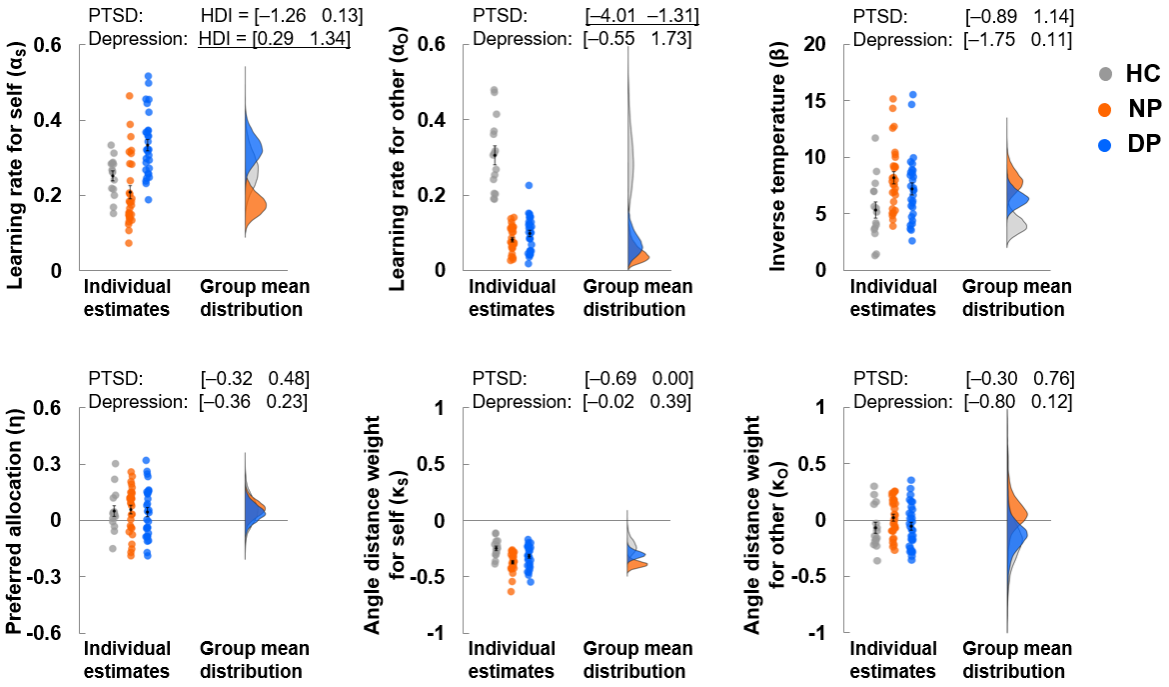

Figure 2.3. Individual-level estimates and group-level mean distributions of model parameters. As shown by Bayesian-estimated regressions for the effects of PTSD and depression on learning parameters, PTSD was associated with decreased other-regarding learning rates [$\mu_{\_slope1} = –2.56$, 95% highest density interval (HDI): [–4.01, –1.31], $P < 0.001$], and depression was associated with increased self-regarding learning rates [$\mu_{\_slope2} = 0.81$, HDI: [0.29, 1.34], $P = 0.001$] after controlling for sex, age, education, and combat exposure (**Table S2.2**). The effects of PTSD and depression on other parameters were not significant. In this figure, the group-level mean distribution for each group and parameter was reconstructed from the estimated intercepts and slopes of regression equations. Pairwise comparisons of learning parameters are reported in **Table S2.3**. Error bars indicate standard errors. Significant effects are underlined. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD.

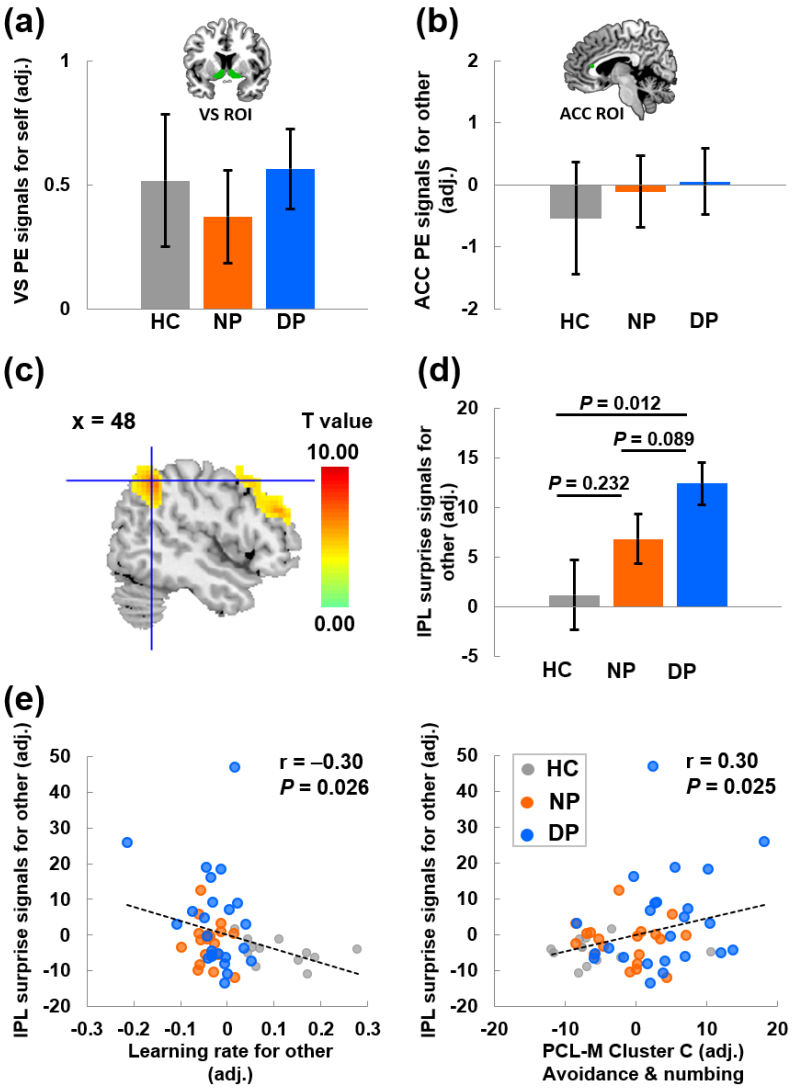

Figure 2.4. Imaging results of self-regarding prediction error (PE) signals, other-regarding PE signals and other-regarding surprise (unsigned PE) signals. (a) Self-regarding PE signals in the ventral striatum (VS) region-of-interest (ROI) were not significantly different between HC, NP, and DP. **(b)** Other-regarding PE signals in the anterior cingulate cortex (ACC) ROI were not significantly different between HC, NP, and DP. **(c)** Activation map for the one-sample t-test of other-regarding surprise signals (P <0.05, FDR-corrected at voxel-level, cluster size ≥ 50 voxels). The peak activation was at the right inferior parietal lobule (IPL; MNI: [48, –51,48]). **(d)** The other-regarding surprise signals in the right IPL ROI showed significant group differences [F(2, 48) = 3.98, P = 0.025]. **(e)** Other-regarding surprise signals in the right IPL ROI were negatively correlated with other-regarding learning rates and positively correlated with avoidance & numbing. Sex, age, education, and combat exposure were controlled for all ROI analyses. Error bars indicate standard errors. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; MNI, Montreal Neurological Institute; NP, nondepressed PTSD.

REFERENCES

Amado-Boccara, I., Gougoulis, N., Littre, M. P., Galinowski, A., & Loo, H. (1995). Effects of antidepressants on cognitive functions: a review. *Neuroscience and Biobehavioral Reviews, 19*(3), 479-493.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Apps, M. A., Lesage, E., & Ramnani, N. (2015). Vicarious reinforcement learning signals when instructing others. *Journal of Neuroscience, 35*(7), 2904-2913.

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron, 90*(4), 692-707.

Armony, J. L., Corbo, V., Clément, M.-H., & Brunet, A. (2005). Amygdala response in patients with acute PTSD to masked and unmasked emotional facial expressions. *American Journal of Psychiatry, 162*(10), 1961-1963.

Barrós-Loscertales, A., Meseguer, V., Sanjuán, A., Belloch, V., Parcet, M., Torrubia, R., & Avila, C. (2006). Striatum gray matter reduction in males with an overactive behavioral activation system. *European Journal of Neuroscience, 24*(7), 2071-2074.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience, 21*(8), 2793-2798.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress, 8*(1), 75-90.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34*(8), 669-673.

Bluhm, R., Frewen, P., Coupland, N., Densmore, M., Schore, A., & Lanius, R. (2012). Neural correlates of self-reflection in post-traumatic stress disorder. *Acta Psychiatrica Scandinavica, 125*(3), 238-246.

Brown, V. M., Zhu, L., Wang, J. M., Frueh, B. C., King-Casas, B., & Chiu, P. H. (2018). Associability-modulated loss learning is increased in posttraumatic stress disorder. *eLife, 7*, e30150.

Bryant, R. A., Kemp, A. H., Felmingham, K. L., Liddell, B., Olivieri, G., Peduto, A., . . . Williams, L. M. (2008). Enhanced amygdala and medial prefrontal activation during nonconscious processing of fear in posttraumatic stress disorder: an fMRI study. *Human Brain Mapping, 29*(5), 517-523.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253-262.

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport, 13*(15), 1975-1978.

Christopoulos, G. I., & King-Casas, B. (2015). With you or against you: social orientation

dependent learning signals guide actions made for others. *Neuroimage, 104*, 326-335.

Cisler, J. M., Bush, K., Steele, J. S., Lenow, J. K., Smitherman, S., & Kilts, C. D. (2015). Brain and behavioral evidence for altered social learning mechanisms among women with assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 63*, 75-83.

Elman, I., Ariely, D., Mazar, N., Aharon, I., Lasko, N. B., Macklin, M. L., . . . Pitman, R. K. (2005). Probing reward function in post-traumatic stress disorder with beautiful facial images. *Psychiatry Research, 135*(3), 179-183.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2007). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders-Patient Edition (With Psychotic Screen) (SCID-I/P (W/ PSYCHOTIC SCREEN), 1/2007 revision)*. Biometrics Research Department, New York State Psychiatric Institute

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping, 39*(7), 2887-2906.

Frewen, P. A., Dean, J. A., & Lanius, R. A. (2012). Assessment of anhedonia in psychological trauma: Development of the Hedonic Deficit and Interference Scale. *European Journal of Psychotraumatology, 3*(1), 8585.

Frueh, B. C., Turner, S. M., Beidel, D. C., & Cahill, S. P. (2001). Assessment of social functioning in combat veterans with PTSD. *Aggression and Violent Behavior, 6*(1), 79-90.

Greenberg, E., & Webster, C. E. (1983). *Advanced econometrics: a bridge to the literature*. New York, NY: John Wiley & Sons.

Hendler, T., & Admon, R. (2016). Predisposing Risk Factors for PTSD: Brain Biomarkers. In C.R. Martin, V.R. Preedy, V.B. Patel (Eds.), *Comprehensive Guide to Post-Traumatic Stress Disorder* (pp. 61-75). Cham, Switzerland: Springer International Publishing.

Hindmarch, I. (1999). Behavioural toxicity of antianxiety and antidepressant agents. *Human Psychopharmacology: Clinical and Experimental, 14*(2), 137-141.

Hopper, J. W., Pitman, R. K., Su, Z., Heyman, G. M., Lasko, N. B., Macklin, M. L., . . . Elman, I. (2008). Probing reward function in posttraumatic stress disorder: expectancy and satisfaction with monetary gains and losses. *Journal of Psychiatric Research, 42*(10), 802-807.

Huys, Q., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology, 8*(3).

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders, 3*(12), 1-16.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 52*(12), 1048-1060.

Killgore, W. D., Britton, J. C., Schwab, Z. J., Price, L. M., Weiner, M. R., Gold, A. L., . . . Rauch, S. L. (2014). Cortico-limbic responses to masked affective faces across PTSD,

panic disorder, and specific phobia. *Depression and Anxiety, 31*(2), 150-159.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford, UK: Oxford University Press.

Lanius, R., Frewen, P., Nazarov, A., & McKinnon, M. C. (2014). A social–cognitive–neuroscience approach to PTSD: Clinical and research perspectives. In U. F. Lanius, S. L. Paulsen, & F. M. Corrigan (Eds.), *Neurobiology and Treatment of Traumatic Dissociation: Towards an Embodied Self* (Vol. 18, pp. 69-80). Springer Publishing Company.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., . . . Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience, 22*(4), 633-641.

Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of Neuroscience, 35*(40), 13720-13727.

Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.

Luce, R. D. (2014). *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lund, M., Foy, D., Sipprelle, C., & Strachan, A. (1984). The Combat Exposure Scale: A systematic assessment of trauma in the Vietnam War. *Journal of Clinical Psychology, 40*(6), 1323-1328.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications, 11*(1), 1-11.

McClintock, C. G. (1972). Social motivation—A set of propositions. *Behavioral Science, 17*(5), 438-454.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4*(1), 1-25.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*(5), 1936-1947.

Myers, C. E., Moustafa, A. A., Sheynin, J., VanMeenen, K. M., Gilbertson, M. W., Orr, S. P., . . . Servatius, R. J. (2013). Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: Empirical data and computational model. *PloS One, 8*(8), 1-13.

Nawijn, L., van Zuiden, M., Frijling, J. L., Koch, S. B., Veltman, D. J., & Olff, M. (2015). Reward functioning in PTSD: a systematic review exploring the mechanisms underlying anhedonia. *Neuroscience and Biobehavioral Reviews, 51*, 189-204.

Nietlisbach, G., & Maercker, A. (2009). Social cognition and interpersonal impairments in trauma survivors with PTSD. *Journal of Aggression, Maltreatment & Trauma, 18*(4), 382-402.

Nietlisbach, G., Maercker, A., Rösler, W., & Haker, H. (2010). Are empathic abilities impaired in posttraumatic stress disorder? *Psychological Reports, 106*(3), 832-844.

Norris, F., & Hamblen, J. (2004). Standardized Self-Report Measures of Civilian Trauma and

PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 63–102). New York, NY: The Guilford Press.

Ousdal, O., Huys, Q., Milde, A., Craven, A., Ersland, L., Endestad, T., . . . Dolan, R. (2018). The impact of traumatic stress on Pavlovian biases. *Psychological Medicine, 48*(2), 327-336.

Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., & Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *American Journal of Psychiatry, 166*(6), 702-710.

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *Journal of Psychiatric Research, 43*(1), 76-87.

Plana, I., Lavoie, M.-A., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *Journal of Anxiety Disorders, 28*(2), 169-177.

Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., & Rumiati, R. (2008). Mentalizing in economic decision-making. *Behavioural Brain Research, 190*(2), 218-223.

Rauch, S. L., Whalen, P. J., Shin, L. M., McInerney, S. C., Macklin, M. L., Lasko, N. B., . . . Pitman, R. K. (2000). Exaggerated amygdala response to masked facial stimuli in posttraumatic stress disorder: a functional MRI study. *Biological Psychiatry, 47*(9), 769-776.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Ross, M. C., Lenow, J. K., Kilts, C. D., & Cisler, J. M. (2018). Altered neural encoding of prediction errors in assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 103*, 83-90.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Schwab, K., Baker, G., Ivins, B., Sluss-Tiller, M., Lux, W., & Warden, D. (2006). The Brief Traumatic Brain Injury Screen (BTBIS): Investigating the Validity of a Self-report Instrument for Detecting Traumatic Brain Injury (TBI) in Troops Returning from Deployment in Afghanistan and Iraq. *Neurology, 66*(5).

Sharp, C., Fonagy, P., & Allen, J. G. (2012). Posttraumatic stress disorder: A social-cognitive perspective. *Clinical Psychology: Science and Practice, 19*(3), 229-240.

Stan Development Team. 2017. *RStan: the R interface to Stan*. R package version 2.16.0. http://mc-stan.org

Stevens, J. S., & Jovanovic, T. (2019). Role of social cognition in post-traumatic stress disorder: A review and meta-analysis. *Genes, Brain and Behavior, 18*(1), e12518.

Sul, S., Tobler, P. N., Hein, G., Leiberg, S., Jung, D., Fehr, E., & Kim, H. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual

differences in prosociality. *Proceedings of the National Academy of Sciences, 112*(25), 7851-7856.

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

Uddin, L. Q., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2006). rTMS to the right inferior parietal lobule disrupts self–other discrimination. *Social Cognitive and Affective Neuroscience, 1*(1), 65-71.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413-1432.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*(Dec), 3571-3594.

Wechsler, D. (2001). *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: The Psychological Corporation.

Self- and Other-Regarding Reinforcement Learning in Post-Traumatic Stress Disorder With and Without Comorbid Depression

SUPPLEMENTARY INFORMATION

Supplementary methods

Participants

Nine participants with missing behavioral data or abnormal performance were excluded. Specifically, one participant was excluded due to missing behavioral data. Four were excluded due to low accuracy in condition 1 or 6 in the probabilistic learning task (more than 3 standard deviations away from the mean accuracy for all participants). Participants were assumed to mainly learn rewards delivered to themselves in these two conditions, so low accuracy may suggest a lack of attention to the task or poor understanding of experimental instructions. Another three participants were excluded due to no choice switching in five or all of the six conditions in the probabilistic learning task. One control participant was excluded due to an abnormal self-regarding learning rate value in model estimation (more than 3 standard deviations away from the mean for controls as well as for all participants).

Behavioral analysis

Model-agnostic analysis

An exploratory regression analysis was conducted to examine how the three groups differed in PTSD and depression's modulation of the influences of self-outcomes and other-outcomes on the choice switching (choice of a different option than the current one) in the next 10 trials. Outcomes for the other person were transformed by multiplying the outcomes received by other and each participant's social preference [Ratio of choosing the high value option for other in condition 3, instead of the social value orientation (SVO) measure, was used as a proxy of social preferences in this analysis due to large discrepancies between the model-estimated parameter γ and SVOs in some participants (**Figure S2.5c**). The ratio could take values from 0 to 1. Before it was multiplied on other-outcomes, 0.5 was subtracted from it so that positive values represented being cooperative and negative values represented being competitive]. In this regression, the dependent variable was the ratio of selecting a different option than the current one in the next 10 trials; the independent variables included self-outcomes, transformed other-outcomes, nondepressed PTSD (NP; relative to controls), depressed PTSD (DP; relative to controls), the interactions between the two PTSD groups and self-outcomes, and the interactions between the two PTSD groups and transformed other-outcomes (**Table S2.1**).

Computational modeling

The six RL models were estimated to fit the participants' choice data. The free parameters of each model were estimated using hierarchical Bayesian analysis (HBA). In HBA, a hierarchical model is constructed to include individual-level

parameters for each participant and group-level parameters that describe the distribution of individual-level parameters. Different levels can inform each other, allowing more precise estimation of individual differences in model parameters (Kruschke & Vanpaemel, 2015). The Stan software package (version 2.16.0; Stan Development Team, 2017a) was used to perform HBA. Stan applies Hamiltonian Monte Carlo (HMC) in its model estimation, which is a Markov chain Monte Carlo sampling algorithm and can efficiently obtain samples for multi-dimensional models with highly correlated parameters (Ahn et al., 2014).

