Universidade do Minho
Escola de Psicologia

Joana Rita da Silva Carvalheiro

Unveiling the impact of acute stress on reinforcement learning: A computational and neuroimaging approach

maio de 2021

Unveiling the impact of acute stress on reinforcement learning:
A computational and neuroimaging approach

Joana Carvalheiro

UMinho | 2021

FCT Fundação para a Ciência e a Tecnologia
POCH
NORTE2020 PROGRAMA OPERACIONAL REGIONAL DO NORTE
PORTUGAL 2020
UNIÃO EUROPEIA Fundo Social Europeu

Universidade do Minho
Escola de Psicologia

Joana Rita da Silva Carvalheiro

Unveiling the impact of acute stress on reinforcement learning: A computational and neuroimaging approach

Tese de Doutoramento
Doutoramento em Psicologia Básica

Trabalho efetuado sob a orientação da
Doutora Ana Seara-Cardoso
e da
Doutora Ana Raquel Mesquita

maio de 2021

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.
Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.
Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição
CC BY

https://creativecommons.org/licenses/by/4.0/

ACKNOWLEDGEMENTS

I would like to thank Ana Seara-Cardoso and Ana Mesquita for their supervision and support throughout the entire Ph.D. I cannot thank Ana Seara-Cardoso enough for sharing her skills and expertise, in particular for the precious writing tips and for all the patience in guiding me through the complexities of fMRI and, but above all for her infinite encouragement, generosity, trust, and kind words. Thanks to Ana Mesquita for all the thoughtful discussions, for being so enthusiastic of this work, and for the kind words and encouragement. It has been a great privilege to work with these wonderful supervisors, who have offered me unconditional support both academically and personally.

I would also like to thank Vasco Conceição for all the discussions on computational modelling and dopamine, among many other intricate topics. Vasco's help and friendship have been determining for this work, and I had great fun working with him.

Thanks to Professora Carla Martins and Professora Adriana Sampaio, from my Ph.D. committee, for the encouragement and for driving me towards the finish line.

To all the colleagues in the Psychological Neuroscience lab, in particular Diana, Carol, Margarida, Paula, Carla, Marlene, Alberto and Augusto, thank you for helping me selflessly on many occasions, but above all for your friendship.

Thanks to Mäel Lebreton for having me at the University of Geneva/Biotech, for sharing his incredible knowledge on reinforcement-learning models and for all the thoughtful discussions. I would also like to thank to all of my friends at Biotech, Anna, Edo, Nahuel, Mi Xue, Seb and Camille, for their enthusiasm, and for giving me such an awesome experience in Geneva.

Thank you to Ana Magalhães for changing my life in the best way possible by encouraging me to apply for this Ph.D., and for making me a more persistent and confident researcher. I also thank Teresa Summavielle for kindly opening me the doors of the Addiction Biology lab (I3S), and João Relvas for encouraging me to do research and inspiring me to always aim higher.

Thanks to my incredible family for their infinite confidence and support, and to all my friends for always being there and taking me on great adventures.

Finally, I thank Fundação para a Ciência e Tecnologia for their doctoral grant award (PD/BD/128467/2017), co-financed by the European Social Fund through the Human Capital Operational Program, North region.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Investigação do impacto do stress agudo na aprendizagem por reforço: Uma abordagem computacional e neuroimagiológica

SUMÁRIO

A aprendizagem por reforço é um processo adaptativo através do qual os indivíduos aprendem a fazer escolhas que levam a recompensas e a evitar escolhas que levam a punições. Este tipo de aprendizagem ocorre através erros de predição, que são sinalizados por respostas dopaminérgicas fásicas quando uma escolha resulta numa recompensa ou punição diferente do esperado. Apesar do stress ser comum no dia-a-dia, e de se pensar que o stress agudo afeta a aprendizagem por reforço e o funcionamento dopaminérgico, ainda pouco se sabe sobre os mecanismos computacionais e neurais através dos quais o stress agudo interfere com a aprendizagem por reforço. Esta tese teve como objetivo investigar o impacto do stress agudo nos mecanismos computacionais e neurais da aprendizagem por reforço, e como é que diferenças individuais podem modular esse impacto.

Esta tese procurou responder a três perguntas de investigação: 1) Como é que o stress agudo afeta a aprendizagem por reforço ao nível comportamental e computacional? [Capítulo 2] 2) O stress agudo afeta a sinalização neural de erros de predição durante a aprendizagem por reforço? [Capítulo 3]; 3) Diferenças individuais em traços de psicopatia modulam o impacto do stress agudo na aprendizagem por reforço [Capítulo 4] e na sinalização neural de erros de predição? [Capítulo 5].

Os resultados desta tese sugerem que 1) o stress agudo piora o desempenho na escolha de recompensas, mas não de punições, e reduz a taxa de aprendizagem a partir de erros de predição positivos; 2) o stress agudo reduz a sinalização de erros de predição, preferencialmente erros de predição positivos, no estriado dorsal durante a aprendizagem por reforço; 3) diferenças individuais em traços de psicopatia modulam o impacto do stress agudo desempenho na escolha de recompensas e na sinalização de erros de predição no estriado dorsal durante a aprendizagem por reforço. Estes resultados são consistentes com a nossa hipótese neurobiológica de disrupções dopaminérgicas induzidas por stress agudo e contribuem para uma compreensão mais completa e coerente do impacto do stress agudo na aprendizagem por reforço.

Palavras-chave: Dopamina; Erro de predição; Punição; Recompensa

Unveiling the impact of acute stress on reinforcement learning: A computational and neuroimaging approach

ABSTRACT

Reinforcement learning is an adaptive process through which individuals learn to make choices that maximise rewards and minimise punishments. This type of learning is driven by prediction errors, which are signalled by phasic dopamine responses when a choice results in a rewarding or punishing outcome different from the expected. Although acute stress is pervasive in modern everyday life and is thought to affect reinforcement learning and dopaminergic functioning, notably, the computational and neural mechanisms by which acute stress affects reinforcement learning remain poorly understood. This thesis set out to investigate how acute stress impacts the computational and neural mechanisms of reinforcement learning, and how individual differences modulate such impact.

This thesis endeavoured to answer three outstanding research questions: 1) How does acute stress impact reinforcement learning at the behavioural and computational levels? [Chapter 2] 2) Does acute stress affect neural signalling of prediction errors during reinforcement learning? [Chapter 3] 3) Do individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning [Chapters 4] and on neural signalling of prediction errors? [Chapter 5].

The findings of this thesis suggest that 1) acute stress impairs behavioural performance towards obtaining rewards, but not towards avoiding punishments, and this performance impairment is accompanied by diminished learning rate for positive prediction errors; 2) acute stress blunts signalling of prediction errors, preferentially positive prediction errors, in the dorsal striatum during reinforcement learning; 3) individual differences in psychopathic traits modulate the impact of acute stress on behavioural performance towards rewards and signalling of prediction errors in the dorsal striatum during reinforcement learning. These findings are consistent with a neurobiological framework of stress-induced dopamine disruptions and contribute to a more complete and coherent understanding of the impact of acute stress on reinforcement learning.

Keywords: Dopamine; Prediction error; Punishment; Reward

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

STATEMENT OF INTEGRITY iv

SUMÁRIO v

ABSTRACT vi

TABLE OF CONTENTS vii

LIST OF FIGURES xii

LIST OF TABLES xiii

CHAPTER 1: GENERAL INTRODUCTION 1

1.1. Reinforcement learning 2

1.1.1. Theoretical background 2

1.1.2. Q-learning model 3

1.1.3. Neural correlates of reinforcement learning 5

1.1.3.1. The prediction error hypothesis of dopamine 6

1.1.3.2. Neural correlates of prediction error in functional imaging human studies 8

1.2. Acute stress 10

1.2.1. The stress response 10

1.2.2. Acute stress-induced altered dopamine release 11

1.3. Reinforcement learning under acute stress 13

1.3.1. Behavioural evidence 13

1.3.2. Putative neural mechanisms 15

1.4. Individual differences in reinforcement learning under acute stress 17

1.5. Summary and current thesis 18

CHAPTER 2: The impact of acute stress on the computational mechanisms of reinforcement learning 22

2.1. Chapter Introduction 23

2.1.1. The current study 26
2.2. Materials and Methods 28
2.2.1. Participants 28
2.2.2. Reinforcement-learning task 28
2.2.3. Acute-stress manipulation 30
2.2.4. Manipulation check 30
2.2.5. Task-performance analyses 31
2.2.6. Computational modelling 31
2.2.6.1. Reinforcement-learning model 31
2.2.6.2. Model fitting, parameter analyses, and model validation 33
2.3. Results 34
2.3.1. Manipulation check 34
2.3.2. Task performance 35
2.3.3. Computational modelling 38
2.4. Discussion 41
2.5. Limitations 45
2.6. Conclusions 46

CHAPTER 3: The impact of acute stress on striatal signalling of prediction errors during reinforcement learning 47

3.1. Chapter Introduction 48
3.1.1. The current study 49
3.2. Materials and Methods 51
3.2.1. Participants 51
3.2.2. Reinforcement-learning task 51
3.2.3. Acute-stress manipulation 53
3.2.4. Task-performance analyses 53
3.2.5. fMRI data acquisition and preprocessing 54
3.2.6. fMRI data analyses 54
3.2.6.1. Primary general linear model 54
3.2.6.2. Subsidiary general linear models 56
3.3. Results 57

3.3.1. Behavioural analyses 57
3.3.1.1. Manipulation check 57
3.3.1.2. Task performance 57
3.3.2. fMRI analyses 59
3.4. Discussion 63
3.5. Conclusions 67

CHAPTER 4: Individual differences in the impact of acute stress on reinforcement learning: The role of psychopathic traits 68

4.1. Chapter Introduction 69
4.1.1. The current study 70
4.2. Materials and Methods 71
4.2.1. Participants 71
4.2.2. Self-Report Psychopathy-Short Form (SRP-SF) 71
4.2.3. Reinforcement-learning task 72
4.2.4. Acute-stress manipulation 73
4.2.5. Task-performance analyses 73
4.3. Results 74
4.4. Discussion 77
4.5. Limitations 79
4.6. Conclusions 80

CHAPTER 5: Individual differences in the impact of acute stress on striatal signalling of prediction errors during reinforcement learning: The role of psychopathic traits 81

5.1. Chapter Introduction 82
5.1.1. The current study 83
5.2. Materials and Methods 84
5.2.1. Participants 84
5.2.2. Self-Report Psychopathy-Short Form (SRP-SF) 84
5.2.3. Reinforcement-learning task 85

5.2.4. Acute-stress manipulation 86
5.2.5. Behavioural analyses 86
5.2.5.1. Self-reported stress levels 86
5.2.5.2. Task-performance analyses 86
5.2.6. fMRI data analyses 87
5.3. Results 89
5.3.1. Behavioural analyses 89
5.3.1.1. Self-reported stress levels and psychopathic traits 89
5.3.1.2. Task-performance analyses 89
5.3.2. fMRI results 91
5.4. Discussion 95
5.5. Conclusions 99

CHAPTER 6: GENERAL DISCUSSION 100
6.1. Overview 101
6.2. Research questions 101
6.2.1. How does acute stress impact reinforcement learning at the behavioural and computational levels? 102
6.2.2. Does acute stress affect neural signalling of prediction errors during reinforcement learning? 104
6.2.3. Do individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning and neural signalling of prediction errors? 106
6.3. Limitations 108
6.4. Implications and future directions 109
6.5. Conclusions 114

REFERENCES 117
APPENDICES 135
Appendix 1: Ethical approval 136
Appendix 2: Supplementary material for Chapter 2 140
Appendix 3: Supplementary material for Chapter 3 157
Appendix 4: Supplementary material for Chapter 4 173

Appendix 5: Supplementary material for Chapter 5 178
Appendix 6: Published paper 185

LIST OF FIGURES

Figure 1.1. Learning from prediction errors 4

Figure 1.2. Phasic dopamine activity encodes prediction error signals 6

Figure 1.3. Model-based fMRI 8

Figure 1.4. Neurochemical waves of the stress response 11

Figure 1.5. Amphetamine increases spontaneous dopamine responses and simultaneously blunts adaptive responses for reward-predicting cues in the striatum 16

Figure 2.1. Reinforcement-learning task 29

Figure 2.2. Manipulation check and task performance 36

Figure 2.3. Model-fitting results 39

Figure 3.1. Reinforcement-learning task 52

Figure 3.2. Manipulation check and task performance 58

Figure 3.3. Effects of acute stress on prediction error signalling in the dorsal striatum 61

Figure 4.1. Effects of acute stress on task performance across psychopathic traits 75

Figure 5.1. Effects of acute stress on task performance across affective-interpersonal traits 90

Figure 5.2. Effects of acute stress on prediction error signalling in the dorsal striatum across affective-interpersonal traits 93

Figure 6.1. Integrated summary of the main findings 115

LIST OF TABLES

Table 4.1. Descriptive statistics (Mean and Standard deviation, Median and Minimum and Maximum values) of Self-Report Psychopathy-Short Form (SRP-SF) total scores and dimensions (affective-interpersonal and lifestyle-antisocial) 72

Table 5.1. Descriptive statistics (Mean and Standard deviation, Median and Minimum and Maximum values) of Self-Report Psychopathy-Short Form (SRP-SF) total scores and dimensions (affective-interpersonal and lifestyle-antisocial) 85

CHAPTER 1: GENERAL INTRODUCTION

Reinforcement learning, acute stress and individual differences

1.1. Reinforcement learning

Reinforcement learning is ubiquitous and critical for adaptive behaviour and has influentially contributed for our understanding of how learning and behaviour changes from moment to moment. Notably, multiple neurobiological evidence has linked reinforcement learning to specific neural correlates with precise computational roles, allowing the investigation of the neural computations underpinning cognition and behaviour. In the following sections, we introduce reinforcement learning (section 1.1.1), formally describe the standard reinforcement-learning model (section 1.1.2), and review evidence on the neural correlates of reinforcement learning (section 1.1.3).

1.1.1. Theoretical background

A major question in psychology and neuroscience concerns how humans, and other animals, select actions in the face of rewards and punishments, and the mechanisms by which rewards and punishments are used by the brain to optimise such action selection. In behavioural psychology, this question has been thoroughly investigated using instrumental (operant) conditioning paradigms. Instrumental conditioning implicates learning to select actions that maximise the probability of rewarding outcomes and minimise the probability of punishing outcomes (Skinner, 1935; Thorndike, 1911). A very simple example of instrumental learning is a cat running around a cage until the cat happens to press a lever, which allows the cat to receive food (Thorndike, 1911). The food is a great reward, and the cat presses the lever again, and again. Conversely, if the cat receives a very unpleasant shock after pressing the lever, it is less likely that the cat presses the lever again. As illustrated in this example, instrumental learning attempts to optimise the consequences of the selected actions in terms of obtained rewards and avoided punishments. This capacity to select actions that maximise rewards and minimise punishments in an often-changing environment is crucial for adaptive behaviour across animal species. More recently, reinforcement learning (Sutton & Barto, 1998) has provided a powerful computational account for this type of conditioned behaviour and its neurobiological correlates.

According to reinforcement learning, individuals learn to gradually select more and more often the actions that maximise rewards and those that minimise punishments by learning the values of the executed actions. Importantly, such learning is driven by prediction errors (Schultz et al., 1997; Sutton & Barto, 1998). Prediction errors — which signal the difference between obtained and expected outcomes — are used to progressively update the values of the executed actions (Maia & Frank, 2011; Schultz et al., 1997; Sutton & Barto, 1998). A positive prediction error occurs when the outcome of an

action is better than expected, increasing the likelihood of repeating that action. A negative prediction error occurs when the outcome of an action is worse than expected, decreasing the likelihood of repeating that action. In the above example, when the cat happens to press the lever and unexpectedly obtains food, a positive prediction error occurs, increasing the value of pressing the lever and therefore it is more likely that the cat presses the lever again; on the other hand, when the cat presses the lever but receives a shock instead, a negative prediction error occurs, decreasing the value of pressing the lever, and thus it is less likely that the cat presses the lever again. Thus, reinforcement learning is an adaptive process which involves using previous experience to improve the outcomes of future actions.

Reinforcement learning provides a biologically and mathematically grounded framework that allows generation of predictions about the computations and neurobiological mechanisms that drive behaviour (Maia & Frank, 2011). By applying mathematical models to behavioural data, reinforcement-learning models differ from (and complement) descriptive models, which only describe observable behaviour, allowing the generation of detailed quantitative hypotheses to test cognitive processes that cannot be directly observed. In other words, reinforcement learning makes clear predictions about learning — an inherently unobservable process — and its underlying hidden variables, such as prediction errors. As different learning processes can lead to similar behaviour, by quantifying these hidden variables, reinforcement-learning models allow a more detailed and dynamic picture of learning.

In the next section, we describe a standard computational algorithm that formalises reinforcement learning.

1.1.2. Q-learning model

Computational modelling involves specifying an algorithm which formalises a cognitive and/or neural process of interest. Here, we describe the standard reinforcement-learning model — Q-learning model (Sutton & Barto, 1998) — which has been robustly supported by numerous behavioural and neurobiological evidence (Maia & Frank, 2011).

Q learning is a mathematical model that simulates reinforcement-based action selection (Sutton & Barto, 1998). The foundational idea of the Q-learning model, based on the Rescorla-Wagner learning rule (Rescorla & Wagner, 1972), is to quantify how error-driven learning occurs after receiving an outcome. That is, the value of a choice option is updated by the difference between the obtained outcome and the expected outcome of that choice. In reinforcement learning, such difference is termed as the prediction error (Figure 1.1). Specifically, the expected value, Q, of the chosen option (say A) is updated every trial, t, according to:

$$Q_A(t+1) = Q_A(t) + \alpha * \delta(t),$$

where δ is the prediction error:

$$\delta(t) = r(t) - Q_A(t),$$

which represents the difference between the obtained outcome, r, and the expected outcome, Q. This prediction error is then used to update the expected value of the choice option for the next trial, $Q(t+1)$, scaled by a free parameter called learning rate, α. The learning rate thus quantifies the extent to which the prediction error updates the current expected value. Higher learning rates correspond to stronger weights of the prediction error for the value update and to faster learning.

In reinforcement-learning models, after values are updated by the prediction error, the next step is to use those values on the next trial and select a new action. This value-action connection is formulated by the softmax rule (Sutton & Barto, 1998). Specifically, when choosing between two options, say A over B, the softmax rule converts action values [Q(A) and Q(B)] to action probabilities [P(A) and P(B)] on a trial-by-trial basis (the higher the value of A on a given trial, the more likely A will be selected) as following:

$$P_A(t) = \frac{e^{[Q_A(t)*\beta]}}{e^{[Q_A(t)*\beta]} + e^{[Q_B(t)*\beta]}},$$

where β is a free parameter known in reinforcement learning as the inverse temperature parameter (Sutton & Barto, 1998). The inverse temperature parameter β controls the amount of noisiness in action selection. A null β leads to random behaviour (i.e., similar action probabilities irrespective of the corresponding action values), whereas the higher the β the more deterministic is the behaviour (i.e., the action with the higher expected value is consistently selected).

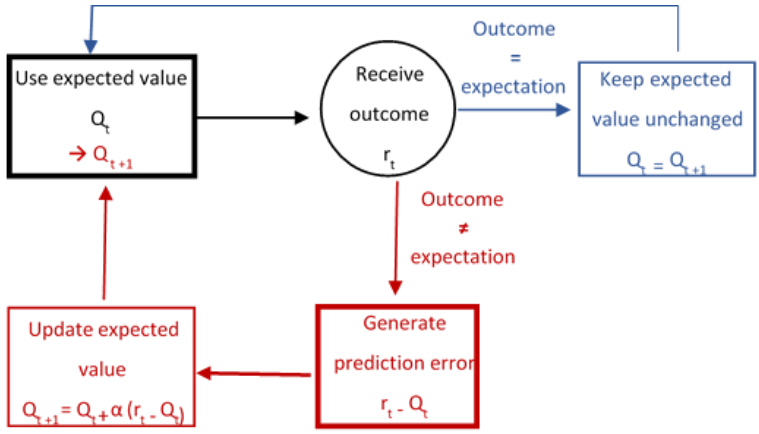


Figure 1.1. Learning from prediction errors. A prediction error occurs when the outcome of an action (reward/punishment) differs from its expected value. A prediction error leads to update of the

expected value, driving behavioural change (red). In contrast, no prediction error is generated when the outcome matches the expected value, and behaviour remains unchanged (blue). Q is expected value, r is obtained outcome, α is learning rate, and t is trial. Adapted from Schultz (2015) with permission of the copyright owner.

Reinforcement-learning models typically have a number of free parameters. In the Q-learning model described above, the free parameters are the learning rate, α, and the inverse temperature parameter, β. Often these parameters characterize factors that are of experimental interest. For instance, the learning rate is considered as the step size of learning, or the speed of learning (Sutton & Barto, 1998). Parameter estimation involves finding the set of parameter values that best account for the observed data in each subject (Daw, 2011; Wilson & Collins, 2019). As a result, one can test predictions about how individual differences or interventions, such as drugs, illness, lesions, or other experimental manipulations, affect those parameters.

The Q-learning model described above corresponds to the standard algorithm used to study reinforcement learning, but reinforcement learning encompasses an entire family of models that assume different hypotheses about the learning and neural process that generate the data, and most models found in the literature are variants of this standard model. Critically, a recommended strategy is to build and implement algorithmic models including parameters that reflect distinct neural mechanisms (Maia & Frank, 2011). For instance, a variant of this Q-learning model including distinct learning rates for positive and negative prediction errors has been widely used to incorporate neurobiological evidence that these two types of learning rely on distinct mechanisms (Frank et al., 2007), as we will address in the next section.

1.1.3. Neural correlates of reinforcement learning

In recent years, reinforcement-learning models, and in particular the standard Q-learning model described above, have been applied to a wide range of neurobiological data (Maia & Frank, 2011; Montague et al., 2006). There is now wealth multidirectional empirical evidence that links the prediction error — the key computation that drives reinforcement learning — to specific neural substrates (Daw & Tobler, 2014; Glimcher, 2011; Pessiglione et al., 2006; Schultz et al., 1997). In the next sections we address how multiple lines of neurobiological evidence support the existence of this key reinforcement learning signal in the brain.

1.1.3.1. The prediction error hypothesis of dopamine

Phasic activity of dopamine neurons is currently thought to encode the prediction error signal that drives reinforcement learning (Daw & Tobler, 2014; Glimcher, 2011; Schultz et al., 1997). Electrophysiological recordings in midbrain dopamine neurons of monkeys during reward-based learning tasks, have convincingly demonstrated that phasic dopamine activity encodes the difference between obtained reward and expected reward — prediction error — which concurs with the basic assumptions of reinforcement learning (Bayer & Glimcher, 2005; Hollerman & Schultz, 1998; Schultz et al., 1997; Tobler et al., 2003, 2005). Specifically, these dopaminergic neurons show a phasic increase in activity, or burst firing, when an unexpected reward is obtained, or when an obtained reward is better than expected, which corresponds to a positive prediction error (Figure 1.2a); the magnitude of this phasic activity correlates with the size of the unexpected reward (Fiorillo et al., 2003; Tobler et al., 2005). Note that prediction errors occur only when outcomes are unexpected, which explains why dopaminergic neurons respond with phasic burst firing to rewards early in training – when rewards are unexpected — but not later in training, after the animal has learned to expect those rewards (Figure 1.2b) (Schultz et al., 1997). Moreover, when an expected reward is omitted, dopamine neurons pause their firing, i.e., a dopamine dip occurs, at the precise moment the reward is expected, which corresponds to a negative prediction error (Figure 1.2c) (Schultz et al., 1997; Tobler et al., 2003).

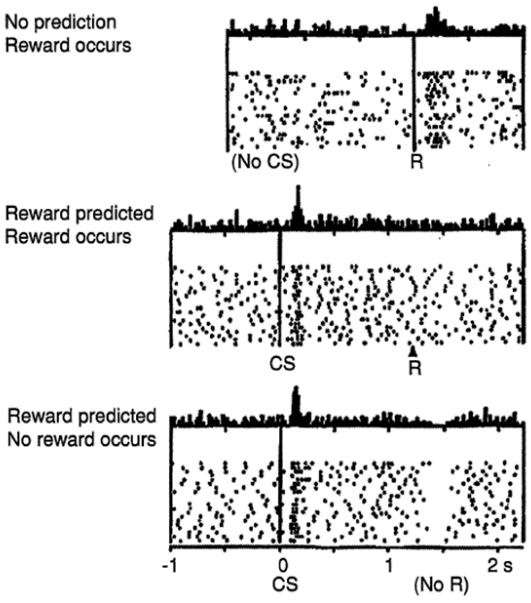


Figure 1.2. Phasic dopamine activity encodes prediction error signals. Firing patterns of dopaminergic midbrain neurons in monkeys performing an instrumental conditioning task. Each plot

shows action potentials (dots) with different rows representing different trials, aligned on the time of the conditioned stimulus (CS) or the reward (R). Histograms show activity summed over the trials plotted below. **(a)** When a reward is unexpectedly obtained a positive prediction error occurs and dopaminergic neurons respond with a phasic burst of firing. **(b)** After learning, the conditioned stimulus predicts reward, and the reward occurs as expected, hence no error in the prediction of reward, and the reward no longer elicits a burst of activity. The phasic burst now occurs at the presentation of the reward-predictive cue. **(c)** When the reward is unexpectedly omitted, dopaminergic neurons showed a precisely timed pause, or dip, in firing, below their baseline firing rate. Reproduced from Schultz et al. (1997), reprinted with permission from AAAS.

Together, these data suggest that dopamine neurons signal the difference between obtained and expected reward: when a reward is better than expected (positive prediction error), dopamine neurons burst fire; when a reward is the same as expected (no prediction error), there is no response; and when reward is worse than expected (negative prediction error), dopamine activity is suppressed. This pattern of dopaminergic responses conforms precisely to the behaviour of the prediction error, formally described in the previous section.

Although the implication of phasic dopaminergic responses in reward learning is well established, the involvement of phasic dopamine activity in punishment learning is not completely resolved yet (Glimcher, 2011; M. Matsumoto & Hikosaka, 2009; Mirenowicz & Schultz, 1996; Palminteri & Pessiglione, 2017; Ungless et al., 2004). Still, there is evidence that omission of an expected punishment, which generates a positive prediction error, might be encoded by burst firing of dopamine neurons; and that the receipt of an unexpected punishment, which generates a negative prediction error, is thought to be signalled by suppression of dopamine activity (H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009; Oleson et al., 2012; Salinas-Hernández et al., 2018; Ungless et al., 2004; Watabe-Uchida et al., 2017). Moreover, studies in rodents using optogenetics, which can manipulate dopamine neurons with temporal and genetic precision to assess their causal effect of learning, have shown that activation or suppression of dopamine neurons can reinforce an action positively or negatively, respectively (Chang et al., 2016; Danjo et al., 2014; Steinberg et al., 2013; Tan et al., 2012; Tsai et al., 2009; Witten et al., 2011). Together, these results suggest that changes in phasic dopamine release may encode the full range of prediction errors necessary for reinforcement learning.

Finally, besides the electrophysiological evidence for the prediction error hypothesis in midbrain dopamine neurons, further studies in rodents have confirmed that phasic changes of dopamine levels in midbrain target brain areas — particularly, in the striatum — also match a prediction error signal (Day et al., 2007; Knutson et al., 2009; Oyama et al., 2010).

In sum, convincing evidence from multiple lines of research indicates that dopaminergic activity can drive reinforcement learning through signalling of prediction error. Although there are alternative theories about the role of dopamine in conditioned behaviour (e.g., Berridge, 2007; Redgrave & Gurney, 2006), to date no alternative has convened as precise and multidirectional empirical evidence as this prediction error theory of dopamine.

1.1.3.2. Neural correlates of prediction error in functional imaging human studies

The prediction error hypothesis of dopamine has motivated a growing number of functional magnetic resonance imaging (fMRI) studies of reinforcement learning in humans (Montague et al., 2006; Wang et al., 2016). The combination of computational models with fMRI — also known as model-based fMRI approaches — provides a powerful framework for testing precise hypotheses about hidden variables that drive reinforcement learning, such as prediction errors, and search for their correlates in the brain (Figure 1.3) (Gläscher & O'Doherty, 2010; O'Doherty et al., 2007).