In our models, parameters for individual participants were assumed to be drawn from group-level distributions, which were shared for all participants. For example, the unconstrained form of individual-level learning rate (α') was set to be from a group-level normal distribution, with normal and half-Cauchy distributions as the priors for its mean ($\mu_{\alpha'}$) and standard deviation ($\sigma_{\alpha'}$), respectively. This unconstrained individual-level learning rate (α') was expressed as the sum of the group mean and the product between the group variance and an individually estimated error parameter (*error*), drawn from a unit normal distribution. This expression is a method of reparameterization to enable more efficient sampling and is referred to as "Matt trick" in *Stan User's Guide and Reference Manual* (Stan Development Team, 2017b). Since the individual-level learning rate (α) should be bounded between 0 and 1, an inverse logit function was applied to convert the unconstrained individual-level learning rate (α') to be within this range.

Therefore, the learning rate was defined as follows:

$$\alpha = 1 / (1+ \exp(-\alpha'))$$

$$\alpha' = \mu_{\alpha'} + \sigma_{\alpha'} * error.$$

To examine how PTSD and depression affected the learning rate, the group-level mean ($\mu_{\alpha'}$) was represented by a regression equation:

$$\mu_{\alpha'} = \mu_{\alpha'\_intercept} + PTSD* \mu_{\alpha'\_slope1} + depression* \mu_{\alpha'\_slope2} + covariate_i* \mu_{\alpha'\_slope\_cov\_i}$$

where *PTSD* and *depression* were dummy coded according to each subject's diagnostic status. To control for sex, age, years of education, and combat exposure, these four variables were included in the regression as $covariate_1$, $covariate_2$, $covariate_3$ and $covariate_4$ (age, years of education and combat exposure were z-scored). The regression slopes represented the effects of independent variables on the group-level mean of learning rates. To determine significance, the 95% credible intervals or highest density intervals (HDI) of these parameters were required to not include 0. We also calculated the proportion of samples in each slope's distribution that had the opposite sign to its mean. It can serve as a significance measure similar to frequentist *P* values (Mack, Preston, & Love, 2020).

The distributions of prior parameters were set as follows:

$$\mu_{\alpha'\_intercept} \sim \text{Normal}\ (0, 0.5)$$

$$\mu_{\alpha'\_slope1} \sim \text{Normal}(0, 2)$$

$$\mu_{\alpha'\_slope2} \sim \text{Normal}(0, 2)$$

$$\mu_{\alpha'\_slope\_cov\_i} \sim \text{Normal}(0, 2)$$

$$\sigma_{\alpha'} \sim \text{Cauchy}(0, 2)$$

$$error \sim \text{Normal}(0, 1).$$

Other parameters were defined similarly. For parameters with only a lower bound of 0 (inverse temperature), the exponential function was used in the transformation: x = exp(x'), in which x represented the constrained parameter and x′ represented the unconstrained parameter. For parameters bounded between –1 and 1, a modified inverse logit function was used: x = 2*[1/ (1+ exp(–x′)) – 0.5].

For each model's estimation, HMC sampling was conducted with four chains. Each chain had 4000 samples, 2500 of which were set as warm up samples and were later discarded, resulting in 10000 samples for each parameter. For the best-fitting model, the samples for all parameters showed good convergence, with the potential scale reduction statistic ($\hat{R}$; Gelman & Rubin, 1992) less than 1. 01. In addition, effective sample sizes (ESS) of model parameters were typically larger than 4000 and all larger than 530, indicating sufficiently low autocorrelation and good mixing of HMC chains.

Imaging analysis

Functional images were obtained using a T2*-weighted echo-planar imaging (EPI) sequence [repetition time (TR) = 2000 ms, echo time (TE) =30 ms, flip angle = 90˚, field of view (FOV) = 220 × 220 mm^2]. Each volume contained 34 interleaved axial slices (matrix = 64 × 64, in-plane spatial resolution = 3.44 × 3.44 mm^2, thickness = 4 mm), which were angled 30˚ with respect to the anterior-posterior commissural line. High-resolution anatomical images were obtained using a T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (TR = 1200 ms, TE = 2.66 ms, flip angle = 8˚, FOV = 245 × 245 mm). The anatomical volume contained 192 axial slices (matrix = 245 × 245, in-plane spatial resolution = 1 × 1 mm^2, thickness = 1 mm).

SPM8 software package (Wellcome Trust Centre for Neuroimaging, London, UK) was used for imaging data preprocessing. Slice timing artifacts of functional images were corrected and then the imaging data were realigned to correct for head movement between scans, and each participants' anatomical scan was coregistered to the mean functional image produced in the realignment stage. The anatomical scans were then segmented and spatial normalization parameter matrices were generated based on the Montreal Neurological Institute (MNI) template. Functional images were transformed into the MNI space using these matrices. Normalized functional data were then spatially smoothed using an isotropic Gaussian filter with a full-width at half-maximum (FWHM) of 8 mm.

In an exploratory analysis, we also examined whether the gray matter volume of the ventral striatum was different between the three groups and had correlations with learning parameters. The VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm8/) for SPM8 was used for preprocessing anatomical data. First, the anatomical images were segmented into gray matter, white matter, and cerebrospinal fluid images based on the tissue probability maps provided in SPM8. Then DARTEL (diffeomorphic anatomical registration through exponentiated Lie algebra) algorithm (Ashburner, 2007) was used to spatially normalize the segmented images into the MNI space. Jacobian modulation was applied to preserve the total amount of gray matter in each voxel. Finally, the resulting gray matter images were smoothed with an 8 mm FWHM isotropic Gaussian filter. A region-of-interest (ROI) analysis was conducted, in which the anatomical mask of the bilateral ventral striatum from the Oxford-GSK-Imanova structural and connectivity striatal atlases (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) was used to extract the gray matter volume from each participant's gray matter image. The gray matter volume of the ROI was compared between the three groups with an analyses of variance (ANOVA); its correlation with learning parameters were also tested. No covariates were included in these exploratory analyses.

Supplementary results

Behavioral results

ANOVA results of model-agnostic performance

In ANOVAs with group and task condition as two factors and after controlling for sex, age and education, the choices of the high value option for self were not significantly different between groups or between task conditions 1, 2, 5, and 6 [group factor: $F(2, 68) = 0.86$, $P = 0.43$; task factor: $F(3, 213) = 1.07$, $P = 0.36$; interaction: $F(6, 213) = 0.68$, $P = 0.67$]. The choices of the high value option for other were not significantly different between groups either, but were marginally higher in condition 3 than in condition 4 [group factor: $F(2, 68) = 0.15$, $P = 0.86$; task factor: $F(1, 71) = 3.67$, $P = 0.060$; interaction: $F(2, 71) = 0.42$, $P = 0.66$]. Response time was comparable between groups, but the task factor was significant [group factor: $F(2, 68) = 0.42$, $P = 0.66$; task factor: $F(5, 355) = 2.33$, $P = 0.043$; interaction: $F(10, 355) = 1.43$, $P = 0.16$]. Post-hoc tests revealed that response time in task condition 4 was marginally longer than that of condition 3 ($t = 2.73$, $P = 0.072$; **Figure S2.2**).

The effects of outcomes on choice switching in the next 10 trials

The regression showed strong effects of both self-outcomes and transformed other-outcomes on choice switching (P's < 0.005). The significant interaction terms *Self-outcomes* Nondepressed PTSD* and *Self-outcomes* Depressed PTSD* indicated higher influences of self-outcomes on choice switching in PTSD groups than in controls (P's < 0.05). The significant interaction terms *Other-outcomes* Nondepressed PTSD* and *Other-outcomes* Depressed PTSD* indicated lower

influences of other-outcomes on choice switching in PTSD groups than in controls (*P*'s < 0.1; **Table S2.1**).

Imaging results

VBM results

The gray matter volume of the ventral striatum ROI was significantly different between the three groups [F(2, 70) = 6.31, *P* = 0.003] and had a negative correlation with the self-regarding learning rate (r = –0.37, *P* = 0.001; **Figure S2.5a, b**).

Supplementary tables

Table S2.1. Summary of the regression analysis testing how PTSD and depression modulated the influences of self-outcomes and other-outcomes on choice switching in the next 10 trials

Independent variable	All six conditions	
	b (SE)	t value
(Intercept)	0.31 (0.01)	50.54***
Self-outcomes	–0.16 (0.01)	–18.73***
Other-outcomes (Transformed)	–0.15 (0.02)	–6.43***
Nondepressed PTSD (NP=1, HC=0)	–0.004 (0.01)	–0.59
Depressed PTSD (DP=1, HC=0)	–0.03 (0.01)	–4.14***
Self-outcomes Nondepressed PTSD*	–0.03 (0.01)	–2.44*
Self-outcomes Depressed PTSD*	–0.03 (0.01)	–2.65**
*Other-outcomes * Nondepressed PTSD*	0.05 (0.03)	1.89†
*Other-outcomes * Depressed PTSD*	0.06 (0.03)	2.16*

Abbreviations: DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD; SE, standard error. *$P < 0.05$, **$P < 0.01$, ***$P < 0.005$, $^{\dagger}P < 0.1$.

Table S2.2. Bayesian-estimated regressions of PTSD and depression effects on learning parameters in the winning model (N = 74)

Parameter	Samples from the double angle distance model			
	Mean (SD)	HDI		*P*
Learning rate for self				
Intercept	–0.99 (0.26)	–1.49	–0.49	0.000
PTSD	–0.55 (0.35)	–1.26	0.13	0.059
Depression	0.81 (0.27)	0.29	1.34	0.001
Sex	–0.05 (0.12)	–0.28	0.18	0.323
Age	–0.21 (0.14)	–0.49	0.06	0.064
Education	0.05 (0.14)	–0.23	0.33	0.360
Combat exposure[a]	0.02 (0.13)	–0.23	0.27	0.454
Learning rate for other				
Intercept	–0.95 (0.42)	–1.77	–0.12	0.013
PTSD	–2.56 (0.68)	–4.01	–1.31	0.000
Depression	0.63 (0.57)	–0.55	1.73	0.125
Sex	0.71 (0.60)	–0.16	2.12	0.067
Age	–0.36 (0.27)	–0.89	0.19	0.087
Education	0.11 (0.32)	–0.59	0.70	0.339
Combat exposure	0.29 (0.30)	–0.28	0.91	0.155
Inverse temperature				
Intercept	1.40 (0.20)	1.14	1.79	0.000
PTSD	0.64 (0.25)	–0.89	1.14	0.005
Depression	–0.22 (0.17)	–1.75	0.11	0.096
Sex	–0.06 (0.07)	–0.20	0.08	0.193
Age	0.17 (0.09)	0.00	0.34	0.024
Education	0.03 (0.09)	–0.15	0.21	0.386
Combat exposure	0.04 (0.09)	–0.13	0.21	0.334
Preferred allocation				
Intercept	0.09 (0.17)	–0.21	0.44	0.283
PTSD	0.09 (0.20)	–0.32	0.48	0.324
Depression	–0.06 (0.15)	–0.36	0.23	0.329
Sex	–0.04 (0.07)	–0.19	0.10	0.287
Age	–0.04 (0.08)	–0.21	0.11	0.326
Education	–0.01 (0.08)	–0.16	0.15	0.443
Combat exposure	–0.12 (0.09)	–0.33	0.05	0.087
Angle distance weight for self				
Intercept	–0.47 (0.16)	–0.75	–0.13	0.006
PTSD	–0.33 (0.18)	–0.69	0.00	0.025
Depression	0.19 (0.10)	–0.02	0.39	0.034
Sex	0.06 (0.04)	–0.02	0.14	0.069
Age	–0.17 (0.05)	–0.27	–0.07	0.000
Education	0.01 (0.06)	–0.10	0.12	0.448
Combat exposure	–0.06 (0.06)	–0.18	0.04	0.126

Angle distance weight for other				
Intercept	–0.19 (0.29)	–0.81	0.38	0.213
PTSD	0.19 (0.26)	–0.30	0.76	0.211
Depression	–0.26 (0.23)	–0.80	0.12	0.093
Sex	0.25 (0.75)	–1.45	2.07	0.223
Age	0.32 (0.18)	0.08	0.74	0.002
Education	–0.07 (0.13)	–0.33	0.14	0.262
Combat exposure	0.18 (0.13)	–0.03	0.51	0.047

Significant effects are underlined.

Abbreviations: HDI, 95% highest density interval; *P*, the proportion of samples in each parameter's posterior distribution that have the opposite sign to its mean. It can serve as a significance measure similar to frequentist *P* values (Mack et al., 2020); SD, standard deviation.

[a] Four NP participants and one DP participant had missing data for combat exposure. In model estimation, the mean combat exposure value in the NP group was assigned to these four NP participants, and similarly, the mean combat exposure value in the DP group was assigned to the DP participant.

Table S2.3. Pairwise comparisons of learning parameters based on Bayesian-estimated regressions in the winning model

Parameter	Samples from the double angle distance model		
	HDI		*P*
Learning rate for self			
HC – NP	–0.02	0.23	0.059
HC – DP	–0.17	0.08	0.207
NP – DP	–0.24	–0.05	0.001
Learning rate for other			
HC – NP	0.11	0.43	0.000
HC – DP	0.09	0.41	0.000
NP – DP	–0.09	0.03	0.125
Inverse temperature			
HC – NP	–6.42	–0.94	0.005
HC – DP	–4.21	0.07	0.029
NP – DP	–0.73	4.00	0.096
Preferred allocation			
HC – NP	–0.48	0.32	0.324
HC – DP	–0.40	0.38	0.438
NP – DP	–0. 23	0.36	0.329
Angle distance weight for self			
HC – NP	–0.76	0.30	0.211
HC – DP	–0.41	0.61	0.400
NP – DP	–0.12	0.81	0.093
Angle distance weight for other			
HC – NP	0.001	0.69	0.025
HC – DP	–0.17	0.49	0.196
NP – DP	–0.39	0.02	0.034

Significant effects are underlined.

Abbreviations: HDI, 95% highest density interval; *P*, the proportion of samples in each parameter's posterior distribution that have the opposite sign to its mean. It can serve as a significance measure similar to frequentist P values (Mack et al., 2020).

Table S2.4. Mean value of trial-by-trial variables in the winning model across all participants (N of trials = 12900)

Variable in the model	Task condition					
	1	**2**	**3**	**4**	**5**	**6**
$\lvert(EV_{S,\,a} + EV_{O,\,a}) - (EV_{S,\,b} + EV_{O,\,b})\rvert$	0.35	0.31	0.31	0.13	0.30	0.31
$\kappa_S * A$	–0.43	–0.46	–0.27	–0.81	–0.46	–0.46
$\kappa_O * A$	–0.04	–0.05	–0.02	–0.07	–0.03	–0.04

Abbreviations: A, angle distance; *a*, option *a*; *b*, option *b*; EV, expected value; κ, weight on angle distance; O, other; S, self.

Table S2.5. Brain regions that show significant other-regarding surprise signals (N = 60)

Region	Voxels	BA	MNI coordinates x y z	T-value
R Inferior parietal lobule R Angular gyrus R Supramarginal gyrus R Superior parietal lobule	542	40/7/39	48 –51 48	6.27
R Middle frontal gyrus R Superior frontal gyrus	377	9/6/46/8	45 39 27	5.64
R Supplementary motor area L Supplementary motor area R Medial superior frontal gyrus	58	8/32	6 21 45	4.41
L Inferior parietal lobule L Superior parietal lobule	132	40/7	–42 –45 42	4.10

Activations were thresholded at $P < 0.05$, false discovery rate (FDR) corrected for multiple comparisons at voxel-level over the whole brain, cluster ≥ 50 voxels.

Abbreviations: BA, Brodmann's area; L, left; MNI, Montreal Neurological Institute; R, right; Regions, brain regions in an activation cluster (Peak voxel was located in the 1st region); T-value, T-value of the peak voxel in a cluster; Voxels, number of voxels in a cluster.

Supplementary figures

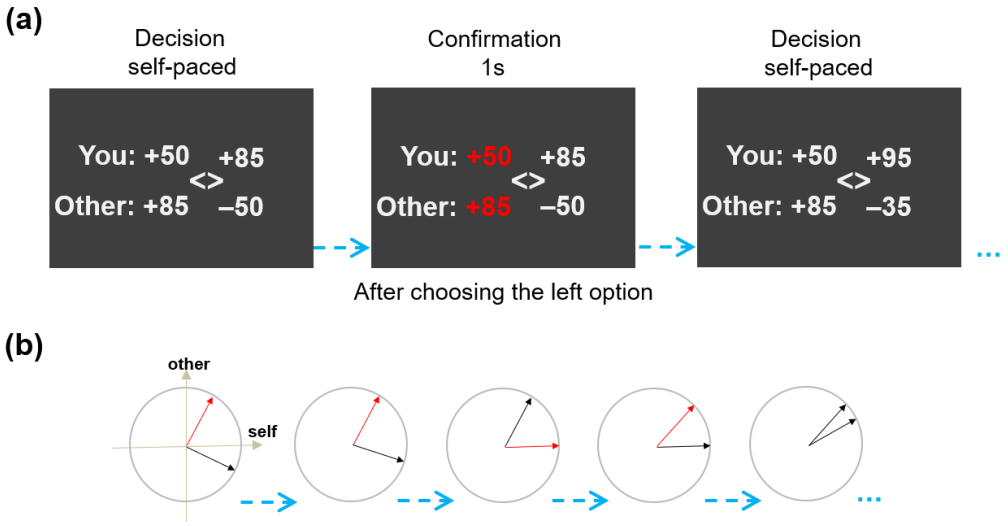

Figure. S2.1. Social value orientation (SVO) assessment. (a) A sequential testing procedure was used to assess the social preference of each participant. In a series of choices, participants indicated their preference between two allocations. Each allocation consisted of gaining or losing points for the participants and an anonymous partner. **(b)** The allocation pairs can be represented by two arrows in a Cartesian coordinate system, where the x-axis represents payoffs for self and the y-axis represents payoffs for other. Note that participants did not see the arrows, but allocations of points. After the participants indicated their preferred arrow (allocation) in each trial, the unchosen arrow would move toward the chosen one in a certain step size in the next trial. When the two arrows were close enough (both the difference between the two outcomes for self and the difference between the two outcomes for other were less than 6), the arrows were considered converged and the average of the two angles was used as the measure of SVO.

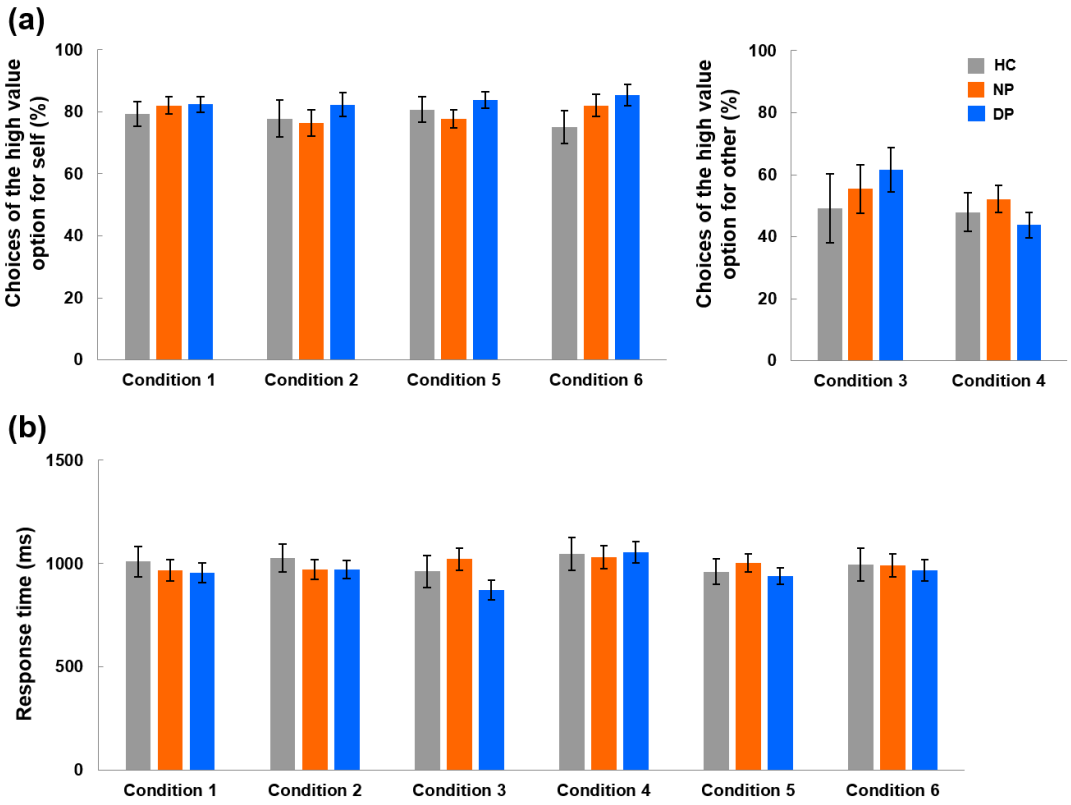

Figure S2.2. Model-agnostic performance in the six task conditions. (a) Choices of the high value option for self were not significantly different between groups or between task conditions 1, 2, 5, and 6 [group factor: $F(2, 68) = 0.86$, $P = 0.43$; task factor: $F(3, 213) = 1.07$, $P = 0.36$; interaction: $F(6, 213) = 0.68$, $P = 0.67$]. Choices of the high value option for other were not significantly different between groups, but the task factor was marginally significant [group factor: $F(2, 68) = 0.15$, $P = 0.86$; task factor: $F(1, 71) = 3.67$, $P = 0.060$; interaction: $F(2, 71) = 0.42$, $P = 0.66$]. **(b)** Response time was not significantly different between groups [group factor: $F(2, 68) = 0.42$, $P = 0.66$; task factor: $F(5, 355) = 2.33$, $P = 0.043$; interaction: $F(10, 355) = 1.43$, $P = 0.16$]. The task factor was significant, and post-hoc tests revealed that response time in task condition 4 was marginally longer than that of condition 3 ($t = 2.73$, $P = 0.072$). Sex, age and education were controlled for in the above ANOVAs. Error bars indicate standard errors. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD.

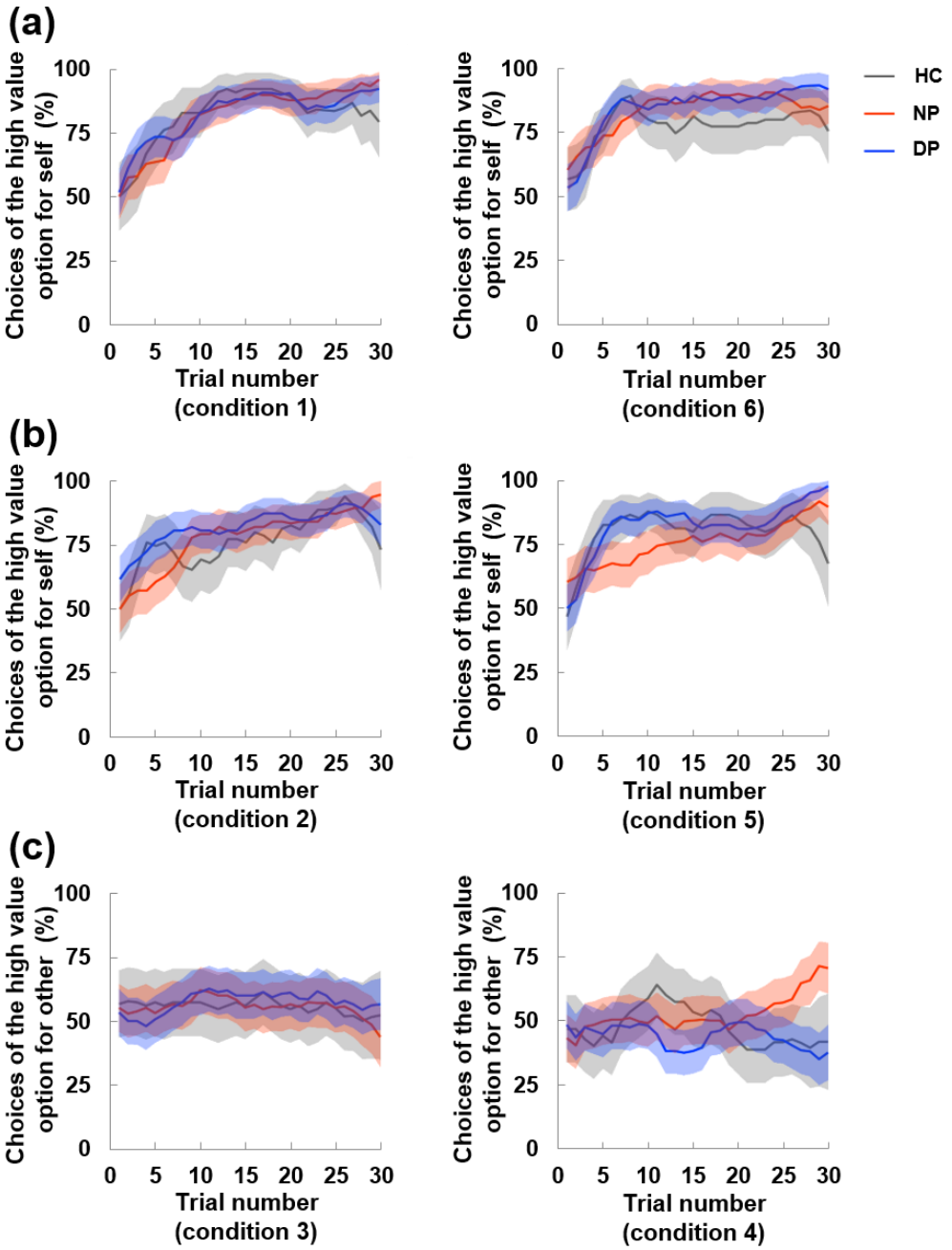

Figure S2.3. Learning curves for the six task conditions. **(a)** In task conditions 1 and 6, participants in all three groups showed learning for self. **(b)** In task conditions 2 and 5, participants in all three groups showed learning for self. **(c)** In task conditions 3 and 4, participants in all three conditions did not show evident learning for other. The learning curves depict the running average (window size = 5; mean ± standard error) of the trial-by-trial proportion of participants in each group that selected the high value option. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD.

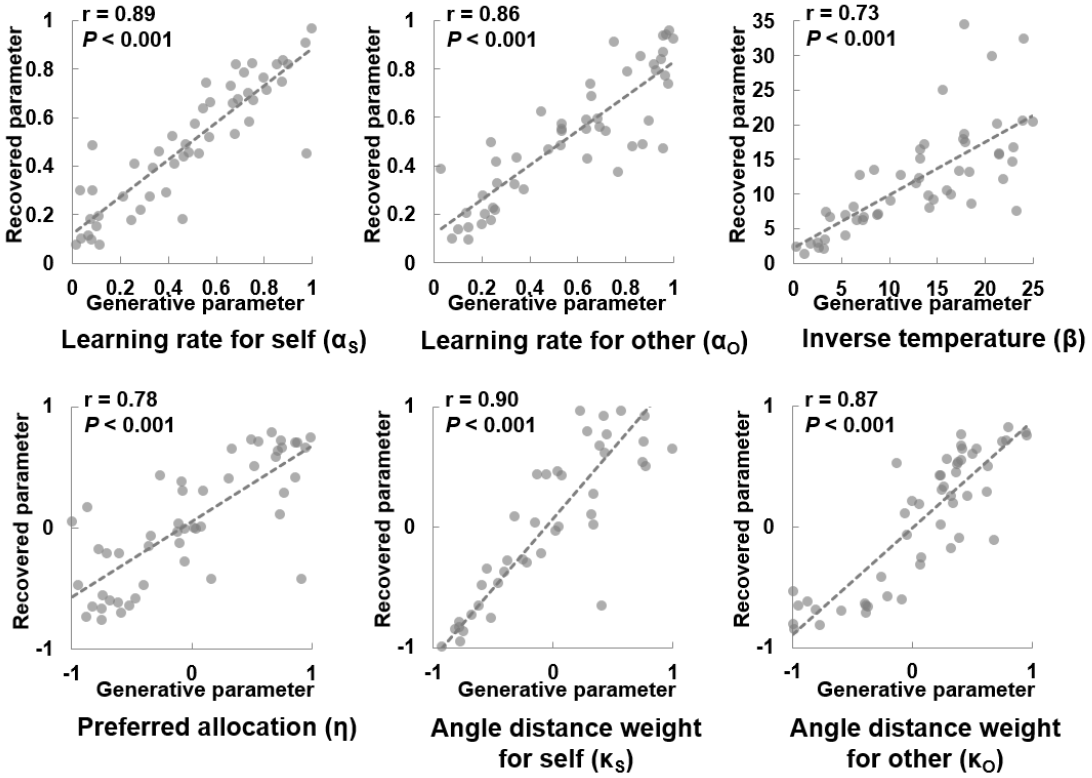

Figure S2.4. Parameter recovery for the winning model. For each parameter, we randomly drew 50 values from a uniform distribution. Specifically, values of learning rates were drawn from U (0, 1); values of inverse temperatures were drawn from U (0, 25); values of preferred allocations and angle distance weights were drawn from U (–1, 1). These generative parameters were randomly assigned to 50 hypothetical participants to generate choice data using the winning model, and then new learning parameters were estimated using these choice data. The values of these learning parameters can be fully recovered with the best-fitting model, as shown by the strong correlations.

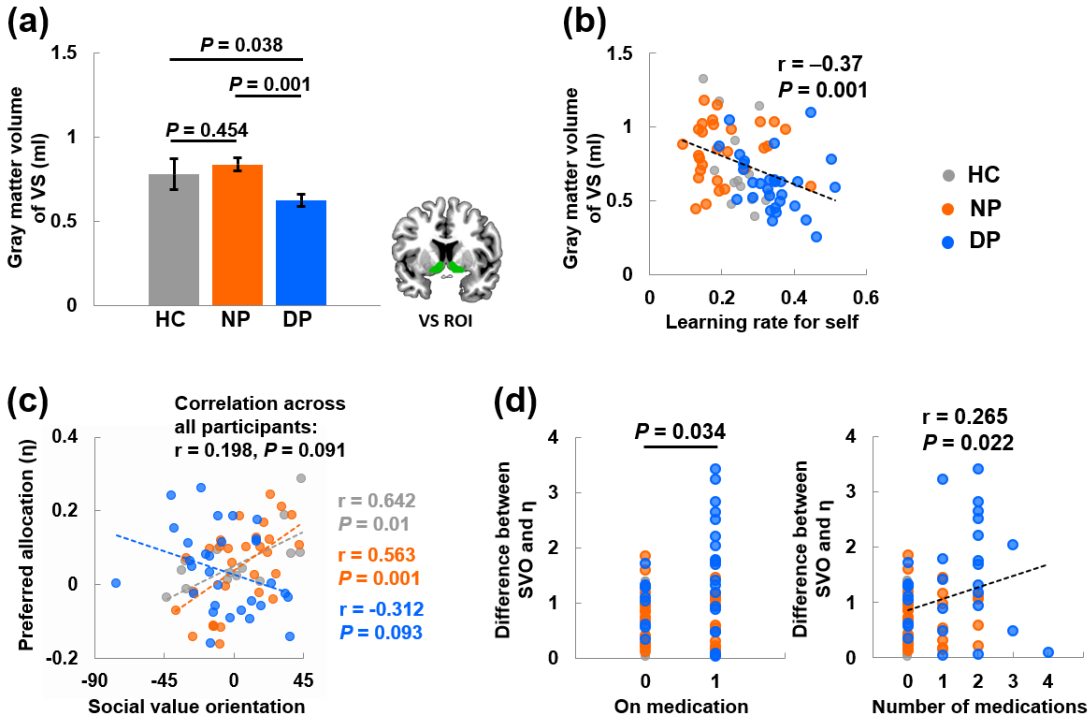

Figure S2.5. The gray matter volume of the ventral striatum (VS) and the association between the social value orientation (SVO) and the preferred allocation. (a) The gray matter volume of VS was significantly different between the three groups [N = 73 (missing anatomical data for one participant); F(2, 70) = 6.31, *P* = 0.003]. **(b)** The gray matter volume of VS and self-regarding learning rates were negatively correlated. **(c)** The SVO measured in a non-learning task and the preferred allocation parameter (η) were positively correlated with each other in HC and NP, but not in DP. **(d)** The discrepancy between SVO and η (both were normalized) was significantly different between participants who were on medication and those who were not [t(41.20) = 2.20, *P* = 0.034]. The discrepancy was also positively correlated with the number of medications. **Abbreviations:** DP, depressed PTSD; HC, healthy controls; NP, nondepressed PTSD.

Supplementary references

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology, 5*, 1-15.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457-472.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford, UK: Oxford University Press.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications, 11*(1), 1-11.

Stan Development Team. 2017a. *RStan: the R interface to Stan*. R package version 2.16.0. http://mc-stan.org

Stan Development Team. 2017b. *Stan Modeling Language Users Guide and Reference Manual*, Version 2.16.0. http://mc-stan.org

Paper III: The Effects of Oxytocin on Self- and Other-Regarding Reinforcement Learning

Shengchuang Feng, George Christopoulos, Pearl H. Chiu, & Brooks King-Casas

ABSTRACT

Oxytocin (OT) is a neuropeptide that has been reported to affect reward learning and perception of oneself and others. Previous studies of the effects of OT administration on reward learning have been restricted to reinforcement learning (RL) about rewards delivered to oneself (self-regarding RL). The effects of OT on RL about rewards delivered to others (other-regarding RL) are poorly examined and understood. The present study aims to investigate the behavioral and neural effects of OT administration on RL for both oneself and unknown others. In this double-blind, placebo (PL)-controlled, within-participant design, we used a probabilistic social learning task with different learning contexts, functional magnetic resonance imaging (fMRI) and neurocomputational analyses to show how intranasal administration of OT would influence learning for self- and other-regarding rewards in 29 healthy adult males. Behaviorally, compared to the PL condition, other-regarding learning rates decreased after OT, but self-regarding learning rates did not change significantly between drug conditions. Participants with cooperative social preferences in the PL condition became significantly less cooperative after OT. For participants with competitive social preferences in the PL condition, they showed a nonsignificant trend of being less competitive after OT. Imaging results revealed a similar pattern as learning rates, with decreased prediction error signals for other in the anterior cingulate cortex and a non-significant change of prediction error signals for self in the ventral striatum. OT's diminishing effects on other-regarding learning may be attributed to its interaction with dopaminergic activity in reward regions. The changes of social preferences may reflect OT's influences on social salience and anxiety in participants with different baseline social preferences. Our study provides new evidence of OT's effects on reward learning and reward perception, suggesting important implications for the use of OT in therapeutic interventions. It also sheds light on the modulating role of context and individual differences in OT's effects and helps to resolve inconsistent findings in the field.

Keywords: oxytocin, self-regarding learning, other-regarding learning, social value orientation, reinforcement learning, prediction error

INTRODUCTION

Oxytocin (OT) is a neuropeptide that has attracted a lot of public attention and is believed by many to be a "love hormone" that can enhance social cognition and prosocial behavior in humans (Bartz, Zaki, Bolger, & Ochsner, 2011; Zik & Roberts, 2015). Despite inconsistent directions of its effects, exogenously administered OT has been clearly shown to modify trust (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005), cooperation (De Dreu et al., 2010; Israel, Weisel, Ebstein, & Bornstein, 2012; Rilling et al., 2012), generosity (Zak, Stanton, & Ahmadi, 2007), and empathy (Aoki et al., 2014; Hurlemann et al., 2010). Reward perception and learning are indispensable components of many social interactions. For example, people consider potential gains and losses when deciding to trust or to cooperate (Balliet, Mulder, & Van Lange, 2011; Bohnet & Zeckhauser, 2004). Therefore, it is possible that OT affects social reward processing and its effects mediate cognitive and behavioral changes observed in previous OT studies.

This hypothesis has gained supporting evidence from both neurophysiological studies of the relationship between oxytocinergic and dopaminergic systems and behavioral studies of OT's effects on reinforcement learning (RL). According to findings from animal studies, endogenous OT is synthesized in the hypothalamus and is projected to brain regions implicated in reward processing, such as the ventral tegmental area and ventral striatum (Baribeau & Anagnostou, 2015; Grinevich, Knobloch-Bollmann, Eliava, Busnelli, & Chini, 2016; Hung et al., 2017). Recent human studies have also found OT projections to the orbitofrontal cortex and anterior cingulate cortex (ACC; Rogers et al., 2018) as well as OT receptors in the cingulate cortex (Boccia, Petrusz, Suzuki, Marson, & Pedersen, 2013), which are brain regions involved in reward processing and social decision-making (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011; Wallis, 2007). Moreover, animal studies have also shown that extraneous injection of OT into the ventral tegmental area can increase extracellular dopamine within the ventral striatum (Melis et al., 2007; Succu et al., 2008), revealing possible mechanisms of extraneous OT's influences on reward processing.

Consistent with these neurophysiological studies, OT has also been shown to modulate social RL in humans (Clark-Elford et al., 2014; Evans, Shergill, & Averbeck, 2010; Hurlemann et al., 2010; Ide et al., 2018). It seems that incorporating social components into the learning tasks is necessary for OT to have a detectable effect. Significant OT effects have been observed when participants were asked to learn monetary reward values of face pairs (Clark-Elford et al., 2014; Evans et al., 2010), number categorization with smiling/angary faces as reinforcers (Hurlemann et al., 2010), and how much to trust an anonymous partner in an iterative trust game (Ide et al., 2018). Once the social components were missing, e.g., when learning number categorization with nonsocial reinforcers (green/red lights), OT's effects became absent (Hurlemann et al., 2010).

These studies mainly examined OT's effects on the processing of rewards delivered to oneself but ignored another important aspect of social RL. There has been a large body of evidence showing that people take account of rewards delivered to others when making decisions, and the extent to which people value others' rewards varies at the individual level (Christopoulos & King-Casas, 2015; Liu et al., 2019; McClintock, 1972; Messick & McClintock, 1968; Sul et al., 2015; Van Lange, 1999). Here we ask if RL will be influenced by OT when non-social reinforcers (e.g., points, money) are delivered to other social partners (other-regarding learning) versus to oneself (self-regarding learning) and if individual differences in social preferences (other-regarding valuation) will modulate OT's effects on learning.

Neuroimaging studies of RL have consistently found the role of the ventral striatum in coding prediction errors (PEs; the difference between received reward and expected reward) in the process of learning self-regarding rewards (Berns, McClure, Pagnoni, & Montague, 2001; Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997). The ACC is suggested to code other-regarding information (Apps, Rushworth, & Chang, 2016), including other's reward probabilities (Lockwood, Apps, Roiser, & Viding, 2015) and other's PEs (Apps, Lesage, & Ramnani, 2015). We also ask whether OT influences self and other-regarding PEs in these brain regions.

To answer these questions, we applied a double-blind, placebo (PL)-controlled, within-participant design, in which healthy adult males performed a probabilistic social learning task with rewards delivered to themselves and an anonymous partner (Christopoulos & King-Casas, 2015) in PL and OT conditions. When participants were performing the task, neuroimaging data were collected with functional magnetic resonance imaging (fMRI). In the data analysis, we used computational modeling to separate different cognitive components of the learning process and to examine the ventral striatum and ACC's coding of PEs in different drug conditions.