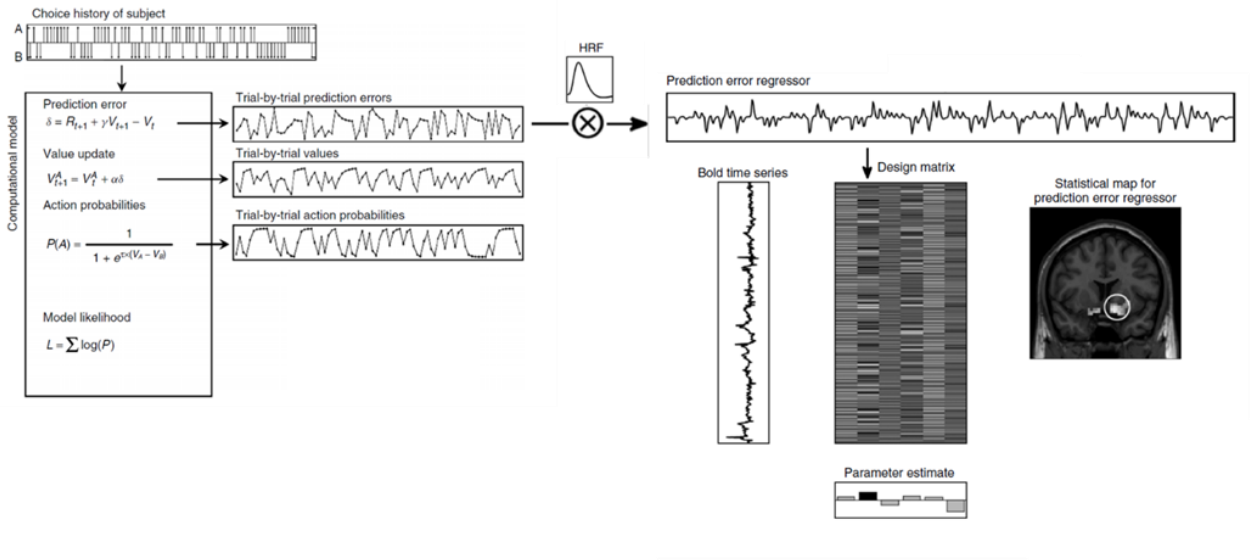


Figure 1.3. Model-based fMRI. An example of a reinforcement-learning model used in combination with fMRI data. **(Left)** This reinforcement-learning model has two free parameters, α and τ (temperature parameter), which are estimated by searching for the set of parameter values that maximise the likelihood of the observed sequence of trial-by-trial choices, given the previously observed outcomes, for each subject. By

fitting the reinforcement-learning to the behavioural data, we can obtain the prediction error, δ, and the expected value of the two actions V(A) and V(B), along with the softmax transformed action probabilities P(A) and P(B), respectively. **(Right)** These hidden variables derived from the model [e.g., prediction errors, modelled at the time of the outcome delivery of each trial] are converted into a time series and convolved with a hemodynamic response function resulting in a parametric modulator regressor in a single-subject fMRI design matrix. Subsequent statistical contrasts for the parameter estimates of the newly created regressor yield a statistical map representing the correlation between activity in a particular BOLD time series voxel and the hidden variable of interest (in this case the prediction error). Adapted from Gläscher & O'Doherty (2010) with permission of the copyright owner.

Specifically, using reinforcement-learning models, it is possible to quantitatively specify the dynamics of prediction error on a moment-to-moment basis during learning tasks, and identify brain regions exhibiting significant correlations with that prediction error signal. A major advantage of model-based fMRI approach compared with more traditional neuroimaging approaches is that it accounts for how a particular cognitive process is implemented in a specific brain area rather than merely determining the location where that process occurs.

Pioneering model-based fMRI studies that searched for prediction errors in the human brain implicated the dorsal striatum (putamen) as well as the ventral striatum (nucleus accumbens), both major dopaminergic targets (McClure et al., 2003; O'Doherty et al., 2003). O'Doherty et al. (2004) then showed that neural correlates of prediction error signals can be dissociated in the dorsal and ventral striatum depending on whether action selection is required to obtain rewards (i.e, instrumental conditioning) or not (pavlovian conditioning). Indeed, numerous studies have now demonstrated correlates of prediction errors in the dorsal and ventral striatum (e.g., Garrison et al., 2013; Pessiglione et al., 2006; Schönberg et al., 2007). It has also been shown that BOLD response of prediction error signals in the dorsal striatum correlates with reward learning efficacy, supporting the idea that striatal prediction error signals are related to behavioural performance in reward-related instrumental learning tasks (Schönberg et al., 2007). Additionally, there is evidence that the human striatum signals not only positive and negative prediction errors during reward learning, but also during punishment learning (Pessiglione et al., 2006; Seymour et al., 2004, 2007).

Although other brain areas have been linked with signalling of prediction errors (Garrison et al., 2013), the striatum is a major midbrain target of dopaminergic influence, and growing evidence from model-based fMRI has directly supported the involvement of striatal dopaminergic system in prediction

error signals (Chowdhury et al., 2013; Diederen et al., 2017; Pessiglione et al., 2006; Schönberg et al., 2010). For example, pharmacological manipulations of the dopamine system (e.g., administration of dopamine agonists or antagonists) during a reinforcement-learning task have been shown to modulate the striatal BOLD response of positive and negative prediction error signals during reward learning, driving altered behavioural performance towards rewards (Pessiglione et al., 2006).

In sum, multiple lines of evidence have linked the key computation of reinforcement learning — the prediction error — to the function of dopaminergic neurons in non-human animals. Importantly, extant neuroimaging evidence indicates that dopamine prediction error signals are present in the human brain – particularly in the striatum – and that these signals conform to a response profile consistent with reinforcement-learning models. Such evidence has contributed for major advances in behavioural neuroscience by supporting the use of reinforcement-learning models to more precisely understand how behaviour and its underpinning neurobiological mechanisms are affected by pharmacological challenges, genetic predispositions, disease, or other experimental manipulations.

1.2. Acute stress

Acute stress is present in day-to-day life and can have a great impact on human functioning. In the following sections, we address the distinct temporal patterns of the stress response in the brain (section 1.2.1), and a putative neural mechanism involved in the acute stress response (section 1.2.2).

1.2.1. The stress response

Stress can be defined as a 'subjective state of sensing potentially adverse changes in the environment' (Joëls & Baram, 2009). Stressors can elicit many physiological and functional changes in the brain in response to perceived changes in the environment (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009). The stress response has classically been characterized by two temporal 'waves' of neurochemical changes (Figure 1.4) (Sapolsky et al., 2000).

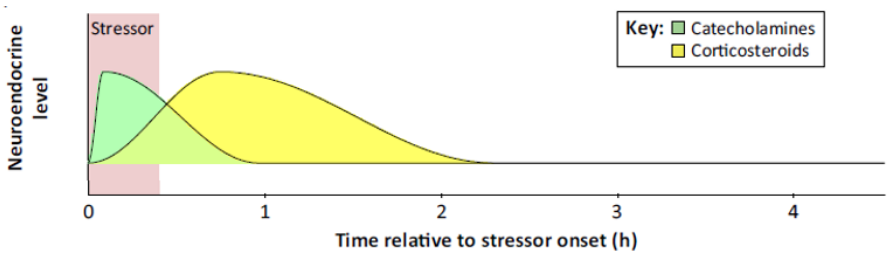


Figure 1.4. Neurochemical waves of the stress response. Following the onset of a stressor, levels of catecholamines (e.g., dopamine and noradrenaline) increase rapidly and normalise not long after stressor offset. Corticosteroid levels in the brain increase more slowly and remain higher for a longer period. Adapted from Hermans et al. (2014) with permission of the copyright owner.

The first wave occurs rapidly and involves prompt increase in the levels of catecholamines, such as dopamine and noradrenaline, which normalise not long after the stressor offset. These rapid and short-lived actions of catecholamines are quickly translated into changes in neural functioning and are thought to promote vigilance and alertness, and to influence cognitive functioning (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009; Sapolsky et al., 2000). A second, and slower, wave involves secretion of glucocorticoids. Cortisol typically has slow, genomic, prolonged actions on neural functioning, although cortisol can also have rapid non-genomic effects that may overlap and interact with catecholaminergic effects in the first wave (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009; Sapolsky et al., 2000). Peak concentrations of cortisol are usually not reached before 15-30 minutes after the stressor onset (de Kloet et al., 2005; Hermans et al., 2014), which suggests a limited role of cortisol in the immediate stress response. Thus, although the stress response is highly preserved across species, it can be greatly influenced by factors, such as the duration of the stressor. Acute stressors occur for short periods of times and cause a rapid release of catecholamines followed by rapid return to baseline levels, whereas chronic stressors occur for longer periods of time and often elicit sustained and/or progressive changes in genomic expression and in brain structure and function (Joëls & Baram, 2009). In addition, individual characteristics, such as age or sex, can determine the pattern and magnitude of the stress response (Joëls & Baram, 2009).

1.2.2. Acute stress-induced altered dopamine release

Several lines of evidence suggest that acute stress enhances striatal dopaminergic functioning. Specifically, studies in humans, using positron emission tomography (PET) (Adler et al., 2000; Booij et

al., 2016; Nagano-Saito et al., 2013; Pruessner et al., 2004; Vaessen et al., 2015), and in rodents, using microdialysis (Abercrombie et al., 1989; Cabib & Puglisi-Allegra, 2012; Puglisi-Allegra et al., 1991), have demonstrated that acute stress increases dopamine release in the striatum. However, PET and microdialysis studies cannot detect the precise mechanism behind increased dopamine release.

Dopamine neurons fire tonically and phasically, leading to tonic dopamine concentrations and spikes in those concentrations called bursts, respectively. Specifically, most dopamine neurons are spontaneously active with firing patterns that range from regular tonic firing to spontaneous burst firing (Grace, 1988; Grace & Bunney, 1984; Sulzer et al., 2016) — bursts that are not time-locked to identified stimuli or events. In addition, unexpected rewarding events, or positive prediction errors, induce brief adaptative burst firing, whereas unexpected adverse events, or negative prediction errors, tend to adaptatively decrease firing activity (cf. section 1.1.3.1). This means that dopamine can be phasically released in a spontaneous, aberrant manner, and that phasic dopamine responses are not always implicated in adaptative signalling of prediction errors (Belujon & Grace, 2015; Covey et al., 2014; Maia & Frank, 2017; Sulzer et al., 2016; Wightman et al., 2007). PET's poor temporal resolution cannot disentangle whether acute stress enhances dopaminergic functioning by increasing tonic dopamine release and/or burst firing over sustained periods. In addition, it has also been debated whether extracellular dopamine levels measured by microdialysis can be increased by burst firing (Wightman et al., 2007). Probing whether acute stress increases burst firing is important due to its role in reinforcement learning.

To detect whether acute stress induces fluctuations in phasic dopamine activity, techniques, such as electrophysiology and voltammetry, have been used in awake (Anstrom et al., 2009; Anstrom & Woodward, 2005) or anesthetized (Valenti et al., 2011) rodents. In two studies, the authors induced acute stress in rodents using restraint stress (Anstrom & Woodward, 2005) and social confrontation (Anstrom et al., 2009), and measured fluctuations in phasic dopamine activity in the mesolimbic pathway using electrophysiology and voltammetry during stress exposure. Interestingly, these studies demonstrated that acute stress was associated with increased spontaneous burst firing of dopamine neurons. These data suggest that striatal dopamine neurons are particularly sensitive to acute stress, to which they respond by increasing burst firing, possibly contributing for the increased extracellular dopamine levels found in PET and microdialysis studies.

In sum, cumulative evidence from human and non-human studies suggests that acute stress increases striatal dopamine release and, particularly, spontaneous phasic dopamine release.

1.3. Reinforcement learning under acute stress

Learning to select actions that lead to rewards and avoid punishments is crucial for adaptive behaviour. According to reinforcement learning, prediction errors drive optimal action selection by gradually updating the value of those actions based on the obtained outcomes. Dopamine plays a key role in prediction error signalling. Curiously, the dopaminergic system is also particularly sensitive to acute stress, raising the question of whether acute stress affects reinforcement learning by disrupting dopaminergic signalling of prediction errors. In this section, we review behavioural evidence concerning the impact of acute stress on reinforcement learning (section 1.3.1) and propose a neurobiological framework of stress-induced dopaminergic disruptions on prediction error signals (section 1.3.2).

1.3.1. Behavioural evidence

Given the pervasiveness of stress in everyday life and its major impact on health and well-being, a growing number of studies have investigated the impact of acute stress on reinforcement learning. Studies that induce acute stress during reinforcement-learning tasks are particularly valuable to assess the effects of acute stress on how individuals learn to obtain rewards and avoid punishments. In these studies, the stressor is experienced within the learning context and the stress-induced emotional state is concurrent with the cognitive processes that operate during the task. The most used paradigm to induce acute stress during a learning task is threat of shock (Robinson et al., 2013), which usually consists of instructing participants that they will receive shocks anytime during the learning task; this is alternated with a control condition, in which participants do not receive any shocks.

In two previous studies (Bogdan et al., 2011; Bogdan & Pizzagalli, 2006), the authors assessed reward learning by indexing response bias (i.e., the systematic preference for the response paired with the more frequent reward) on a signal detection task incorporating probabilistic monetary rewards under threat of shock and under control conditions. Both studies indicated that response bias was reduced in the treat-of-shock condition, suggesting impaired reward learning under acute stress. Moreover, Berghorst et al. (2013) used a probabilistic stimulus selection task to assess whether threat of shock affected learning from rewarding and/or punishing outcomes. The authors found that threat of shock decreased the probability of approaching a learned reward, but did not affect avoidance of learned punishments, particularly in stress-reactive individuals (i.e., individuals who showed higher cortisol responses in the threat-of-shock condition). More recently, Paret et al. (2020) showed that threat of shock impaired reward learning during a reversal learning task. Collectively, these studies

suggest that individuals perform more poorly when learning to maximise rewards whilst exposed to an acute stressor.

However, a number of studies have also investigated the effects of stress on learning adopting very different paradigms from threat of shock, such as the Trier social stress test or the (socially evaluated) cold-pressor test, which involve relatively long latencies between the induction of stress and the behavioural learning task (McRae et al., 2006). Specifically, the Trier social stress test classically requires participants to make an interview-style presentation, or a public speech, that can be followed by a surprise mental arithmetic test, in front of an interview panel who do not provide encouragement (Allen et al., 2017; Kirschbaum et al., 1993); the cold-pressor test consists of a physical stressor that requires participants to immerse their hand into an ice water container, combined, or not, with a social evaluation component (socially evaluated cold-pressor test) (Bullinger et al., 1984; Schwabe et al., 2008). These stress manipulations are not compatible with simultaneous performance of the learning task, and the impact of stress on learning is usually assessed only around 10-20 minutes after the stress manipulation has occurred, so that cortisol levels peak during the learning task (Byrne et al., 2019; Cremer et al., 2021; de Berker et al., 2016; Ehlers & Todd, 2017; Kruse et al., 2018; Lighthall et al., 2013; Otto et al., 2013; Petzold et al., 2010; Schwabe & Wolf, 2009). However, it is not uncommon to find decreasing cortisol levels during task performance, probably due to suppressive actions of glucocorticoids implicated in acute stress recovery (Sapolsky et al., 2000). Moreover, it is dubious whether participants are actually perceiving stress at the time of learning, as the stressor is experienced outside of the learning context (Schwabe et al., 2012). Overall, studies that have induced acute stress using such paradigms before a reinforcement-learning task, show more mixed results than those that have induced acute stress concurrently with the task. Whereas some of studies that induced acute stress before the learning task indicate impaired reward learning after exposure to an acute stressor (Cremer et al., 2021; de Berker et al., 2016; Morris & Rottenberg, 2015), others suggest enhanced reward learning (Byrne et al., 2019; Lighthall et al., 2013). It is possible that methodological differences in the timing and type of stress induction, as well as in the used reinforcement-learning tasks, contribute for these discrepant results (Porcelli & Delgado, 2017).

Despite the discrepancies described above, extant evidence mostly points to a negative effect of acute stress on reward learning. Evidence for an effect of acute stress on punishment learning is scarcer. To our knowledge, studies that induced acute stress concurrently with the learning task, using threat of shock, did not provide evidence for stress-induced effects on punishment learning relative to a control condition (Aylward et al., 2019; Berghorst et al., 2013; Binti Affandi et al., 2021). Only two

studies, in which stress was induced before the learning task, indicated impairments in punishment learning (de Berker et al., 2016; Petzold et al., 2010), but possibly due to a stress-induced general deficit in action-learning, which affects both punishment and reward learning, rather than a valence-specific deficit on punishment learning (de Berker et al., 2016).

In sum, extant evidence suggests that acute stress affects reinforcement learning — acute stress seems to impair learning to obtain rewards and, possibly, learning to avoid punishments.

1.3.2. Putative neural mechanisms

One putative mechanism behind the impact of acute stress on reinforcement learning is that acute stress disrupts striatal dopaminergic functioning, which plays a key role in prediction error signalling. Specifically, acute stress increases spontaneous phasic dopamine release — phasic responses that do not occur in response to identified stimuli or events. Such exaggerated, aberrant spontaneous dopamine release is thought to blunt adaptive phasic dopamine responses that signal unexpected events, or prediction errors (Grace, 2016; Maia & Frank, 2017). It has been speculated that spontaneous phasic dopamine release leads to background fluctuations in dopamine levels, which makes adaptative phasic dopamine responses that signal prediction errors less easily differentiated, resulting in a low signal-to-noise ratio (Sulzer et al., 2016). Indeed, there is empirical evidence for a link between exaggerated spontaneous phasic dopamine release and blunted adaptive phasic dopamine responses (Daberkow et al., 2013; Werlen et al., 2020). Data from rodent studies indicate that amphetamine administration increases spontaneous phasic dopamine release and blunts adaptative phasic dopamine release that signal positive prediction errors in the striatum in response to a reward-predicting cue (Figure 1.5), leading to impaired behavioural responses towards rewards.

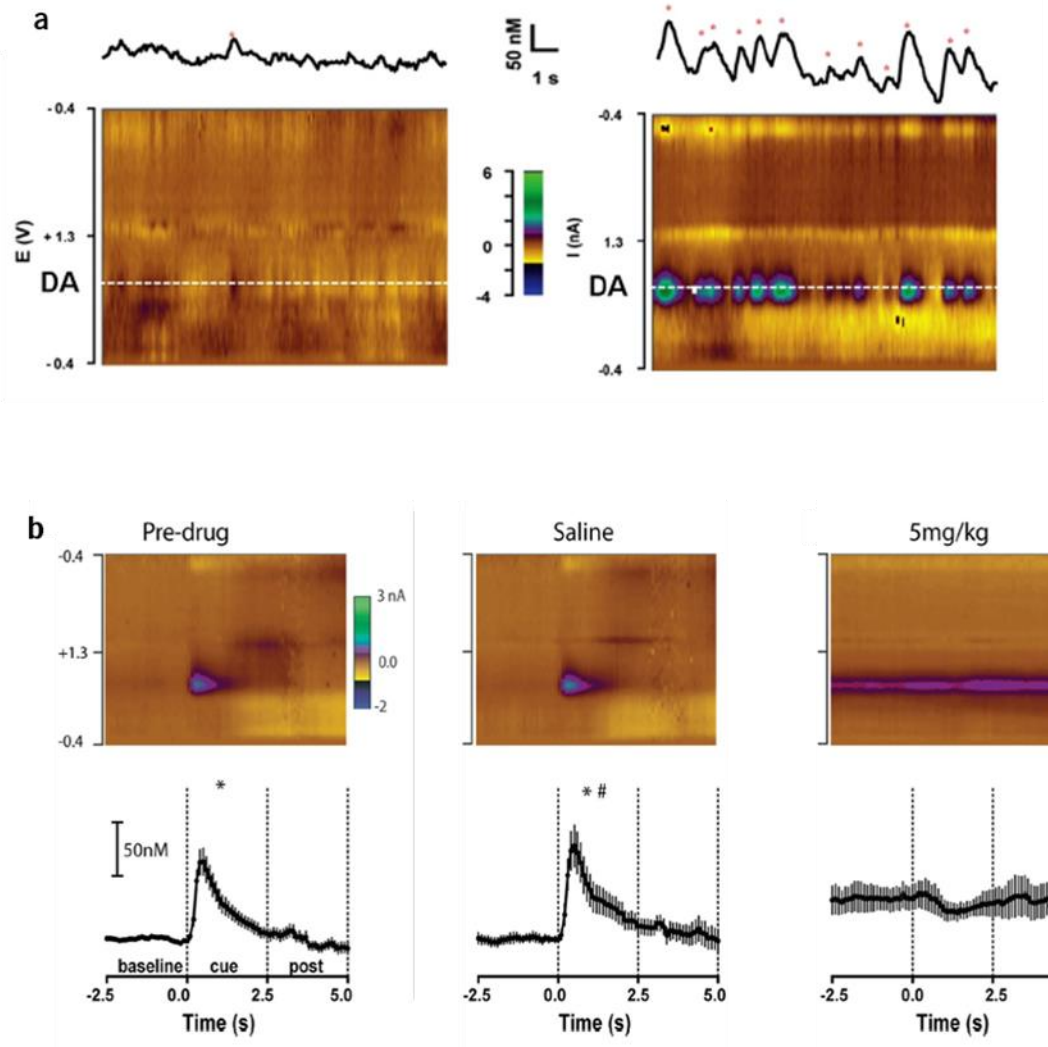


Figure 1.5. Amphetamine increases spontaneous dopamine responses and simultaneously blunts adaptive responses for reward-predicting cues in the striatum. **(a)** Amphetamine administration (right) increases the number of spontaneous phasic dopamine responses (red asterisks) relative to the predrug state (left). **(b)** A reward-predicting cue (time 0) elicits a phasic response in the predrug and saline conditions (two panels on the left), but amphetamine blunts this adaptative phasic dopamine response (right panel). Adapted from Daberkow et al. (2013) with permission of the copyright owner.

The current evidence for a relation between increased spontaneous dopamine release and blunted positive prediction errors seems convincing, and such evidence has even been applied to the interpretation of experimental findings in psychiatric disorders, such as schizophrenia (Maia & Frank, 2017). Additionally, it is also possible that aberrant spontaneous dopamine release impairs the detection of phasic dopamine dips implicated in negative prediction errors (Maia & Frank, 2017). Thus,

we propose that stress-induced spontaneous phasic dopamine release may blunt positive prediction errors that signal unexpected rewards, resulting in impaired reward learning. Additionally, but more tentatively, stress-induced spontaneous dopamine release might also blunt negative prediction errors that signal unexpected punishments, leading to deficits in punishment learning. However, direct evidence of stress blunting prediction error signals in the human brain during reinforcement learning is still very scarce (Cremer et al., 2021; Treadway et al., 2017).

The proposed neurobiological framework of stress-induced dopaminergic disruptions can account for behavioural data which indicate that acute stress impairs reward learning, and possibly punishment learning. Notably, the computational and neural mechanisms by which acute stress affects reinforcement learning remain poorly understood.

1.4. Individual differences in reinforcement learning under acute stress

Humans show great variability when performing learning tasks (Cohen, 2008; Patzelt et al., 2018; Schad et al., 2020; Schönberg et al., 2007). Reinforcement-learning models provide a particularly useful framework to investigate those individual differences (Cohen, 2008; Patzelt et al., 2018; Wuensch et al., 2021). Although reinforcement learning assumes that learning the value of actions is underpinned by the same computational mechanisms across humans and across species, by including free parameters, such as the learning rate, reinforcement-learning models allow those parameters to vary across subjects and can thus account for important interindividual variability (Cohen, 2008; Lebreton et al., 2019).

Reinforcement learning is thought to be affected by acute stressors (Porcelli & Delgado, 2017), but likely not to the same extent between individuals, which might be related to individual differences in personality traits (Edelstein et al., 2010; Pruessner et al., 1997; Soliemanifar et al., 2018; Suridjan et al., 2012). Personality traits, such as psychopathic traits, have been associated with altered reinforcement-learning processes (R. J. R. Blair, 2013a) and stress responsivity (Thompson et al., 2014). Therefore, psychopathic traits might constitute a source of individual differences that modulate the extent to which reinforcement learning is affected by acute stress.

Psychopathy is a personality construct that involves a constellation of personality and behavioural traits, such as lack of empathy, manipulativeness, impulsiveness and antisocial behaviour (R. J. R. Blair et al., 2005; Hare, 2003; Hare & Neumann, 2008). Psychopathic traits have been associated with cognitive deficits, which may ultimately result in maladaptive behaviour (R. J. R. Blair, 2019). Indeed, empirical evidence from clinical and community populations indicates that psychopathy

is associated with difficulties in reinforcement learning, particularly when this is indexed through passive avoidance (R. J. R. Blair et al., 2004; De Brito et al., 2013; Finger et al., 2011; Newman & Kosson, 1986; Newman & Schmitt, 1998) and reversal learning (Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Gregory et al., 2015; Mitchell et al., 2002) paradigms. Specifically, individuals with psychopathy show difficulties in passive avoidance tasks that require the formation of stimulus–punishment associations and in response reversal tasks, which requires flexible adjustment of behaviour when the previously learned reward and/or punishment-related contingencies are reversed. This could partially explain why individuals with psychopathy often engage in behaviours that resulted in rewards in the past, despite the other possible negative outcomes of those same behaviours. Furthermore, psychopathic traits have been associated with blunted stress responsivity (Cima et al., 2008; Holi et al., 2006; House & Milligan, 1976; Johnson et al., 2015; Lidberg et al., 1978; Loney et al., 2006; Thompson et al., 2014). The influential low-fear hypothesis (Lykken, 1957) proposes that psychopathy is linked to a reduced capacity to experience fear and to deficits in automatic threat detection and responsivity (Hoppenbrouwers et al., 2016). Other influential model of psychopathy, the response-modulation hypothesis, proposes that individuals with psychopathy show blunted responding to threatful stimuli due to difficulties in reallocating attention to environmental stimuli once they are focused on a goal, such as seeking rewards (Baskin–Sommers et al., 2012; Hamilton & Newman, 2018; Newman, 1998; Newman et al., 2010; Newman & Baskin-Sommers, 2012; Newman & Kosson, 1986; Smith & Lilienfeld, 2015). As such, individuals with high psychopathic traits might be more insensitive to stimuli that typically induce acute stress which, in turn, disrupt reinforcement learning processes. Notably, little is known about whether and how psychopathic traits can modulate the impact of acute stress on reward and punishment learning and on prediction error signals, which drive these types of learning.

Investigating whether and how individual differences in psychopathic traits can account for behavioural and neural variations in reinforcement learning under acute stress, may contribute for our understanding of the variability in how individuals learn from the rewarding and punishing outcomes of their choices in stressful situations.

1.5. Summary and current thesis

In summary, although reinforcement learning is critical for adaptative behaviour, and acute stress is thought to affect this type of learning, some questions remain unanswered. Notably, the computational and neural mechanisms by which acute stress affects reinforcement learning remain

poorly understood. Finally, little is known about how individual differences, and in particular individual differences in psychopathic traits, can account for behavioural and neural variations in the impact of acute stress on reinforcement learning. This thesis identifies three outstanding questions regarding the impact of acute stress on reinforcement learning and investigates these questions empirically.

1) **How does acute stress impact reinforcement learning at the behavioural and computational levels?**

In **Chapter 2** of this thesis, we describe a study conducted to examine the impact of acute stress on reinforcement learning at the behavioural and computational levels. Our aim was two-fold. First, we aimed to investigate how acute stress impacted behavioural performance during reward and punishment learning. Second, we aimed to inspect the computational mechanisms behind the effects of acute stress on reward and punishment learning. Given that extant evidence suggests that acute stress affects reward learning, and possible to a larger extent than punishment learning, we predicted that acute stress would impair reward learning and, more tentatively punishment learning. Relatedly, given the putative roles of phasic dopamine responses on prediction errors signalling and of acute stress on aberrant spontaneous phasic dopamine release, we predicted that acute stress would decrease the learning rate for positive prediction errors and, more tentatively, the learning rate for negative prediction errors. We developed a novel stress manipulation to induce acute stress whilst healthy male participants performed a well-established reinforcement-learning task involving monetary gains and losses. To examine the impact of acute stress on reward and punishment learning we assessed participants' behavioural task performance. To further unravel the computational mechanism(s) affected by acute stress, we fitted a biologically inspired reinforcement-learning model to participants' trial-by-trial data to estimate their learning rates for positive and negative prediction errors.

2) Does acute stress affect neural signalling of prediction errors during reinforcement learning?