METHODS

Participants

Thirty-five healthy male participants [age range: 18-45, mean age ± standard deviation (SD): 26.83 ± 7.18 years] were recruited. Participants were excluded if they had any current or past DSM-IV Axis I disorders (verified by the full Structured Clinical Interview for DSM-IV Axis I disorders; First, Spitzer, Gibbon, & Williams, 2007), any other physical illnesses, or on any medication that affects brain function. Other exclusion criteria were: younger than 18 or older than 55; history of seizure disorder, stroke, or head injury resulting in more than 10 minutes of unconsciousness or with neurological sequelae; hormone disorder; history of electroconvulsive therapy or chemotherapy for cancer; and fMRI contraindications. All participants were English speaking and had normal or corrected-to-normal vision. Written informed consent was acquired from them before their participation. The study was approved by the Institutional Review Boards at Virginia Tech and Baylor College of Medicine.

One participant was excluded due to being on medication. Four were excluded due to incomplete data and one more was excluded due to low accuracy in learning for oneself (more than 3 SD away from the mean accuracy), resulting in 29 participants for behavioral data (age range: 18-45, mean age ± SD: 26.72 ± 7.27 years). Three more participants were excluded for having excessive head motion within the scanners (cumulative translation > 5 mm and rotation > 5˚) and one more was excluded as an outlier (more than 3 SD away from the mean of other-PE signals in the ACC) in the imaging analysis, resulting in 25 participants in the imaging data (age range: 18-45, mean age ± SD: 26.28 ± 7.39 years).

Study design

A double-blind, PL-controlled, within-participant design was applied in this study. Participants self-administered nasal sprays with 24 International Units of OT (Syntocinon-spray; Novartis, Switzerland) or with PL that contained all ingredients as in the OT sprays except for OT. The two treatment conditions (OT or PL) were separated by at least one week and the order was randomized across participants. All participants were asked to refrain from alcohol 24 hours before testing and not intake caffeine on the testing day.

Social preference assessment

Before the first treatment condition, each participant's social value orientation (SVO; the preference for allocating rewards between oneself and others) was assessed through a non-learning task with a sequential testing procedure (Christopoulos & King-Casas, 2015; Luce, 2014), in which participants serially made preference choices between two allocations of points for themselves and an anonymous partner (e.g., [Self: +50, Other: +85] vs. [Self: +85, Other: –50]; **Figure S3.1a**). Since each allocation can be represented by an arrow with a certain angle (**Figure S3.1b**), we can see choices as being made between two arrows. After each choice, the chosen arrow (allocation) was retained as one option in the next trial, and the unchosen arrow moved toward the chosen one with a certain step size. Based on this procedure, the two arrows gradually converged to each participant's preferred allocation. This assessment was repeated three times, with different initial allocations. The mean of the three measurements was used to represent a participant's SVO (in degrees) and can serve as an index of his social preference.

Experimental task

Fifty minutes after the nasal sprays, participants went into the fMRI scanner and performed a probabilistic social learning task with points for themselves and an anonymous partner, the same as in Christopoulos & King-Casas (2015). Prior to the scanning, participants were instructed to collect as many points as they could. They were also informed that: (i) they would never meet the other person or know each other's identity; (ii) the participant and the other person would be paid based on the outcomes of a random subset of the task trials; and (iii) the other person would not

perform a task that could influence the participant's payoffs.

Given that the allocations between oneself and the other person can be represented as arrows with a certain angle, there are four kinds of arrows/allocations positioned in the four quadrants of a Cartesian coordinate system (x-axis for self and y-axis for other): Quadrant I [self-gain & other-gain], Quadrant II [self-loss & other-gain], Quadrant III [self-loss & other-loss], and Quadrant IV [self-gain & other-loss]. To examine the learning of different allocations, the task was designed to consist of six conditions or contexts, each including two of the four allocations (**Figure 3.1**). In each condition, participants could learn the reward contingencies of two abstract patterns by making choices between them. One pattern was associated with an 80% probability of a certain allocation and the other one was associated with an 80% probability of another allocation. The exact values for different allocations in each trial were randomly sampled from a uniform discrete distribution with a mean of +70 or –70 and a range of 20. There were 30 trials for each condition, which also formed one block. The order in which the six conditions/blocks were presented was pseudorandomized across participants. Between two blocks, there was an instruction screen to remind the participant that there would be "new sets of symbols".

The procedure of one trial was as follows: at the start, a fixation cross was shown at the center of the screen for 1s plus a value randomly selected from an exponential distribution with a mean of 1, truncated at 6. Then two abstract patterns were shown on the screen, allowing the participant to make a choice via a scanner-compatible button box within 3 s. The positions of the two patterns were randomized. After the participant's response, that chosen pattern was framed for 0.5 s plus a value also randomly selected from an exponential distribution with a mean of 1 and truncated at 6. Following this jittered confirmation, the outcomes for self and other were displayed for 2 s. The self- and other-outcomes were also randomly positioned to be above or under the center of the screen.

Behavioral analysis

Model-agnostic analysis

For the probabilistic social learning task, choices of the high value option for self and other and response time in all six conditions were calculated for each participant. The effects of OT on these performance measures were tested using Analyses-of-variance (ANOVA) with drug and task conditions as two factors. Trial-by-trial learning curves were also plotted for each condition. The learning curves depict the running average of the trial-by-trial proportion of participants that selected the high value option in each condition.

To test whether dispositional social preferences would interact with OT's effects on learning, nine participants with SVOs higher than 7 and 10 participants with SVOs lower than –9 were included in the cooperative and competitive groups, respectively. ANOVAs with drug and group as two factors and choices of the high value option for

self as the dependent variables were conducted for conditions 1 and 6, in which presumably there was mainly learning for self. ANOVAs with drug and group as two factors and choices of the SVO-congruent option for other as the dependent variables were conducted for conditions 3 and 4, in which presumably there was mainly learning for other. For participants with cooperative SVOs, "choices of the SVO-congruent option" were the selections of the high value option for other, whereas for participants with competitive SVOs, "choices of the SVO-congruent option" were the selections of the low value option for other. Learning curves were also plotted separately for the two groups in conditions 1, 6, 3, and 4.

As a verification of the ANOVA results, exploratory logit regression analyses were conducted in each condition to examine OT's modulation of the effects of self-outcomes and other-outcomes on switching choices in the next trial. Outcomes for the other person were transformed by multiplying the other-outcomes and each participant's SVO measure (in tangents). In each logit regression, the dependent variable was switching to a different option in the next trial than the current trial; the independent variables included self-outcomes, transformed other-outcomes, drug conditions, the interaction between drug and self-outcomes, and the interaction between drug and transformed other-outcomes (**Table S3.1**).

Computational modeling

To disentangle various cognitive components in the learning process, we fitted participants' choice data to different RL models. The basic form of an RL model (Sutton & Barto, 1998) describes an updating rule (Rescorla & Wagner, 1972) of the expected value for a certain option:

$$EV_t = EV_{t-1} + \alpha^* (V_{t-1} - EV_{t-1}). \qquad \text{(Equation 3.1)}$$

In this equation, EV_t and EV_{t-1} denote the expected value of an option at time/trial t and t–1, respectively; V_{t-1} denotes the actual value received by a learner after selecting that option at time t–1; $(V_{t-1} - EV_{t-1})$ is termed PE, denoting the difference between the received value and the expected value at time t–1. α is learning rate and can take values between 0 and 1. The higher it is, the faster the learner updates the expected value of an option and relies more on most recent information versus past reward history.

When multiple options are present, the learner can update the expected value for each option after it is chosen and delivers rewards. The learner's probability of choosing a certain option at each time point can be modeled with a standard softmax function (Luce, 1959):

$$P_{a,t} = \frac{\exp(\beta^* EV_{a,t})}{\exp(\beta^* EV_{a,t}) + \exp(\beta^* EV_{b,t})}$$

(Equation 3.2)

where $P_{a,t}$ denotes the probability of choosing option *a* out of two options (*a* and *b*) at time t; β is inverse temperature and can be used as a measure of the noisiness or randomness of the learner's choices. It can take values equal to or larger than 0. The higher the inverse temperature is, the less random the choices are and the more likely the learner would choose the option with the highest expected value.

To account for learning for oneself and others, the gamma model (Equation 3.3) was proposed by Christopoulos & King-Casas (2015). In this model, the expected value for self (EV_S) and the expected value for other (EV_O) are updated through the learning rate (α_S) and PEs for self and the learning rate (α_O) and PEs for other, respectively. In a learning task with each option associated with both self- and other-outcomes, the expected value of one option is calculated by summing up these two expected values. To represent the social preference of the learner, the outcomes for the other person are weighted by a gamma parameter, which is a discrete variable that can take the values –1, 0, and 1 in Christopoulos & King-Casas (2015)'s original paper. A competitive person would have the gamma as –1, meaning that he/she sees others' gains as his/her losses. A cooperative participant's gamma would be 1 in that he/she sees others' gains as his/her gains. A gamma of 0 suggests that the person is indifferent to others' outcomes. Given the fact that the SVO measure is a continuous variable and the SVOs of most participants in the present study were between –45° and +45° (The tangents of these two degrees are –1 and 1), γ was set as a continuous variable and bounded between –1 and 1.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * (V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * (\gamma * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \qquad \text{(Equation 3.3)}$$

Based on the gamma model, we developed the angle distance model (Equation 3.4). This model takes account of the difference between the learner's preferred allocation and the received outcomes in each trial, both of which can be represented by arrows (**Figure 3.1c**). This difference is termed angle distance (A) and the units are radians. Instead of a single γ in the gamma model, the angle distance model uses the preferred allocation (η), the angle distance, and a weight on angle distance (κ) to dynamically transform outcomes received by the other person to subjective values for the learner. Here, η is the parameter representing the learner's social preference and was set as a continuous variable bounded between –1 and 1.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * (V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * ((\eta + \kappa * A_{t-1}) * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \qquad \text{(Equation 3.4)}$$

It is possible that the learner's perception of outcomes for self is similarly influenced by the angle distance. Therefore, we also tested the following double angle distance model (Equation 3.5) with the outcome transformation mechanisms applied to both self- and other-outcomes.

$$EV_{S,t} = EV_{S,t-1} + \alpha_S * ((1 + \kappa_S * A_{t-1}) * V_{S,t-1} - EV_{S,t-1})$$
$$EV_{O,t} = EV_{O,t-1} + \alpha_O * ((\eta + \kappa_O * A_{t-1}) * V_{O,t-1} - EV_{O,t-1})$$
$$EV_t = EV_{S,t} + EV_{O,t} \quad \text{(Equation 3.5)}$$

There are also models that take account of inequality of rewards between self and other. The Fehr-Schmidt model (Fehr & Schmidt, 1999) is a well-known model of this kind, and it can be written as follows for an RL task:

$$EV_t = EV_{t-1} + \alpha * (V_{S,t-1} - w_a * (V_{S,t-1} - V_{O,t-1}) - EV_{t-1}) \quad \text{if} \quad V_{S,t-1} - V_{O,t-1} > 0$$
$$EV_t = EV_{t-1} + \alpha * (V_{S,t-1} - w_b * (V_{O,t-1} - V_{S,t-1}) - EV_{t-1}) \quad \text{if} \quad V_{S,t-1} - V_{O,t-1} < 0$$
$$w_b \geq w_a \geq 0;\ w_a < 1 \quad \text{(Equation 3.6)}$$

where w_a measures participants' compassion or guilt when self is better off than other and w_b measures envy when self is worse off than other.

Another inequality model (Equation 3.7) proposed by Van Lange (1999) takes account of cooperation and egalitarianism in perceiving outcomes for self and other. It has three weight parameters (w_S, w_O, and $w_{|S-O|}$) respectively multiplied on self-outcomes, other-outcomes, and the differences between self- and other-outcomes.

$$EV_t = EV_{t-1} + \alpha * (w_S * V_{S,t-1} + w_O * V_{O,t-1} + w_{|S-O|} * |V_{S,t-1} - V_{O,t-1}| - EV_{t-1}) \quad \text{(Equation 3.7)}$$

The aforementioned basic RL model with no learning for other and the five other-regarding RL models were fitted using the participants' choice data. Hierarchical Bayesian analysis (performed with the Stan software package, version 2.16.0; Stan Development Team, 2017) was used to estimate the six models [see **SUPPLEMENTARY INFORMATION** (**SI**) for more details]. As part of the model estimation, OT effects on the group-level mean (μ) for each learning parameter were tested using a regression equation:

$$\mu = \mu_{\_intercept} + OT * \mu_{\_slope} \quad \text{(Equation 3.8)}$$

in which *OT* was a dummy variable coded according to each subject's drug conditions, and its effect on the group-level mean was represented by the regression slope. After model estimation, model-fit indices, including the integrated Bayesian information criterion (iBIC; Huys et al., 2012), widely applicable information criterion (WAIC; Watanabe, 2010), and leave-one-out cross-validation information criterion (LOOIC; Vehtari, Gelman, & Gabry, 2017), were calculated for all six models.

For the winning model, the 95% credible intervals or highest density intervals (HDI) of the regression slopes were used to determine the significance of OT's effects (Significant effects require that HDI does not overlap zero; Kruschke & Vanpaemel, 2015). We also calculated the proportion of samples in each slope's distribution that had the opposite sign to its mean. It can serve as a significance measure similar to frequentist *P* values (Mack, Preston, & Love, 2020), although it was not used for the statistical inferences in the present study. As for the individual-level parameters, their

estimates were extracted to generate PEs for imaging analysis and to test the correlations between learning parameters and other variables.

After the best-fitting model was selected, we conducted a parameter recovery analysis to verify that the parameters in the winning model could be reliably recovered and did not have the parameter identification problem (The parameter identification problem occurs when choice data can be generated with more than one set of parameters; Greenberg & Webster, 1983). For each parameter, we randomly drew 50 values from a uniform distribution. These generative parameters were randomly assigned to 50 hypothetical participants to generate choice data using the winning model, and then new learning parameters were estimated using these choice data. High correlations between the generative and recovered parameters indicate that the estimation method can reliably capture the true parameter values.

Imaging analysis

The functional and anatomical imaging was conducted on 3 T Siemens Trio MR scanners (Siemens, Munich, Germany) at Baylor College of Medicine. SPM8 software package (Wellcome Trust Centre for Neuroimaging, London, UK) was used for imaging data preprocessing and analysis (see **SI** for details of image acquisition and preprocessing).

To examine brain regions involved in self- and other-regarding PEs, two first-level general linear models (GLMs) were constructed for each participant using an event-related analysis procedure. In each GLM, the prepossessed functional imaging data were set as a dependent variable and the task events convolved with a hemodynamic response function (HRF) were set as independent variables. For both GLMs, the events were the same, including the presentation of two options before choices, choices, confirmation of choices, and outcomes. Trial-by-trial self-PEs $[(1+\kappa_S * A_t) * V_{S,t} - EV_{S,t}]$ and other-PEs $[(\eta + \kappa_O * A_t) * V_{O,t} - EV_{O,t}]$ calculated from the best-fitting RL model were respectively entered into the two GLMs as a parametric modulator of the outcome event. Six head motion parameters were also included in the GLMs as regressors of no interest. Standard linear regression and parametric modulation analyses were performed to obtain beta maps for self- and other-PEs.

In the region-of-interest (ROI) analysis, the anatomical mask of bilateral ventral striatum from the Oxford-GSK-Imanova structural and connectivity striatal atlases (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) was used as the self-PE ROI. Given the ACC's role in tracking other's rewards (Apps et al., 2016), the other-PE ROI was defined as a sphere with a radius of 6 mm, centered at MNI (Montreal Neurological Institute template) coordinates [5 31 12], which are the middle point of the two peak voxels from Lockwood et al. (2015; MNI: [8,32,12], involved in processing other's reward probabilities) and Apps et al. (2015; MNI: [2, 30, 12], involved in processing other's PEs). Self- and other-PE signals for each participant were extracted from the ROIs in the corresponding beta maps and were compared between drug conditions with paired t-tests.

RESULTS

Behavioral results

Learning curves in the six conditions

Learning curves in conditions 1, 2, 5, and 6 of the probabilistic social learning task showed increasing trends of choosing the high value option for self, indicating the acquisition of reward contingencies for self. However, in conditions 3 and 4, where the participants were supposed to learn the rewards delivered to others, only rather flat learning curves were found. In all the six conditions, there was no evident divergence between learning curves in the two drug conditions (**Figure S3.2**).

Nonetheless, when we looked at the learning curves of the one-third of the participants with the highest SVO measures (cooperative), a clear trend of increasing choices of the high value option for other was found in condition 3 after PL. But after OT administration, their learning curve became flat. For the one-third of the participants with the lowest SVO measures (competitive), the learning curve showed a decreasing trend of choosing the high value option for other in condition 3 after OT, suggesting their preference for the unknown partner to lose points. Similar to the cooperative participants, this trend was flattened after OT. No similar divergences of learning curves for other were present in condition 4 (**Figure 3.2b**). The learning curves for self did not diverge for the two groups in either condition 1 or 6 (**Figure 3.2a**).

ANOVA results of model-agnostic performance

Consistent with the learning curves, results of ANOVAs without considering participants' SVOs showed that the choices of the high value option for self were not different between drug conditions or between task conditions 1, 2, 5, and 6 [drug factor: $F(1, 28) = 0.44$, $P = 0.51$; task factor: $F(3, 84) = 0.73$, $P = 0.54$; interaction: $F(3, 84) = 0.23$, $P = 0.88$]. The choices of the high value option for other were not different between drug conditions or between task conditions 3 and 4 [drug factor: $F(1, 28) = 0.17$, $P = 0.68$; task factor: $F(1, 28) = 1.89$, $P = 0.18$; interaction: $F(1, 28) = 0.12$, $P = 0.74$]. Response time was not significantly different between drug conditions, but there was a main effect of task conditions [drug factor: $F(1, 28) = 0.67$, $P = 0.42$; task factor: $F(5, 140) = 5.67$, $P < 0.001$; interaction: $F(1, 28) = 1.70$, $P = 0.14$]. Post-hoc tests revealed that response time in task condition 4 was significantly longer than all the other conditions ($|t| > 3.44$, $P < 0.01$; **Figure S3.3**).

When separately looking at cooperative and competitive participants (highest one-third and lowest one-third on the SVO measure), both groups made more choices of the SVO-congruent option for other under PL than OT treatment in condition 3 [drug factor: $F(1, 17) = 4.84$, $P = 0.042$; group factor: $F(1, 17) = 0.67$, $P = 0.42$; interaction: $F(1, 17) = 0.07$, $P = 0.80$]. No significant effects were found in condition 4 [drug factor: $F(1, 17) = 0.73$, $P = 0.41$; group factor: $F(1, 17) = 0.09$, $P = 0.76$; interaction: $F(1, 17) = 0.07$, $P = 0.80$; **Figure S3.4b**].

The effects of outcomes on choice switching in the next trial

The logit regressions in conditions 1, 2, 5, and 6 showed highly strong effects of self-outcomes on choice switching in the next trial (z's < –6.90, p's < 0.001). Conditions 2 and 6 showed the effects of transformed other-outcomes on choice switching (z's < –2.42, $P < 0.05$). No significant interactions between drug and self-outcomes were present in these four conditions. The only significant interaction between drug and transformed other-outcomes in these four conditions was found in condition 6 (z = 2.84, $P = 0.005$).

In condition 3, the transformed other-outcomes robustly affected choice switching in the next trial (z = –6.21, $P < 0.001$), and the interaction between drug and other-outcomes was also highly significant (z = 3.78, $P < 0.001$), indicating decreased learning for other after OT. In condition 4, the effect of transformed other-outcomes on choice switching was marginally significant (z = –1.82, $P = 0.069$), but no interaction was found (**Table S3.1**).

Computational modeling results

The model comparison result indicated that the double angle distance model was the winning model, which had the smallest model-fit indices (**Table 3.1**). The highly significant correlations between generative and recovered parameters in parameter recovery suggested that the winning model and the estimation method could reliably capture the true values of the parameters (**Figure S3.5**).

Bayesian-estimated regressions from this model demonstrated a significant effect of OT on the group-level mean of learning rates for other [$\mu_{slope} = –0.93$, HDI: [–1.98, –0.04], $P = 0.020$], showing that learning rates for other decreased after OT. The effects of OT on other model parameters were not significant **(Figure 3.3; Table S3.2**).

The individual estimates of preferred allocations (η) in the PL condition had a robust positive correlation with the SVO measure (r = 0.48, $P = 0.009$; **Figure 3.4a**). When dividing the participants into two groups based on preferred allocations in the PL condition, the positive η group (N = 17) showed decreased preferred allocations after OT [t(16) = –2.68, $P = 0.017$]. The negative η group (N = 12) only showed a nonsignificant trend of increased preferred allocations after OT [t(11) = 1.15, $P = 0.27$; **Figure 3.4b**]. We also tested whether individual differences in η modulated OT's effect on the learning rate for other, and found that OT decreased learning rates for other in both positive and negative η groups [t(16) = –4.75, $P < 0.001$; t(11) = –5.65, $P < 0.001$], and the decreased values of learning rates were not different between these two groups [t(27) = 0.67, $P = 0.51$].

Imaging results

The ROI analysis showed that the self-PE signals in the ventral striatum were significantly larger than 0 in both PL [t(24) = 3.10, $P = 0.005$] and OT [t(24) = 2.49, $P = 0.020$] conditions, but no significant difference was observed between the two

conditions [t(24) = –0.04, *P* = 0.80]. For the ACC ROI, the mean of other-PE signals in the PL condition was slightly above 0 but did not reach significance [t(24)=1.57, *P* = 0.13]. After OT, the other-PE signals in the ACC decreased significantly, as shown by a paired t-test [t(24) = –2.19, *P* = 0.039; **Figure 3.4c**).

DISCUSSION

The present study investigated the behavioral and neural effects of OT on learning of rewards delivered to oneself and others. Using a neurocomputational approach, we demonstrated that both the behavioral learning of non-social rewards for others and the related other-regarding learning signals in the ACC were attenuated after OT administration, whereas neither behavioral nor neural learning of non-social rewards for oneself was affected by OT. These results are in line with findings that OT could only influence learning with social components (Evans et al., 2010; Hurlemann et al., 2010).

Bartz et al. (2011) suggested taking account of contexts and individual differences when examining the effects of OT in the social domain. Therefore, we should consider these two factors in the analysis of a social learning task. As the model-agnostic results showed, when the participants were always gaining points and had the chance to learn others' rewards (as in the condition 3 of our learning task), the cooperative individuals would learn to select the high value option for other, while the competitive individuals would learn to select the low value option for other. However, when the participants were always losing points (as in condition 4), despite the chance of learning others' rewards, neither cooperative or competitive participants would show distinct learning. These results demonstrated a clear interaction between learning contexts and learners' social preferences.

The angle distance model and the double angle distance model were constructed to capture both the context-dependent learning as well as individual differences in social preferences. In the angle distance model, the other-outcomes are transformed as a function of both individual-level preferred allocations and trial-by-trial angle distances. This notion is supported by a recent study (Liu et al., 2019), in which participants were shown to use their individual preferred allocations as a reference point for rating potential self-other allocations in a reward evaluation task. The double angle distance made an improvement by applying a similar dynamic transformation to self-outcomes. Not only had model comparisons shown its superiority over other candidate models, but additional support is also provided by the robust correlation between the preferred allocations estimated from this model and the SVO measures (The same correlation was only marginally or not significant for other models with the preference parameter). Moreover, the double angle distance model can help to explain some model-agnostic findings. If we derive the differences between expected values of the two options in the six conditions from this winning model, we can see that the average difference in condition 4 (0.11) is much smaller than that in other conditions (all larger than 0.33), which might be the reason for the lack of learning in

condition 4 (**Table S3.3**). This smaller difference in expected values also makes it more difficult to choose between the two options, which is reflected by the longer response time in condition 4 relative to other conditions (**Figure S3.3**).

When we applied this double angle distance model to test OT's effects on reward processing, decreased learning rates for other were observed after OT administration. Model-agnostic results were consistent with this effect. The learning curves for other in condition 3 became flat in both cooperative and competitive participants after OT. The logit regression results also revealed that, in task condition 3, other-outcomes transformed by SVO could predict choice switching in the next trial much better for the PL condition than the OT condition. All these results showed converging evidence for the attenuated other-regarding learning as a result of extraneous OT administration. Moreover, the decreased learning rates for other could be found in both participants with positive and negative preferred allocations in the PL condition, suggesting that OT's effect on other-regarding learning is independent of individual differences in social preferences. These results lend a complement to Ide et al. (2018)'s finding that OT decreased learning rates for social rewards delivered to oneself, providing strong evidence for their hypothesis that OT attenuates social reward learning.

The imaging results of other-regarding PE signals in the ACC provide more supportive evidence for this hypothesis as well as the ACC's role in tracking rewards received by others (Apps et al., 2016; Chang, Gariépy, & Platt, 2013). The decreased PE signals in the ACC can be attributed to the interaction between OT and the dopaminergic system. Exogenous administration of OT has previously been shown to elevate concentrations of extracellular (tonic) dopamine in reward-related regions (Kohli et al., 2019; Sanna, Argiolas, & Melis, 2012; Young, Liu, Gobrogge, Wang, & Wang, 2014), and enhanced extracellular dopamine is associated with diminished burst-firing (phasic) activity of dopamine neurons (Grace, 2016), which is necessary for coding PEs (Montague et al., 1996; Schultz et al., 1997). Presumably, exogenous administration of OT may lead to the suppression of phasic dopaminergic activity in the ACC, which in turn causes attenuated coding of PEs for other-related reward information.

Apart from its influences on reward learning, OT could also affect social preferences. As shown in our results, participants with positive/cooperative preferred allocations became less cooperative after OT. For participants with negative/competitive preferred allocations, they showed a nonsignificant trend of being less competitive after OT. These results provide a complement to the findings of Liu et al. (2019), which showed OT's effects of increasing cooperation in competitors and individualists, and a nonsignificant trend of decreasing cooperation in prosocials. Results from these studies can be accounted for with OT's effects of social salience promotion (Shamay-Tsoory & Abu-Akel, 2016) and anxiety reduction (Bartz & Hollander, 2006; Bartz et al., 2011). Participants with competitive social preferences might see the anonymous/unfamiliar other as unfriendly (with a baseline of high

salience of unfamiliarity) and feel more anxiety. The administration of OT could alleviate this feeling and decrease their competitiveness. Participants with cooperative social preferences might see the anonymous/unfamiliar other as friendly and feel no anxiety, but OT could increase the salience of unfamiliarity and results in lower prosociality (Declerck, Boone, & Kiyonari, 2010). This increased unfamiliarity may also be a reason of decreased other-regarding learning after OT in that the participants would care less about the well-being of the anonymous partner.

In summary, the present study has three important implications. Firstly, it provides a generic social reward learning model that can effectively account for learning contexts and individual differences in social preferences (Bartz et al., 2011). This model can also serve as a tool for hypothesis testing in future studies of self- and other-regarding reward learning. Secondly, it reveals the behavioral and neural mechanisms of OT's effects on the perception and learning of other-regarding rewards, which confirmed several important hypotheses from previous studies and may help to resolve inconsistent findings in the field. Thirdly, our results show that OT attenuates other-regarding learning and decreases prosociality in some individuals, which entails clinicians' caution when using OT as a therapeutic intervention for mental disorders. In spite of the implications, some limitations should also be noted. In our learning task, the variations of the four outcomes were relatively small and could not cover all angles in a Cartesian coordinate system. Future studies should test reward processing for more types of outcomes (e.g., self gaining 0 points). Another limitation is that the social partner in the task was an anonymous person. Whether social familiarity would modulate OT's effects was not specifically examined. As for the participants, we only recruited males; therefore, the effects of OT on learning for others in females are yet to be explored in future research.

TABLES AND FIGURES

Table 3.1. Model-fit indices of candidate models

Model	iBIC	LOOIC	WAIC
Basic RL model (Only self-outcomes included)	8897.5	8975.4	8974.1
Gamma model	8106.3	8216.1	8213.3
Angle distance model	7724.2	7886.9	7879.6
Double angle distance model	7715.9	7876.9	7868.9
Fehr-Schmidt model	7901.6	8053.7	8047.6
Van Lange model	7856.2	8002.5	7993.5

The best-fitting model is underlined.

Abbreviations: iBIC, integrated Bayesian information criterion; LOOIC, leave-one-out cross-validation information criterion; WAIC, widely applicable information criterion.

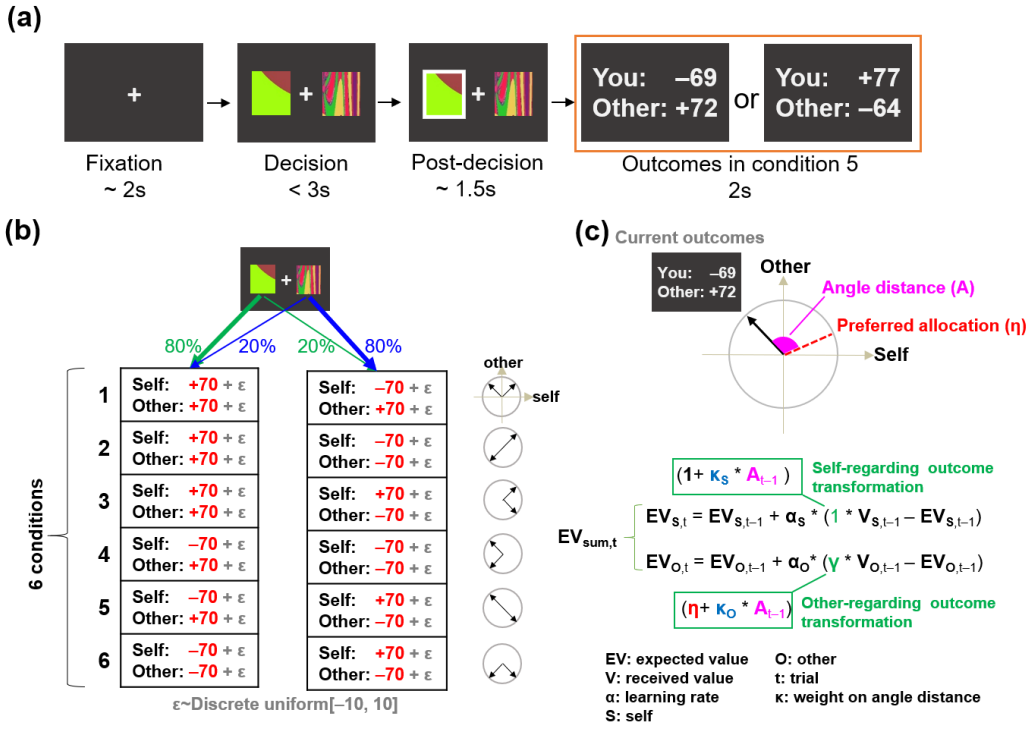

Figure 3.1. Probabilistic social learning task and the double angle distance model. (a) One example trial from condition 5 of all six conditions in the probabilistic social learning task. Participants chose between two abstract patterns, viewed the outcomes of the choice, and learned over time which was the better option. Their choices affected the payoffs for themselves and an anonymous partner. **(b)** Participants made 30 choices in each condition and learned the contingencies between the patterns and outcomes. The two possible outcomes in each condition can be represented by two vectors/arrows with certain angles in a Cartesian coordinate system with x-axis for self and y-axis for other. In conditions 1 and 6, other is always winning or losing, so it can be assumed that there is mainly learning for self. On the contrary, in conditions 3 and 4, self is always winning or losing, so presumably there is mainly learning for other. In conditions 2 and 5, we can assume that there is learning for both self and other. **(c)** The double angle distance model is adapted from the gamma model (Christopoulos & King-Casas, 2015), in which the expected values for self (EV_S) and other (EV_O) are independently updated as in the basic RL model. In the updating of EV_O, the reward (V_O) delivered to others is transformed by γ. Our new model takes account of the angle between a learner's preferred allocation and the outcomes in each trial, both of which can be represented by vectors. Hence, the other-regarding transformation parameter γ can be substituted by $(\eta + \kappa * A_{t-1})$, in which η is the preferred allocation and A_{t-1} is the angle distance between the two vectors representing η and outcomes at trial t–1. Parameter κ is the weight representing the extent to which the angle distance can influence the other-regarding preference transformation. The self-regarding transformation parameter is 1 in the gamma model, and can be modified similarly as γ.

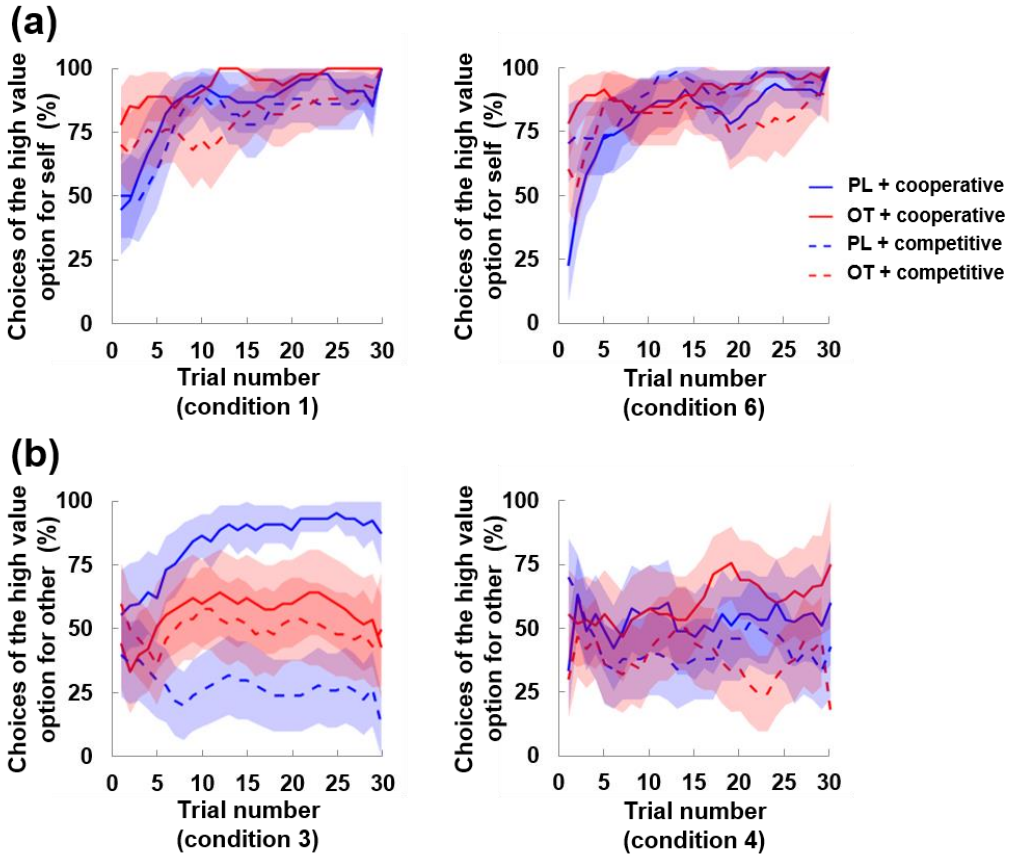

Figure 3.2. Learning curves for self and other in participants with cooperative and competitive social preferences. (a) In task conditions 1 and 6, cooperative and competitive participants in both drug conditions showed similar learning for self. **(b)** In task condition 3 and after placebo (PL) administration, cooperative participants showed learning of the high value option for other and competitive participants showed learning of the low value option for other. The learning in condition 3 was eliminated after oxytocin (OT). In task condition 4, the learning curves did not show learning for other in either group after PL or OT. The cooperative group included 9 participants with social value orientation measures (SVOs) higher than 7. The competitive group included 10 participants with SVOs lower than –9. The learning curves depict the running average (window size = 5; mean ± standard error) of the trial-by-trial proportion of participants in each group and drug condition that selected the high value option.

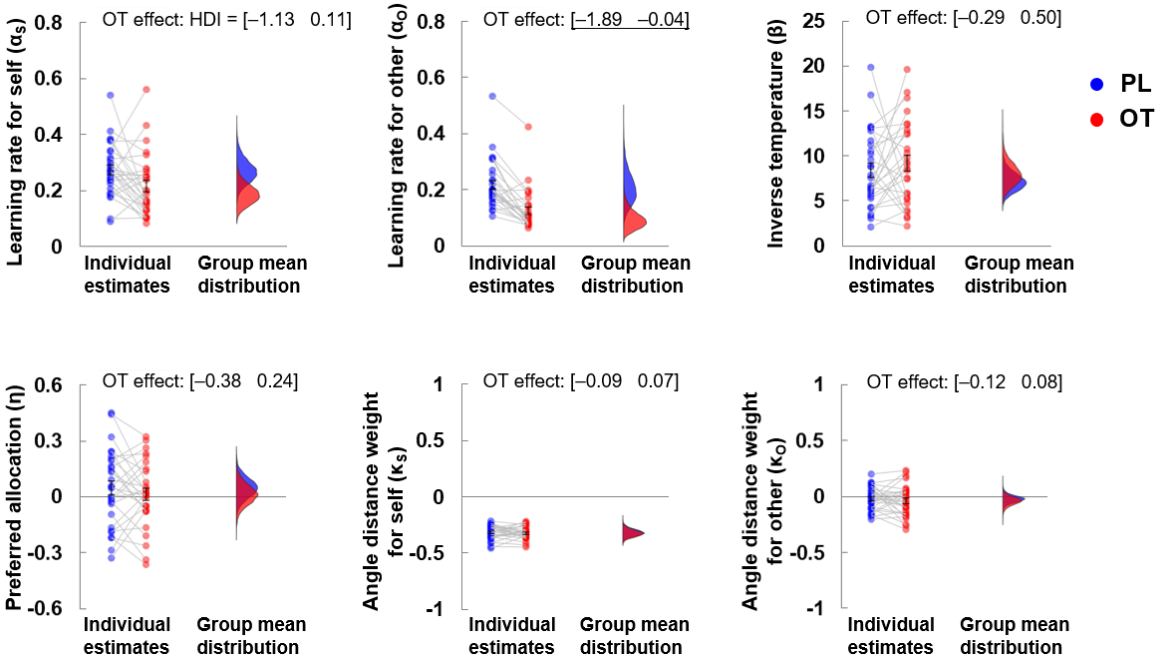

Figure 3.3. Individual-level estimates and group-level mean distributions of model parameters. Bayesian-estimated regressions for the effects of oxytocin (OT) on learning parameters showed that learning rates for other decreased after OT administration relative to placebo [PL; μ_slope = –0.93, 95% highest density interval (HDI): [–1.98, –0.04], $P = 0.020$]. Other parameters were not significantly different between drug conditions (**Table S3.2**). Significant OT effects are underlined. Error bars indicate standard errors.