In **Chapter 3** of this thesis, we describe a study conducted to investigate how acute stress affects neural signalling of the key computational mechanism of reinforcement learning — the prediction error. Extant behavioural evidence suggests that acute stress disrupts reward learning and, possibly, punishment learning. Thus, and given that acute stress is thought to increase spontaneous phasic dopamine release, which in turn might blunt adaptive phasic dopamine responses that signal prediction

errors, we predicted that acute stress would blunt prediction error signals in the striatum during reward learning. Additionally, we explored whether and how acute stress would impact signalling of prediction errors during punishment learning. Inside the MRI scanner, male participants performed an adapted version of the reinforcement-learning task described in Chapter 2, whilst under acute and in control conditions. Using a similar reinforcement-learning task and stress manipulation, we attempted to replicate the behavioural findings of Chapter 2. Importantly, we conducted model-based fMRI analyses to examine whether and how acute stress blunted signalling of prediction errors in the striatum, regressing trial-wise prediction errors, estimated by a well-established reinforcement-learning model (the same model as in Chapter 2), against neural activity in the dorsal striatum and nucleus accumbens. Finally, we conducted subsidiary fMRI analyses to explore whether acute stress blunted preferentially positive or negative prediction errors during reward and punishment learning.

3) Do individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning and neural signalling of prediction errors?

In **Chapter 4,** we describe a study conducted to examine whether and how individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning. As in Chapter 2 we had examined the impact of acute stress on reinforcement learning at the behavioural level, in this study we asked whether and how individual differences in psychopathic traits could modulate such impact. To answer this question, we used the same data set from Chapter 2 and examined whether and how psychopathic traits were associated with variations in the impact of acute stress on reward and punishment learning.

In **Chapter 5**, we describe a study conducted to identify the neural mechanisms by which psychopathic traits account for variations in the impact of acute stress on reinforcement learning. As in Chapter 3 we had investigated the impact of acute stress on striatal signalling of prediction errors, in this study we aimed to assess whether and how individual differences in psychopathic traits could modulate such impact. Using the same data set from Chapter 3, we performed model-based fMRI analyses to probe whether and how psychopathic traits were associated with variations in prediction error signals under acute stress.

In short, this thesis set out to investigate how acute stress impacts reinforcement learning and its underlying computational and neural mechanisms, and how individual differences in psychopathic traits modulate such impact. On a first stage, we set out to investigate the effects of acute stress on

reinforcement learning at the behavioural and computational levels **(Chapter 2)**; and on a second stage, we investigated the neural mechanisms underlying the impact of acute stress on reinforcement learning **(Chapter 3)**; finally, we examined whether and how individual differences in psychopathic traits modulated the effects of acute stress on reinforcement learning at the behavioural level **(Chapter 4)** and the neural mechanisms behind such behavioural variations **(Chapter 5)**. In **Chapter 6**, we summarise the findings from these empirical studies and discuss possible implications.

CHAPTER 2

The impact of acute stress on the computational mechanisms of reinforcement learning

2.1. Chapter Introduction

As set out in the introduction of this thesis, learning to choose options that lead to rewards and to avoid those that result in punishments, i.e., reinforcement learning, is crucial for adaptive behaviour. Situational factors, such as stress, can have deleterious effects on the ability to make the best choices and learn from them (Porcelli & Delgado, 2017). Stress is present in our day-to-day life, but, notably, how acute stress affects reward and punishment learning remains largely unknown. A growing body of evidence suggests that acute stress impairs reward learning (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Cremer et al., 2021; de Berker et al., 2016; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), but less is known about the impact of acute stress on punishment-avoidance behaviour (de Berker et al., 2016; Petzold et al., 2010). More critically, there is even less evidence on the mechanisms that underlie the behavioural effects of acute stress on reward and punishment learning. In the current chapter, we describe a study where we employed a computational reinforcement-learning framework to investigate the impact of acute stress on reward and punishment learning in men[1].

In the past decades, the use of computational modelling approaches to describe behaviour-brain relationships in healthy humans has played an influential role on cognitive science (Daw & Frank, 2009; Frank, 2015; Huys et al., 2016; Maia, 2015; Maia & Frank, 2011). Computational models, such as reinforcement-learning models, are built and implemented to capture very specific cognitive and neural mechanisms, thus linking different levels of analysis, from cognitive and behavioural phenomena to neurobiological mechanisms (Chater, 2009; Daw & Frank, 2009; Nair et al., 2020). Reinforcement-learning models are considered extremely useful tools to investigate the neural computations underpinning cognition and behaviour (Daw, 2011; Daw & Frank, 2009; Huys et al., 2016; Maia & Frank, 2011), and can thus provide a mechanistic framework to disentangle the effects of acute stress on reward and punishment learning (Luksys & Sandi, 2011).

According to reinforcement-learning theory, individuals learn to gradually select more and more often the actions that maximise rewards and those that minimise punishments by learning the values of the executed actions, and such learning is driven by prediction errors (Maia & Frank, 2011; Schultz et al., 1997; Sutton & Barto, 1998). Specifically, prediction errors — which signal the difference between obtained and expected outcomes — are used to progressively update the values of the executed actions

[1] This chapter is published as Carvalheiro, J., Conceição, V., Mesquita, A. Seara Cardoso, A. (2021). Acute stress impairs reward learning in men. *Brain and Cognition*, 147. doi: 10.1016/j.bandc.2020.105657. See Appendix 6 for the reproduction of the published article.

(Collins & Frank, 2014; Maia & Frank, 2011; Schultz et al., 1997; Sutton & Barto, 1998). Prediction errors can be positive or negative. Positive prediction errors occur when outcomes are better than expected, and thus can signal unexpected rewards or avoidance of expected punishments (Daw & Tobler, 2014; Schultz et al., 1997; Watabe-Uchida et al., 2017). Negative prediction errors occur when outcomes are worse than expected, thus signalling unexpected punishing outcomes or omission of expected rewards (Daw & Tobler, 2014; Schultz et al., 1997; Watabe-Uchida et al., 2017). This means that positive and negative prediction errors can be both present in reward and punishment learning. Interestingly though, recent evidence indicates that in simple reinforcement-learning tasks, reward learning should be mostly driven by positive prediction errors (Lefebvre et al., 2017; Palminteri & Pessiglione, 2013, 2017), while punishment learning seems to be largely driven by negative prediction errors (Palminteri & Pessiglione, 2013, 2017). Studies that do not consider the positive and negative valence of prediction errors when modelling reward and/or punishment learning often assume that, during reward learning, obtaining unexpected rewards (positive prediction errors) is coded similarly to the omission of expected rewards (negative prediction errors), and that, during punishment learning, obtaining unexpected punishments (negative prediction errors) is coded similarly to the omission of expected punishments (positive prediction errors). However, extant evidence suggests that the brain encodes positive and negative prediction errors differently (Frank, 2005; Maia & Frank, 2011; O'Doherty et al., 2003; Schultz et al., 1997). Thus, to better understand the mechanisms underlying reward and punishment learning, computational models could benefit from the integration of such neurobiological evidence and model positive and negative prediction errors separately. Importantly, reinforcement-learning models can capture how quickly unexpected positive and negative outcomes are integrated over time through distinct learning rates for positive and negative prediction errors, respectively (Frank et al., 2007). Relatedly, blunted signalling of positive and negative prediction errors can be captured by reinforcement-learning models as reduced learning rates.

Dopaminergic functioning plays a key role in prediction-error-based learning (Daw & Tobler, 2014; Glimcher, 2011; Maia & Frank, 2011; Pessiglione et al., 2006; Schultz et al., 1997). Prediction-error signals are known to be encoded in the phasic activity of dopamine neurons (Schultz et al., 1997). Specifically, phasic bursts of dopaminergic neurons are thought to adaptively encode positive prediction errors, whereas dopamine dips have been associated with the adaptive encoding of negative prediction errors (Daw & Tobler, 2014; Maia & Frank, 2011; Schultz et al., 1997; Watabe-Uchida et al., 2017). However, phasic dopamine responses do not seem to be always adaptive, and there is evidence that dopamine can be phasically released in an aberrant spontaneous manner (Belujon & Grace, 2015;

Grace, 1988; Maia & Frank, 2017; Sulzer et al., 2016). Studies with non-human male animals suggest that acute stress increases aberrant spontaneous phasic dopamine release in the mesolimbic pathway, which originates in the ventral tegmental area and projects to terminal regions such as the striatum (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011). Crucially, exaggerated, aberrant spontaneous dopamine release seems to reduce adaptive phasic bursts that signal positive prediction errors in the striatum (Daberkow et al., 2013; Maia & Frank, 2017; Werlen et al., 2020). Additionally, though more speculative (Maia & Frank, 2017), aberrant spontaneous dopamine release might also block the effects of dopamine dips (Frank, 2005; Frank & O'Reilly, 2006) needed to signal negative prediction errors. Still, the extent to which stress-induced aberrant spontaneous dopamine release blunts signalling of positive and/or negative prediction errors remains poorly understood.

Stress exerts various effects on cognitive functioning (de Kloet et al., 2005), namely via fast-initial release of catecholamines and slow-delayed release of cortisol (de Kloet et al., 2005; Joëls & Baram, 2009). Thus, one important factor when studying the impact of acute stress on learning is to consider whether learning is tested immediately after the onset of the stressor when catecholamines levels are higher or later when cortisol levels peak (Hermans et al., 2014; Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; Porcelli & Delgado, 2017; Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012). A significant number of stress and learning studies have considered only slower effects of cortisol in the stress response, adopting designs which may involve relatively longer latencies between stress induction and task solving (Byrne et al., 2019; Cremer et al., 2021; de Berker et al., 2016; Ehlers & Todd, 2017; Kruse et al., 2018; Lighthall et al., 2013; Otto et al., 2013; Paul et al., 2019; Petzold et al., 2010). In those studies, participants are often exposed to a stress manipulation, such as the Trier social stress test or the (socially evaluated) cold pressor test, and the impact of stress on learning is assessed around 10-20 minutes after the stress manipulation has occurred, so that cortisol levels peak during the learning task. Such stress inductions can combine a social evaluation component, unpredictability, and uncontrollability, which are known to produce robust increases in cortisol levels (McRae et al., 2006). However, in those paradigms, it is less clear if individuals are actually perceiving stress at the time of learning, as the stressor is experienced outside of the learning context (McRae et al., 2006; Schwabe et al., 2012). Moreover, it is not unlikely that stress recovery occurs during the learning task (see Cremer et al., 2021; de Berker et al., 2016; Paul et al., 2019; Radenbach et al., 2015; Schwabe & Wolf, 2009, for examples of studies that show decreasing cortisol levels during task solving) due to suppressive actions of glucocorticoids (Hermans et al., 2014; Sapolsky et al., 2000).

Fast-acting stress-induced catecholaminergic effects, such as increased release of dopamine (Abercrombie et al., 1989; Anstrom et al., 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Nagano-Saito et al., 2013; Pruessner et al., 2004), can influence brain function in regions known to be critically associated with learning processes. Thus, by exposing participants to an acute stressor whilst they perform a learning task, one can conceivably capture the impact of stress-induced catecholamine (e.g., dopamine) release on learning. However, one potential drawback of using acute stressors, such as threat of shock, time pressure or noise, during a cognitive task, is that it can make it difficult to disentangle whether cognitive impairments occur as a result of stress or because the stressor acted as distractor (Robinson et al., 2013). Indeed, the first wave of the stress response can trigger a sensory hypervigilant state and the redirection of attention to stressors, conceivably causing distraction in tasks that require selective attention (Hermans et al., 2014). Importantly though, distractibility seems to be more robustly induced by unpredictable changes (Parmentier, 2014; Parmentier & Hebrero, 2013; Sussman et al., 2003). Therefore, one possible alternative to minimise the potential distractor effect of an acute stressor during a cognitive task could be to minimise the unpredictability of that stressor.

2.1.1. The current study

In this study, we chose to use an acute stressor (with reduced unpredictability) during a reinforcement-learning task, to better understand whether and how acute stress affects reward and/or punishment learning, very likely via putative dopaminergic mechanisms, which can be captured by computational models such as the one we used (Frank et al., 2007). The aim of the present study was twofold. First, to investigate how acute stress impacted behavioural performance during reward and punishment learning. Second, to inspect the computational mechanisms behind the effects of acute stress on reward and punishment learning. Given the putative roles of phasic dopamine responses on prediction errors signalling and of acute stress on aberrant spontaneous phasic dopamine release, we hypothesised that acute stress would impair reward learning and, more tentatively punishment learning. Relatedly, we hypothesised that acute stress would decrease the learning rate for positive prediction errors and, more tentatively, the learning rate for negative prediction errors.

To test those hypotheses, we used a well-established reinforcement-learning task involving monetary gains and losses (Pessiglione et al., 2006) combined with a novel acute stress manipulation. Using this task combined with pharmacological manipulations of the dopaminergic system, Pessiglione et al. (2006) showed that dopamine-related drugs modulate prediction errors expressed in the striatum

during reward learning. The stress manipulation consisted of exposing participants to an uncontrollable sound (i.e., participants could not put an end to it) whilst they performed the reinforcement-learning task. We chose to use a repetitive alarm sound as stressor, because alarms are known to activate the nervous sympathetic system (e.g., increase heart rate and skin conductance), involved in the earlier phase of the stress response (Hall et al., 2016; Kuorinka & Korhonen, 1981; Westman & Walters, 1981). Repetitive, uncontrollable sounds for long periods of time seem to be particularly effective in increasing dopamine release (Arnsten & Goldman-Rakic, 1998) and levels of adrenocorticotropic hormone (Arguelles et al., 1962; Breier et al., 1987). To minimise the potential distractor effect of the stressor, we used a sound that was always constant and repetitive, and therefore predictable, as unexpected changes in sound sequences robustly induce distraction and negatively affect performance in an unrelated task (Hughes, 2014; Parmentier, 2014; Parmentier et al., 2008). To further minimise unpredictability, we also instructed participants before they started the task, that they would hear an alarm, not associated with the task itself or with their performance, during blocks cued with a warning signal and a red background (Hughes, 2014; Parmentier & Hebrero, 2013; Sussman et al., 2003) . Although the stressor was not unpredictable, it was uncontrollable (i.e., participants could not put an end to it). Uncontrollable auditory stimuli can elevate stress responses (Arguelles et al., 1962; Breier et al., 1987; Rylander, 2004; Westman & Walters, 1981) and may disrupt dopaminergic mechanisms (Arnsten & Goldman-Rakic, 1998) and cognitive functioning (Glass et al., 1971). To check the success of the acute-stress manipulation, we collected self-report stress levels at the end of each block of the task and measured skin conductance response (SCR) rate throughout the task. Then, we inspected how acute stress altered reward and punishment learning during the reinforcement-learning task using both classical statistical analyses and computational-model-based analyses of participants' behavioural data. For the latter analyses, we fitted participants' choices with a previously established, biologically inspired, reinforcement-learning model (Frank et al., 2007), which has been extensively used to investigate the cognitive and behavioural impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans (Diederen et al., 2017; Doll et al., 2011; Frank et al., 2007; Frank & Fossella, 2011; Grogan et al., 2017; Rutledge et al., 2009). The fitted reinforcement-learning model allowed us to examine the effects of acute stress on learning rates for positive and negative prediction errors. In this study, we included only male participants due to females' hormonal-dependent variations on stress responsivity, as well as on reward and punishment learning (Diekhof & Ratnayake, 2016; Dreher et al., 2007; Ossewaarde et al., 2010).

2.2. Materials and Methods

2.2.1. Participants

Sixty-two healthy male participants (age range = 18 – 35; *M* = 21.9, *SD* = 3.7) were recruited at University of Minho, Portugal. Due to similarities in study design, we used the study of Pezold et al. (2010) as a reference to determine our minimum sample size. To allow for possible data exclusions due to technical problems and/or nonresponders in skin conductance data, participant enrollment in our study was stopped once we reached a sample size two-fold the sample size analysed in Petzold et. al (2010). Data collection continued until all participants enrolled up to that point were tested. Data collection was stopped before data analyses, independently of the statistical significance of the data. Four participants were excluded from skin-conductance analyses due to poor signal quality. No participants were *a priori* excluded from any other data analyses, although we conducted confirmatory analyses excluding potential outliers to ensure that our results were not driven by extreme values.

All participants provided their informed consent before the experimental session. All experimental procedures were approved by the Ethics Committee of University of Minho.

2.2.2. Reinforcement-learning task

After a short practice (12 trials), to familiarize participants with the task timings and response keys, participants completed four blocks of an adapted version of a well-established reinforcement-learning task (Pessiglione et al., 2006) (Figure 2.1). Each block included three pairs of abstract stimuli, and each pair of stimuli was presented 40 times, making 120 trials in total per block. Different abstract stimuli were used in each block. On a given trial, each pair of stimuli was associated with a valence: one pair of stimuli was associated with gains (gain 0.5€ or no change), a second pair associated with losses (loss 0.5€ or no change), and a third pair associated with neutral, or non-financial outcomes (look at a 0.5€ coin or no change; for a depiction of neutral trials see Figure S1a in Appendix 2). The pairs of stimuli were presented in a pseudo-randomized order within each block. The outcome probabilities were reciprocally 0.8 and 0.2 for the stimuli in each of the three pairs. That is, in gain trials, one stimulus was associated with a probability of 0.8 of winning 0.5€ and with a probability of 0.2 of winning nothing ("correct" stimulus), and the other stimulus was associated with a probability of 0.2 of winning 0.5€ and with a probability of 0.8 of winning nothing ("incorrect" stimulus); in loss trials, one stimulus was associated with a probability of 0.8 of losing 0.5€ and with a probability of 0.2 of losing nothing ("incorrect" stimulus), and the other stimulus was associated with a probability of 0.2 of

losing 0.5€ and with a probability of 0.8 of losing nothing ("correct" stimulus); in neural trials, one stimulus was associated with a probability of 0.8 of looking at 0.5€ coin and with a probability of 0.2 of nothing, and the other stimulus was associated with a probability of 0.2 of looking at 0.5€ coin and with a probability of 0.8 of nothing (note that neutral trials were not associated with financial outcomes). These outcome probabilities remained constant within each block. On each trial, one pair was randomly presented on the screen, with one stimulus from the pair above and the other below a central fixation cross (the stimuli position was counterbalanced across trials). Participants were instructed to choose between the two visual stimuli displayed on the computer screen to maximise payoffs. Missing choices occurred when participants did not press the response keys within 2000 ms (0.38% missing choices: 40 in the stress condition and 72 in the control condition, in a total of 29760 trials across all participants) and were signalled with a "Missed" message (no other outcome was provided). Missing choices were excluded from data analyses. Before starting the task, participants were informed that they would be paid the amount of money obtained during a randomly selected block, although they all left with the same fixed compensation (15€) for their participation. The experiment was programmed and presented with Cogent 2000 (http://www.vislab.ucl.ac.uk/cogent.php) implemented in MATLAB R2015a (MathWorks).

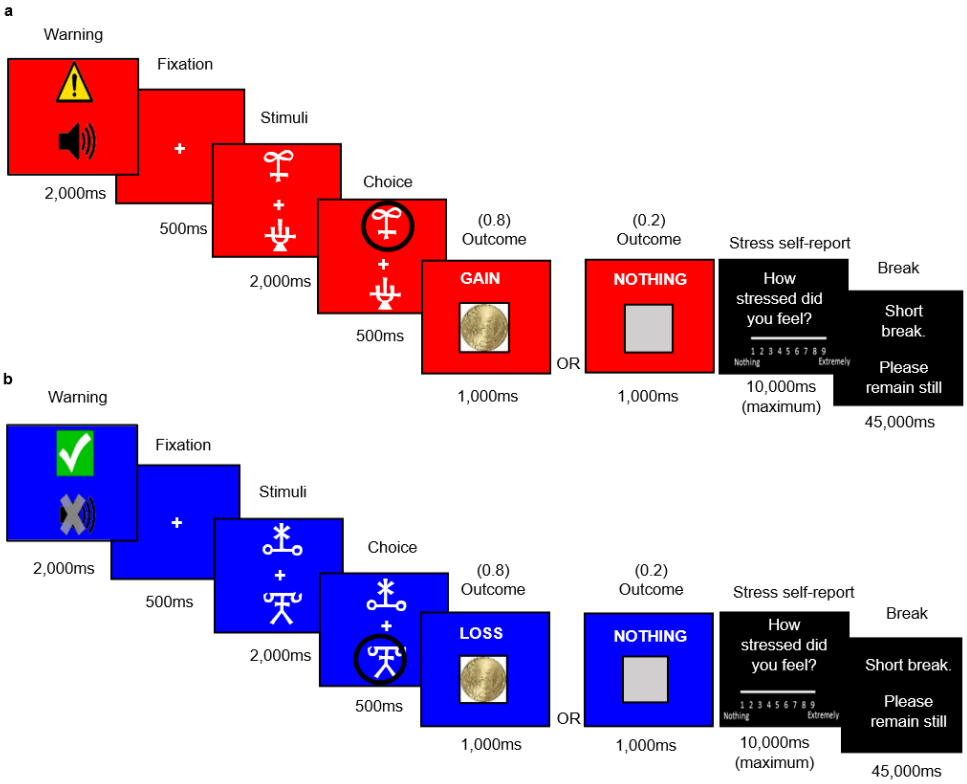


Figure 2.1. Reinforcement-learning task. On each trial of the task, participants had to choose either the upper or the lower of two abstract visual stimuli presented on a computer screen and to

subsequently observe the obtained outcome, whilst under acute stress and under control conditions. **(a)** In the depicted stress-condition example, the chosen stimulus was associated with a probability of 0.8 of winning 0.5€ and with a probability of 0.2 of winning nothing. The other (not chosen) stimulus was associated with a probability of 0.8 of winning nothing and with a probability of 0.2 of winning 0.5€. **(b)** In the depicted control-condition example, the chosen stimulus was associated with a probability of 0.8 of losing 0.5€ and with a probability of 0.2 of losing nothing. The other (not chosen) stimulus was associated with a probability of 0.8 of losing nothing and with a probability of 0.2 of losing 0.5€. Participants completed a total of four blocks, consisting of an alternation between two stress and two control blocks. To assess stress responses, self-reported stress levels were collected at the end of each block and skin conductance was measured throughout the task.

2.2.3. Acute-stress manipulation

During the experimental session, participants performed two blocks of the reinforcement-learning task whilst exposed to a stressor (i.e., stress condition; Figure 2.1a) and two blocks without the stressor (i.e., control condition; Figure 2.1b). To elicit stress responses we exposed participants to a predictable, but uncontrollable auditory stimulus, a constant alarm ("Annoying modern office building alarm.wav", retrieved from freesound.org, and programmed to loop uninterruptedly), played through the same set of over-ear headphones (GOODIS, model GWH4093, with the volume set to the maximum), with the sound volume adjusted in the laptop to level 27 (in a scale from 0-100) for all participants. We conducted a brief pilot to qualitatively confirm that this volume was tolerable, yet stressful. Stress blocks were signalled by a warning sign and a red background (Figure 2.1a), and control blocks were signalled by a safe sign and blue background (Figure 2.1b). Stress and control blocks were administered alternately and in counterbalanced order. Each block of the task lasted 8 minutes, and the whole task took around 36 minutes (including warnings, breaks between blocks, and stress ratings with variable durations). The experiment was conducted in a soundproof room to avoid interference due to environmental noise, and between 12 pm and 6 pm to minimise diurnal variability in stress responses.

2.2.4. Manipulation check

Stress levels were assessed by asking participants at the end of each block to rate how stressed they felt during that block on a scale of 1 (nothing) to 9 (extremely). To further assess the impact of the acute stressor on autonomic responses, we acquired skin conductance responses (SCRs) using BIOPAC

MP150 and a pair of finger electrodes. Electrodes were attached to participants' left index and ring fingers; the gain was set to 5, the low pass filter to 10 Hz, and the high pass filters to DC. Recordings were performed using Acqknowledge 4.4. Data were acquired at 200 Hz, downsampled to 62 Hz, and smoothed with a median filter in order to remove outliers. Each participant's SCRs were detected using a threshold of 0 μs and a rejection rate of 10% (Kim et al., 2004), meaning that we excluded SCR with an amplitude smaller that 10% of the maximum SCR amplitude in each subject, rather than using an arbitrary threshold that would not account for inter-individual differences in SCRs. The SCR rate was calculated by dividing the number of SCRs detected in each block by the duration of that block (in minutes).

Self-reported stress levels and SCR rates were analysed using repeated-measures analyses of variance (ANOVAs), with condition (stress and control) and block (1 and 2) as within-subject factors, and post-hoc paired *t*-tests. ANOVAs effect sizes are reported as eta-squared, η^2, and post-hoc paired *t*-tests' effect sizes are reported as Cohen's *d* and 95% confidence intervals. We further conducted non-parametric Wilcoxon signed-rank tests, which are more robust to outliers, to confirm the results from post-hoc paired *t*-tests. Additionally, to confirm that our findings were robust to extreme values, we repeated the SCR rate analyses excluding participants with abnormally large SCR rates. Statistical analyses were conducted using JASP 0.9.

2.2.5. Task-performance analyses

To examine the impact of acute stress on choice performance during the reinforcement-learning task, we applied a generalized linear mixed-effects (glme) model to participants' choice data (with correct and incorrect choices coded as 1 and 0, respectively). We used a "logit" link function to account for the binomial distribution of the data. As predictor variables in the glme model we included condition (stress or control), valence (gains or losses), block number (1 or 2), and trial number (1 to 40), and the interaction of interest (condition × valence; see Table S2.1 in Appendix 2 for a full description of the glme model). The glme included a fixed intercept, as well as random intercepts for each participant. Inclusion of by-subject random slopes for all predictor variables (trial, block, condition, and valence) did not change the significance and interpretation of the condition × valence interaction, and thus we report the results from the glme model with the simplest random effects structure. We fitted the glme model to the behavioural data using MATLAB's *fitglme* function and conducted post-hoc analyses via contrast matrices using the MATLAB's *coefTest* function. To assess the robustness of our findings, we also

tested the significance of the interaction of interest (condition × valence) using confirmatory likelihood ratio tests (Daw, 2011) between the aforementioned full glme model and two nested models, which assumed equal performance in stress and control conditions during either gain or loss trials, through the MATLAB's *lratiotest* function. Additionally, we repeated the analyses excluding participants that performed below chance levels, which is indicative that participants did not learn to perform the task correctly and may additionally reflect non-compliance with the experimental setting.

To confirm that the choice probabilities estimated by the glme model showed a close correspondence with the actual observed choices, we used MATLAB's *predict* function. Then, we assessed the Pearson's correlation between the percentage of actual "correct" choices (i.e., choice of the stimuli associated with a probability of 0.8 of winning or a probability of 0.2 of losing) for each participant in each condition (averaged across blocks) and the percentage of "correct" choices as estimated by the glme model. We also performed confirmatory Spearman's correlations, which are more robust to outliers.

2.2.6. Computational modelling

2.2.6.1. Reinforcement-learning model

We modelled participants' trial-by-trial behaviour in the stress and control conditions using a reinforcement-learning framework (Sutton & Barto, 1998) that has been extensively used to investigate the behavioural and neural impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans (Diederen et al., 2017; Doll et al., 2011; Frank et al., 2007; Frank & Fossella, 2011; Grogan et al., 2017; Rutledge et al., 2009). Importantly, the fitted model included separate learning rates for positive (α^+) and negative (α^-) prediction errors, to account both for the differential firing of dopaminergic neurons for positive and negative prediction errors (Daw & Tobler, 2014; Schultz et al., 1997) and the differential effects of dopamine onto the plasticity of the corticostriatal synapses implicated in action-value learning (Frank & O'Reilly, 2006; Maia & Conceição, 2017; Maia & Frank, 2011; Möller & Bogacz, 2019). This model also included the inverse temperature parameter (β) which controls the stochasticity of choice selection, or the decision noise (Daw, 2011; Sutton & Barto, 1998), as detailed below.

In the context of our experimental study, the specific reinforcement-learning model used (Frank et al., 2007) assumes that each participant gradually learns the value of choosing a given stimulus (say A or B) from a given pair of stimuli (here, "gain" or "loss" stimuli pairs, as the "neutral" pair of stimuli

always yielded null monetary outcomes) as a function of the outcome that was obtained on that trial following stimulus selection. Specifically, each expected pair-stimulus value, or Q-value, was initialized to zero, and for each trial, t, within that pair of stimuli, the value of the chosen stimulus (say A was chosen) was updated according to:

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

where δ was the prediction error:

$$\delta(t) = r(t) - Q_A\,(t),$$

where $r(t)$ was 0.5 for winning, 0 for obtaining nothing, and -0.5 for losing. The learning rate, α , was given by:

$$\alpha = \begin{cases} \alpha^+, & if\ \delta(t) > 0 \\ \alpha^-, & if\ \delta(t) < 0 \end{cases}',$$

where α^+ and α^- were the learning rates for positive and negative prediction errors, respectively (Frank et al., 2007).

The probability of choosing one stimulus over another (say A over B) was given by the *softmax* equation:

$$P_A(t) = \frac{e^{[Q_A(t)*\beta]}}{e^{[Q_A(t)*\beta]} + e^{[Q_B(t)*\beta]}},$$

where the β parameter, or inverse temperature, controlled the amount of decision noise. The lower the β, the higher the decision noise (i.e., decisions are made more at random).

2.2.6.2. Model fitting, parameter analyses, and model validation

We fitted the reinforcement-learning model to the trial-by-trial choice data from each participant in each condition. Model fitting involved estimating the values of the parameters (α^+, α^-, and β) that best accounted for the respective trial-by-trial choices in each condition.

We estimated the best-fitting model parameters (α^+, α^-, and β) for each subject in each condition using maximum *a posteriori* estimation (Daw, 2011). Specifically, to optimise model parameters, we drew the learning rates from Beta distributions [*Beta* (1.1, 1.1)] and the inverse temperature from a Gamma distribution [*Gamma* (1.2, 5)] (Palminteri et al., 2015). We then used the MATLAB's *fmincon* function, initialized at 100 random starting points of the parameter space, to search for the parameter values that minimised the negative log posterior of the observed sequence of choices, given the previously observed outcomes, with respect to different settings of the model parameters (Daw, 2011).

To assess the effects of acute stress on the parameters estimated by the reinforcement-learning model, we conducted repeated-measures ANOVAs with condition (stress and control) and valence (positive and negative) as within-subject factors, and post-hoc paired *t*-tests. ANOVAs' effect sizes are reported as eta-squared, η^2, and post-hoc paired *t*-tests effect sizes are reported as Cohen's *d* and 95% confidence intervals. We also confirmed all results from post-hoc tests using non-parametric tests (Wilcoxon signed-rank tests), which are more robust to outliers. Additionally, we repeated all analyses excluding participants that performed below chance levels to test whether the significance of the results remained unchanged. These statistical analyses were conducted using JASP 0.9.

To validate the used reinforcement-learning model, we computed trial-by-trial choice probabilities for all participants using the best-fitting set of parameters in each condition. The actual observed choices and outcomes were used to update the choice probabilities. To assess whether the choice probabilities estimated by the reinforcement-learning model (for each subject, the choice probabilities were averaged across gains or loss trials in each condition) followed the same pattern as the actual observed choices, we conducted Pearson's correlations and confirmatory Spearman's correlations for both conditions, using the respective mean percentage of correct choices.

To further validate the robustness of our model-fitting procedure, we examined the capacity of recovering subject-condition-specific parameters using simulated data sets. Specifically, we simulated the task-choice behaviour of 62 virtual participants using the parameter values that we had estimated for each of the 62 participants in each condition. We ran 100 simulations. Then, for each simulation, we fitted the model to the virtual participants' data to estimate new (recovered) parameters. Finally, we tested the correlations between the original and these recovered parameters using Pearson's and confirmatory Spearman's correlations.

2.3. Results

2.3.1. Manipulation check

First, we confirmed that the acute stress manipulation successfully elicited stress responses in the participants. Self-reported stress levels differed significantly between conditions, $F(1, 61) = 107.67$, $p < .001$, $\eta^2 = 0.64$, as participants reported higher levels of stress in the stress condition ($M = 5.16$, $SEM = 0.21$) than in the control condition ($M = 3.31$, $SEM = 0.20$), $t(61) = 10.38$, $p < .001$, $d = 1.32$, 95% confidence interval (CI) = [1.49, 2.20] (Figure 2.2a). SCR rate also differed significantly between conditions, $F(1,57) = 20.61$, $p < .001$, $\eta^2 = 0.27$, being higher in the stress condition ($M = 2.89$, *SEM*

= 0.27) than in the control condition (M = 2.46, SEM = 0.22), t(57) = 4.54, p < .001, d = 0.57, 95% CI = [0.24, 0.62] (Figure 2.2b). The condition × block interactions were non-significant for the self-reported stress levels, F(1, 61) = 0.004, p = .95, η^2 = 0, and for the SCR rate, F(1,57) = 0.46, p = .50, η^2 = 0.0080, suggesting that both the self-reported stress levels and the SCR rate remained stable across blocks within conditions. Furthermore, confirmatory non-parametric Wilcoxon signed-rank tests replicated the effects of acute stress on self-reported stress levels, when comparing the stress condition with the control condition, Z = 6.29, p < .001, and SCR rate, Z = 4.08, p < .001. Additionally, to check whether our results were robust to extreme values, we identified two participants with abnormally large SCR rate and reanalysed the data without those participants. The significance of the results remained unchanged (main effect of condition: F(1, 55) = 19.47, p < .001, η^2 = 0.26), with higher SCR rate in the stress condition than in the control condition, t(55) = 4.41, p < .001, d = 0.59, 95% CI = [0.21, 0.56], even after excluding the two participants.

In sum, these results suggest that the acute stress manipulation successfully elicited stress responses.

2.3.2. Task performance

After confirming that self-reported stress levels and SCR rate were augmented in the stress condition, we examined the impact of acute stress on choice performance during the reinforcement-learning task (Figure 2.2c) using a glme model, which accounted for the binomial distribution of the data (correct or incorrect responses) and included as predictor variables condition (stress or control), valence (gains or losses), block number (1 or 2), trial number (1 to 40), and the interaction of interest (condition × valence). We found a significant condition × valence interaction, β = -0.19, p = .018, 95% CI = [-0.34, -0.031] (Figure 2.2d), and post-hoc analyses revealed that under stress, relative to the control condition, participants performed significantly worse when seeking monetary gains, F(1, 19755) = 12.87, p < .001, but not when avoiding losses, F(1, 19755) = 0.14, p = .71. As performance below chance levels might be indicative of non-compliance with the experimental setting, we also inspected whether each participant's behavioural performance across gains and losses was below chance levels (i.e., less than 50% of correct choices) in the stress or control conditions. One participant performed below chance levels in the stress condition, and two participants performed below chance levels in the control condition. Thus, we repeated the aforementioned analyses excluding these three participants. We found that the condition × valence interaction was still significant after

excluding participants that did not learn how to perform the task, β = -0.28, p < .001, 95% CI = [-0.44, -0.12]. We also confirmed the robustness of our findings using likelihood ratio tests. We found that the full glme model (which assumed different performance towards gains between the stress and control conditions) had a significantly better fit than a model that assumed no differences in performance towards gains between conditions, $\chi^2(1)$ = 12.80, p < .001, and that the same full model did not have a significantly better fit than a model that assumed no differences in performance towards losses between the stress and control conditions χ^2 (1) = 0.20, p = .66. These analyses indicate that behavioural performance towards gains, but not losses, significantly differed between the stress and control conditions.

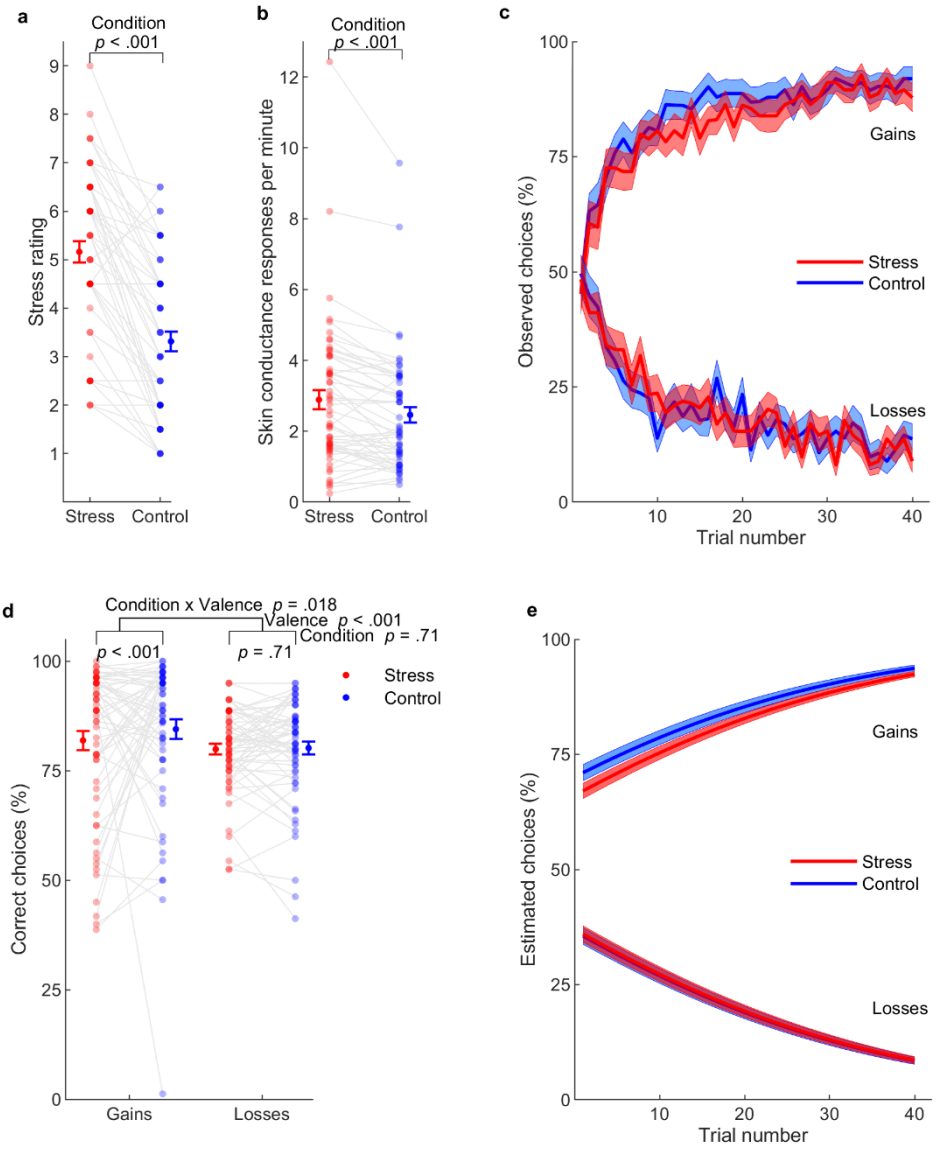


Figure 2.2. Manipulation check and task performance. (a) Participants (n = 62) reported higher stress levels in the stress condition (red) compared with the control condition (blue). **(b)** Skin

conductance response rate (n = 58) was significantly higher in the stress condition than in the control condition. **(c)** Learning curves represent the trial-by-trial percentage of participants (n = 62) who chose the "correct" gain stimulus (associated with a probability of 0.8 of winning 0.5€; upper part of the graph) and the "incorrect" loss stimulus (associated with a probability of 0.8 of losing 0.5€; lower part of the graph), in the stress and control conditions. **(d)** Participants performed significantly worse when seeking monetary gains, but not when avoiding monetary losses, in the stress condition relative to the control condition. The reported *p*-values are from the generalized linear mixed-effects model that included each participant's trial-by-trial choices and respective post-hoc tests. **(e)** The choices estimated by the generalized linear mixed-effects model captured the evolution of the actual observed choices during the reinforcement-learning task (to compare the choices estimated by the model with the actual observed choices, compare the overlap of the curves between the stress and control conditions depicted here with the overlap of the curves depicted in Figure. 2.2c). In graphs a, b, and d, connected dots represent data points from the same participant, and more transparent (opaque) dots represent less (more) overlapping data points; the error bar displayed on the side of the scatter plots indicate the sample mean ± standard error of the mean. In panels c and e, each central line represents the mean, and the filled area represents the ± standard error of the mean.

In sum, acute stress selectively impaired choice performance towards monetary gains during the reinforcement-learning task. As an additional check, we confirmed that the choices estimated by the glme model showed close correspondence with the observed choices across trials in both conditions, Pearson's $r > 0.65$, $p < .001$, Spearman's $r > 0.52$, $p < .001$ (Figure 2.2e, see Figure S2.2 in Appendix 2 for scatterplots).

Additionally, given that acute stress is thought to increase aberrant spontaneous phasic dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011), which, in turn, may lead to augmented learning and behavioural responding for neutral stimuli (Maia & Frank, 2017; Roiser, Howes, Chaddock, Joyce, & McGuire, 2013), we performed an exploratory analysis of participants' choices for the neutral stimuli pairs (the pairs not associated with financial outcomes; for further details see Table S2.2 and "Analyses of neutral trials" in the "Supplementary Analyses" section of the Appendix 2). We found that, within the pairs of neutral stimuli, acute stress increased behavioural responding towards the stimuli that more often yielded as outcome a coin with no financial value (the high-probability "look" stimuli) relative to the stimuli that yielded no outcome at all (Figure S2.1b in Appendix 2). This tentative finding seems consistent with the idea that acute stress might bias

behavioural responding for neutral stimuli due to augmented aberrant spontaneous phasic dopamine release.

2.3.3. Computational Modelling

To further probe the nature of the effects of acute stress on reward and punishment learning, we fitted a biologically inspired reinforcement-learning model (Frank et al., 2007) to participants' trial-by-trial choices (see subsection 2.2.6.1 in "Materials and Methods" for a full description of the model). The fitted model included separate learning rates for positive (α^+) and negative (α^-) prediction errors, to account for the differential firing of dopaminergic neurons for positive and negative prediction errors (Schultz et al., 1997). The model also included the inverse temperature parameter, β, which controls the decision noise (Daw, 2011; Sutton & Barto, 1998). We then analysed the best-fitting model parameters using ANOVAs.

These analyses revealed a main effect of condition on learning rates, $F(1, 61) = 4.69$, $p = .034$, $\eta^2 = 0.071$, but a non-significant condition × valence interaction, $F(1, 61) = 2.13$, $p = .15$, $\eta^2 = 0.034$ (Figure 2.3a). For completeness, we performed paired *t*-tests. Despite not formally well-justified (due to the lack of a significant interaction), such paired *t*-tests yielded results consistent with those from task-performance analyses and those from ensuing computational modelling analyses (cf. analyses that excluded participants who performed below chance levels and analyses of $\alpha^{\pm} * \beta$ below). *t*-tests revealed that α^+ was significantly lower in the stress condition ($M = 0.40$, $SEM = 0.033$) than in the control condition ($M = 0.51$, $SEM = 0.038$), $t(61) = -2.25$, $p = .028$, $d = -0.29$, 95% CI = [-0.21, -0.013], while α^- was not significantly different between the stress ($M = 0.25$, $SEM = 0.025$) and control conditions ($M = 0.27$, $SEM = 0.027$), $t(61) = -0.72$, $p = .47$, $d = -0.092$, 95% CI = [-0.098, 0.046] (Figure 2.3a). Additionally, as for task-performance analyses, we repeated all statistical analyses excluding the three participants that performed below chance levels. Exclusion of those three participants revealed that the condition × valence interaction reached significance, $F(1, 58) = 4.61$, $p = 0.036$, $\eta^2 = 0.074$), such that α^+ was significantly lower in the stress condition than in the control condition, $t(58) = -2.48$, $p = .016$, $d = -0.32$, 95% CI = [-0.22, -0.024], and α^- did not differ significantly between conditions $t(58) = -0.16$, $p = .88$, $d = -0.021$, 95% CI = [-0.073, 0.063]. Thus, our findings suggest that acute stress decreases α^+.

The parameter β, which controls the amount of noise in choice selection, also did not differ significantly between the stress (M = 9.52, SEM = 0.49) and control (M = 9.53, SEM = 0.51) conditions, t(61) = -0.016, p = .99, d = 0.0020, 95% CI = [-0.98, 0.99] (Figure 2.3b).

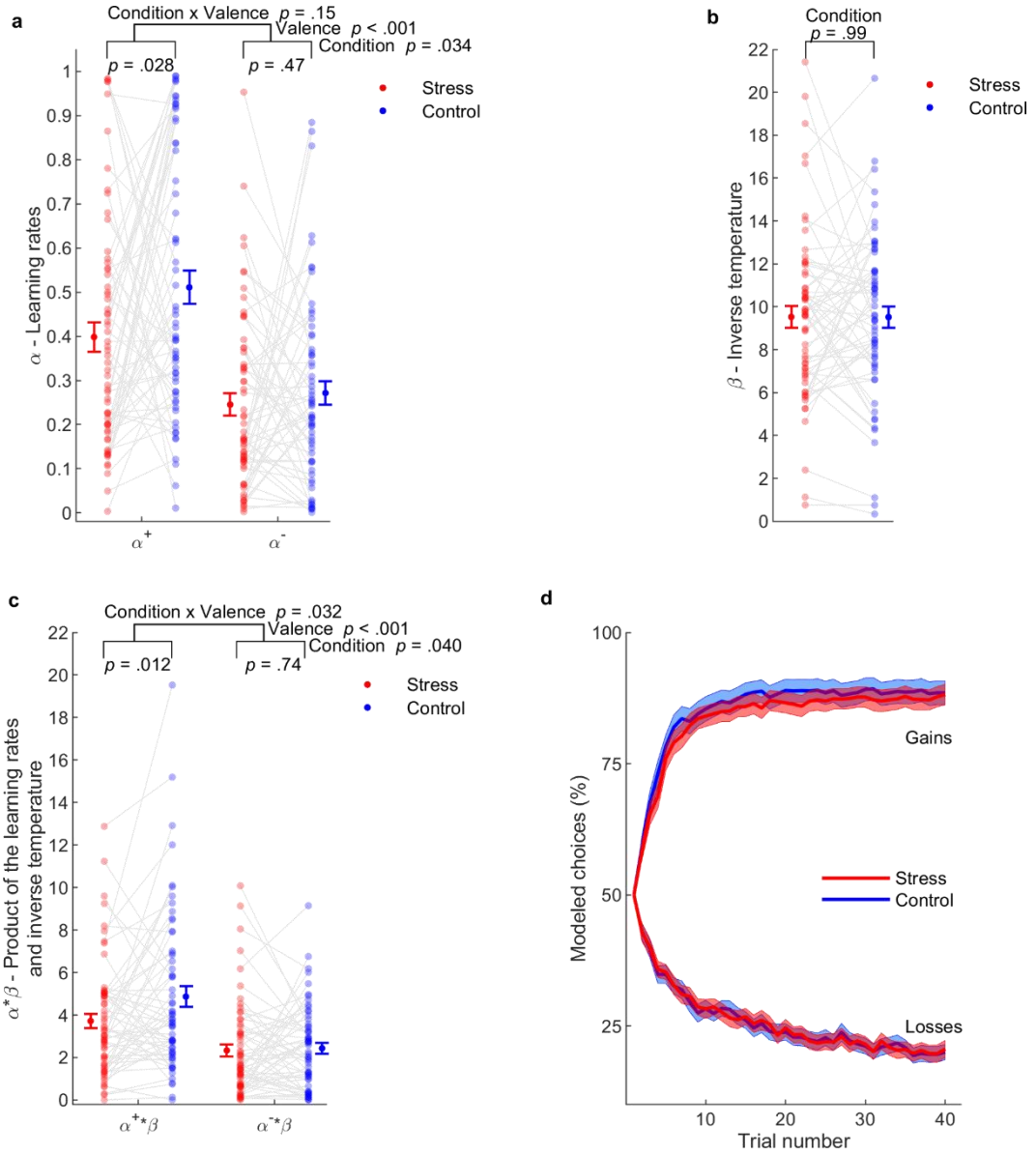


Figure 2.3. Model-fitting results. (a) Participants (n = 62) had lower learning rates, α^+ and α^-, in the stress condition (red) compared with the control condition (blue). Paired t-tests revealed a significantly lower learning rate for positive prediction errors (α+) in the stress condition comparatively to the control condition. **(b)** The inverse temperature (β) did not differ between conditions. **(c)** $\alpha^+ * \beta$, but not $\alpha^- * \beta$, was significantly reduced in the stress condition compared with the control condition. In panels a, b and c, connected dots represent data points from the same

participant, and more transparent (opaque) dots represent less (more) overlapping data points. The error bar displayed on the side of the scatter plots indicate the sample mean ± standard error of the mean. **(d)** The probabilities of choosing the "correct" gain (upper part of the graph) and the "incorrect" loss (lower part of the graph) stimuli estimated by the reinforcement-learning model followed the same pattern of the actual observed choices (compare the overlap of the curves between the stress and control conditions depicted here with the overlap of the curves that represent the actual observed choices depicted in Figure 2.2c). Each central line represents the mean, and the filled area represents the ± standard error of the mean.

We further tested whether stress had a differential effect on reward compared with punishment learning by analysing the products between each learning rate, $\alpha^{\pm}$, and the inverse temperature, β. In reinforcement-learning models, $\alpha^{\pm}$ and β tend to be inversely coupled (Daw, 2011; Supplementary Material of Maia & Conceição, 2017) because $\alpha^{\pm}$ multiply by state-action values and the state-action values themselves are multiplied by β to compute choice probabilities (see equations in subsection 2.2.6.1). As a result, the parameters viewed separately can have larger estimation errors, while their product tends to be more reliably estimated (Daw, 2011; Schönberg et al., 2007; Zhang et al., 2020), and thus better recovered. Note that the products $\alpha^{\pm} * \beta$ control how strongly the outcomes impact subsequent choice preferences (Schönberg et al., 2007). Statistical analyses revealed a significant condition × valence interaction, $F(1, 61) = 4.85$, $p = 0.032$, $\eta^2 = 0.074$, meaning that acute stress significantly decreased $\alpha^{+} * \beta$ (stress: $M = 3.72$, $SEM = 0.34$, control: $M = 4.87$, $SEM = 0.48$), $t(61) = -2.58$, $p = .012$, $d = -0.33$, 95% CI = [-2.04, -0.26], while not significantly affecting $\alpha^{-} * \beta$ (stress: $M = 2.33$, $SEM = 0.28$, control: $M = 2.44$, $SEM = 0.25$), $t(61) = -0.33$, $p = .74$, $d = -0.042$, 95% CI = [-0.71, 0.51] (Figure 2.3c). Exclusion of the three participants that performed below chance levels did not change the significance of the results for the parameter β, $t(58) = 0.15$, $p = .88$, $d = 0.020$, 95% CI = [-0.94, 1.10], nor for $\alpha^{\pm} * \beta$ condition × valence interaction, $F(1, 58) = 4.45$, $p = 0.039$, $\eta^2 = 0.071$. Notably, the previous results were fully replicated when comparing the estimated parameters between the stress and control conditions using non-parametric Wilcoxon signed-rank, rather than parametric, post-hoc tests (α^{+}: Z = -2.22, $p = .026$; α^{-}: Z = -0.17, $p = .86$; β: Z = -0.11, $p = .91$; $\alpha^{+} * \beta$: Z = -2.19, $p = .028$; $\alpha^{-} * \beta$: Z = -0.35, $p = .73$).

To validate the used reinforcement-learning model, we confirmed that the probability of choices estimated under the reinforcement-learning model had a close correspondence with the actual observed choices across trials in both conditions, Pearson's $r > 0.76$, $p < .001$, Spearman's $r \geq 0.70$, $p < .001$

(Figure 2.3d, see Figure S2.3 in Appendix 2 for scatterplots). Next, to further validate that our previous model-fitting results were reliable, we conducted parameter-recovery analyses. Those parameter-recovery analyses demonstrated that the results of our model-fitting procedure were robust both in the stress and control conditions, Pearson's $r > 0.67$, $p < .001$, Spearman's $r > 0.69$, $p < .001$ (see Figure S2.4 in Appendix 2). Additionally, Bayesian model averaging analyses (Hoeting et al., 1999) using both the aforementioned, neurobiologically inspired reinforcement-learning model (with separate learning rates for positive and negative prediction errors) and a nested, alternative candidate model (with a single learning rate) provided further support for the robustness of the parameters estimated by the model with separate learning rates (see Figure S2.5 and "Model comparison and Bayesian model averaging" in the "Supplementary Analyses" section of Appendix 2).

Taken together, these computational findings indicate that under acute stress participants incorporated positive prediction errors at a lower rate, which seems to explain why acute stress impaired behavioural performance towards monetary gains during the reinforcement-learning task.

2.4. Discussion

Acute stress is present in day-to-day life, and people recurrently need to make choices and learn from the rewarding or punishing outcomes of those choices whilst under stress. In this study we investigated whether and how acute stress impacted reward and punishment learning in men using a reinforcement-learning framework. Given the putative roles of phasic dopamine responses on prediction errors signalling and of acute stress on aberrant spontaneous phasic dopamine release, we hypothesised that acute stress would impair reward learning and, more tentatively punishment learning. Relatedly, we hypothesised that acute stress would decrease the learning rate for positive prediction errors and, more tentatively, the learning rate for negative prediction errors. We found that acute stress impaired behavioural performance towards monetary gains, but not losses, and that this impaired performance could be explained by a decreased learning rate for positive prediction errors.

Our finding that acute stress impaired reward-seeking performance is consistent with several previous studies which found reduced reward responsivity under acute stress (Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), particularly in high stress-reactive individuals (Berghorst et al., 2013). At a first sight, however, some of the extant literature may seem equivocal, possibly due to critical methodological differences related to stress operationalization (Porcelli & Delgado, 2017). A significant number of studies have investigated the effects of stress on learning using different paradigms, such as the Trier social stress test (Boyle et al.,

2019; Cremer et al., 2021; Kruse et al., 2018; Petzold et al., 2010; Radenbach et al., 2015), or the (socially evaluated) cold pressor test (Byrne et al., 2019; de Berker et al., 2016; Ehlers & Todd, 2017; Glienke et al., 2015; Lighthall et al., 2013; Otto et al., 2013; Paul et al., 2019), in which acute stress is induced before the learning task. In these paradigms, stress induction precedes any learning processes, thus the stress-induced emotional state may be less concurrent with the cognitive processes that operate during the task. As such, it is unclear whether these paradigms probe the effects of acute stress or of recovery from stress on learning (Hermans et al., 2014). In contrast, in our study, we induced stress during the learning task, and we used relatively short task blocks to avoid the slow cortisol response from peaking and subsequent recovery within blocks. The timing of the stress induction relative to the task is an important factor to take into account when comparing studies, as early-fast catecholamine release and delayed-slow glucocorticoids actions might account for different effects on brain function (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009). Critically, studies that induce stress before the learning task may suggest divergent behavioural results from ours (Byrne et al., 2019; Lighthall et al., 2013), whereas studies that have induced stress during the learning task (as ours) seem to be in agreement with our findings of impaired reward learning during acute-stress exposure (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Paret & Bublatzky, 2020). Therefore, cross-study comparisons suggest that individuals may perform more poorly when learning to maximise rewards whilst exposed to an acute stressor.

In this study we further inspected the computational mechanisms behind impaired reward learning under acute stress. Using a reinforcement-learning model with separate learning rates for positive and negative prediction errors (Frank et al., 2007), we found that acute stress reduced the rate at which participants learned from positive prediction errors. A reduced learning rate for positive predictions errors means that under acute stress individuals learned more slowly about unexpected rewards and therefore took longer to adapt their behaviour on the basis of the more recent rewarding outcomes of their choices. Other studies have successfully applied this reinforcement-learning model to analyse data from distinct tasks to describe the genetic (Doll et al., 2011; Frank et al., 2007) and neural (Diederen et al., 2017; Lefebvre et al., 2017; Niv et al., 2012) correlates of cognition and behaviour. However, to the best of our knowledge, no studies had attempted to model how acute stress affects cognition, particularly reward and punishment learning, using such model. Another reinforcement-learning model commonly used in the stress literature includes distinct learning rates for gains and losses regardless of the valence of the prediction error (Aylward et al., 2019; Treadway et al., 2017), meaning that learning from rewards in gain trials (or punishments in loss trials) is modelled the

same way as learning from omission of rewards in gain trials (or omission of punishments in loss trials, respectively), which may be misleading and not be so well supported by neurobiological evidence as a model that assumes distinct learning rates based on the prediction errors' valence (as the one we used) (Maia & Frank, 2011; O'Doherty et al., 2003; Schultz et al., 1997). In simple reinforcement-learning tasks, such as ours, reward learning seems to mostly occur via positive prediction errors (Lefebvre et al., 2017). But, given the structure of our reinforcement-learning model, stress-induced impaired reward learning could have conceivably been underpinned by a higher learning rate for negative prediction errors — which could have led to unlearning of the correct stimulus — and/or by lower inverse temperature parameter (β) — which could have been reflected in choices more at random. The fact that our data indicated that the learning rate for negative prediction errors was not higher in the stress condition and that the inverse temperature parameter did not differ between conditions, but that the learning rate for positive prediction errors was decreased in the stress condition, suggest that acute stress specifically impaired reward learning via disruption of positive prediction errors. Thus, our findings contribute to a better mechanistic understanding of how acute stress may impact reward learning.