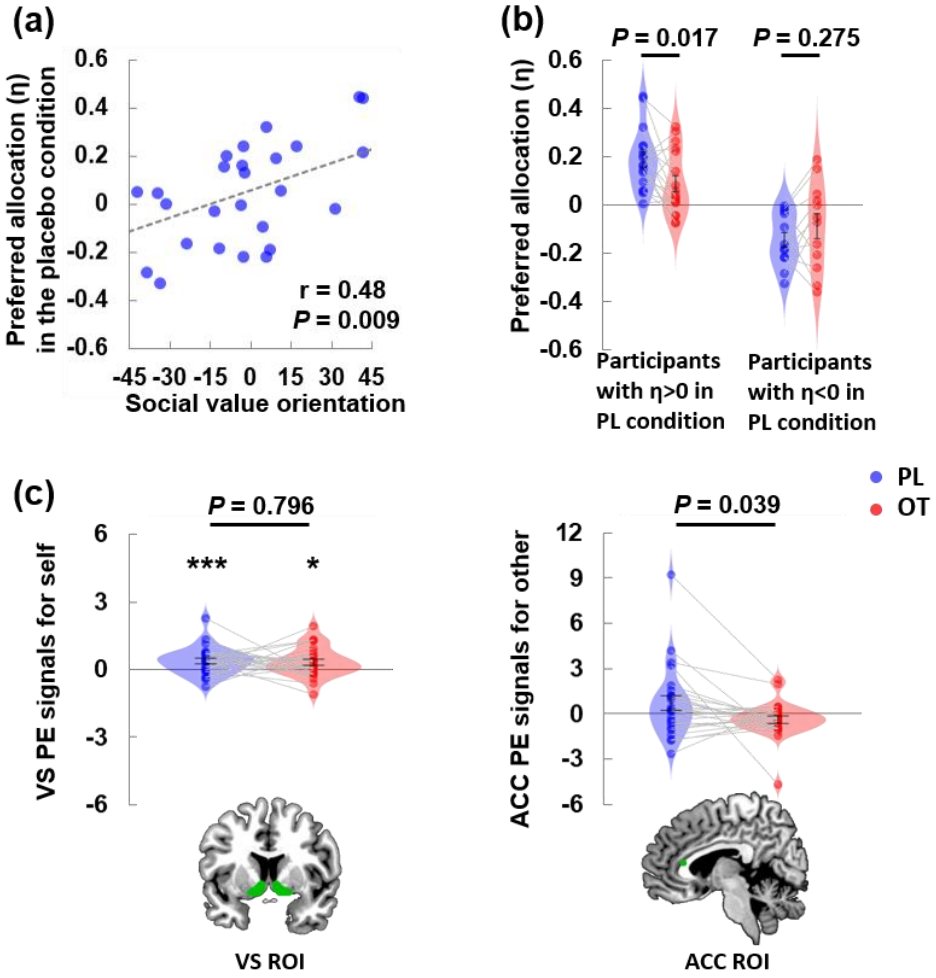

Figure 3.4. Behavioral results of the preferred allocation and neuroimaging results of self- and other-regarding prediction error (PE) signals. (a) The preferred allocation (η) in the placebo (PL) condition was positively correlated with the social value orientation measured in a non-learning task. **(b)** For participants with positive preferred allocations in the PL condition, oxytocin (OT) decreased the values of preferred allocations. For participants with negative preferred allocations in the PL condition, they showed a weak trend of increased preferred allocations after OT. **(c)** Self- and other-PE signals in the ventral striatum (VS) and anterior cingulate cortex (ACC) regions-of-interest (ROIs). The other-PE signals in the ACC decreased after OT. Error bars indicate standard errors. $^{*}P < 0.05$, $^{***}P < 0.005$.

REFERENCES

Aoki, Y., Yahata, N., Watanabe, T., Takano, Y., Kawakubo, Y., Kuwabara, H., . . . Suga, M. (2014). Oxytocin improves behavioural and neural deficits in inferring others' social emotions in autism. *Brain, 137*(11), 3073-3086.

Apps, M. A., Lesage, E., & Ramnani, N. (2015). Vicarious reinforcement learning signals when instructing others. *Journal of Neuroscience, 35*(7), 2904-2913.

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron, 90*(4), 692-707.

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4), 594–615.

Baribeau, D. A., & Anagnostou, E. (2015). Oxytocin and vasopressin: linking pituitary neuropeptides and their receptors to social neurocircuits. *Frontiers in Neuroscience, 9*, 335.

Bartz, J. A., & Hollander, E. (2006). The neuroscience of affiliation: forging links between basic and clinical research on neuropeptides and social behavior. *Hormones and Behavior, 50*(4), 518-528.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience, 21*(8), 2793-2798.

Boccia, M., Petrusz, P., Suzuki, K., Marson, L., & Pedersen, C. (2013). Immunohistochemical localization of oxytocin receptors in human brain. *Neuroscience, 253*, 155-164.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4), 467-484.

Chang, S. W., Gariépy, J.-F., & Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience, 16*(2), 243.

Christopoulos, G. I., & King-Casas, B. (2015). With you or against you: social orientation dependent learning signals guide actions made for others. *Neuroimage, 104*, 326-335.

Clark-Elford, R., Nathan, P. J., Auyeung, B., Voon, V., Sule, A., Müller, U., . . . Baron-Cohen, S. (2014). The effects of oxytocin on social reward learning in humans. *International Journal of Neuropsychopharmacology, 17*(2), 199-209.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408-1411.

Declerck, C. H., Boone, C., & Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information. *Hormones and Behavior, 57*(3), 368-374.

Evans, S., Shergill, S. S., & Averbeck, B. B. (2010). Oxytocin decreases aversion to angry faces in an associative learning task. *Neuropsychopharmacology, 35*(13), 2502.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2007). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders-Patient Edition (With Psychotic Screen) (SCID-I/P (W/ PSYCHOTIC SCREEN), 1/2007 revision)*. Biometrics Research

Department, New York State Psychiatric Institute

Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience, 17*(8), 524-532.

Greenberg, E., & Webster, C. E. (1983). *Advanced econometrics: a bridge to the literature*. New York, NY: John Wiley & Sons.

Grinevich, V., Knobloch-Bollmann, H. S., Eliava, M., Busnelli, M., & Chini, B. (2016). Assembling the puzzle: pathways of oxytocin signaling in the brain. *Biological Psychiatry, 79*(3), 155-164.

Hung, L. W., Neuner, S., Polepalli, J. S., Beier, K. T., Wright, M., Walsh, J. J., . . . Dölen, G. (2017). Gating of social reward by oxytocin in the ventral tegmental area. *Science, 357*(6358), 1406-1411.

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., . . . Maier, W. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience, 30*(14), 4999-5007.

Huys, Q., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology, 8*(3).

Ide, J. S., Nedic, S., Wong, K. F., Strey, S. L., Lawson, E. A., Dickerson, B. C., . . . Mujica-Parodi, L. R. (2018). Oxytocin attenuates trust as a subset of more general reinforcement learning, with altered reward circuit functional connectivity in males. *Neuroimage, 174*, 35-43.

Israel, S., Weisel, O., Ebstein, R. P., & Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology, 37*(8), 1341-1344.

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience, 9*(7), 940-947.

Kohli, S., King, M. V., Williams, S., Edwards, A., Ballard, T. M., Steward, L. J., . . . Fone, K. C. (2019). Oxytocin attenuates phencyclidine hyperactivity and increases social interaction and nucleus accumben dopamine release in rats. *Neuropsychopharmacology, 44*(2), 295-305.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford, UK: Oxford University Press.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., . . . Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience, 22*(4), 633-641.

Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of Neuroscience, 35*(40), 13720-13727.

Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.

Luce, R. D. (2014). *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications, 11*(1), 1-11.

McClintock, C. G. (1972). Social motivation—A set of propositions. *Behavioral Science, 17*(5), 438-454.

Melis, M. R., Melis, T., Cocco, C., Succu, S., Sanna, F., Pillolla, G., . . . Argiolas, A. (2007). Oxytocin injected into the ventral tegmental area induces penile erection and increases extracellular dopamine in the nucleus accumbens and paraventricular nucleus of the hypothalamus of male rats. *European Journal of Neuroscience, 26*(4), 1026-1035.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4*(1), 1-25.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*(5), 1936-1947.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rilling, J. K., DeMarco, A. C., Hackett, P. D., Thompson, R., Ditzen, B., Patel, R., & Pagnoni, G. (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology, 37*(4), 447-461.

Rogers, C. N., Ross, A. P., Sahu, S. P., Siegel, E. R., Dooyema, J. M., Cree, M. A., . . . Albers, H. E. (2018). Oxytocin-and arginine vasopressin-containing fibers in the cortex of humans, chimpanzees, and rhesus macaques. *American Journal of Primatology, 80*(10), e22875.

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron, 70*(6), 1054-1069.

Sanna, F., Argiolas, A., & Melis, M. R. (2012). Oxytocin-induced yawning: sites of action in the brain and interaction with mesolimbic/mesocortical and incertohypothalamic dopaminergic neurons in male rats. *Hormones and Behavior, 62*(4), 505-514.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Shamay-Tsoory, S. G., & Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biological Psychiatry, 79*(3), 194-202.

Stan Development Team. 2017. *RStan: the R interface to Stan*. R package version 2.16.0. http://mc-stan.org

Succu, S., Sanna, F., Cocco, C., Melis, T., Boi, A., Ferri, G. L., . . . Melis, M. R. (2008). Oxytocin induces penile erection when injected into the ventral tegmental area of male rats: role of nitric oxide and cyclic GMP. *European Journal of Neuroscience,*

28(4), 813-821.

Sul, S., Tobler, P. N., Hein, G., Leiberg, S., Jung, D., Fehr, E., & Kim, H. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National Academy of Sciences, 112*(25), 7851-7856.

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

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413-1432.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30*, 31-56.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*(Dec), 3571-3594.

Young, K. A., Liu, Y., Gobrogge, K. L., Wang, H., & Wang, Z. (2014). Oxytocin reverses amphetamine-induced deficits in social bonding: evidence for an interaction with nucleus accumbens dopamine. *Journal of Neuroscience, 34*(25), 8499-8506.

Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PloS One, 2*(11).

Zik, J. B., & Roberts, D. L. (2015). The many faces of oxytocin: implications for psychiatry. *Psychiatry Research, 226*(1), 31-37.

The Effects of Oxytocin on Self- and Other-Regarding Reinforcement Learning

SUPPLEMENTARY INFORMATION

Supplementary methods

Behavioral analysis

Computational modeling

The six RL models were estimated to fit the participants' choice data. The free parameters of each model were estimated using hierarchical Bayesian analysis (HBA). In HBA, a hierarchical model is constructed to include individual-level parameters for each participant and group-level parameters that describe the distribution of individual-level parameters. Different levels can inform each other, allowing more precise estimation of individual differences in model parameters (Kruschke & Vanpaemel, 2015). The Stan software package (version 2.16.0; Stan Development Team, 2017a) was used to perform HBA. Stan applies Hamiltonian Monte Carlo in its model estimation, which is a Markov chain Monte Carlo sampling algorithm and can efficiently obtain samples for multi-dimensional models with highly correlated parameters (Ahn et al., 2014).

In our models, parameters for individual participants were assumed to be drawn from group-level distributions, which were shared for all participants. For example, the unconstrained form of individual-level learning rate (α') was set to be from a group-level normal distribution, with normal and half-Cauchy distributions as the priors for its mean ($\mu_{\alpha'}$) and standard deviation ($\sigma_{\alpha'}$), respectively. This unconstrained individual-level learning rate (α') was expressed as the sum of the group mean and the product between the group variance and an individually estimated error parameter (*error*), drawn from a unit normal distribution. This expression is a method of reparameterization to enable more efficient sampling and is referred to as "Matt trick" in *Stan User's Guide and Reference Manual* (Stan Development Team, 2017b). Since the individual-level learning rate (α) should be bounded between 0 and 1, an inverse logit function was applied to convert the unconstrained individual-level learning rate (α') to be within this range. Therefore, the learning rate was defined as follows:

$$\alpha = 1 / (1+ \exp(-\alpha'))$$

$$\alpha' = \mu_{\alpha'} + \sigma_{\alpha'} * error.$$

To examine how oxytocin (OT) affected the learning rate, the group-level mean ($\mu_{\alpha'}$) was represented by a regression equation:

$$\mu_{\alpha'} = \mu_{\alpha'\_intercept} + OT * \mu_{\alpha'\_slope}$$

where *OT* was dummy coded according to each participant's drug conditions. The regression slope represented the effect of OT on the group-level mean of learning rates. To determine significance, the 95% credible intervals or highest density intervals (HDI) of these parameters were required to not include 0. We also calculated

the proportion of samples in each slope's distribution that had the opposite sign to its mean. It can serve as a significance measure similar to frequentist *P* values (Mack, Preston, & Love, 2020).

The distributions of prior parameters were set as follows:

$$\mu_{\alpha'\_intercept} \sim \text{Normal}\ (0, 0.5)$$

$$\mu_{\alpha'\_slope} \sim \text{Normal}\ (0, 2)$$

$$\sigma_{\alpha'} \sim \text{Cauchy}\ (0, 2)$$

$$error \sim \text{Normal}\ (0, 1).$$

Other parameters were defined similarly. For parameters with only a lower bound of 0 (inverse temperature), the exponential function was used in the transformation: $x = \exp(x')$, in which x represented the constrained parameter and x' represented the unconstrained parameter. For parameters bounded between –1 and 1, a modified inverse logit function was used: $x = 2 * [1 / (1 + \exp(-x')) - 0.5]$.

For each model's estimation, HMC sampling was conducted with four chains. Each chain had 4000 samples, 2500 of which were set as warm up samples and were discarded, resulting in 10000 samples for each parameter. For the best-fitting model, the samples for all parameters showed good convergence, with the potential scale reduction statistic ($\hat{R}$; Gelman & Rubin, 1992) less than 1. 01. In addition, effective sample sizes (ESS) of model parameters were all larger than 1732, indicating sufficiently low autocorrelation and good mixing of HMC chains.

Imaging analysis

Functional images were obtained using a T2*-weighted echo-planar imaging (EPI) sequence [repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90˚, field of view (FOV) = 220 × 220 mm^2]. Each volume contained 34 interleaved axial slices (matrix = 64 × 64, in-plane spatial resolution = 3.44 × 3.44 mm^2, thickness = 4 mm), which were angled 30˚ with respect to the anterior-posterior commissural line. High-resolution anatomical images were obtained using a T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (TR = 1200 ms, TE = 2.66 ms, flip angle = 8˚, FOV = 245 × 245 mm). The anatomical volume contained 192 axial slices (matrix = 245 × 245, in-plane spatial resolution = 1 × 1 mm^2, thickness = 1 mm).

SPM8 software package (Wellcome Trust Centre for Neuroimaging, London, UK) was used for imaging data preprocessing. Slice timing artifacts of functional images were corrected and then the imaging data were realigned to correct for head movement between scans, and each participants' anatomical scan was coregistered to the mean functional image produced in the realignment stage. The anatomical scans were then segmented and spatial normalization parameter matrices were generated based on the Montreal Neurological Institute (MNI) template. Functional images were

transformed into the MNI space using these matrices. Normalized functional data were then spatially smoothed using an isotropic Gaussian filter with a full-width at half-maximum parameter of 8 mm.

Supplementary tables

Table S3.1. Summary of logit regression analyses testing how oxytocin (OT) modulated the influences of self-outcomes and other-outcomes on choice switching in the next trial for the six task conditions.

Independent variable	**Condition 1**		**Condition 2**		**Condition 3**	
	b (SE)	**z value**	**b (SE)**	**z value**	**b (SE)**	**z value**
(Intercept)	–1.19 (0.09)	–12.72***	–1.38 (0.10)	–14.06***	1.60 (1.97)	0.81
Self-outcomes	–1.29 (0.13)	–9.81***	–1.25 (0.14)	–8.81***	–5.01 (2.81)	–1.78†
Other-outcomes (Transformed)	–0.33 (0.25)	–1.36	–0.63 (0.26)	–2.42*	–1.85 (0.30)	–6.21***
Drug (OT=1, PL=0)	–0.02 (0.13)	–0.16	0.23 (0.13)	1.70†	–2.64 (2.57)	–1.03
*Self-outcomes*Drug*	0.25 (0.19)	1.32	0.25 (0.19)	1.32	4.43 (3.65)	1.21
*Other-outcomes *Drug*	–0.50 (0.35)	–1.42	0.23 (0.36)	0.65	1.44 (0.38)	3.78***
Independent variable	**Condition 4**		**Condition 5**		**Condition 6**	
	b (SE)	**z value**	**b (SE)**	**z value**	**b (SE)**	**z value**
(Intercept)	–2.14 (1.32)	–1.61	–1.24 (0.09)	–13.42***	–1.31 (0.10)	–13.58***
Self-outcomes	–2.54 (1.86)	–1.36	–0.91 (0.13)	–6.90***	–1.15 (0.14)	–8.45***
Other-outcomes (Transformed)	–0.35 (0.19)	–1.82†	–0.06 (0.24)	–0.26	–0.61 (0.25)	–2.43*
Drug (OT=1, PL=0)	1.10 (1.89)	0.58	0.08 (0.13)	0.60	–0.11 (0.14)	–0.77
*Self-outcomes*Drug*	1.72 (2.66)	0.65	–0.30 (0.19)	–1.60	0.01 (0.19)	0.08
*Other-outcomes *Drug*	–0.10 (0.28)	–0.38	–0.09 (0.34)	–0.26	1.03 (0.36)	2.84**

Abbreviations: PL, placebo; SE, standard error. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.005$, $^{\dagger}P < 0.1$.

Table S3.2. Bayesian-estimated regressions of oxytocin (OT)'s effects on learning parameters in the winning model (N = 29)

Parameter	Samples from the double angle distance model			
	Mean (SD)	HDI		*P*
Learning rate for self				
Intercept	–1.02 (0.21)	–1.43	–0.58	0.000
OT effect	–0.48 (0.32)	–1.13	0.11	0.055
Learning rate for other				
Intercept	–1.38 (0.35)	–2.07	–0.68	0.000
OT effect	–0.93 (0.47)	–1.89	–0.04	0.020
Inverse temperature				
Intercept	1.96 (0.14)	1.68	2.24	0.000
OT effect	0.11 (0.20)	–0.29	0.50	0.293
Preferred allocation				
Intercept	0.10 (0.11)	–0.12	0.31	0.185
OT effect	–0.07 (0.16)	–0.38	0.24	0.321
Angle distance weight for self				
Intercept	–0.32 (0.03)	–0.37	–0.25	0.000
OT effect	–0.01 (0.04)	–0.09	0.07	0.426
Angle distance weight for other				
Intercept	–0.02 (0.03)	–0.09	0.05	0.249
OT effect	–0.02 (0.05)	–0.12	0.08	0.369

Significant effects are underlined.

Abbreviations: HDI, 95% highest density interval; *P*, the proportion of samples in each parameter's posterior distribution that have the opposite sign to its mean. It can serve as a significance measure similar to frequentist P values (Mack et al., 2020); SD, standard deviation.