Given that the (quantifiable) parameters from the reinforcement-learning model that we used seem to reflect specific dopaminergic-related neural mechanisms (e.g., Frank et al., 2007; Frank & O'Reilly, 2006) our computational findings may shed light on putative neural mechanisms underlying the impact of acute stress on cognition and behaviour. Specifically, our computational findings seem broadly consistent with the proposed neurobiological account of dopaminergic neurons functioning under acute stress. Acute stress is thought to induce aberrant spontaneous dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Valenti et al., 2011); stress-induced spontaneous dopamine release, in turn, may disrupt the adaptive striatal phasic-burst dopamine responses that signal positive prediction errors (Daberkow et al., 2013; Grace, 2016; Maia & Frank, 2017; Werlen et al., 2020), which would explain the stress-induced impairment of reward learning. While our computational results suggest that acute stress disrupts striatal responses to prediction errors during reward learning (Huys et al., 2013), we cannot disregard, however, that acute stress may also affect reward learning via prefrontal cortex disturbances (Arnsten, 2009; Lighthall et al., 2013; Otto et al., 2013).

Acute stress impaired reward learning, but we found no evidence for an effect of acute stress on punishment learning. Computational-modelling analyses did not provide evidence that acute stress affected the learning rate for negative prediction errors to the same extent that it affected the learning

rate for positive prediction errors, which might explain why behavioural performance in losses — which might rely more on negative prediction errors compared with performance towards gains (Palminteri & Pessiglione, 2013, 2017) — was not impaired under acute stress.

According to a long-standing influential loss aversion framework (Kahneman & Tversky, 1979), losses can have more debilitative potential than gains; therefore, as an adaptive strategy, it is possible that individuals may be more attuned to losses than to gains (Lejarraga et al., 2012; Yechiam, 2019; Yechiam & Hochman, 2013), explaining why punishment learning was spared under acute stress. However, recent evidence suggests that individuals learn gain associations better than loss associations in reinforcement-learning tasks despite symmetrical task structure and symmetrical outcome probabilities (Lin et al., 2020) (as in our task). Indeed, our behavioural data indicated that participants performed better when learning to seek gains than when learning to avoid losses in both the stress and control conditions. In addition, our computational data indicated that the learning rate for positive prediction errors was significantly higher than the learning rate for negative prediction errors, which suggests underestimation of negative outcomes relative to positive outcomes (see also Lefebvre et al., 2017). Thus, our data suggest that it is unlikely that acute stress spared the mechanisms involved in punishment learning, but not in reward learning, due to heightened attention towards losses relative to gains.

One potential neurocognitive explanation for such lack of effect of acute stress on loss avoidance is that D2 dopamine receptors, which mediate punishment learning (Frank & O'Reilly, 2006; Maia & Frank, 2011), are already mostly activated at baseline dopamine levels, so their activation might be affected by decreases, but less so by increases, in dopamine levels (Maia & Conceição, 2017; Möller & Bogacz, 2019). Finally, non-dopaminergic mechanisms may also be involved in punishment learning (Boureau & Dayan, 2011; Moran et al., 2018; Palminteri & Pessiglione, 2017), which may partially explain why previous studies using the same reinforcement-learning task also did not find significant effects of pharmacological manipulations of the dopaminergic system on punishment learning (Eisenegger et al., 2014; Pessiglione et al., 2006).

We found preliminary evidence that acute stress biased behavioural responding to neutral stimuli, as the choice of the high-probability "look" stimuli — compared to the stimuli with a high probability of yielding "nothing" — was augmented under stress compared to the control condition (for a detailed discussion about this finding, see Appendix 2). Although very tentative, this finding might be of relevance as excessive, aberrant spontaneous dopamine release is thought to underlie increased behavioural responding and aberrant learning for neutral stimuli (Maia & Frank, 2017; Roiser et al.,

2013). Still, whether acute stress promotes aberrant valuation of neutral stimuli via dopaminergic disturbances remains unknown and should be further investigated.

2.5. Limitations

We induced acute stress in participants, using a repetitive and uncontrollable sound, whilst they completed a reinforcement-learning task. We exposed participants to the sound during the task to ensure that acute stress was contingent on the learning processes, but at the expense of possibly confounding the induction of stress with distraction. To minimise this potential confound, we used a sound that was always constant and repetitive, as unpredictable changes in sound sequences induce distraction more robustly (Hughes, 2014; Parmentier, 2014; Parmentier et al., 2008; Sussman et al., 2003). Moreover, and also aiming to reduce the unpredictability of the sound, before starting the task participants were instructed that they would hear a sound during blocks cued with a warning signal and a red background (Hughes, 2014; Parmentier & Hebrero, 2013; Sussman et al., 2003). Besides the experimental precautions we took to avoid that the stressor acted as a distractor, our data indicated a selective effect of the task manipulation in gains. If the sound was acting as a distractor, rather than as a stressor, it would be more likely to find instead a general behavioural impairment; that is, worse performance towards both gains and losses, rather than a selective effect in gains. Relatedly, if the manipulation had acted mostly as a distractor, it would be expected that participants behaved more at random. For example, previous data indicate that when participants are asked to generate novel combinations of numbers whilst exposed to an auditory distractor, they generate combinations more at random than participants who were not exposed to the auditory distractor, suggesting that enhanced distraction during a cognitive task may enhance the ability to generate randomness (Tsushima & Nakayama, 2010). Thus, if participants were distracted by our auditory stressor, we would have expected them to make choices more at random, which would likely be captured in a reduction of the inverse temperature parameter — reflecting increased random behaviour — in the stress condition. Although our data and experimental design make it unlikely that the observed effect on reward learning was due to distractibility rather than stress, additional control conditions should be added in future studies to experimentally disentangle distraction from stress.

Finally, only men were included in this study, to avoid the potential confounding effects of menstrual-cycle-dependent variation on stress responsivity (Ossewaarde et al., 2010) and reward and punishment processing (Diekhof et al., 2020; Dreher et al., 2007). Our finding that acute stress disrupts reward learning in men seems to be in line with previous reports showing that acute stress

disrupts reward-seeking behaviour in women (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), but further studies are needed to assess whether acute stress has the same computational effects on reward and punishment learning in men and women.

2.6. Conclusions

In this chapter, we investigated how acute stress impacted reward and punishment learning using a reinforcement-learning task and computational modelling. Our data indicated that acute stress impaired choice performance towards obtaining monetary rewards, but not towards avoiding punishments. These behavioural findings are in line with previous empirical work suggesting the acute stress impairs reward learning, and possibly to a larger extent than punishment learning. Importantly, to unravel the mechanism(s) underlying such impairment, we fitted a reinforcement-learning model to participants' trial-by-trial choices. Computational modelling indicated that under acute stress participants learned more slowly from positive prediction errors — i.e., when the outcomes were better than expected. That is, acute stress reduced how quickly male adults integrated the unexpected rewarding outcomes of their choices over time. Our results seem consistent with a neurobiological framework of stress-induced dopaminergic disturbances and can contribute to a better understanding of the computational mechanisms that underlie the deleterious impact of acute stress on reward learning.

In the next chapter, Chapter 3, we will present a study investigating the neural mechanisms underlying the impact of acute stress on reinforcement learning. We will extent the current experimental design to functional magnetic resonance imaging (fMRI) in order to examine the impact of acute stress on striatal signalling of prediction errors during reward and punishment learning.

CHAPTER 3

The impact of acute stress on striatal signalling of prediction errors during reinforcement learning

3.1. Chapter Introduction

In the previous chapter, Chapter 2, we examined the impact of acute stress on reinforcement learning at the behavioural and computational levels. In the current chapter we employed model-based fMRI to assess the impact of acute stress on neural signalling of prediction errors — the key computational mechanism that drives reinforcement learning. More specifically, in the current chapter we extended our experimental design to fMRI to further test whether acute stress blunts signalling of prediction error in the striatum during reward learning, and also during punishment learning[2].

Reinforcement learning provides a powerful neurocomputational account of how individuals learn to maximise rewards and minimise punishments (Maia & Frank, 2011; Sutton & Barto, 1998). Prediction errors — the difference between an experienced and an expected outcome — are used to progressively update the values of the executed actions driving reinforcement learning (Maia & Frank, 2011; Schultz et al., 1997; Sutton & Barto, 1998). Positive prediction errors indicate that outcomes are better than expected, and negative prediction errors indicate that outcomes are worse than expected (Schultz et al., 1997). Prediction errors can therefore be used to learn which actions are advantageous or disadvantageous. For example, when an action results in an outcome that is better than expected, a positive prediction error occurs, and the value of the action is increased, leading to increased likelihood of selecting that action in the future. Prediction error signals are thought to be encoded in the phasic activity of dopamine neurons (Schultz et al., 1997). Extant evidence indicates that brain areas with dense dopaminergic projections, such as the dorsal striatum and the nucleus accumbens, show activity correlated with prediction errors (O'Doherty et al., 2004; Pessiglione et al., 2006; Schönberg et al., 2007; Valentin & O'Doherty, 2009) and that prediction error signals in the dorsal striatum correlate with behavioural performance efficacy in a reward-based task (Schönberg et al., 2007). Indeed, the striatal dopaminergic system seems to be critical for prediction-error-based reward learning (Daw & Tobler, 2014; Glimcher, 2011; Maia & Frank, 2011; Pessiglione et al., 2006).

The striatal dopaminergic system also seems to be particularly sensitive to acute stress (Booij et al., 2016; Cabib & Puglisi-Allegra, 2012; Pruessner et al., 2004). Acute stress elicits a myriad of physiological and functional changes in the brain in response to perceived adverse changes in the environment (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009), including increased dopamine release in the striatum (Abercrombie et al., 1989; Booij et al., 2016; Cabib & Puglisi-Allegra, 2012; Hermans et al., 2014; Joëls & Baram, 2009; Pruessner et al., 2004; Vaessen et al., 2015).

[2] This chapter is submitted for publication and available as a preprint in bioRxiv, Carvalheiro, J., Conceição, V., Mesquita, A. Seara Cardoso, A. (2021). Acute stress blunts prediction error signals in the dorsal striatum during reinforcement learning. *bioRxiv.* doi: 10.1101/2021.02.11.430640

Specifically, studies with non-human male animals suggest that acute stress increases aberrant spontaneous phasic dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011). Such exaggerated, aberrant spontaneous dopamine release is thought to blunt adaptive phasic dopamine responses that signal positive prediction errors (Daberkow et al., 2013; Maia & Frank, 2017; Werlen et al., 2020) and, more tentatively, negative prediction errors (Maia & Frank, 2017). Thus, stress-induced dopamine aberrant release may lead to impairments in reward learning, and more speculatively in punishment learning. Indeed, growing behavioural data suggest that acute stress impairs reward learning (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; de Berker et al., 2016; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), whereas the impact of stress on punishment learning remains more equivocal (de Berker et al., 2016; Petzold et al., 2010). In line with this behavioural evidence, we showed in the previous chapter, Chapter 2, that exposure to an acute stressor during a reinforcement-learning task resulted in impaired performance towards obtaining rewards, but not towards avoiding punishments.

Extant neural evidence on how acute stress directly affects prediction error signals in the human striatum during reward learning is still scarce (Cremer et al., 2021; Robinson et al., 2013), but indirect neural evidence indicates that women who show the greatest increase in interleukin-6 (an inflammatory marker) in response to a stressor also show the greatest reduction in signalling of prediction errors in the nucleus accumbens during reinforcement learning (Treadway et al., 2017). Moreover, we showed in the previous chapter, Chapter 2, using computational modelling, that acute stress decreases the learning rate for positive prediction errors (i.e., how quickly better-than-expected outcomes are integrated over time), which seems to be consistent with the idea that acute stress might impair reward learning by blunting neural signalling of prediction errors.

3.1.1. The current study

In this study, we aimed to investigate the impact of acute stress on striatal prediction error signalling during reinforcement learning. As mentioned above, extant literature and our data from the previous chapter suggest that acute stress disrupts reward learning to a larger extent than punishment learning. Thus, given the putative impact of acute stress on aberrant phasic dopamine release, and the role of adaptive phasic dopamine responses on prediction error signalling, we predicted that 1) acute stress would impair reward learning and, relatedly, that 2) acute stress would blunt prediction error signals in the striatum during reward learning. Additionally, given that striatal dopamine prediction error signals are also implicated in punishment learning (Oleson et al., 2012; Palminteri & Pessiglione, 2017;

Seymour et al., 2004, 2007; Watabe-Uchida et al., 2017), we explored whether and how acute stress would impact punishment learning. Finally, we explored whether acute stress would preferentially blunt positive or negative prediction error signals during reward and punishment learning.

Thirty-seven male participants completed an adapted version of a well-established reinforcement-learning task involving monetary gains and losses (Pessiglione et al., 2006) inside the MRI scanner (i.e., the task described in Chapter 2 was adapted to fMRI) whilst under acute stress and control conditions (Figure 3.1). This reinforcement-learning task disentangles reward from punishment learning and has been used to assess fluctuations in dopamine-prediction errors signals; using this task, combined with pharmacological manipulations of the dopaminergic system, Pessiglione et al. showed that dopamine-related drugs modulate prediction errors expressed in the striatum during reward (but not during punishment) learning. During half of the task blocks, participants were exposed to an uncontrollable stressor, a constant alarm (i.e., stress condition), which was previously shown in Chapter 2 to be effective in increasing self-reported stress levels and skin conductance responses rate. These blocks were alternated with blocks without the stressor (i.e., control condition). To check the success of the acute-stress manipulation, we collected self-reported stress levels at the end of each block. Participants who reported to be non-responsive to the stress manipulation (i.e., who did not report higher stress levels in the stress condition than in the control condition) were excluded from the main analyses. This resulted in a final pool of 23 participants for behavioural and neuroimaging data analyses. For completeness, we also analysed the data from the total sample, which yielded findings for the impact of acute stress on reward learning consistent with those from the analyses of the aforementioned subsample of interest (see Appendix 3 for analyses and results concerning the total sample).

To test whether acute stress impaired reward learning, we inspected the impact of acute stress on task performance. Next, to examine whether and how acute stress blunted signalling of prediction errors in the striatum, we used trial-wise prediction errors, estimated by the reinforcement-learning model described in Chapter 2 (Frank et al., 2007), as parametric modulators of striatal — dorsal striatum and nucleus accumbens — activity at the time of the outcomes in gain (i.e., reward learning) and loss (i.e., punishment learning) trials (Pessiglione et al., 2006).

3.2. Materials and Methods

3.2.1. Participants

We scanned a total of 42 right-handed male participants with no reported history of neurological or psychiatric disorders. One participant was excluded due to incidental findings and 4 participants were excluded due to technical problems during the scanning session. We assessed whether the stress manipulation increased stress levels by comparing the self-reported stress levels of the remaining 37 participants that completed the task in the stress and control conditions. Self-reported stress levels were higher in the stress condition than in the control condition in 23 participants. Thus, we report results from data analyses of these 23 participants (age range = 18 – 29 years; *M* = 23.0 years, *SD* = 3.3 years). For completeness, we also analysed the data from the total sample (n = 37); those analyses can be found in Appendix 3.

All participants provided their informed consent before the experimental session. All experimental procedures were approved by the Ethics Committee of Hospital of Braga.

3.2.2. Reinforcement-learning task

After a short practice (12 trials) outside the scanner to familiarise participants with the task timings and response keys, participants completed four blocks of an adapted version of a well-established reinforcement-learning task (Pessiglione et al., 2006) whilst inside the scanner (Figure 3.1). The task was divided in two runs, each run consisting of a stress block and a control block. Stress and control blocks were administered alternately and in a counterbalanced order across the two runs. Each block included three pairs of abstract stimuli, and each pair of stimuli was presented 24 times, totalling 72 trials per block and resulting in 288 trials per participant. In each trial, participants were asked to choose between two abstract visual stimuli so as to maximise payoffs. Each pair of stimuli was associated with a valence: one pair of stimuli was associated with gains (+0.5€ or no change), a second pair was associated with losses (-0.5€ or no change), and a third pair was associated with neutral, or non-monetary outcomes (look at a 0.5€ coin or no change). New abstract stimuli were used in each block. Each pair of stimuli was associated with a valence: one pair of stimuli was associated with gains (gain 0.5€ or no change), a second pair was associated with losses (loss 0.5€ or no change), and a third pair was associated with neutral, or non-financial outcomes (look at a 0.5€ coin or no change). The outcome probabilities were reciprocally 0.75 and 0.25 for the stimuli in each of the three pairs, and these outcome probabilities remained constant within each block. On each trial, one pair was randomly

presented on the MRI screen, with one stimulus from the pair on the left and the other on the right of a central fixation cross (the stimuli position was counterbalanced across trials). Missing choices occurred when participants did not press the response keys within 2000 ms (total of 0.20% missing choices: 8 in the stress condition and 5 in the control condition, in a total of 6624 trials) and were signalled with a "Missed" message (no other outcome was provided). Missing choices were not considered for behavioural data analyses. Participants took around 5.47 minutes, on average, to complete each block of the task. Before starting the task, participants were informed that they would be paid the amount of money obtained during their most profitable block, although they all left with the same fixed compensation (15€) for their participation. The experiment was programmed and presented with Cogent 2000 (http://www.vislab.ucl.ac.uk/cogent.php) implemented in MATLAB R2015a (MathWorks).

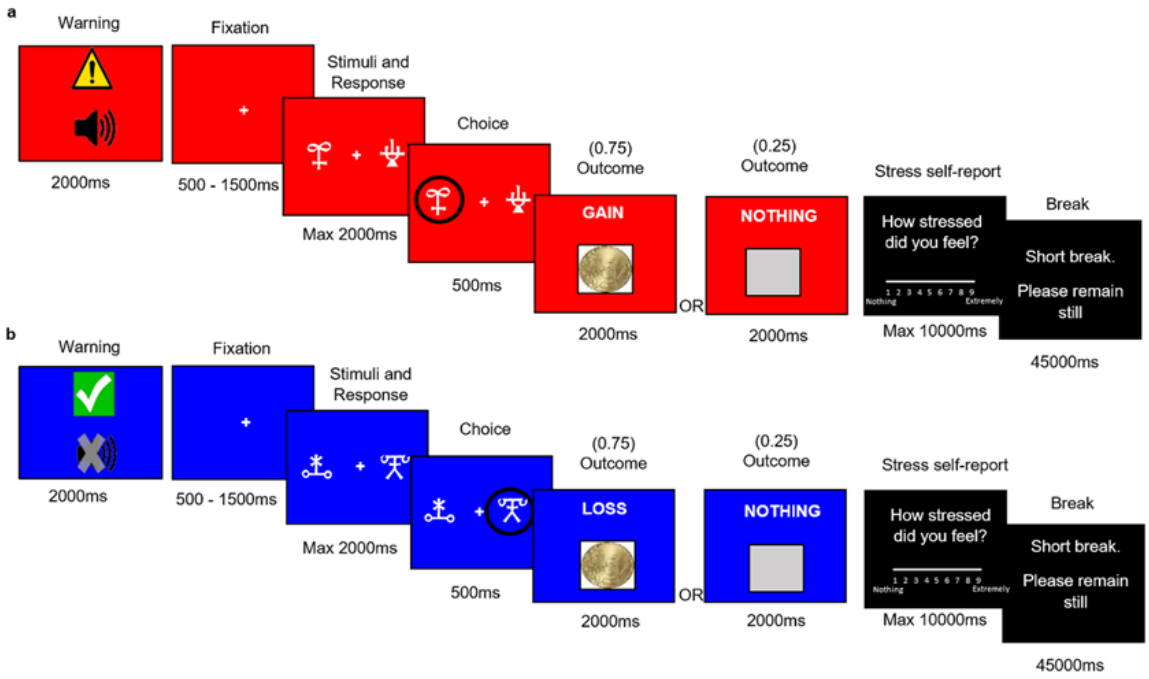


Figure 3.1. Reinforcement-learning task. Inside the scanner, participants chose between two abstract visual stimuli and observed the outcome of their choice, whilst under acute stress and under control conditions. **(a)** In the depicted stress-condition trial, the chosen stimulus was associated with a probability of 0.75 of winning 0.5€ and with a probability of 0.25 of winning nothing. The other (not chosen) stimulus was associated with a reciprocate probability of 0.75 of winning nothing and a 0.25 probability of winning 0.5€. **(b)** In the depicted control-condition example, the chosen stimulus was associated with a probability of 0.75 of losing 0.5€ and with a probability of 0.25 of losing nothing. The other (not chosen) stimulus was associated with a reciprocate probability of 0.75 of losing nothing and with a 0.25 probability of losing 0.5€. Participants completed a total of four blocks (two

stress and two control blocks), consisting of an alternation between stress and control blocks. To assess stress responses, self-reported stress levels were collected at the end of each block.

3.2.3. Acute-stress manipulation

During the scanning session, participants performed two blocks of the reinforcement-learning task whilst exposed to a stressor (i.e., stress condition) and two blocks without the stressor (i.e., control condition), as in Chapter 2. To elicit stress responses, we exposed participants to a predictable, but uncontrollable auditory stimulus: a constant alarm ("Annoying modern office building alarm.wav", retrieved from freesound.org, and programmed to loop *uninterruptedly*), played through the scanner with the volume set to the maximum. This uncontrollable sound was always constant and repetitive, to minimise the potential entanglement between stress and distraction, as evidence suggests that unpredictable changes in sound sequences seem to induce distraction more robustly (Hughes, 2014; Parmentier, 2014; Parmentier et al., 2008) (further details on the reasons behind the choice of this stressor can be found in Chapter 2). Stress blocks were further signalled by a warning sign and a red background (Figure 3.1a), and control blocks were signalled by a safe sign and blue background (Figure 3.1b).

Stress levels were assessed by asking participants at the end of each block to rate how stressed they felt during that block on a scale of 1 (nothing) to 9 (extremely). We showed in the previous chapter, Chapter 2, that this stress manipulation increased self-reported stress levels and skin conductance responses rate in men.

3.2.4. Task-performance analyses

To examine the impact of acute stress on behavioural choice performance during the reinforcement-learning task, we applied a generalized linear mixed-effects (glme) model to participants' trial-by-trial choice data (with correct and incorrect choices coded as 1 and 0, respectively). We used a "logit" link function to account for the binomial distribution of the data. We included as predictor variables in the glme model: condition (stress or control), valence (gain or loss), block number (1 or 2), trial number (1 to 24), and the interaction of interest (condition × valence). The glme included a fixed intercept, as well as random intercepts for each participant; inclusion of by-subject random slopes for all predictor variables (trial, block, condition, and valence) did not change the significance and interpretation of the condition × valence interaction, and thus we report the results from the glme model

with the simplest random effects structure. We fitted the glme model to the behavioural data using MATLAB's *fitglme* function and performed planned post-hoc analyses via contrast matrices using MATLAB's *coefTest* function.

3.2.5. fMRI data acquisition and preprocessing

A Siemens Verio 3T MRI scanner at the Clinical Academic Center – Braga with a 32-channel head coil was used to acquire a 5.5 min 3D T1-weighted anatomical scan and multislice T2*-weighted echo planar images (EPIs) with BOLD contrast. The T2* EPI sequence used the following acquisition parameters: field of view = 200 x 200 mm, matrix size = 66 x 66 mm, interleaved slice order acquisition, 42 slices with slice thickness of 3 mm with no gap between slices, flip angle of 60°, echo time of 22 ms, and repetition time of 2000 ms. Functional task-related data were acquired in two runs, separated by a short break during which participants remained inside the scanner in the same position. Fieldmaps were acquired for use in the unwarping stage of data preprocessing. Imaging data were analysed using SPM12 (www.fil.ion.ucl.ac.uk/spm). Data preprocessing followed a standard sequence: the first five volumes were discarded, and data were realigned to the sixth volume, unwarped using a fieldmap (normalized to the Montreal Neurological Institute, MNI, template), and coregistered to the participant's own anatomical image. The anatomical images were normalized using a unified segmentation procedure (Ashburner & Friston, 2005), combining segmentation, bias correction, and calculation of the wrapping or distortions needed to map the anatomical image into Montreal Neurological Institute space (i.e., deformation fields), and then applying these warps to the EPI data. The voxel size was resampled to 1.5 × 1.5 × 1.5 mm. Last, a Gaussian kernel of 8 mm FWHM was applied to smooth the images spatially.

3.2.6. fMRI data analyses

3.2.6.1. Primary general linear model

The primary fMRI analyses were based on a single general linear model, as in previous studies that used a similar reinforcement-learning task (Kumar et al., 2018; Pessiglione et al., 2006; Treadway et al., 2017). Each trial was modelled as having two time points: stimuli and outcome onsets. Note that, although our analyses focused on the prediction errors at the onset of outcomes, the onsets of stimuli were also modelled, to account for likely shared variance between BOLD signals at the time of the stimuli and outcomes. Separate regressors were created for the 6 types of stimuli [2 conditions

(stress/control) × 3 valences (gain/loss/neutral)] and the 6 types of outcomes [2 conditions (stress/control) × 3 valences (gain/loss/neutral)] in each run (see Figure S3.5 in Appendix 3 for an example of a first-level design matrix); the regressors were modelled as stick functions and convolved with SPM's canonical hemodynamic response function (Pessiglione et al., 2006). Each time point was regressed with a parametric modulator, separately for gain and loss trials: stimuli onset was modulated by the value of the chosen option, $Q_{chosen}(t)$; and, importantly, outcome onset was modulated by the prediction error, $\delta(t)$. Such values and predictions errors were estimated trial-wise using a well-established reinforcement-learning model (Frank et al., 2007), as comprehensively described in Chapter 2. Briefly, in this reinforcement-learning model, the value of the chosen stimulus, Q_{chosen}, is updated on each trial, t, according to the following learning rule: $Q_{chosen}(t+1) = Q_{chosen}(t) + \alpha * \delta(t)$. The prediction error, δ, is the difference between the actual and the expected outcome: $\delta(t) = r(t) - Q_{chosen}(t)$, where the reinforcement r is either 0.5, 0, or -0.5. The used reinforcement-learning model included separate learning rates for positive (α^+) and negative (α^-) prediction errors to account for the differential signalling of positive and negative prediction errors (Maia & Frank, 2011; Schultz et al., 1997). The reinforcement-learning model also included the inverse temperature parameter, β, which accounts for randomness in choice selection (see Appendix 3 for a detailed description of the reinforcement-learning model). Values and prediction errors were estimated using the parameters $\alpha^{\pm}$ and β estimated for each subject in each condition and used as separate parametric modulators of neural activity at the time of stimuli and outcomes, respectively, either in gain or loss trials, in each condition.

We also included an additional regressor to model missed trials, when participants did not select one of the two symbols and there was no outcome. For participants with visible headmotion in a particular scan (scans with >1 mm or 1° movement relative to the next) an extra regressor was included. Those images were removed and replaced with an image created by interpolating the two adjacent images to prevent distortion of the between-subjects mask (seven participants with visible headmotion; less than 1% of the total time series for each of them). Six headmotion parameters modelled the residual effects of headmotion. Data were high-pass filtered at 128s to remove low-frequency drifts, and the statistical model included an AR(1) autoregressive function to account for autocorrelations intrinsic to the fMRI time series.

Our primary analyses focused on prediction errors at outcome. First-level contrast images were calculated by applying appropriate linear contrasts to the parametric modulators of interest — prediction errors — and were entered into second-level analyses. Second-level one-sample *t*-tests were conducted

for each contrast using the summary-statistics approach to random-effects analysis. Regions of interest (ROI) analyses in the dorsal striatum and nucleus accumbens were conducted using an initial threshold of $p < .001$ (uncorrected) and responses were considered significant if they survived voxel-level small-volume family-wise error correction (SVC-FWE) at $p < .05$. The *a priori* ROIs — dorsal striatum and nucleus accumbens — were anatomically defined using masks. Specifically, a bilateral mask for the dorsal striatum was defined using a conjunction of the left and right putamen and caudate from the automated anatomical labelling (AAL) atlas. A bilateral mask for the nucleus accumbens was defined using a conjunction of the left and right nucleus accumbens from the Individual Brain Atlases using Statistical Parametric Mapping (IBASPM). As the bilateral nucleus accumbens mask had a slight overlap with the dorsal striatum mask, we subtracted the mask of the nucleus accumbens from the dorsal striatum mask. The atlases and the conjunctions were implemented using the WFU PickAtlas Toolbox in SPM12. Individual BOLD estimates (i.e., regression slopes) of prediction error parametric modulators were extracted from significantly activated clusters using the MarsBaR toolbox (Brett M. et al., 2002).

3.2.6.2. Subsidiary general linear models

To better understand the impact of acute stress on prediction error signals, we generated two subsidiary general linear models. Note that these two subsidiary models did not include any parametric modulators, as our purpose was to visualise how the BOLD response varied along different magnitudes of prediction errors.

For the first subsidiary model, we split prediction errors into four equally sized bins. The boundaries of the bins did not differ significantly between the stress and control conditions, in gain (all $p > .071$, paired *t*-tests) nor in loss (all $p > .13$, paired *t*-tests) trials (Table S3.2 in Appendix 3). Specifically, this first subsidiary general linear model included separate regressors for trials corresponding to each bin, in each valence (gain and loss) and condition (stress and control), modelled at the stimuli and outcome onset (as in the primary general linear model) resulting in thirty-two regressors, plus regressors for neutral trials in each condition, missing trials (if applicable) and headmotion (and visible headmotion, if applicable), for each run.

For the second subsidiary model, we split prediction errors into negative and positive. This model included separate regressors for trials corresponding to negative and positive prediction errors, in each valence (gain and loss) and condition (stress and control), modelled at the stimuli and outcome onset (as in the primary general linear model) resulting in sixteen regressors, plus regressors for neutral

trials in each condition, missing trials (if applicable) and headmotion (and visible headmotion, if applicable), for each run.

In both subsidiary models, the BOLD estimates at the outcome onset (when prediction errors occur) were extracted from the significant dorsal striatum cluster identified in the primary general linear model using the MarsBaR toolbox (Brett M. et al., 2002).

3.3. Results

3.3.1. Behavioural analyses

3.3.1.1. Manipulation check

First, we computed the difference in self-reported stress levels between the stress and control conditions in the total sample (n = 37). Twenty-three participants reported higher stress levels in the stress condition than in the control condition (Figure 3.2a). Then, we conducted analyses of variance (ANOVAs) with condition (stress and control) and block (1 and 2) as within-subject factors in those 23 participants (see Appendix 3 for manipulation-check analyses of the total sample). Self-reported stress levels differed significantly between conditions, $F(1,22) = 69.28$, $p < .001$, $\eta^2 = 0.76$, (Figure 3.2b), but there was no main effect of block, $F(1,22) = 0.008$, $p = .93$, $\eta^2 = 0$, and the condition × block interaction was also non-significant, $F(1,22) = 1.21$, $p = .28$, $\eta^2 = 0.052$. This suggests that self-reported stress levels increased with the acute stress manipulation and remained stable across blocks within each condition for those participants. Participants whose self-reported levels of stress did not increase with the acute-stress induction were excluded from the following analyses (see Appendix 3 for total sample analyses).

3.3.1.2. Task performance

To assess whether acute stress impaired reward learning, we examined the impact of acute stress on choice performance during the reinforcement-learning task (Figure 3.2c). We used a generalized linear mixed-effects (glme) model, which accounted for the binomial distribution of the trial-by-trial data (correct or incorrect responses). The glme model included condition (stress or control), valence (gain or loss), block number (1 or 2), trial number (1 to 24), and the interaction of interest (condition × valence) as predictor variables. We found a significant condition × valence interaction, $\beta = -0.39$, $p = .0038$, 95% confidence interval (CI) = [-0.66, -0.13]) (Figure 3.2d). Planned post-hoc analyses showed that under stress, comparatively to the control condition, participants performed significantly

worse when learning to obtain gains, $F(1, 4400) = 20.23$, $p < .001$), but not when learning to avoid losses, $F(1, 4400) = 0.32$, $p = .57$). In sum, acute stress selectively impaired choice performance towards monetary gains during the reinforcement-learning task.

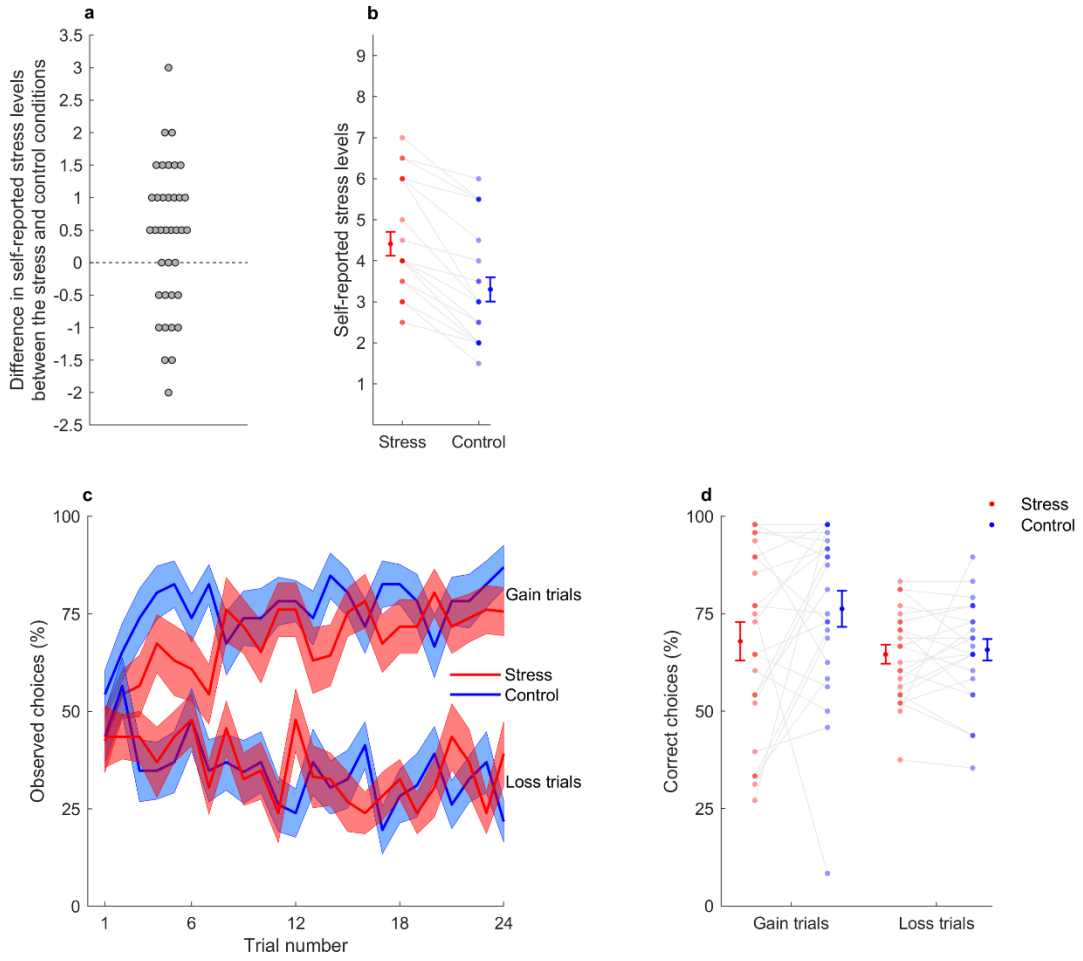


Figure 3.2. Manipulation check and task performance. (a) Difference in self-reported stress levels between the stress and control conditions (averaged across blocks). Each grey dot represents a participant (n = 37). Participants who reported higher stress levels in the stress than in the control condition correspond to the dots above the horizontal dashed line (n = 23)**. (b)** Self-reported stress levels in stress (red) and control (blue) conditions (averaged across blocks) in the pool of participants that reported higher stress levels in the stress than in the control condition (n = 23). **(c)** Learning curves represent the trial-by-trial percentage of participants (n = 23) who chose the correct gain stimulus (associated with a probability of 0.75 of winning 0.5€; upper part of the graph) and the incorrect loss stimulus (associated with a probability of 0.75 of losing 0.5€; lower part of the graph), in the stress and control conditions. Each central line represents the mean and each filled area the ± standard error of the mean. **(d)** Percentage of correct choices per participant (n = 23) in gain and

loss trials, across the stress and control conditions (averaged across blocks). In graphs b and d, connected dots represent data points from the same participant, and more transparent (opaque) dots represent less (more) overlapping data points; error bars displayed on the sides of those scatter plots indicate the mean ± standard error of the mean.

3.3.2. fMRI analyses

To examine the impact of acute stress on prediction error signalling in the striatum during reinforcement learning, we generated a primary fMRI general linear model that included prediction errors as parametric modulators of BOLD response in the striatum (dorsal striatum and nucleus accumbens) at the time of the outcomes in gain and loss trials, in the stress and control conditions (see Materials and Methods and Figure S3.5 in Appendix 3 for further details on the primary general linear model). Prediction errors were estimated in the stress and control conditions using the reinforcement-learning model described in Chapter 2, which has been extensively used to investigate the behavioural and neural impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans (Diederen et al., 2017; Doll et al., 2011; Frank et al., 2007; Frank & Fossella, 2011; Grogan et al., 2017; Rutledge et al., 2009) (see Appendix 3 for computational modelling methods and results). Parametric analyses incorporating prediction errors allow a more precise estimation of how brain activity fluctuates during learning compared to examination of outcome-associated activation alone (O'Doherty et al., 2007).

As expected, we observed that BOLD response in the dorsal striatum and nucleus accumbens — regions consistently shown to respond to unexpected rewards and punishments (Garrison et al., 2013; Pessiglione et al., 2006; Schönberg et al., 2007; Seymour et al., 2004, 2007) — scaled parametrically with the magnitude of prediction errors at the time of the outcomes in gain and loss trials in the stress and/or control conditions. Specifically, we identified a positive parametric modulation of prediction errors in the dorsal striatum bilaterally (i.e., the magnitude of the prediction errors correlated positively with BOLD response in this region on a trial-by-trial basis), during gain and loss trials, both in the stress and control conditions [all $Z > 4.12$, $p < .05$, voxel-level small-volume family-wise error corrected (SVC-FWE)]. We also found a positive parametric modulation of prediction errors in the nucleus accumbens during gain trials, both in the stress and control conditions, and during loss trials in the control condition (all $Z > 3.63$, $p < .05$, SVC-FWE; see Table S3.1 in Appendix 3 for whole-brain and all SVC-FWE results).

After confirming that striatal activity scaled parametrically with the magnitude of prediction errors, we inspected the impact of acute stress on prediction error signalling in the striatum. To examine whether acute stress would blunt signalling of prediction errors in the striatum during reward learning, first we tested the main effect of stress using the control > stress contrast for the parametric modulators of prediction errors at the time of the outcomes delivered across gain and loss trials in each condition. A significant main effect would mean that acute stress decreased prediction error signals across gain and loss trials. Second, we tested the contrast for the condition (stress or control) × valence (gain or loss) interaction. A significant interaction would mean that acute stress affected prediction error signals differently in gain and loss trials.

The control > stress contrast showed a significant main effect of stress on the parametric modulation of prediction errors in the dorsal striatum, [$x = 32$, $y = 0$, $z = 12$], $Z = 4.08$, $k = 26$, $p = .040$ SVC-FWE (Figure 3.3a), meaning that prediction error signals were decreased in the stress condition compared with the control condition, both in gain and loss trials (Figure 3.3b). Confirmatory one sample *t*-tests comparing the parameter estimates (i.e., the regression slopes from the primary general linear model) extracted from the identified dorsal striatum cluster against zero, indicated that the parametric modulation of prediction errors was significantly higher than zero in the control condition, both for gain and loss trials (both $p < .023$), but not in the stress condition (both $p > .33$) (Figure 3.3b). We did not observe any significant responses for the parametric modulation of prediction errors in the nucleus accumbens for the control > stress contrast (nor for the inverse contrast, stress > control).

For the contrast that tested the condition × valence interaction, we did not find any significant activations in the dorsal striatum nor in the nucleus accumbens, indicating that the effect of acute stress on prediction error signals was not dependent on the valence of the trial. This non-significant interaction, together with the significant main effect described above, indicates that acute stress blunted prediction errors in the dorsal striatum both in gain and loss trials.

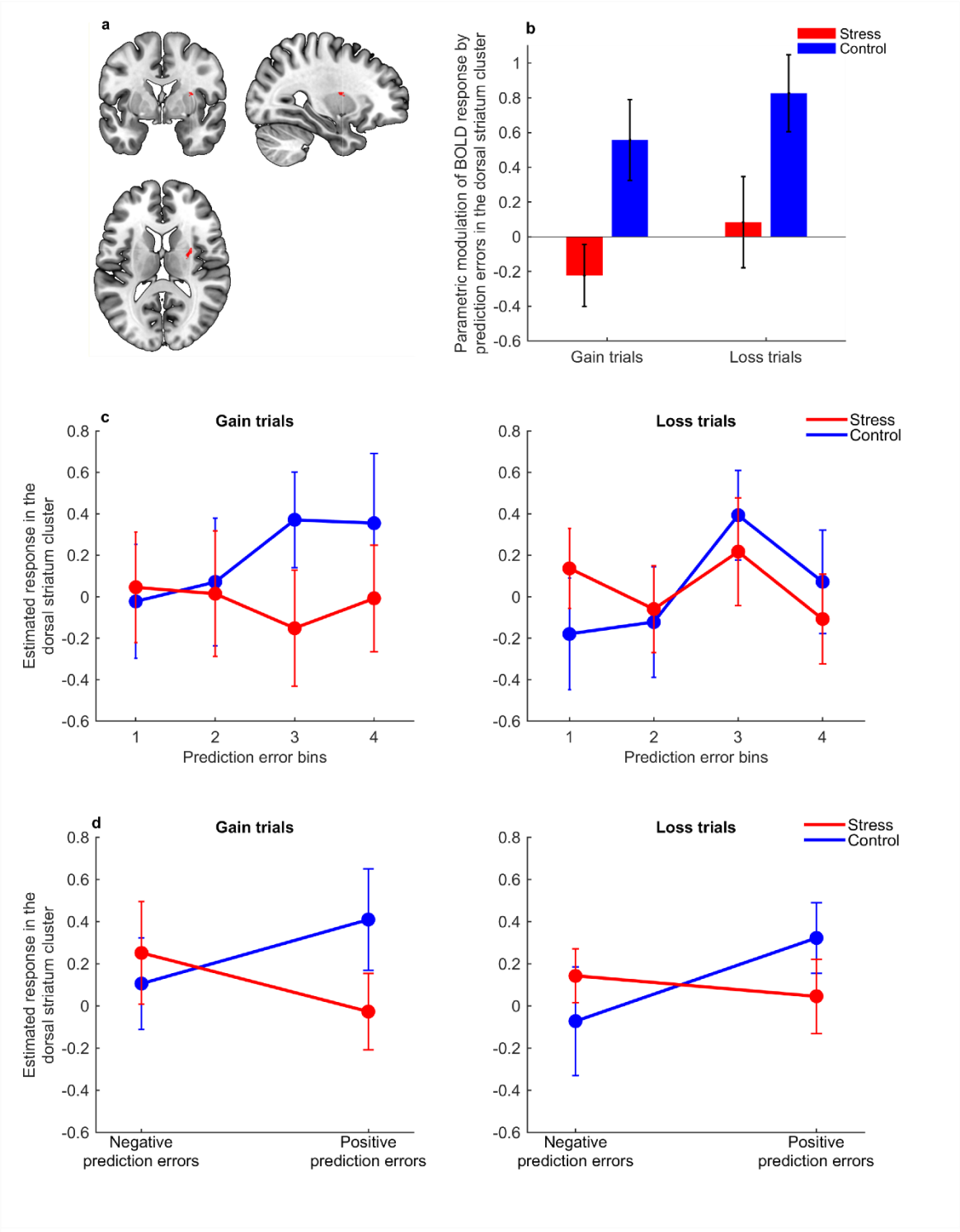


Figure 3.3. Effects of acute stress on prediction error signalling in the dorsal striatum. **(a)** Cluster in the dorsal striatum where the modulation of prediction errors at the time of the outcome was significantly decreased in the stress condition compared with the control condition (SVC-FWE, $p < .05$). **(b)** Bars depict parameter estimates (i.e., regression slopes) for the BOLD response at the dorsal striatum cluster [from (a)] modulated by trial-by-trial prediction errors, in gain

and loss trials, across the stress (red) and control (blue) conditions (n = 23). **(c, d)** Graphs illustrate the modulation of BOLD response by prediction errors at the time of the outcome in gain (left) and loss (right) trials, in the dorsal striatum cluster identified in the primary general linear model [from (a)], in the stress (red) and control (blue) conditions (n = 23). BOLD response estimates from the dorsal striatum cluster were extracted for each participant. Error bars indicate the mean ± standard error of the mean. In **(c)**, data for illustrative graphs were derived from a subsidiary model where trials were divided into quartiles of magnitude of prediction errors (with the lowest and highest magnitudes corresponding to bins 1 and 4, respectively). In **(d)**, data for illustrative graphs were obtained from a second subsidiary model where trials were divided into negative and positive prediction errors.

To illustrate the effect of acute stress on the parametric modulation of prediction errors in the dorsal striatum, we conducted a subsidiary general linear model from which we extracted estimates of BOLD response in the dorsal striatum cluster identified in the primary general linear model (cluster represented in Figure 3.3a) across trials of different categories of prediction error magnitudes. Specifically, in this subsidiary model, trials from each condition (i.e., stress and control) and valence (i.e., gain and loss) were further divided into four bins corresponding to quartiles of magnitude of prediction errors (see Table S3.2 in Appendix 3 for a description of each bin). BOLD response estimates at the outcome onset were extracted from the dorsal striatum cluster for each subject (see Material and Methods for full description) and plotted to illustrate the variation in the BOLD response in the dorsal striatum cluster along the magnitude of prediction errors. The blunting main effect of stress (both in gain and loss trials) on prediction error signals in the dorsal striatum seemed to be mostly driven by decreased signalling of prediction errors of higher magnitude (Figure 3.3c). It is important to note that the 1st bin and 4th bins roughly corresponded to negative and positive prediction errors, respectively (Table S3.2 in Appendix 3), which suggests that acute stress mostly decreased positive prediction error signals.

To further explore whether acute stress had preferentially blunted signalling of positive prediction errors during reward and punishment learning, we conducted an additional parametric modulation model similar to the primary general linear model, but this time we modelled positive and negative prediction errors separately for either gain or loss trials. We did not find any parametric modulation of the dorsal striatum or nucleus accumbens response by positive nor negative prediction errors, very possibly due to reduced variance within each parametric modulator. Therefore, and for

completeness, we performed a second subsidiary general linear model (see Material and Methods for full description). In this subsidiary model, trials from each condition (i.e., stress or control) and valence (i.e., gain or loss) were further divided according to the prediction error valence (i.e., positive or negative). We extracted estimates of BOLD response in the dorsal striatum cluster (cluster represented in Figure 3.3a) at the outcome onset, when positive or negative prediction errors occurred, and performed an ANOVA on those estimates derived from the subsidiary model. We found a significant condition × prediction error valence interaction, $F(1,22) = 8.84$, $p = .007$, $\eta^2 = 0.29$ (Figure 3.3d), indicating that acute stress differently affected positive and negative prediction errors, both in gain and loss trials (the condition × prediction error valence × trial valence interaction was non-significant, $F(1,22) = 0.047$, $p = .83$, $\eta^2 = 0.0020$. Post-hoc planned comparisons were non-significant, but inspection of effect sizes suggested that acute stress decreased positive prediction error signals (paired *t*-tests in gain trials: $t(22) = -1.36$, p = .19, Cohen's $d = -0.28$; in loss trials: $t(22) = -1.12$, p = .27, Cohen's $d = -0.23$) to a larger extent than negative prediction errors (paired *t*-tests in gain trials: $t(22) = 0.48$, $p = .64$, Cohen's $d = 0.10$; in loss trials: $t(22) = 0.86$, $p = .40$, Cohen's $d = 0.18$) (Figure 3.3d), in line with the previous subsidiary analysis of prediction error bins depicted in Figure 3.3c.

In sum, we found that the BOLD response in the dorsal striatum scaled parametrically with the magnitude of prediction errors during reward and punishment learning, both under acute stress and control conditions. More importantly, we found that signalling of prediction errors in the dorsal striatum was blunted by acute stress, with subsidiary data suggesting that acute stress mostly decreased positive prediction errors signals.

3.4. Discussion

Previous studies have found that acute stress impacts reinforcement learning (Porcelli & Delgado, 2017). Yet, the mechanisms that underlie the impact of acute stress on reinforcement learning are still poorly understood. Acute stress is ubiquitous in modern day-to-day life. Understanding whether and how acute stress influences our ability to learn from the rewarding and punishing outcomes of our choices might be important for the development of interventions that target the debilitating effects of acute stress. In this study, we investigated whether and how acute stress impacts reinforcement learning using behavioural and model-based fMRI data. Acute stress alters human striatal dopaminergic functioning (Booij et al., 2016; Cabib & Puglisi-Allegra, 2012; Pruessner et al., 2004; Vaessen et al., 2015), and striatal dopaminergic functioning plays a key role in signalling of prediction errors (Daw & Tobler, 2014; Glimcher, 2011; Maia & Frank, 2011; Pessiglione et al., 2006) — the

result of a positive or negative difference between obtained and expected outcomes — which drive reward and punishment learning. Thus, we set out to investigate the impact of acute stress on striatal signalling of prediction error signals during reward and punishment learning.

In line with extant literature (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Paret & Bublatzky, 2020), we replicated the finding from the previous chapter, Chapter 2, that acute stress impairs performance towards rewards in a reinforcement-learning task. Importantly, we also found that such behavioural impairment was accompanied by blunted signalling of prediction errors in the dorsal striatum. Additionally, even though we did not find evidence of the impact of acute stress on punishment learning at the behavioural level, neural data indicated that acute stress also blunted prediction error signals in the dorsal striatum during punishment learning. Subsidiary analyses indicated that acute stress may blunt preferentially positive prediction error signals, which might explain the differential impact of acute stress in reward and punishment learning.

Our finding that acute stress blunted signalling of prediction errors — and mostly positive prediction errors — in the dorsal striatum during reward learning is consistent with a neurobiological framework of stress-induced dopamine disruptions. Prediction errors are encoded in phasic activity of dopaminergic neurons (Daw & Tobler, 2014; Glimcher, 2011; Schultz et al., 1997). Phasic bursts of dopaminergic neurons are thought to adaptively encode positive prediction errors, whereas dopamine dips have been associated with the adaptive encoding of negative prediction errors (Daw & Tobler, 2014; Schultz et al., 1997). However, phasic dopamine responses do not seem to be always adaptive, and there is evidence that dopamine can be phasically released in an aberrant spontaneous manner (Grace, 2016; Maia & Frank, 2017; Sulzer et al., 2016; Wightman et al., 2007). Relatedly, studies with non-human male animals suggest that acute stress induces aberrant spontaneous dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011). It is therefore possible that, if acute stress increases aberrant spontaneous phasic dopamine release, then phasic dopamine release that signals positive prediction errors following burst firing of dopaminergic neurons is less easily differentiated from background fluctuations in dopamine levels, resulting in a low signal to noise ratio (Grace, 2016; Sulzer et al., 2016). In addition, if there is increased aberrant spontaneous release of dopamine, then less dopamine may be available to be released from dopaminergic neurons when positive prediction errors occur (Sulzer et al., 2016). Thus, stress-induced aberrant dopamine release may indirectly or directly blunt positive prediction errors that signal unexpected rewards (Daberkow et al., 2013; Maia & Frank, 2017; Werlen et al., 2020), resulting in impaired reward learning. Furthermore, we previously showed in Chapter 2, using computational modelling, that acute stress

decreases the learning rate for positive prediction errors, which is in striking agreeement with our neuroimaging data. Our finding that acute stress blunts prediction errror signals — and preferentially positive prediction errors — in the dorsal striatum during reward learning is thus consistent with a neurobiological framework of stress-induced striatal dopaminergic disturbances and might explain why individuals have difficulties in updating their behaviour in response to unexpected rewards when under acute stress.

In this study, we also explored how acute stress affected punishment learning, although the behavioural evidence for an effect of acute stress on punishment learning seems less robust than for an effect of stress on reward learning (Porcelli & Delgado, 2017). In line with the findings reported in the previous chapter, Chapter 2, using the same stress manipulation and an adapted version of the same reinforcement-learning task in an independent sample, we found no evidence of a behavioural impairment of acute stress on punishment learning. This suggests that punishment learning may not be affected by acute stress to the same extent as reward learning is. However, our neuroimaging data showed a main effect of stress on the signalling of prediction errors in the dorsal striatum, indicating that acute stress also decreased prediction error signals during punishment learning. Further subsidiary analyses suggest that the main effect of stress on prediction errors seem to be mostly explained by decreased signalling of positive prediction errors. It is thus possible that acute stress compromises the ability to use dopamine phasic bursts that signal positive prediction errors but not to use dopamine dips that encode negative prediction errors. Indeed, empirical evidence suggests that aberrant spontaneous dopamine release decreases striatal adaptive phasic dopamine responses that signal positive prediction errors (Daberkow et al., 2013; Werlen et al., 2020), and, only more speculatively, negative prediction errors (Maia & Frank, 2017). Although positive prediction errors can also occur during punishment learning, in simple reinforcement-learning tasks, such as ours, punishment learning seems to be largely driven by negative prediction errors (Palminteri & Pessiglione, 2017). Consequently, stress-induced disruptions in positive prediction errors might not necessarily be reflected in impaired learning from punishments. Finally, non-dopaminergic mechanisms may also be involved in punishment learning (Bayer & Glimcher, 2005; Boureau & Dayan, 2011), which might partially explain why previous studies using similar reinforcement-learning tasks also did not find effects of pharmacological manipulations of the dopaminergic system on punishment learning (Eisenegger et al., 2014; Pessiglione et al., 2006).

We found evidence for stress-induced blunted prediction errors in the dorsal striatum, but not in the nucleus accumbens. The dorsal striatum has been associated with reward-based action selection (Lohrenz et al., 2007; O'Doherty et al., 2004; Schönberg et al., 2007) — and is thus thought to play a

key role in instrumental learning tasks, such as ours, by maintaining information about action-contingent response-reward associations to guide future choices based on the outcomes of past ones — whereas the ventral portion of the striatum, the nucleus accumbens, has been more implicated in classical conditioning (O'Doherty et al., 2004). By blunting prediction error signals in the dorsal striatum, acute stress may thus impair learning of stimulus-response-reward associations, which are crucial to perform our reinforcement-learning task; yet, it is possible that acute stress affects distinct regions of the striatum, and the functions they support, differently (Robinson et al., 2013; Treadway et al., 2017). Further studies are required to better understand the impact of acute stress on the neurocomputational mechanisms of classical and instrumental punishment learning and could focus not only on the striatum, but also on other brain areas, such as the anterior insula or habenula, which are known to play a key role in punishment avoidance (Garrison et al., 2013; Pessiglione et al., 2006; Salas et al., 2010).

Some limitations of this research should be noted. In this study, we induced acute stress in participants, using a repetitive and uncontrollable sound, whilst they completed a reinforcement-learning task. Such stressor was previously validated outside the scanner, in Chapter 2, using self-reported stress levels and concomitant skin conductance responses. In the current study, the stressor increased self-reported stress levels, although to a lesser extent than outside the scanner. One potential explanation for this discrepancy is that, inside the scanner, the stressor was not as salient as it was outside the scanner. We used additional visual cues as warning signals and coloured backgrounds, in an attempt to amplify the effects of stress manipulation. Importantly, the manipulation seemed to be particularly effective in a proportion of participants, based on their self-reported stress levels. In this study we used a self-report measure to capture the 'subjective state of sensing potentially adverse changes in the environment' (Joëls & Baram, 2009). Although we cannot assume that the same processes governing physiological stress explain subjective feelings of stress (Ali et al., 2017; Campbell & Ehlert, 2012), the subjective emotional experience of stress plays an important role in the stress response (Birk, 2020). Due to technical limitations, we did not analyse physiological responses in this study [but see Chapter 2 for a skin conductance measure obtained for the same stress manipulation]; still, inclusion of such variables in future research could be valuable.

To avoid the potential confounding effects of menstrual-cycle-dependent variation on stress responsivity (Ossewaarde et al., 2010) as well as on reward and punishment learning (Diekhof & Ratnayake, 2016; Dreher et al., 2007), only men were included in this study, as in the study described in Chapter 2. Our behavioural findings seem to be in line with previous reports showing that acute

stress disrupts reward learning in women (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), but further studies are needed to assess whether acute stress affects the same neurocomputational mechanisms of reinforcement learning in both men and women. Furthermore, given that our stress manipulation did not increase stress levels in all participants, future studies should explicitly account for individual differences, and for the modulatory role of those individual differences on the neural mechanisms that underlie reinforcement learning under acute stress.