Table S3.3. Mean value of trial-by-trial variables in the winning model across all participants (N of trials = 10066)

Variable in the model	Task condition					
	1	**2**	**3**	**4**	**5**	**6**
$\lvert(EV_{S,\,a} + EV_{O,\,a}) - (EV_{S,\,b} + EV_{O,\,b})\rvert$	0.35	0.33	0.33	0.11	0.33	0.33
κ_S * A	–0.40	–0.41	–0.24	–0.76	–0.41	–0.41
κ_O * A	–0.03	–0.04	–0.03	–0.07	–0.04	–0.04

Abbreviations: A, angle distance; *a*, option *a*; *b*, option *b*; EV, expected value; κ, weight on angle distance; O, other; S, self.

Supplementary figures

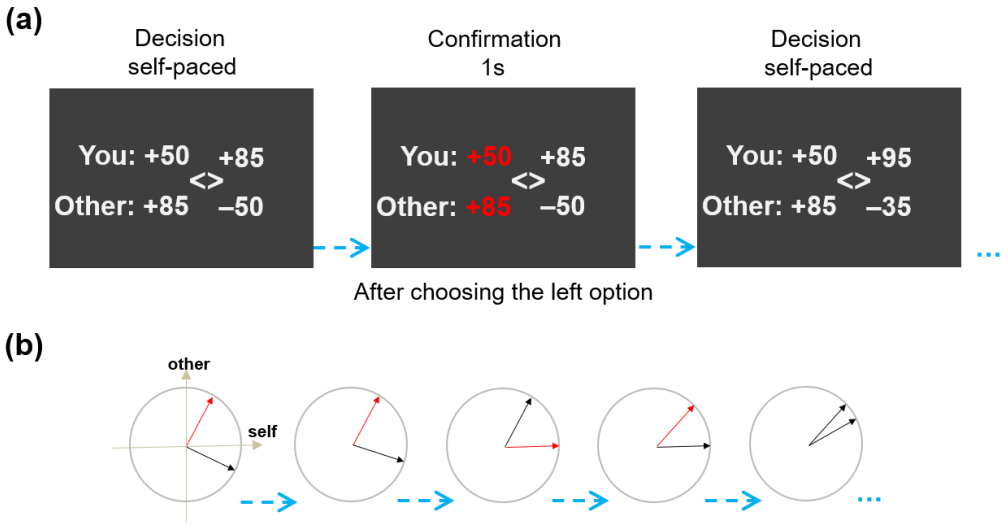

Figure. S3.1. Social value orientation (SVO) assessment. (a) A sequential testing procedure was used to assess the social preference of each participant. In a series of choices, participants indicated their preference between two allocations. Each allocation consisted of gaining or losing points for the participants and an anonymous partner. **(b)** The allocation pairs can be represented by two arrows in a Cartesian coordinate system, where the x-axis represents payoffs for self and the y-axis represents payoffs for other. Note that participants did not see the arrows, but allocations of points. After the participants indicated their preferred arrow (allocation) in each trial, the unchosen arrow would move toward the chosen one in a certain step size in the next trial. When the two arrows were close enough (both the difference between the two outcomes for self and the difference between the two outcomes for other were less than 6), the arrows were considered converged and the average of the two angles was used as the measure of SVO.

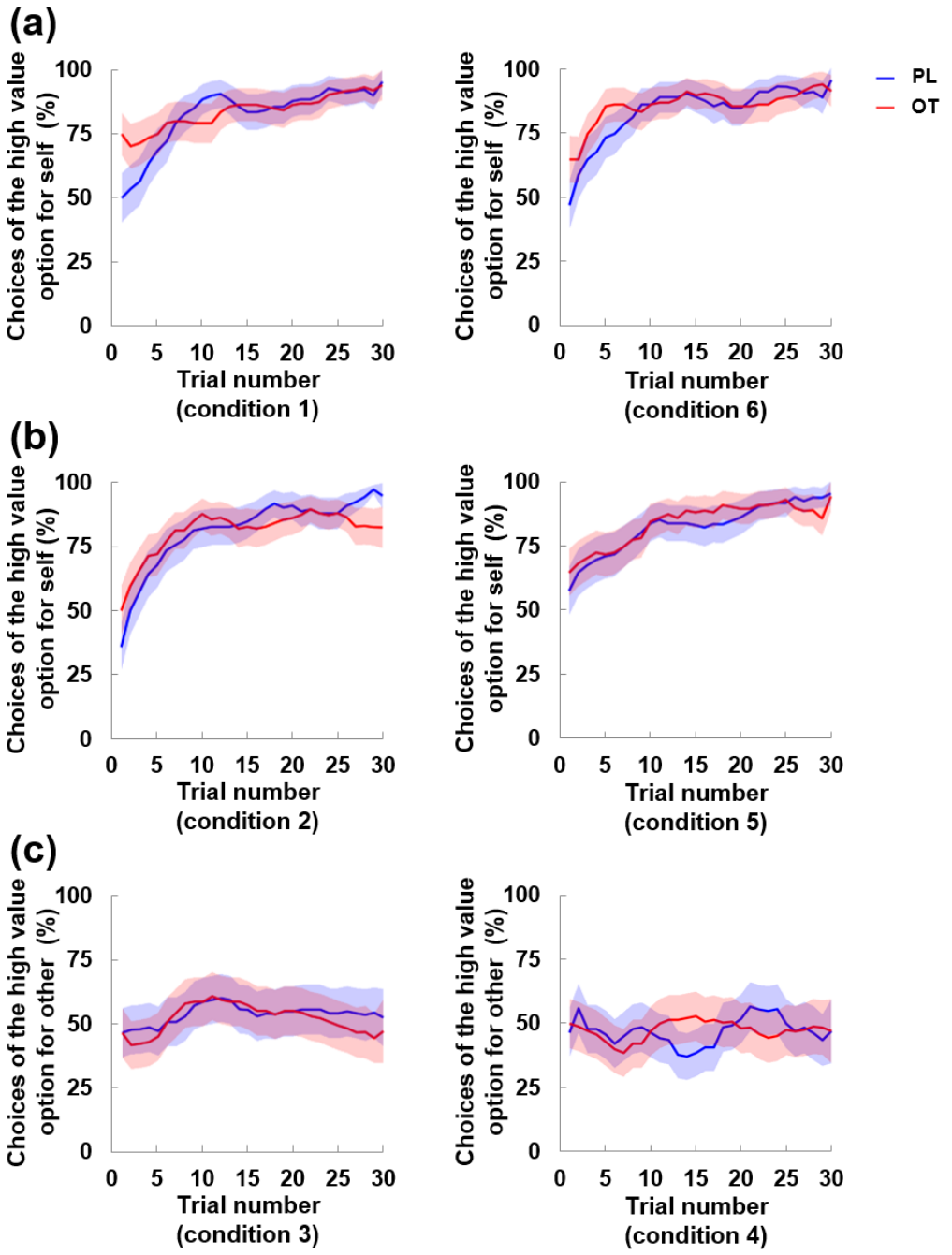

Figure S3.2. Learning curves for the six task conditions. (a) In task conditions 1 and 6, participants in both drug (oxytocin, OT; placebo, PL) conditions showed similar learning for self. **(b)** In task conditions 2 and 5, participants in both drug conditions showed similar learning for self. **(c)** In task conditions 3 and 4, participants in both drug conditions did not show learning for other.

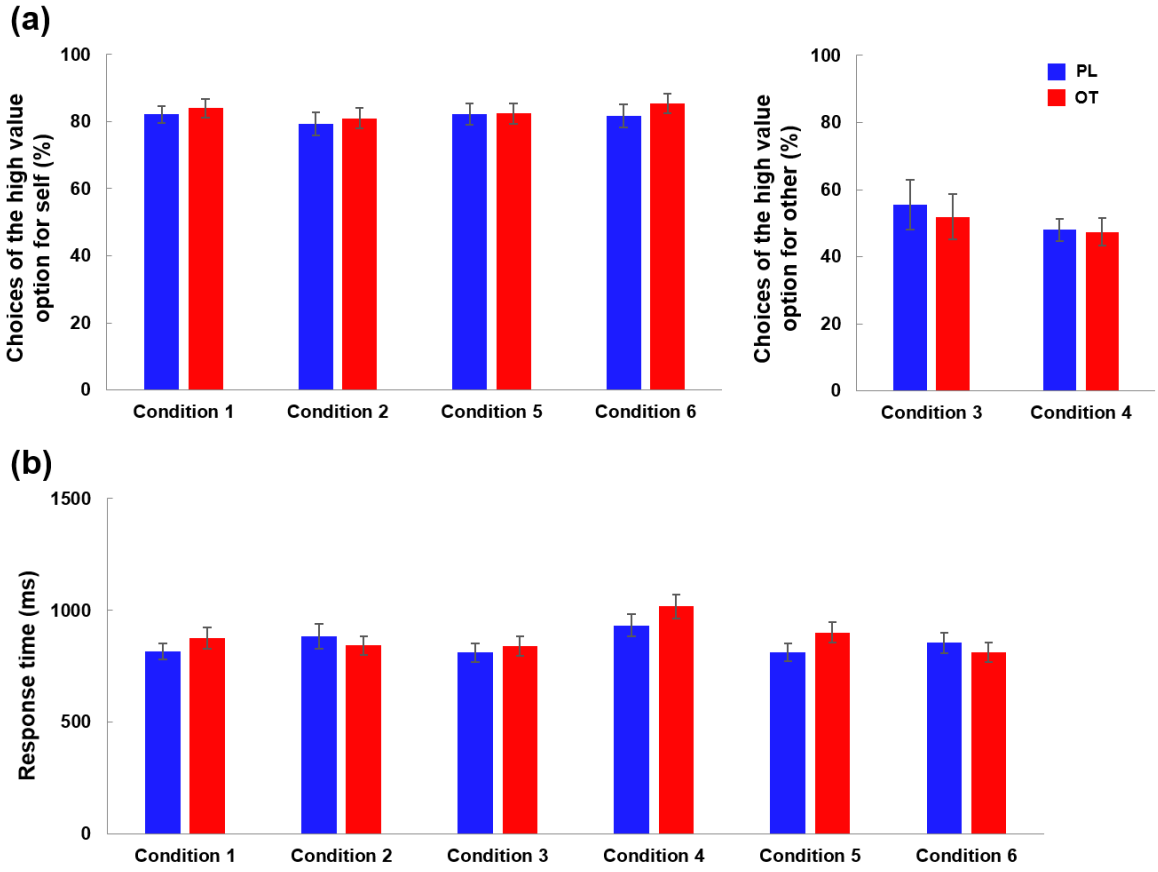

Figure S3.3. Model-agnostic performance in the six task conditions. (a) Choices of the high value option for self were not significantly different between drug (oxytocin, OT; placebo, PL) conditions or between task conditions 1, 2, 5, and 6 [drug factor: $F(1, 28) = 0.44$, $P = 0.51$; task factor: $F(3, 84) = 0.73$, $P = 0.54$; interaction: $F(3, 84) = 0.23$, $P = 0.88$]. Choices of the high value option for other were not significantly different between drug conditions or between task conditions 3 and 4 [drug factor: $F(1, 28) = 0.17$, $P = 0.68$; task factor: $F(1, 28) = 1.89$, $P = 0.18$; interaction: $F(1, 28) = 0.12$, $P = 0.74$]. **(b)** Response time was not significantly different between drug conditions, but there was a main effect of task conditions [drug factor: $F(1, 28) = 0.67$, $P = 0.42$; task factor: $F(5, 140) = 5.67$, $P < 0.001$; interaction: $F(1, 28) = 1.70$, $P = 0.14$]. Post-hoc tests revealed that response time in task condition 4 was significantly longer than all the other conditions ($|t| > 3.44$, $P < 0.01$). Error bars indicate standard errors.

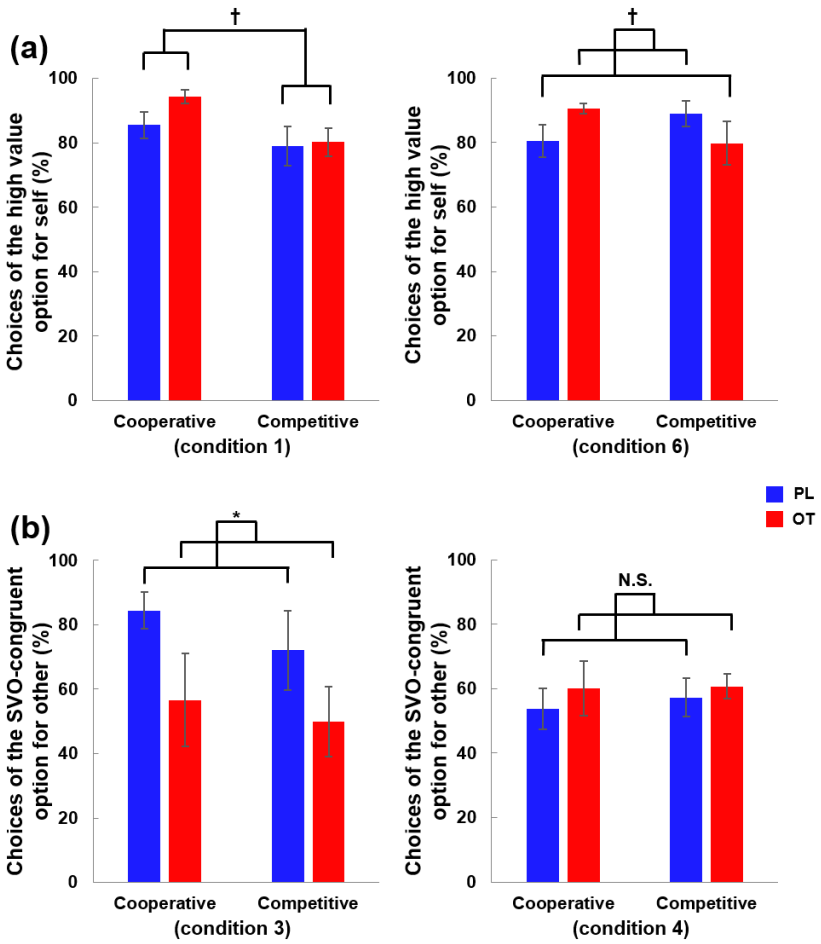

Figure S3.4. Choices for self and other in participants with cooperative and competitive social preferences. (a) In task condition 1, cooperative participants had a marginally higher ratio of choosing the high value option for self than competitive participants [drug factor: F(1, 17) = 1.63, P = 0.22; group factor: F(1, 17) = 4.06, P = 0.060; interaction: F(1, 17) = 1.00, P = 0.33]. In task condition 6, there was a marginally significant interaction between social value orientations (SVOs) and drug (oxytocin, OT; placebo, PL) conditions [drug factor: F(1, 17) = 0.01, P = 0.99; group factor: F(1, 17) = 0.08, P = 0.78; interaction: F(1, 17) = 3.04, P = 0.099]. **(b)** In condition 3, participants made more choices of the SVO-congruent option for other under PL than under OT treatment [drug factor: F(1, 17) = 4.84, P = 0.042; group factor: F(1, 17) = 0.67, P = 0.42; interaction: F(1, 17) = 0.07, P = 0.80]. In condition 4, no effects of drug or SVO were found [drug factor: F(1, 17) = 0.73, P = 0.41; group factor: F(1, 17) = 0.09, P = 0.76; interaction: F(1, 17) = 0.07, P = 0.80]. SVO-congruent option is the high other-value option for cooperative people, but the low other-value option for competitive people. The cooperative group included nine participants with SVOs higher than 7. The competitive group included 10 participants with SVOs lower than –9. Error bars indicate standard errors. $^{*}P < 0.05$, $^{\dagger}P < 0.1$, N.S.= not significant.

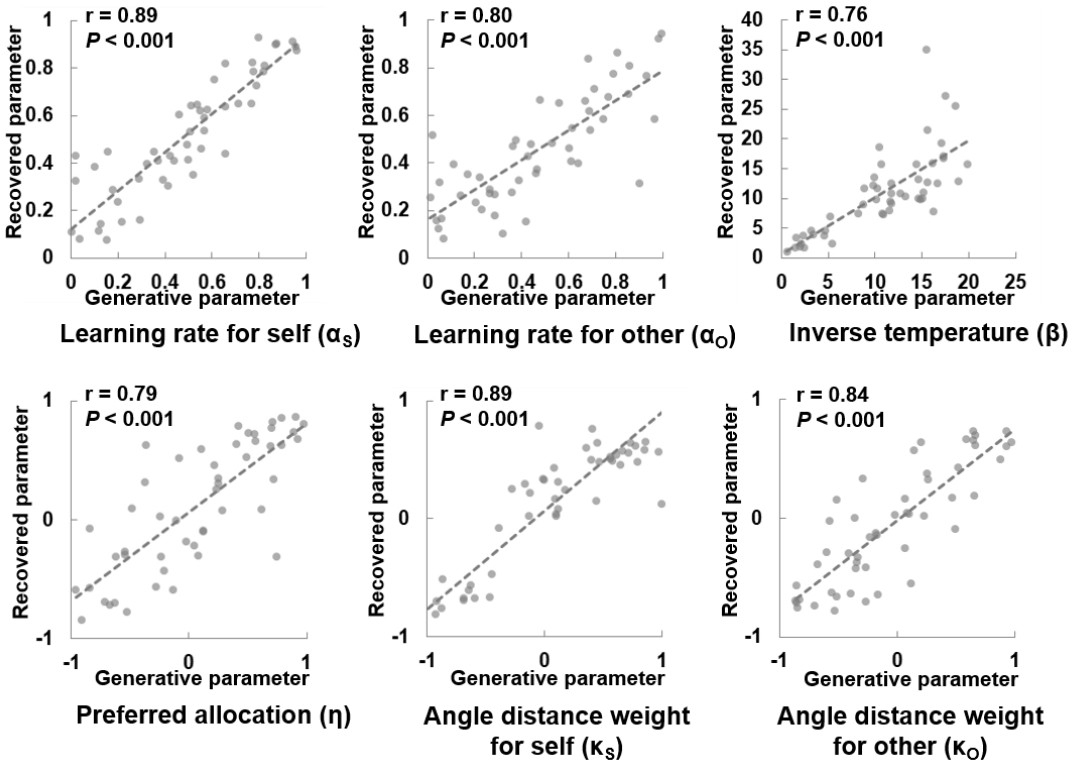

Figure S3.5. Parameter recovery for the winning model. For each parameter, we randomly drew 50 values from a uniform distribution. Specifically, values of learning rates were drawn from U (0, 1); values of inverse temperatures were drawn from U (0, 25); values of preferred allocations and angle distance weights were drawn from U (–1, 1). These generative parameters were randomly assigned to 50 hypothetical participants to generate choice data using the winning model, and then new learning parameters were estimated using these choice data. The values of these learning parameters can be fully recovered with the best-fitting model, as shown by the strong correlations.

Supplementary references

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology*, *5*, 1-15.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-472.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford, UK: Oxford University Press.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications, 11*(1), 1-11.

Stan Development Team. 2017a. *RStan: the R interface to Stan*. R package version 2.16.0. http://mc-stan.org

Stan Development Team. 2017b. *Stan Modeling Language Users Guide and Reference Manual*, Version 2.16.0. http://mc-stan.org

DISCUSSION

Summary of studies

This set of studies aimed to (i) unveil common and distinct self-regarding RL disturbances in depression, addiction, and anxiety, (ii) examine self- and other-regarding RL in PTSD with and without comorbid depression, and (iii) test whether OT can modify self- and other-regarding RL.

In Paper I, a systematic review revealed that learning inflexibility (possibly impaired model-based & model-free learning) and inconsistent choices were the similarities across all these disorders. As for the differences, depression was associated with decreased learning rates. Addiction was characterized by a higher weight to rewards versus punishments. Anxiety was featured by less optimism, hypersensitivity to punishments, and enhanced Pavlovian biases. Neurally, both depression and addiction were associated with impaired self-regarding learning signals in key reward regions. Anxiety showed impaired learning signals in some reward regions and also demonstrated heightened neural responses to unexpected feedback.

In Paper II, we found that PTSD patients, with or without depression, all showed decreased behavioral other-regarding learning compared to healthy controls. Neurally, other-regarding surprise (magnitude of PEs) signals in the right IPL were higher in depressed PTSD patients than in controls, with nondepressed PTSD patients in between. The other-regarding surprise signals in the right IPL also had a positive correlation with a measure of avoidance & numbing across all participants. Self-regarding learning was not affected by PTSD, but comorbid depression was associated with increased learning for self.

In Paper III, we found that, behaviorally, other-regarding learning decreased after OT relative to the PL condition, but self-regarding learning did not change significantly between drug conditions. Participants with cooperative social preferences in the PL condition became significantly less cooperative after OT. For participants with competitive social preferences in the PL condition, they showed a nonsignificant trend of being less competitive after OT. Imaging results revealed a similar pattern as the behavioral learning, with decreased PE signals for other in the ACC and non-significant change of PE signals for self in the ventral striatum.

General discussion

For all the reviewed studies in Paper I and the two studies in Paper II and III, k-armed bandit tasks (Sutton & Barto, 1998) or their variants were implemented. These tasks represent a simplified version of an environment, in which a learner tries to acquire the contingencies between options and outcomes despite uncertainties, with the goal of maximizing received rewards and/or minimizing received punishments. In a typical two-armed bandit task, participants choose between two options with different reward probabilities, and the payoffs after each choice reinforce the participants' actions to

select the advantageous option in the future. To test specific hypotheses, researchers can modify the structure of the task or add certain components. For example, in the Iowa gambling task (Bechara, Damasio, Damasio, & Anderson, 1994), a variant of the four-armed bandit task, choosing an option will yield gains and losses at the same time, enabling direct testing of the learner's propensity to seek rewards when facing potential punishments. In order to study other-regarding learning, we implemented a social learning task devised by Christopoulos & King-Casas (2015) in Paper II and III, which is a two-armed bandit task with outcomes for the learner and an anonymous partner at the same time. In this task, we could test how participants learned differently for options with different self-other allocations.

As mentioned above, the k-armed bandit tasks can provide useful approaches to studying the mechanisms of reward learning, but RL models are also necessary for further dissecting underlying learning components. The reviewed studies in Paper I applied various forms of RL models to disentangle different cognitive components in the learning process, such as using a model parameter to represent attention to rewards versus punishments in the Iowa Gambling task (Busemeyer & Stout, 2002). In Paper II and III, to explain participants' choices in our social learning task, we also developed a novel RL model that has superior explanatory power to previous other-regarding models (Christopoulos & King-Casas, 2015; Fehr & Schmidt, 1999; Van Lange, 1999). It can take account of both the learning contexts (structure of available reward allocations) and individual differences in social preferences. According to this model, the learner estimates the trial-by-trial discrepancies between his/her preferred reward allocation and the actual self-other allocations shown in the outcomes. The rewards delivered to the learner and others are transformed as a function of both the individual-level preferred allocation and the trial-by-trial discrepancies. This notion is supported by a recent study (Liu et al., 2019), in which participants were found to use their individual preferred allocations as a reference point for rating potential self-other allocations in a reward evaluation task. Using this model, we can disentangle social preferences and other-regarding learning, and uncover the mechanisms of outcome transformation in self- and other-regarding learning.

With the aid of the k-armed bandit tasks and RL models, in the present work, we could identify self- and other-regarding learning deficits in mental disorders. In Paper I, we found inconsistent self-regarding choices across depression, addiction, and anxiety, which was reflected by decreased inverse temperatures in multiple studies. We could also detect unique reward learning deficits in a certain type of disorder. For example, learning rates were found to be decreased for patients in several of the reviewed depression studies, suggesting impaired self-regarding reward learning. Moreover, most of the addiction studies with a parameter measuring the relative attention to rewards versus punishments (Busemeyer & Stout, 2002) showed increased reward bias in patients, reflecting disrupted self-regarding reward perception and addicts' hypersensitivity to drugs despite potential harms (Robinson & Berridge, 2001).

In addition to these behavioral learning deficits, disrupted neural learning signals have also been detected. Specifically, the depression and addiction studies in Paper I showed decreased neural encoding of self-regarding PE signals in key reward regions for the patient groups. But some anxiety studies demonstrated a different neural pattern, with enhanced self-regarding PE signals in anxious patients, suggesting hypersensitivity to unexpected rewards (Cisler et al., 2015; Hauser et al., 2017; Murray et al., 2019). Among these studies, Cisler et al. (2015) detected increased PE signals in the TPJ for PTSD patients when faces were used as self-regarding reinforcers. In Paper II, veterans with PTSD were also found to have heightened other-regarding surprise signals in the IPL. Both the TPJ and IPL are brain regions implicated in processing other-regarding social information (Frith & Frith, 2006; Mar, 2011; Saxe & Kanwisher, 2003; Uddin, Molnar-Szakacs, Zaidel, & Iacoboni, 2006; Van Overwalle & Baetens, 2009), and their heightened responses provide neural evidence for the hypersensitivity to unexpected social rewards (self-regarding and other-regarding) in PTSD. The heightened responses may disrupt the participants' attention to and incorporation of new reward information, which in turn may result in impaired behavioral RL, as evidenced by decreased learning rates in PTSD patients in both Cisler et al. (2015) and Paper II. These results suggest that PTSD might be associated with RL disturbances in both self- and other-regarding social reward learning and are in line with previous PTSD literature showing heightened neural responses to emotional facial stimuli (Armony, Corbo, Clément, & Brunet, 2005; Bryant et al., 2008; Killgore et al., 2013; Rauch et al., 2000).

The k-armed bandit tasks and RL models also helped us to examine OT's effects on self- and other-regarding learning. In Paper III, we demonstrated that OT could attenuate both behavioral and neural other-regarding learning and reduce cooperativeness for individuals with cooperative social preferences. This pattern is not surprising in that previous OT studies have demonstrated that OT's effectiveness is dependent on contexts and individual differences (Bartz et al., 2011). The participants in this study were making decisions for an anonymous partner, and the salience of the anonymity might be increased by OT (Shamay-Tsoory & Abu-Akel, 2016). This could lead to less concern of the participants for their partners' well-being and the decreased other-regarding learning may ensue. The decreased cooperativeness in some participants may also be a result of the increased anonymity. Given that Paper II and some other anxiety studies showed heightened neural responses to unexpected social rewards (Cisler et al., 2015) or emotional facial stimuli (Armony et al., 2005; Bryant et al., 2008; Killgore et al., 2013; Rauch et al., 2000), our finding of the decreased other-PE signals after OT administration provides supportive evidence for OT's therapeutic potential. This is consistent with previous findings that OT could dampen the amygdala's responses to aversive facial stimuli (Domes et al., 2007; Kirsch et al., 2005; Labuschagne et al., 2010). Nonetheless, the detrimental effects of OT on other-regarding reward processing suggested by Paper III should not be ignored, which entails clinicians' caution when using OT as a therapeutic intervention for mental disorders.

Limitations and future directions

Together, this set of studies help to provide new evidence of disrupted self- and other-regarding RL in mental disorders, and reveal behavioral and neural mechanisms of OT's effects on self- and other-regarding RL. In spite of the contributions, some limitations of these studies should also be noted. Firstly, due to the lack of other-regarding RL research on mental disorders, our review (Paper I) only focused on self-regarding RL. Future studies are encouraged to investigate other-regarding learning in mental disorders. Future literature reviews are also encouraged to summarize reward learning disturbances in other types of mental disorders, such as autism, schizophrenia, Parkinson's disease, etc. Secondly, many of the reviewed studies had participants with comorbid disorders and the results may reflect combined effects of more than one mental illness. Similarly, our PTSD study (Paper II) did not include a depression-only group, preventing us from testing PTSD and depression as two separate factors. Future empirical studies are recommended to use factorial designs to investigate the effects of separate disorders as well as their combined effects on RL. Thirdly, the participants in our PTSD study (Paper II) and OT study (Paper III) were predominantly males; therefore, our findings are yet to be replicated in females in future research. Moreover, OT has been found to have mixed effects as a potential therapeutic intervention and may pose detrimental influences on other-regarding learning. Future studies should systematically examine factors that may affect the magnitude and direction of OT's effectiveness for mental disorders. Alternative interventions, such as transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) should also be tested for their effectiveness in normalizing altered behavioral and neural reward learning.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Apps, M. A., Lesage, E., & Ramnani, N. (2015). Vicarious reinforcement learning signals when instructing others. *Journal of Neuroscience, 35*(7), 2904-2913.

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron, 90*(4), 692-707.

Armony, J. L., Corbo, V., Clément, M.-H., & Brunet, A. (2005). Amygdala response in patients with acute PTSD to masked and unmasked emotional facial expressions. *American Journal of Psychiatry, 162*(10), 1961-1963.

Aylward, J., Valton, V., Ahn, W.-Y., Bond, R. L., Dayan, P., Roiser, J. P., & Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nature Human Behaviour, 3*(10), 1116-1123.

Baribeau, D. A., & Anagnostou, E. (2015). Oxytocin and vasopressin: linking pituitary neuropeptides and their receptors to social neurocircuits. *Frontiers in Neuroscience, 9*, 335.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences, 15*(7), 301-309.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170-176.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience, 21*(8), 2793-2798.

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

Bryant, R. A., Kemp, A. H., Felmingham, K. L., Liddell, B., Olivieri, G., Peduto, A., . . . Williams, L. M. (2008). Enhanced amygdala and medial prefrontal activation during nonconscious processing of fear in posttraumatic stress disorder: an fMRI study. *Human Brain Mapping, 29*(5), 517-523.

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, *107*(32), 14431-14436.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253-262.

Chang, S. W., Gariépy, J.-F., & Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience, 16*(2), 243.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817-869.

Christopoulos, G. I., & King-Casas, B. (2015). With you or against you: social orientation dependent learning signals guide actions made for others. *Neuroimage, 104*, 326-335.

Cisler, J. M., Bush, K., Steele, J. S., Lenow, J. K., Smitherman, S., & Kilts, C. D. (2015). Brain and behavioral evidence for altered social learning mechanisms among women with assault-related posttraumatic stress disorder. *Journal of Psychiatric Research, 63*, 75-83.

Clark-Elford, R., Nathan, P. J., Auyeung, B., Voon, V., Sule, A., Müller, U., . . . Baron-Cohen, S. (2014). The effects of oxytocin on social reward learning in humans. *International Journal of Neuropsychopharmacology, 17*(2), 199-209.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision Making, Affect, and Learning: Attention and Performance XXIII* (pp. 3-38). Oxford, UK: Oxford University Press.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408-1411.

de Oliveira, D. C., Zuardi, A. W., Graeff, F. G., Queiroz, R. H., & Crippa, J. A. (2012). Anxiolytic-like effect of oxytocin in the simulated public speaking test. *Journal of Psychopharmacology*, *26*(4), 497-504.

Domes, G., Heinrichs, M., Gläscher, J., Büchel, C., Braus, D. F., & Herpertz, S. C. (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biological Psychiatry*, *62*(10), 1187-1190.

Evans, S., Shergill, S. S., & Averbeck, B. B. (2010). Oxytocin decreases aversion to angry faces in an associative learning task. *Neuropsychopharmacology, 35*(13), 2502.

Fehr, E., & Krajbich, I. (2014). Social preferences and the brain. In *Neuroeconomics* (pp. 193-218). Elsevier.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817-868.

Fladung, A.-K., Grön, G., Grammer, K., Herrnberger, B., Schilly, E., Grasteit, S., . . . von Wietersheim, J. (2009). A neural signature of anorexia nervosa in the ventral striatal reward system. *American Journal of Psychiatry, 167*(2), 206-212.

Forbes, E. E., Hariri, A. R., Martin, S. L., Silk, J. S., Moyles, D. L., Fisher, P. M., . . . Axelson, D. A. (2009). Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *American Journal of Psychiatry, 166*(1), 64-73.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

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

Grinevich, V., Knobloch-Bollmann, H. S., Eliava, M., Busnelli, M., & Chini, B. (2016). Assembling the puzzle: pathways of oxytocin signaling in the brain. *Biological Psychiatry, 79*(3), 155-164.

Hauser, T. U., Iannaccone, R., Dolan, R., Ball, J., Hättenschwiler, J., Drechsler, R., . . . Brem, S. (2017). Increased fronto-striatal reward prediction errors moderate decision

making in obsessive–compulsive disorder. *Psychological Medicine, 47*(7), 1246-1258.

Heinrichs, M., Baumgartner, T., Kirschbaum, C., & Ehlert, U. (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biological Psychiatry*, *54*(12), 1389-1398.

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational dysfunctions in anxiety: Failure to differentiate signal from noise. *Biological Psychiatry, 82*(6), 440-446.

Hung, L. W., Neuner, S., Polepalli, J. S., Beier, K. T., Wright, M., Walsh, J. J., . . . Dölen, G. (2017). Gating of social reward by oxytocin in the ventral tegmental area. *Science, 357*(6358), 1406-1411.

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., . . . Maier, W. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience, 30*(14), 4999-5007.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, 3(12), 1-16.

Ide, J. S., Nedic, S., Wong, K. F., Strey, S. L., Lawson, E. A., Dickerson, B. C., . . . Mujica-Parodi, L. R. (2018). Oxytocin attenuates trust as a subset of more general reinforcement learning, with altered reward circuit functional connectivity in males. *Neuroimage, 174*, 35-43.

Israel, S., Weisel, O., Ebstein, R. P., & Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology, 37*(8), 1341-1344.

Killgore, W. D., Britton, J. C., Schwab, Z. J., Price, L. M., Weiner, M. R., Gold, A. L., . . . Rauch, S. L. (2014). Cortico-limbic responses to masked affective faces across PTSD, panic disorder, and specific phobia. *Depression and Anxiety, 31*(2), 150-159.

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., ... & Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *Journal of Neuroscience*, *25*(49), 11489-11493.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport, 12*(17), 3683-3687.

Koch, S. B., van Zuiden, M., Nawijn, L., Frijling, J. L., Veltman, D. J., & Olff, M. (2014). Intranasal oxytocin as strategy for medication-enhanced psychotherapy of PTSD: Salience processing and fear inhibition processes. *Psychoneuroendocrinology, 40*, 242-256.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676.

Labuschagne, I., Phan, K. L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., ... & Nathan, P. J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology*, 35(12), 2403-2413.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593-622.

Li, J., McClure, S. M., King-Casas, B., & Montague, P. R. (2006). Policy adjustment in a dynamic economic game. *PloS One, 1*(1), 1-11.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., . . . Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience, 22*(4), 633-641.

Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of Neuroscience, 35*(40), 13720-13727.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103-134.

McClintock, C. G. (1972). Social motivation—A set of propositions. *Behavioral Science, 17*(5), 438-454.

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of FMRI. *The Neuroscientist*, *10*(3), 260-268.

McGregor, I. S., & Bowen, M. T. (2012). Breaking the loop: oxytocin as a potential treatment for drug addiction. *Hormones and Behavior, 61*(3), 331-339.

Melis, M. R., Melis, T., Cocco, C., Succu, S., Sanna, F., Pillolla, G., . . . Argiolas, A. (2007). Oxytocin injected into the ventral tegmental area induces penile erection and increases extracellular dopamine in the nucleus accumbens and paraventricular nucleus of the hypothalamus of male rats. *European Journal of Neuroscience, 26*(4), 1026-1035.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4*(1), 1-25.

Murray, G. K., Knolle, F., Ersche, K. D., Craig, K. J., Abbott, S., Shabbir, S. S., . . . Bullmore, E. T. (2019). Dopaminergic drug treatment remediates exaggerated cingulate prediction error responses in obsessive-compulsive disorder. *Psychopharmacology, 236*(8), 2325-2336.

Nietlisbach, G., Maercker, A., Rösler, W., & Haker, H. (2010). Are empathic abilities impaired in posttraumatic stress disorder? *Psychological Reports, 106*(3), 832-844.

Ormel, J., VonKorff, M., Ustun, T. B., Pini, S., Korten, A., & Oldehinkel, T. (1994). Common mental disorders and disability across cultures: results from the who collaborative study on psychological problems in general health care. *JAMA, 272*(22), 1741-1748.

Palgi, S., Klein, E., & Shamay-Tsoory, S. (2017). The role of oxytocin in empathy in PTSD. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(1), 70-75.

Pearce, J. M. (2013). *Animal Learning and Cognition: An Introduction*: Psychology press.

Plana, I., Lavoie, M.-A., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *Journal of Anxiety Disorders, 28*(2), 169-177.

Plichta, M. M., & Scheres, A. (2014). Ventral–striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: A meta-analytic review of the fMRI literature. *Neuroscience and Biobehavioral Reviews, 38*, 125-134.

Rauch, S. L., Whalen, P. J., Shin, L. M., McInerney, S. C., Macklin, M. L., Lasko, N. B., . . . Pitman, R. K. (2000). Exaggerated amygdala response to masked facial stimuli in

posttraumatic stress disorder: a functional MRI study. *Biological Psychiatry, 47*(9), 769-776.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Robinson, T. E., & Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction, 96*(1), 103-114.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences, 30*(5), 203-210.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Shamay-Tsoory, S. G., & Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biological Psychiatry, 79*(3), 194-202.

Sharp, C., Fonagy, P., & Allen, J. G. (2012). Posttraumatic stress disorder: A social-cognitive perspective. *Clinical Psychology: Science and Practice, 19*(3), 229-240.

Slattery, D. A., & Neumann, I. D. (2010). Oxytocin and major depressive disorder: experimental and clinical evidence for links to aetiology and possible treatment. *Pharmaceuticals, 3*(3), 702-724.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology, 43*(2), 476-493.

Stevens, J. S., & Jovanovic, T. (2019). Role of social cognition in post-traumatic stress disorder: A review and meta-analysis. *Genes, Brain and Behavior, 18*(1), e12518.

Succu, S., Sanna, F., Cocco, C., Melis, T., Boi, A., Ferri, G. L., . . . Melis, M. R. (2008). Oxytocin induces penile erection when injected into the ventral tegmental area of male rats: role of nitric oxide and cyclic GMP. *European Journal of Neuroscience, 28*(4), 813-821.

Sul, S., Tobler, P. N., Hein, G., Leiberg, S., Jung, D., Fehr, E., & Kim, H. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National Academy of Sciences, 112*(25), 7851-7856.

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

Uddin, L. Q., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2006). rTMS to the right inferior parietal lobule disrupts self–other discrimination. *Social Cognitive and Affective Neuroscience, 1*(1), 65-71.

Uekermann, J., Channon, S., Lehmkämper, C., Abdel-Hamid, M., Vollmoeller, W., & Daum, I. (2008). Executive function, mentalizing and humor in major depression. *Journal of the International Neuropsychological Society, 14*(1), 55-62.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron, 96*, 348-354.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584.

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. *Journal of Anxiety Disorders, 37*, 71-77.

Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders, 132*(1-2), 104-111.

Zald, D. H., & Treadway, M. T. (2017). Reward processing, neuroeconomics, and psychopathology. *Annual Review of Clinical Psychology, 13*, 471-495.

Zik, J. B., & Roberts, D. L. (2015). The many faces of oxytocin: implications for psychiatry. *Psychiatry Research, 226*(1), 31-37.