3.5. Conclusions

This chapter described a neuroimaging study where we investigated the impact of acute stress on striatal signalling of prediction errors during reinforcement learning using model-based fMRI. We employed an adapted version of the reinforcement-learning task and the same stress manipulation used in the previous chapter, Chapter 2. We were able to confirm the impairing effect of acute stress on performance towards obtaining monetary rewards, but not towards avoiding losses, in line with the behavioural findings reported in Chapter 2. Importantly, we found that acute stress blunted prediction error signals in the dorsal striatum during reward and punishment learning. Subsidiary analyses suggested that acute stress preferentially blunted signalling of positive prediction errors. This might explain why individuals learned worse from the rewarding, but not from the punishing, outcomes of their choices when under acute stress. These behavioural and neural findings are consistent with a neurobiological framework of stress-induced dopamine disruptions. In sum, this chapter extends behavioural evidence that acute stress impairs reward learning, possibly to a larger extent than punishment learning. Importantly, it suggests a neurocomputational mechanism by which acute stress may impair reward learning. Our findings can thus contribute to a better understanding of the neural mechanisms that underlie the deleterious impact of acute stress on reward learning.

In the previous (Chapter 2) and present (Chapter 3) chapters, we noticed a substantial inter-individual variability in the extent to which acute stress impaired reward learning and blunted signalling of prediction errors in the dorsal striatum. In the next chapters, Chapter 5 and 6, we will describe two studies in which we used the data sets from Chapters 2 and 3, respectively, to examine whether and how individual differences in psychopathic traits modulated the impact of acute stress on reinforcement learning and underlying striatal signalling of prediction errors.

CHAPTER 4

Individual differences in the impact of acute stress on reinforcement learning: The role of psychopathic traits

4.1. Chapter Introduction

In the previous chapters, Chapters 2 and 3, we found that acute stress impaired reward learning, but seemingly not to the same extent in all individuals. Such variations in the extent to which acute stress impairs reward learning might be related to individual differences in personality traits. Personality traits, such as psychopathic traits, have been associated with altered reinforcement-learning processes (R. J. R. Blair, 2005) and stress responsivity (Thompson et al., 2014), and can thus constitute a source of individual differences that modulate the extent to which reinforcement learning is affected by acute stress. However, whether and how acute stress affects reinforcement learning in individuals with different levels of psychopathic traits is not well-understood. In this chapter, we used the data set from Chapter 2 to examine whether and how individual differences in psychopathic traits modulated the impact of acute stress on reward and punishment learning[3].

Psychopathy is a personality disorder characterised by abnormal affective-interpersonal traits, such as lack of empathy and manipulativeness, and by deviant lifestyle-antisocial characteristics, such as impulsiveness and antisocial behaviour (R. J. R. Blair et al., 2005; Hare, 2003; Hare & Neumann, 2008). Extant data suggests that psychopathic traits are continuously distributed in the general population (Neumann & Hare, 2008) and that individual differences in psychopathic traits in the general population can be mapped onto dysfunctional behavioural and neural processes systematically reported in individuals with a diagnosis of psychopathy (Seara-Cardoso & Viding, 2015).

Evidence from clinical populations suggests that psychopathy is associated with atypical reinforcement learning (K. S. Blair et al., 2006; R. J. R. Blair, 2013b, 2017, 2019; R. J. R. Blair et al., 2005; Brazil et al., 2009; Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Finger et al., 2011; Gregory et al., 2015; Newman & Kosson, 1986; Von Borries et al., 2010; White et al., 2013), which may ultimately result in maladaptive behaviour (R. J. R. Blair, 2007, 2017; Brazil et al., 2011). Reinforcement learning implicates that individuals learn to select actions that maximise rewards and minimise punishments by gradually learning the values of those actions (Sutton & Barto, 1998). This requires the ability to learn from past actions: choices that result in rewards should be repeated, whereas choices that result in punishments should be avoided. Growing behavioural (K. S. Blair et al., 2006; R. J. R. Blair et al., 2004; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Newman & Kosson, 1986; Oba et al., 2019) and neural (R. J. R. Blair, 2013b, 2019; Finger et al., 2011; Gregory et al., 2015; Pujara et al., 2014; Von Borries et al., 2010;

[3] This chapter is under peer-review as Carvalheiro, J., Conceição, V., Mesquita, A. Seara Cardoso, A. (2021). Psychopathic traits and reinforcement learning under acute stress.

White et al., 2013) data suggest that this ability to learn from past actions, and consequently choose between stimuli associated with rewards and/or punishments, is dysfunctional in psychopathy. This suggests that individuals with high psychopathic traits might have difficulties in learning to adjust their behaviours based on the rewarding and punishing outcomes of their choices. Despite the extant evidence, the extent to which different levels of psychopathic traits in the general population are associated with dysfunctional reinforcement learning remains more equivocal (Brazil, Hunt, et al., 2013; Oba et al., 2019).

Individuals often need to make choices, and learn from their outcomes, in non-ideal situations, such as when under acute stress. In line with previous literature (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Paret & Bublatzky, 2020), we showed in Chapters 2 and 3 that exposure to an acute stressor during a reinforcement-learning task resulted in impaired performance towards obtaining rewards. Although it is possible that acute stress impairs punishment learning (de Berker et al., 2016; Petzold et al., 2009), we did not detect any effects of acute stress towards avoiding losses. Previous research indicates that individuals with high levels of psychopathic traits present blunted stress reactivity (Cima et al., 2008; Holi et al., 2006; House & Milligan, 1976; Johnson et al., 2015; Lidberg et al., 1978; Loney et al., 2006; Thompson et al., 2014). The influential low-fear hypothesis theorises that psychopathy is linked to a reduced ability to experience fear (Lykken, 1957) and, in particular, to deficits in automatic threat detection and responsivity (Hoppenbrouwers et al., 2016). Other studies propose that the blunted responding to threatful stimuli in individuals with psychopathy results from a deficit in their ability to reallocate attention to environmental stimuli once they are focused on a goal, such as obtaining rewards (Baskin–Sommers et al., 2012; Newman, 1998; Newman et al., 2010; Newman & Baskin-Sommers, 2012). As such, individuals with high psychopathic traits might be more insensitive to stimuli that typically induce acute stress which, in turn, disrupt reinforcement learning processes. Yet, little is known about whether psychopathic traits can interact with acute stress on reward and punishment learning.

4.1.1. The current study

In this study, we aimed to investigate the extent to which psychopathic traits modulated the impact of acute stress on reinforcement learning. We used the data set from Chapter 2, originally collected to assess how acute stress affects reward and punishment learning. Participants performed a well-established reinforcement-learning task (Pessiglione et al., 2006) involving monetary gains (i.e., rewards) and losses (i.e., punishments) whilst under acute stress and in control conditions, as

previously described in Chapter 2, and completed the Self-Report Psychopathy-Short Form [(Paulhus et al., 2016); Portuguese version (Seara-Cardoso et al., 2020)]. To provide a more comprehensive understanding of the role of psychopathic traits in modulating the impact of acute stress on reward and punishment learning, we also explored the contribution of each dimension of psychopathy (affective-interpersonal and lifestyle-antisocial) (Hare, 2003; Hare & Neumann, 2008).

Considering that psychopathic traits have been associated with both abnormal reinforcement learning and atypical stress responsivity, we hypothesised that psychopathic traits would modulate the impact of acute stress on reward and punishment learning. Specifically, given that acute stress is thought to impair reward learning, and possibly punishment learning, and that psychopathy has been characterized by reduced stress responsivity, we expected that the impairing effect of acute stress on reward and punishment learning would decrease as psychopathic traits increased.

4.2. Materials and Methods

4.2.1. Participants

Sixty-two male healthy participants were recruited at University of Minho, as previously described in Chapter 2. No participants were excluded from data analyses, although we conducted confirmatory analyses excluding potential outliers to ensure that our results were not driven by extreme values. Participants provided informed consent before the experimental session. All experimental procedures were approved by the Ethics Committee of University of Minho.

4.2.2. Self-Report Psychopathy-Short Form (SRP-SF)

The Self-Report Psychopathy-Short Form [SRP-SF; (Paulhus et al., 2016); Portuguese version (Seara-Cardoso et al., 2020)] is a 29-item questionnaire assessing psychopathic traits. The instrument follows the same factor structure as the Psychopathy Checklist–Revised (Hare, 2003) and assesses four facets of psychopathy — affective, interpersonal, lifestyle, antisocial — which can also be modelled in terms of the traditional two-factor dimensions of psychopathy: affective-interpersonal and lifestyle-antisocial. Items are answered on a 5-point Likert scale that ranges from 1 (disagree strongly) to 5 (agree strongly). Scores of each dimension are obtained by summing up the corresponding individual item scores. The affective-interpersonal dimension taps the affective aspects of psychopathy, such as compromised empathy or lack of guilt and concern about others (e.g., “I never feel guilty over hurting others” and “People sometimes say that I’m cold-hearted”), and interpersonal and dissocial

characteristics, such as pathological lying and manipulation (e.g., "I would get a kick out of scamming someone", and "I have pretended to be someone else in order to get something"). The lifestyle-antisocial dimension relates to impulsive (e.g., "I admit that I often mouth off without thinking" and "I've often done something dangerous just for the thrill of it") and antisocial behaviours (e.g., "I have threatened people into giving me money, clothes, or makeup" and "I was convicted of a serious crime"). The antisocial subscale includes eight items, but only seven items are considered, as the item "gang activity" is omitted in community samples due to its low variability, as is the case of the current sample. The other subscales are also composed of seven items, resulting in 14 items in each dimension. Therefore, SRP-SF scores were computed from a total of 28 items.

Overall, the distributions of the levels of psychopathic traits measured by the SRP-SF in the current sample were consistent with other published studies on the SRP-SF with large normative samples (Gordts et al., 2017; Seara-Cardoso et al., 2020) (Table 4.1; see also Figure S4.1 in Appendix 4 for histograms with the distribution of psychopathic traits in our sample). Calculation of the Chronbach's alpha showed adequate reliability of the SRP-SF total scale ($\alpha = 0.87$) and of both dimension subscales (affective-interpersonal, $\alpha = 0.81$; lifestyle-antisocial, $\alpha = 0.75$). The two dimensions were significantly correlated ($r_s = 0.65$, $p < .001$).

Table 4.1. Descriptive statistics (Mean and Standard deviation, Median and Minimum and Maximum values) of Self-Report Psychopathy-Short Form (SRP-SF) total scores and dimensions (affective-interpersonal and lifestyle-antisocial).

	Total	**Dimensions**	
		Affective-interpersonal	**Lifestyle-antisocial**
Mean (SD)	53.73 (12.74)	29.00 (7.84)	24.73 (5.99)
Median [Min, Max]	53 [34, 95]	27.5 [16, 54]	25 [15, 41]

4.2.3. Reinforcement-learning task

Participants completed four blocks of an adapted version of a well-established reinforcement-learning task (Pessiglione et al., 2006), as comprehensively described in Chapter 2. Briefly, each block included three pairs of abstract stimuli, and each pair of stimuli was presented 40 times, resulting in 120 trials per block and 480 trials per participant. In each trial, participants were asked to choose between two abstract visual stimuli so as to maximise gains and minimise losses. Each pair of stimuli

was associated with a valence: one pair of stimuli was associated with gains (+0.5€ or nothing), a second pair was associated with losses (-0.5€ or nothing), and a third pair was associated with neutral, or non-monetary outcomes (look at a 0.5€ coin or nothing). The two outcome probabilities were reciprocally 0.8 and 0.2 for the stimuli in each of the three pairs. Given that neutral trials were not associated with monetary outcomes, and thus there were no correct/incorrect responses during neutral trials, we did not analyse participants performance during this type of trial, meaning that our analyses included 2/3 of the trials, i.e., gain and loss trials. Further details on the task can be found in Chapter 2.

4.2.4. Acute-stress manipulation

During the experimental session, participants performed two blocks of the reinforcement-learning task whilst exposed to a stressor, a sound alarm (i.e., stress condition), and two blocks without the stressor (i.e., control condition). Stress and control blocks were administered alternately and in counterbalanced order, as described in detail in Chapter 2. In Chapter 2, we confirmed that participants from this data set reported higher stress levels and exhibited augmented skin-conductance response (SCR) rate in the stress condition compared with the control condition.

To assess how psychopathic traits were associated with stress responsivity, we correlated psychopathic traits with the difference in self-reported stress levels and SCR rate (averaged across blocks) between the stress and control conditions. We performed Spearman correlations, which are more robust to extreme values. To assess whether the correlations for each dimension were significantly different from each other, we applied the Meng's et al., 1992 approach using the "cocor" package in R (http://www.r-project.org).

4.2.5. Task-performance analyses

To examine how psychopathic traits modulated the impact of acute stress on choice performance during the reinforcement-learning task, we applied generalized linear mixed-effects (glme) models to participants' trial-by-trial choice data (with correct and incorrect choices coded as 1 and 0, respectively). We used a "logit" link function to account for the binomial distribution of the data. As predictor variables in the glme models, we included individual scores of psychopathic traits (continuous SRP-SF scores), condition (0 or 1 for control and stress conditions, respectively), valence (0 or 1 for loss and gain trials, respectively), block number (0 or 1, for blocks 1 and 2, respectively), and trial number

(1 to 40), and tested the interactions between SRP-SF scores, condition, and valence. Block and trial numbers were included as covariates of noninterest to control for potential temporal effects. We included by-subject random intercepts and by-subject random slopes for all predictor variables (condition, valence, block, and trial).

We generated a glme model to assess the interaction between SRP-SF total scores, condition, and valence, and two other glme models to assess the interaction between SRP-SF scores in each dimension of psychopathy (affective-interpersonal and lifestyle-antisocial), condition, and valence. We fitted the glme models to the behavioural data using the "lme4" package in R (http://www.r-project.org), and we used the "predict" function to obtain the logit-glme-model-estimated probability of choosing the "correct" stimulus (i.e., the stimulus associated with a 0.8 probability of winning in gain trials and the stimulus associated with a 0.2 probability of losing in loss trials). To further understand the nature of the interactions, we performed simple slope and Johnson-Neyman analyses using the "interactions" package in R. Simple slope analyses tested the significance of the associations between SRP-SF scores and behavioural performance during gain or loss trials in each condition, and Johnson-Neyman analyses detected the ranges of values of SRP-SF scores for which the interaction effect was significant ($p < .05$). To test for potential differences between the associations involving each of the psychopathy dimensions, we compared the regression coefficients (i.e., slopes) of the interactions obtained from the analysis of each dimension (Paternoster et al., 1998).

As performance below chance could be indicative of non-compliance with the experimental setting, we inspected whether the percentage of correct answers across gain and loss trials was lower than 50% in the stress or control conditions. We repeated the analyses excluding the three participants who performed below chance levels. Additionally, we identified one participant with "extreme" (± 3 standard deviations from the mean) SRP-SF scores (SRP-SF total = 95) and repeated the analyses excluding that participant to confirm that the results were not driven by extreme values. Given the two distinct criteria (i.e., performance below chance and extreme SRP-SF scores), we performed these additional checks in two independent steps.

4.3. Results

First, we inspected how total psychopathic traits modulated the impact of acute stress on behavioural performance towards monetary gains and losses (see Figures S4.2 and S4.3 in Appendix 4 for a distribution of the percentage of correct choices in gain and loss trials, respectively, across psychopathic traits). We found a significant interaction between SRP-SF total scores, condition, and

valence, β = 0.028, p < .001, 95% confidence interval (CI) = [0.012, 0.042] (Figure 4.1a). To better understand the nature of this interaction, we performed simple slope and Johnson-Neyman analyses. Simple slope analyses did not show any significant results, p > .22. However, Johnson-Neyman analyses detected that the effect of condition was significant in gain trials, but not in loss trials, when SRP-SF total scores were lower than 54.07 (representing 56.45% of the sample). This indicates that individuals with lower total psychopathic traits performed significantly worse towards obtaining gains in the stress condition than in the control condition, whereas in individuals with higher total psychopathic traits this difference in performance was absent (Figure 4.1a).

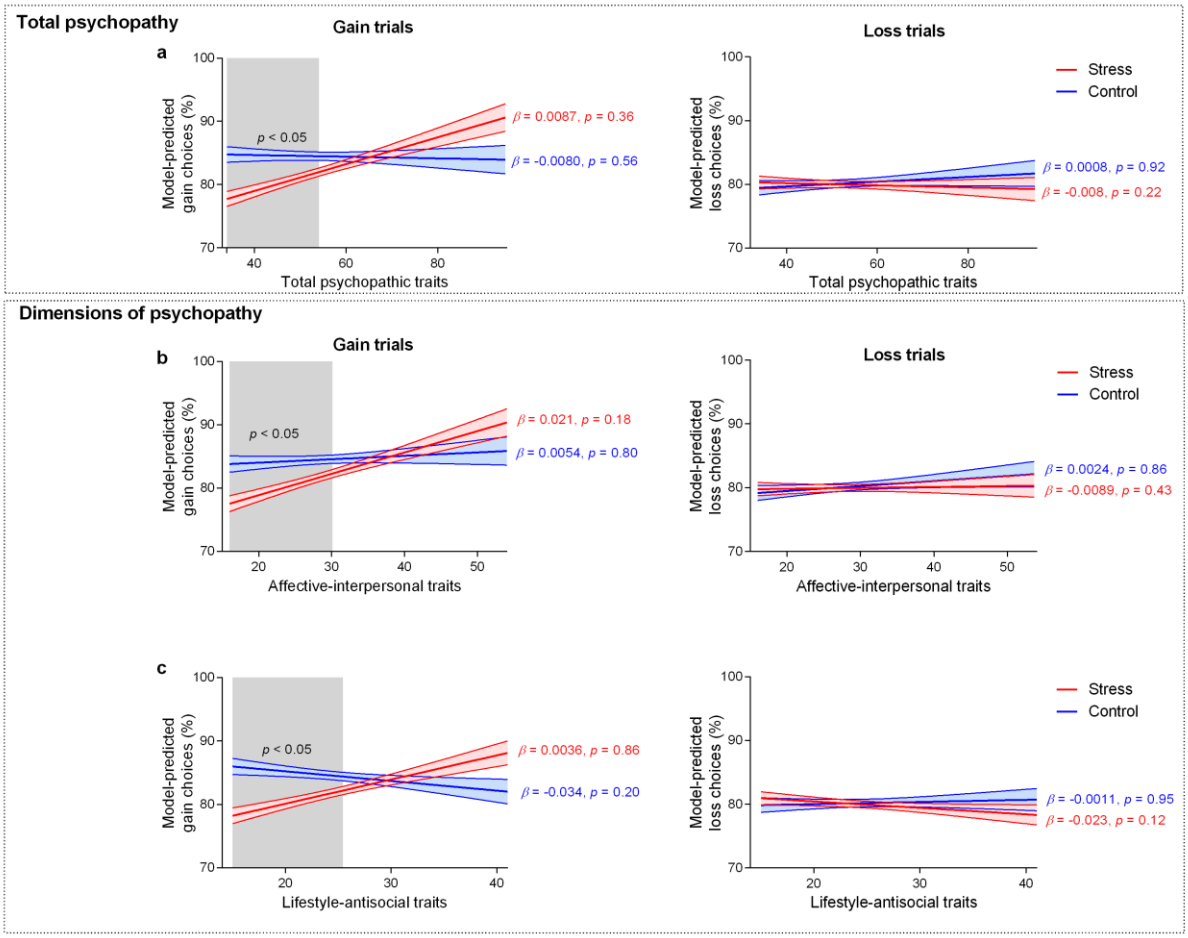


Figure 4.1. **Effects of acute stress on task performance across psychopathic traits.** Percentage of correct gain (left) and loss (right) choices in the stress (red) and control (blue) conditions estimated by the generalized linear mixed-effects models across Self-Report Psychopathy-Short Form (SRP-SF) **(a)** total, **(b)** affective-interpersonal, and **(c)** lifestyle-antisocial scores (b and c correspond to the SRP-SF dimensions). The central lines represent the linear fits, and the shading around those central lines represents 95% confidence intervals. The slopes (β) and p-values adjacent

to the lines were obtained through simple slope analyses for each condition (stress and control) and valence (gains and losses). Shaded gray areas were obtained through Johnson-Neyman analyses and identify the ranges of values of SRP-SF scores for which the interaction is significant ($p < .05$).

Second, to better understand the extent to which each dimension of SRP-SF — affective-interpersonal and lifestyle-antisocial — modulated the impact of acute stress on task performance in gain and loss trials, we conducted new glme models including the SRP-SF scores on each dimension as predictor variables. We found a significant interaction between affective-interpersonal traits, condition, and valence, $\beta = 0.033$, $p = .0075$, 95% CI = [0.00089, 0.058] (Figure 4.1b). Simple slope analyses were all non-significant, $p > .18$. However, Johnson-Neyman analyses detected that the slope of condition was significant in gain trials, but not in loss trials, when affective-interpersonal scores were lower than 30.12 (representing 56.45 % of the sample). We also found a significant interaction between lifestyle-antisocial traits, condition, and valence, $\beta = 0.060$, $p < .001$, 95% CI = [0.029, 0.090] (Figure 4.1c). Again, simple slope analyses were all non-significant, $p > .11$, but Johnson-Neyman analyses detected that the effect of condition was significant in gain trials, but not in loss trials, when lifestyle-antisocial scores were lower than 25.44 (representing 56.45% of the sample). The above interactions for affective-interpersonal and lifestyle-antisocial traits were not significantly different, $Z = -1.35$, $p = .18$. These results indicate that individuals with low traits in both the affective-interpersonal and lifestyle-antisocial dimensions of psychopathy performed significantly worse towards obtaining gains in the stress condition than in the control condition, but in individuals with higher levels of these traits such stress-induced impairment was not observed (Figures 4.1b and 4.1c).

Given that three participants performed below chance levels, we repeated the aforementioned analyses excluding those three participants to ensure that our results were robust. After excluding those participants, the significance of the interactions (all interactions between SRP-SF scores, condition, and valence; $p < .038$) and the interpretation of Johnson-Neyman regions of significance remained unchanged (the effect of condition was significant in gain trials when SRP-SF total scores < 62.21, affective-interpersonal scores < 31.00, and lifestyle-antisocial scores < 29.23). Moreover, we excluded one participant who scored very high on the SRP-SF (i.e., more than 3 standard deviations above from the mean; SRP-SF total score = 95); once again, the interpretation of the interactions (all interactions between SRP-SF scores, condition, and valence; $p < .048$) and Johnson-Neyman regions of significance remained unchanged (the effect of condition was significant in gain trials when SRP-SF total scores < 54.43, affective-interpersonal scores < 29.51, and lifestyle-antisocial scores < 25.30).

For completeness, we assessed how different levels of psychopathic traits were associated with stress responsivity. For this, we used the difference in self-reported stress levels and skin conductance response (SCR) rate between the stress and control conditions as a proxy for stress responsivity. We identified a trend for a negative association between SRP-SF total scores and self-reported stress responsivity, although the correlation was non-significant, r_s = -0.22, p = .090 (Figure S4.4 in Appendix 4). Inspection of the dimensions of psychopathy showed a significant negative association between lifestyle-antisocial and self-reported stress responsivity, r_s = -0.23, p = .019, but not between affective-interpersonal traits and self-reported stress responsivity, r_s = -0.15, p = .25. However, these two correlations for the dimensions of psychopathy were not significantly different, Z = -1.42, p = .16. We did not find any associations between psychopathic traits and the SCR rate, -0.074 < r_s < -0.012, all p < .58.

4.4. Discussion

In this study, we investigated whether and how individual differences in psychopathic traits modulated the impact of acute stress on reinforcement learning. We used a multidimensional measure of psychopathy and a reinforcement-learning task combined with a stress induction, to examine how psychopathic traits modulated the impact of acute stress on both reward and punishment learning. We found that psychopathic traits modulated the impact of acute stress on behavioural performance towards monetary rewards, but not punishments. More specifically, acute stress impaired behavioural performance towards rewards in individuals with lower levels of psychopathic traits, but such impairing effect of acute stress was not observed in individuals with higher levels of psychopathic traits. That is, as levels of psychopathic traits increased, the impairing effect of acute stress on reward learning decreased. This effect was observed across the affective-interpersonal and lifestyle-antisocial dimensions of psychopathy, suggesting a modulatory role of global psychopathic traits on reward learning under acute stress.

According to the response-modulation hypothesis, psychopathic individuals demonstrate difficulties in reallocating attention to environmental stimuli once they are focused on a goal, such as seeking rewards (Baskin–Sommers et al., 2012; Hamilton & Newman, 2018; Newman, 1998; Newman et al., 2010; Newman & Baskin-Sommers, 2012; Newman & Kosson, 1986; Smith & Lilienfeld, 2015). Whereas in healthy individuals, attention is shifted periodically to attend to environmental stimuli, such as threatful stimuli or stressors, it has been proposed that individuals with psychopathy might be unable to shift their attention to such peripheral stimuli unrelated to their main goal, particularly if that goal

involves obtaining rewards. This model could explain, at least partially, why the impairing effect of acute stress on reward learning decreased as levels of psychopathic traits increased.

Interestingly, our data suggest a selective modulatory effect of psychopathic traits on the impact of acute stress on reward learning, but not punishment avoidance learning. The effects of acute stress on punishment learning are generally less well-understood (Porcelli & Delgado, 2017). Previous behavioural studies (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Cremer et al., 2021; de Berker et al., 2016; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), together with our findings of Chapters 2 and 3, suggest that acute stress impairs reward learning, but the evidence for stress-induced deficits in punishment learning is scarcer (de Berker et al., 2016; Petzold et al., 2010). Reward learning has been robustly associated with dopaminergic functioning (Daw & Tobler, 2014; Glimcher, 2011; Pessiglione et al., 2006; Schultz et al., 1997), and it is known that acute stress increases dopamine release (Adler et al., 2000; Cabib & Puglisi-Allegra, 2012; Pruessner et al., 2004), which in turn may affect this type of learning. Psychopathic traits have been associated with dopaminergic disturbances (Buckholtz et al., 2010; Yildirim & Derksen, 2015). For example, Buckholtz et al. (2010) showed that psychopathic traits in a community sample were associated with a hyper-reactive dopaminergic reward system. It is possible that this enhanced dopaminergic functioning interacts with stress-induced dopamine release thus leading to a modulation of the effect of stress in reward learning. The involvement of dopaminergic functioning on punishment learning is more poorly understood (Glimcher, 2011; Pessiglione et al., 2006), suggesting that stress-induced dopamine release might not affect punishment learning to the same extent as it affects reward learning. Our results are consistent with this hypothesis; if reward learning is impaired by acute stress via dopaminergic mechanisms, that might explain why psychopathic traits selectively modulate the effect of acute stress on reward learning.

Our results did not provide support for an association between psychopathy scores and performance towards obtaining rewards and/or avoiding punishments in the control condition. This is interesting and may seem at odds with previous behavioural literature, which suggests that psychopathy is characterized by deficits in reward and, more markedly, in punishment learning (K. S. Blair et al., 2006; R. J. R. Blair et al., 2005; R. J. R. Blair & Mitchell, 2009). However, closer inspection of these studies shows important differences in the type of learning deficits identified. Previous studies found robust evidence of learning deficits in passive avoidance of punishments (R. J. R. Blair et al., 2004; De Brito et al., 2013; Finger et al., 2011; Newman & Kosson, 1986; Newman & Schmitt, 1998), a form of classic conditioning which requires the formation of stimulus–punishment associations but no action

execution, and in reversal learning (Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Gregory et al., 2015; Mitchell et al., 2002), which occurs when previously learned reward/punishment contingencies are reversed. There is less evidence of impaired stimulus – response associations in the early, gradual learning of actions that lead to rewards/punishments (Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Mitchell et al., 2002), which is the type of learning assessed by our task. Thus, it is possible that psychopathy is not characterized by deficits in all types of reinforcement learning (R. J. R. Blair et al., 2005).

Previous evidence indicates that, even at relatively low levels, psychopathic traits can be pathological (Neumann & Hare, 2008) and associated with atypical behavioural and neural processes (Seara-Cardoso & Viding, 2015). Studies focused on the normative variation in psychopathic traits can thus provide important insights into the broader phenotypic variability found across the dimensional distribution of psychopathic traits (Guay et al., 2007). Using a normative sample, we were able to capture broader variability in psychopathic traits, although we cannot exclude the possibility that additional associations between psychopathic traits, reward and/or punishment learning and acute stress could have emerged in a sample that included individuals at the higher end of the distribution.

4.5. Limitations

Reduced stress responsivity in psychopathy has been typically indexed by levels of cortisol (Cima et al., 2008; Holi et al., 2006; Johnson et al., 2015; Loney et al., 2006; Thompson et al., 2014). Although our task design precluded the measurement of cortisol levels, we assessed whether psychopathic traits were associated with differences in self-reported stress levels and SCR rate between stress and control conditions. Our findings indicated that the difference in self-reported stress levels tended to decrease as psychopathic traits increased but we did not find evidence for an association between psychopathic traits and the difference in SCR rate. This is at odds with previous literature (Arnett, 1997; Boucsein, 2012; Lorber, 2004). However, skin conductance hypoactivity for negative stimuli in psychopathy seems to appear more clearly for tonic skin conductance levels or for the amplitude of event-specific skin conductance responses (e.g., response to punishment stimuli), rather than for the rate of non-specific phasic SCRs (Arnett, 1997; Boucsein, 2012), as measured during our task. Future studies in both clinical and non-clinical populations should assess how stress responsivity (subjective and physiological) modulates the impact of acute stress on reinforcement learning across the spectrum of psychopathy.

In this study, we included only male participants due to females' hormonal-dependent variations on stress responsivity (Ossewaarde et al., 2010), as well as on reward and punishment learning (Diekhof & Ratnayake, 2016; Dreher et al., 2007). Further studies are needed to assess whether women with high levels of psychopathic traits are also characterised by a diminished impairing effect of acute stress on reward learning.

Finally, it remains debatable whether the putative protective role of psychopathic traits against the impairing effect of acute stress on reward, but not punishment-avoidance, learning is associated with maladaptive behaviours under acute stress. Future studies could assess how stress-induced task performance variability across the spectrum of psychopathy relates to real-life behaviour, by collecting measures of criminal behaviour or substance use, for example, which are often associated with psychopathy. Computational modelling studies in large community and clinical samples could further contribute for a better understanding of altered reinforcement-learning mechanisms, and how acute stress interacts with those mechanisms, in psychopathy.

4.6. Conclusions

This chapter described a study where we investigated whether and how individual differences in psychopathic traits modulated the impact of acute stress on reinforcement learning. In Chapter 2, we found that acute stress impaired reward learning, but not punishment learning. Using the same data set, we found that psychopathic traits modulated that impact of acute stress on reward learning, but not on punishment learning. The impairing effect of acute stress on reward learning decreased as levels of psychopathic traits increased, such that individuals with lower levels of psychopathic traits were more negatively affected by acute stress when learning to obtain rewards than individuals with higher levels of these traits. Thus, our data suggest that psychopathic traits may protect against the deleterious impact of acute stress on reward learning and that impaired reinforcement learning associated with psychopathy may depend not only on the type of learning but also on situational factors, such as acute stress.

This chapter highlights the substantial variability in how acute stress impacts reward learning, and the importance of assessing individual differences that may contribute for such variability. In the next chapter, Chapter 5, we will describe a study where we employed model-based fMRI to further investigate the neural mechanisms behind such modulatory effects of psychopathic traits on reward learning under acute stress.

CHAPTER 5

Individual differences in the impact of acute stress on striatal signalling of prediction errors during reinforcement learning: The role of psychopathic traits

5.1. Chapter Introduction

The study described in Chapter 3 suggests that acute stress impairs reward learning by blunting signalling of prediction errors in the dorsal striatum. In the previous chapter, Chapter 4, we described a behavioural study where we examined how psychopathic traits modulated the impact of acute stress on reinforcement learning. We found that as psychopathic increased, the impairing effect of acute stress on reward learning decreased. It is thus possible that psychopathic traits modulate the impact of acute stress on striatal signalling of prediction errors, but this is yet unknown. In this chapter, we employed behavioural and model-based fMRI analyses to the data set used in Chapter 3 to investigate whether and how individual differences in psychopathic traits modulated the impact of acute stress on striatal signalling of prediction errors during reward and punishment learning.

Humans show great variability when making choices and learning from the outcomes of their choices. Examining how individual differences shape this learning ability is important to understand why some individuals make better, or poorer, choices than others. Reinforcement learning is a particularly useful framework to investigate how individual differences modulate the way by which integration of rewarding and punishing outcomes drive choice behaviour (Cohen, 2008; Patzelt et al., 2018; Wuensch et al., 2021). During reinforcement learning, individuals learn to select the actions that maximise rewards and/or minimise punishments by gradually learning the values of those actions (Sutton & Barto, 1998). This type of learning is driven by the prediction error — the difference between an experienced and an expected outcome (Schultz et al., 1997; Sutton & Barto, 1998) — which has been robustly associated with striatal BOLD activity in the human brain (Montague et al., 2006; O'Doherty et al., 2004; Pessiglione et al., 2006; Schönberg et al., 2007; Valentin & O'Doherty, 2009).

Acute stress is thought to affect reinforcement learning (Porcelli & Delgado, 2017). In Chapter 3, we demonstrated that acute stress blunts prediction error — in particular positive prediction error — signals in the dorsal striatum, which may explain why individuals learn worse from the unexpected rewarding outcomes of their choices when under acute stress. However, the extent to which acute stress impacts reinforcement learning — and, possibly, underlying neural signalling of prediction errors — seems to vary between individuals, suggesting that individual differences modulate such impact. In Chapter 4, we showed that psychopathic traits may constitute a source of individual differences that modulate the impact of stress on reinforcement learning, such that the performance of individuals with higher levels of these traits during reward learning seems to be less impaired by the presence of acute stress. This modulatory effect of psychopathic traits does not seem to be observed in learning to avoid

punishments. Yet, whether and how psychopathic traits modulate the impact of acute stress on the key neurocomputational mechanism that drives reward and punishment learning — the prediction error — remains unclear.

Psychopathy is a personality disorder characterised by abnormal affective-interpersonal traits, such as lack of empathy, guilt and manipulativeness, and by deviant lifestyle-antisocial characteristics, such as impulsiveness, poor behavioural control and antisocial behaviour (R. J. R. Blair et al., 2005; Hare, 2003; Neumann & Hare, 2008; Viding et al., 2014). Psychopathic traits have been associated with emotional deficits, such as reduced response to distress cues (R. J. R. Blair, 2019; Patrick, 1994) and blunted stress responsiveness (Cima et al., 2008; Holi et al., 2006; House & Milligan, 1976; Johnson et al., 2015; Lidberg et al., 1978; Loney et al., 2006; Thompson et al., 2014), and it has been tentatively suggested that high levels of these traits might constitute a resilience factor against the deleterious impact of stress on behaviour (Anestis et al., 2017; Sandvik et al., 2015). Individuals with high levels of these traits typically engage in maladaptive behaviours that seem resist to modification by its consequences (R. J. R. Blair et al., 2005). Indeed, experimental evidence has associated psychopathic traits with difficulties in reinforcement learning, particularly in response reversal — that is, in flexible adjustment of behaviour when previously learning reward/punishment contingencies are unexpectedly reversed (Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Von Borries et al., 2010) — and in the formation of stimulus–punishment associations (Newman & Kosson, 1986; Newman & Schmitt, 1998). These difficulties might partially explain why individuals with psychopathy often persevere in behaviours that resulted in past rewards, despite other possible punishing consequences of those same behaviours. Thus, given the putative interplay between psychopathic traits, altered stress responsiveness, and reinforcement learning, psychopathic traits may constitute a source of individual differences that shape the extent to which acute stress affects reinforcement learning and underlying neural mechanisms.

5.1.1. The current study

In this study, we aimed to investigate whether and how individual differences in psychopathic traits modulated the impact of acute stress on striatal signalling of prediction errors during reinforcement learning. Given that acute stress is thought to impair reward learning by blunting striatal signalling of prediction errors (Chapter 3), and that the impairing effect of acute stress on reward learning decreases the levels of psychopathic traits increase (Chapter 4), we predicted that the blunting effect of acute stress on striatal prediction error signals during reward learning would decrease as

psychopathic traits increased. Additionally, given that the role of psychopathic traits on punishment learning under acute stress remains more equivocal (Chapter 4), we also assessed whether psychopathic traits would modulate the impact of acute stress on striatal prediction error signals during punishment learning.

We used the data set of Chapter 3, originally collected to assess how acute stress affects prediction error signalling during reward and punishment learning. We analysed data from 35 participants who completed the Self-Report Psychopathy-Short Form [(Paulhus et al., 2016); Portuguese version (Seara-Cardoso et al., 2020)] and performed a well-established fMRI reinforcement-learning task (Pessiglione et al., 2006) involving monetary gains (i.e., rewards) and losses (i.e., punishments) inside the scanner, whilst under acute stress and control conditions. First, we examined how self-reported stress levels varied across different levels of psychopathic traits. Second, we examined behavioural task performance, to assess whether and how psychopathic traits modulated the impact of acute stress on reward and punishment learning. Finally, and more importantly, we conducted model-based fMRI analyses to investigate whether and how psychopathic traits modulated the impact of acute stress on striatal prediction error signals during reward learning, and also during punishment learning.

5.2. Materials and Methods

5.2.1. Participants

We scanned a total of 42 right-handed male participants with no reported history of neurological or psychiatric disorders, as previously described in Chapter 3. One participant was excluded due to an incidental finding and 4 participants were excluded due to technical problems during the scanning session. Due to time constraints, two participants did not complete the questionnaire Self-Report Psychopathy-Short Form after the scanning session and were therefore excluded from the analyses. Thus, in the present study we analysed data of 35 participants (age range = 18 – 29; M = 22.91, SD = 3.44) from the Chapter 3 data set.

5.2.2. Self-Report Psychopathy-Short Form (SRP-SF)

Psychopathic traits were assessed with the Self-Report Psychopathy-Short Form [SRP-SF; (Paulhus et al., 2016); Portuguese version (Seara-Cardoso et al., 2020)], a 29-item questionnaire designed to measure psychopathic traits in non-institutionalised samples. The instrument follows the same factor structure as the Psychopathy Checklist – Revised (PCL-R) (Hare, 2003) and can be

modelled in terms of the traditional two-factor dimensions of psychopathy: affective-interpersonal and lifestyle-antisocial. Further details of the SRP-SF subscales can be found in the previous chapter, Chapter 4.

Overall, the distributions of the SRP-SF scores obtained in the current sample (Table 5.1; see Figure 5.1 in Appendix 5 for histograms with the distribution of psychopathic traits in our sample) did not differ significantly from the distribution of SRP-SF total and dimension scores reported in the previous chapter (Kolmogorov-Smirnov test, $p > .91$), and were consistent with other published studies on the SRP-SF with larger samples (Gordts et al., 2017; Seara-Cardoso et al., 2020). Calculation of the Chronbach's alpha showed adequate reliability of the SRP-SF total scale ($\alpha = 0.87$) and both dimensions (affective-interpersonal, $\alpha = 0.83$; lifestyle-antisocial, $\alpha = 0.75$). The two dimensions were significantly correlated ($r_s = 0.58$, $p < .001$).

No participant was beyond three standard deviations from the mean of each SRP-SF measure and no univariate outliers were identified in any SRP-SF measure.

Table 5.1. Descriptive statistics (Mean and Standard deviation, Median and Minimum and Maximum values) of Self-Report Psychopathy-Short Form (SRP-SF) total scores and dimensions (affective-interpersonal and lifestyle-antisocial).

	Total	**Dimensions**	
		Affective-interpersonal	**Lifestyle-antisocial**
Mean (SD)	52.80 (13.07)	28.57 (8.72)	24.23 (6.02)
Median [Min, Max]	50 [28, 80]	27 [13, 45]	25 [14, 42]

5.2.3. Reinforcement-learning task

Participants completed four blocks of an adapted version of a well-established reinforcement-learning task (Pessiglione et al., 2006) inside the scanner, as previosuly described in Chapter 3. Briefly, each block included three pairs of abstract stimuli, and each pair of stimuli was presented 24 times (new pairs of stimuli were used in each block). Each pair of stimuli was associated with a valence: one pair of stimuli was associated with gains (gain 0.5€ or no change), a second pair was associated with losses (loss 0.5€ or no change), and a third pair was associated with neutral, or non-financial outcomes (look at a 0.5€ coin or no change). The outcome probabilities were reciprocally 0.75 and 0.25 for the stimuli in each of the three pairs. On each trial, participants were asked to choose between two abstract

visual stimuli so as to maximise gains and minimise losses. Further details of the task can be found in Chapter 3.

5.2.4. Acute-stress manipulation

During the scanning session, participants performed two blocks of the reinforcement-learning task whilst exposed to a stressor — a predictable, but uncontrollable, sound alarm— and two blocks without the stressor. Stress and control blocks were administered alternately and in counterbalanced order. Further details of the stress manipulation can be found in Chapter 3.

5.2.5. Behavioural analyses

5.2.5.1. Self-reported stress levels

To assess whether self-reported stress levels differed between the stress and control conditions, we conducted an analysis of variance (ANOVAs) with condition (stress and control) and block (1 and 2), to account for potential temporal effects, as within-subject factors. More importantly, to assess whether psychopathic traits were associated with stress responsivity, we correlated (Spearman's correlations) psychopathic traits with the difference in self-reported stress levels between the stress and control conditions. To assess whether the correlations for each dimension were significantly different from each other, we applied the Meng's et al., 1992 approach using the "cocor" package in R (http://www.r-project.org).

5.2.5.2. Task-performance analyses

To examine how psychopathic traits modulated the impact of acute stress on choice performance during the reinforcement-learning task, we applied generalized linear mixed-effects (glme) models to participants' trial-by-trial choice data (with correct and incorrect choices coded as 1 and 0, respectively). We used a "logit" link function to account for the binomial distribution of the data. As predictor variables in the glme models, we included individual scores of psychopathic traits (continuous SRP-SF scores), condition (0 or 1 for control and stress conditions, respectively), valence (0 or 1 for loss and gain trials, respectively), block number (0 or 1, for blocks 1 and 2, respectively), and trial number (1 to 24), and tested the interactions between SRP-SF scores, condition, and valence. Block and trial number were included as predictors to control for potential temporal effects (i.e., covariates of noninterest). We included a fixed intercept as well by-subject random intercepts and by-subject random

slopes for condition, valence, block, and trial. We generated a glme model to assess the interaction between SRP-SF total scores, condition, and valence, and two other glme models to assess the interaction between SRP-SF scores in each dimension of psychopathy (affective-interpersonal and lifestyle-antisocial), condition, and valence. We fitted the glme models to the behavioural data using the "lme4" package in R.

To further understand the nature of the interactions, we performed post-hoc simple slope and Johnson-Neyman analyses using the "interactions" package in R. Simple slope analyses tested the significance of the associations between SRP-SF scores and behavioural performance during gain or loss trials in each condition. To assess whether the slopes differed significantly between conditions, we defined contrast matrices and used the "multcomp" package in R to apply those contrast matrices to the glme. Johnson-Neyman analyses detected the ranges of values of SRP-SF scores for which the interaction effect was significant ($p < .05$).

Additionally, as performance below chance could be indicative of non-compliance with the experimental setting, we inspected whether the percentage of correct answers across gain and loss trials was below chance levels in the stress or control conditions. We identified four participants that performed below chance levels, and repeated the analyses excluding those participants.

5.2.6 fMRI data analyses

fMRI data acquisition and preprocessing followed a standard sequence as comprehensively described in Chapter 3.

The primary fMRI analyses were based on a single general linear model, as in previous studies that employed a similar reinforcement-learning task (Kumar et al., 2018; Pessiglione et al., 2006; Treadway et al., 2017), and as previously described in Chapter 3. Briefly, each trial was modelled as having two time points, stimuli and outcome onsets. Each time point was regressed with a parameter modulator, separately for gain and loss trials: stimuli onset was modulated by the chosen option value; and outcome onset was modulated by the prediction error. The value of the chosen stimulus and predictions errors were estimated trial-wise using a well-established reinforcement-learning model (Frank et al., 2007). Further details on the regressors of the primary general linear model can be found in Chapter 3.

First-level contrast images were calculated by applying appropriate linear contrasts to the parametric modulators of interest — prediction errors — and were entered into second-level analyses. Second-level one-sample *t*-tests were conducted for each contrast using the summary-statistics

approach to random-effects analysis. To identify associations between brain responses and psychopathic traits, total and dimension scores of the SRP-SF were entered separately as a regressor of interest in these second level models.

Regions of interest (ROI) analyses in the dorsal striatum and nucleus accumbens were conducted using an initial whole-brain threshold of $p < .001$ (uncorrected) and responses were considered significant if they survived voxel-level small volume family-wise error correction (SVC-FWE) at $p < .05$. ROIs — dorsal striatum and nucleus accumbens — were anatomically defined using masks from the Automated anatomical labelling atlas and Individual brain atlases.

To further understand the nature of the modulatory effects of psychopathic traits on prediction error signalling under acute stress during gain and loss trials, we extracted individual parameter estimates for the modulation of BOLD response by prediction errors from the significant cluster with the strongest activation using the MarsBaR toolbox (Brett M. et al., 2002) and defined a linear mixed-effects model which included the interaction of SRP-SF scores, condition (stress and control), and valence (gain and loss trials), as well as a fixed intercept and by-subject random intercepts and by-subject random slopes for condition, and valence. We further conducted post-hoc simple slope and Johnson-Neyman analyses using the "interactions" package in R. Simple slope analyses tested the significance of the associations between SRP-SF scores and the parametric modulation of BOLD response by prediction errors during gain and loss trials in each condition. Johnson-Neyman analyses detected the ranges of values of SRP-SF scores for which the effect of condition was significant ($p < .05$) on the BOLD response modulation by prediction errors during gain and/or loss trials.

To further illustrate how psychopathic traits were associated with the modulation of BOLD response by prediction errors, we generated a subsidiary general linear model where trials of each condition (stress and control) and valence (gain and loss trials) were further split into three equally sized bins according to the magnitude of prediction errors. Note that this subsidiary model did not include parametric modulators, as our purpose was to visualise how the BOLD response varied along different magnitudes of prediction errors. BOLD estimates at the outcome onset for each bin were extracted from the significant cluster with the strongest activation identified in the primary parametric modulation model described above using MarsBaR toolbox (Brett M. et al., 2002). We plotted BOLD estimates for each bin, both in gain and loss trials, and in the stress and control conditions, for participants with low and high psychopathic traits. Participants were split into low and high psychopathic traits based on the regions of significance identified by the Johnson-Neyman analyses described above.

5.3. Results

5.3.1. Behavioural analyses

5.3.1.1. Self-reported stress levels and psychopathic traits

First, we examined whether self-reported stress levels differed between the stress and control conditions. Self-reported stress levels were higher in the stress than in the control condition, however this difference did not reach significance, $F(1,34) = 3.22$, $p = .081$. We further assessed whether different levels of psychopathic traits were associated with stress responsivity. For this, we used the difference in self-reported stress levels between the stress and control conditions as a proxy for self-reported stress responsivity. The correlation between SRP-SF total scores and self-reported stress responsivity was non-significant, $r_s = -0.13$, $p = .46$. Inspection of the dimensions of psychopathy identified a weak negative association between affective-interpersonal traits and self-reported stress responsivity, although it was non-significant, $r_s = -0.24$, $p = .16$ (Figures S5.2a and S5.2b in Appendix 5). We did not identify an association between lifestyle-antisocial traits and self-reported stress responsivity, $r_s = 0.041$, $p = .81$ (Figure S5.2c in Appendix 5). The associations between each dimension scores and self-reported stress levels did not differ significantly, $Z = -1.76$, $p = .079$.

5.3.1.2. Task-performance analyses

Then, we assessed whether psychopathic traits modulated the impact of acute stress on behavioural performance towards obtaining monetary gains and avoiding monetary losses. We found a significant SRP-SF total scores × condition × valence interaction, $\beta = 0.053$, $p < .001$, 95% confidence intervals (CI) = [0.038, 0.071]. Post-hoc simple slope analyses indicated that the association between SRP-SF total scores and performance differed significantly between the stress and control conditions during gain trials, $Z = 3.87$, $p < .001$ (Figure S5.3a in Appendix 5), but not during loss trials, $Z = 0.017$, $p = .99$ (Figure S5.4a in Appendix 5); in gains trials, SRP-SF total scores were positively associated with performance in the stress condition, $\beta = 0.029$, $p = .051$, but not in the control condition, $\beta = -0.019$, $p = .11$, whereas in loss trials SRP-SF total scores were positively associated with performance both in the stress, $\beta = 0.015$, $p = .053$, and control conditions, $\beta = 0.019$, $p = .053$. To better understand the nature of this interaction, we performed Johnson-Neyman analyses to detect the range of SRP-SF total scores where the effect of condition was significant in gain and/or loss trials. We found that the effect of condition was significant in gain trials, when SRP-SF total scores were lower than 46.14 (representing

28.57 % of the sample) and higher than 61.70 (representing 22.86% of the sample). No regions of significance were found in loss trials.

To better understand the extent to which each dimension of psychopathy — affective-interpersonal and lifestyle-antisocial — modulated the impact of acute stress on task performance, we conducted separate generalized linear-mixed effects (glme) models to test the interaction between scores of each SRP-SF dimension, condition, and valence. We identified a significant affective-interpersonal scores × condition × valence interaction, $\beta = 0.066$, $p < .001$, 95% CI = [0.037, 0.095] (Figure 5.1). Post-hoc simple slope analyses indicated that the association between affective-interpersonal scores and performance differed significantly between the stress and control conditions during gain trials, $Z = 3.76$, $p < .001$ (Figure 5.1a, and Figure S5.3b in Appendix 5), but not during loss trials, $Z = 0.42$, $p = .67$ (Figure 5.1b, and Figure S5.4b in Appendix 5); in gain trials, affective-interpersonal scores were positively associated with performance in the stress condition, $\beta = 0.051$, $p = .021$, but not in the control condition, $\beta = -0.019$, $p = .32$, whereas in loss trials SRP-SF affective-interpersonal scores were positively associated with both in the stress, $\beta = 0.034$, $p = .027$, and control, $\beta = 0.021$, $p = .10$, conditions. Further Johnson-Neyman analyses showed that the effect of condition in gain trials was significant when affective-interpersonal scores were lower than 23.96 (representing 37.14 % of the sample) and higher than 34.27 (representing 28.57 % of the sample) (Figure 5.1a). No effects of condition were detected in loss trials (Figure 5.1b).

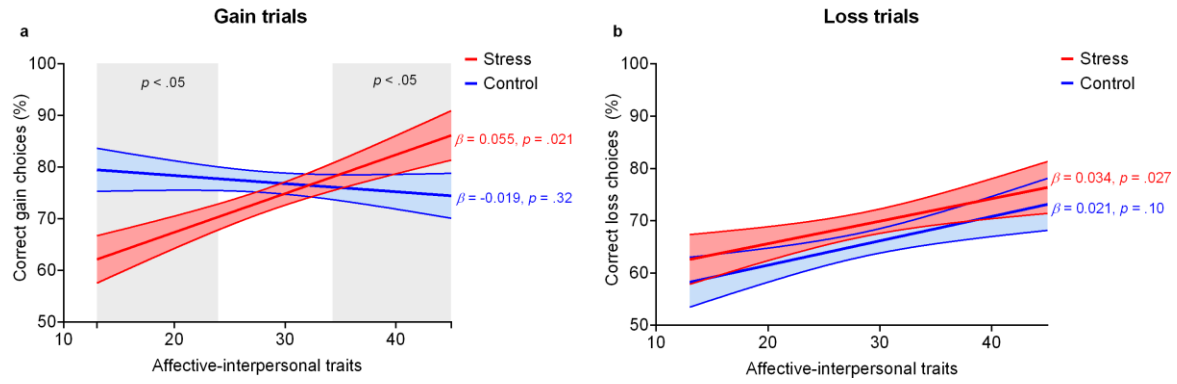


Figure 5.1. Effects of acute stress on task performance across affective-interpersonal traits. (a) Percentage of correct gain choices (i.e., choices of the stimulus associated with a higher probability of winning) in the stress (red) and control (blue) conditions across affective-interpersonal traits **(c)** Percentage of correct loss choices (i.e., choices of the stimulus associated with a lower probability of losing) in the stress and control conditions across affective-interpersonal traits. The central lines represent the linear fits, and the shading around those central lines represent 95%

confidence intervals. The slopes (β) and p-values adjacent to the lines were obtained through simple slope analyses for each condition and valence. The shaded gray areas correspond to Johnson-Neyman regions of significance ($p < .05$).

The lifestyle-antisocial scores × condition × valence interaction was also significant, $\beta = 0.12$, $p < .001$, 95% CI = [0.079, 0.16]. Post-hoc simple slope analyses indicated that the association between lifestyle-antisocial scores and performance differed significantly between the stress and control conditions during gain trials, $Z = 3.20$, $p = .0014$ (Figure S5.3c in Appendix 5), but not during loss trials, $Z = -1.21$, $p = .23$ (Figure S5.4c in Appendix 5); in gain trials lifestyle-antisocial scores were positively associated with performance towards gains in the stress condition, $\beta = 0.036$, $p = .25$, but negatively associated with this variable in the control condition, $\beta = -0.049$, $p = .052$, whereas in loss trials lifestyle-antisocial scores were positively associated with performance both in the stress, $\beta = 0.015$ $p = .039$, and control conditions, $\beta = 0.044$, $p = .013$. Further Johnson-Neyman analyses showed that the effect of condition in gain trials was significant when lifestyle-antisocial scores were lower than 20.83 (representing 28.57 % of the sample) and higher than 30.54 (representing 14.29 % of the sample). No effects of condition were detected in losses trials.

Additionally, given that four participants performed below chance levels, we repeated the aforementioned analyses excluding those four participants. The affective-interpersonal scores × condition × valence interaction remained significant, $\beta = 0.033$, $p = .033$, 95% CI = [0.0029, 0.064], and the interpretation of simple slope and Johnson-Neyman analyses did not change (significant effect of condition in gain trials when affective-interpersonal traits were lower than 22.41, corresponding to 31.43 % of the sample, and higher than 44.30, corresponding to 2.86% of the sample). The interactions for SRP-SF total scores and lifestyle-antisocial dimension became non-significant ($p > .062$) after excluding the four participants.

Overall, our results suggest that psychopathic traits, in particular affective-interpersonal traits, modulate the impact of acute stress on performance towards obtaining rewards, but not towards avoiding punishments; the impairing effect of acute stress on performance towards obtaining rewards decreases as levels of psychopathic traits increase.

5.3.2. fMRI results

After confirming that psychopathic traits modulated the effect of acute stress on behavioural performance towards obtaining gains, but not towards avoiding losses, we examined the neural

mechanisms behind such association. We focused our fMRI analyses on the associations between psychopathic traits and modulation of BOLD response by prediction errors in the dorsal striatum and nucleus accumbens — brain regions consistently shown to encode prediction errors (the results for simple associations between psychopathic traits and parametric modulation in the striatum by prediction errors can be found in Table S5.1 in Appendix 5).

To test whether and how psychopathic traits modulated the impact of acute stress on striatal signalling of prediction errors, we used the stress > control contrast on the parametric modulators of prediction errors in gain and loss trials across the stress and control conditions. We further tested the interaction between condition (stress/control) and valence of the trial (gain/loss) to probe whether psychopathic traits modulated the impact of acute stress on striatal signalling of prediction errors differently in gain and loss trials.

The stress > control contrast identified a positive association between affective-interpersonal scores and the dorsal striatum parametric modulation by prediction errors in the stress condition compared with the control condition, [x = 30, y = -5, z = 12], Z = 4.21, k = 38, p = .022, SVC-FWE; [x = 33, y = 3, z = 0], Z = 4.08, k = 84, p = .036, SVC-FWE; [x = 23, y = 21, z = 3], Z = 4.01, k =12, p = .046, SVC-FWE (Figure 5.2a). We did not find a significant association for SRP-SF total scores nor for lifestyle-antisocial scores. As an additional check, after excluding the four participants who performed below chance levels, we still detected a dorsal striatum cluster where the positive association between affective-interpersonal scores and parametric modulation of prediction errors in the stress condition compared with the control condition was at trend level, [x = 33, y = 3, z = 0], Z = 3.97, k = 89, p = .054, SVC-FWE. We did not observe any significant responses for the stress > control contrast (nor for the inverse contrast) in the nucleus accumbens.

The contrast for the condition × valence interaction did not identify significant activations in the dorsal striatum, nor in the nucleus accumbens. This non-significant interaction together with the effects found for the stress > control contrast described above, indicate that affective-interpersonal traits were positively associated with increased parametric modulation of prediction errors in the dorsal striatum under stress compared with the control condition, both during gain and loss trials.

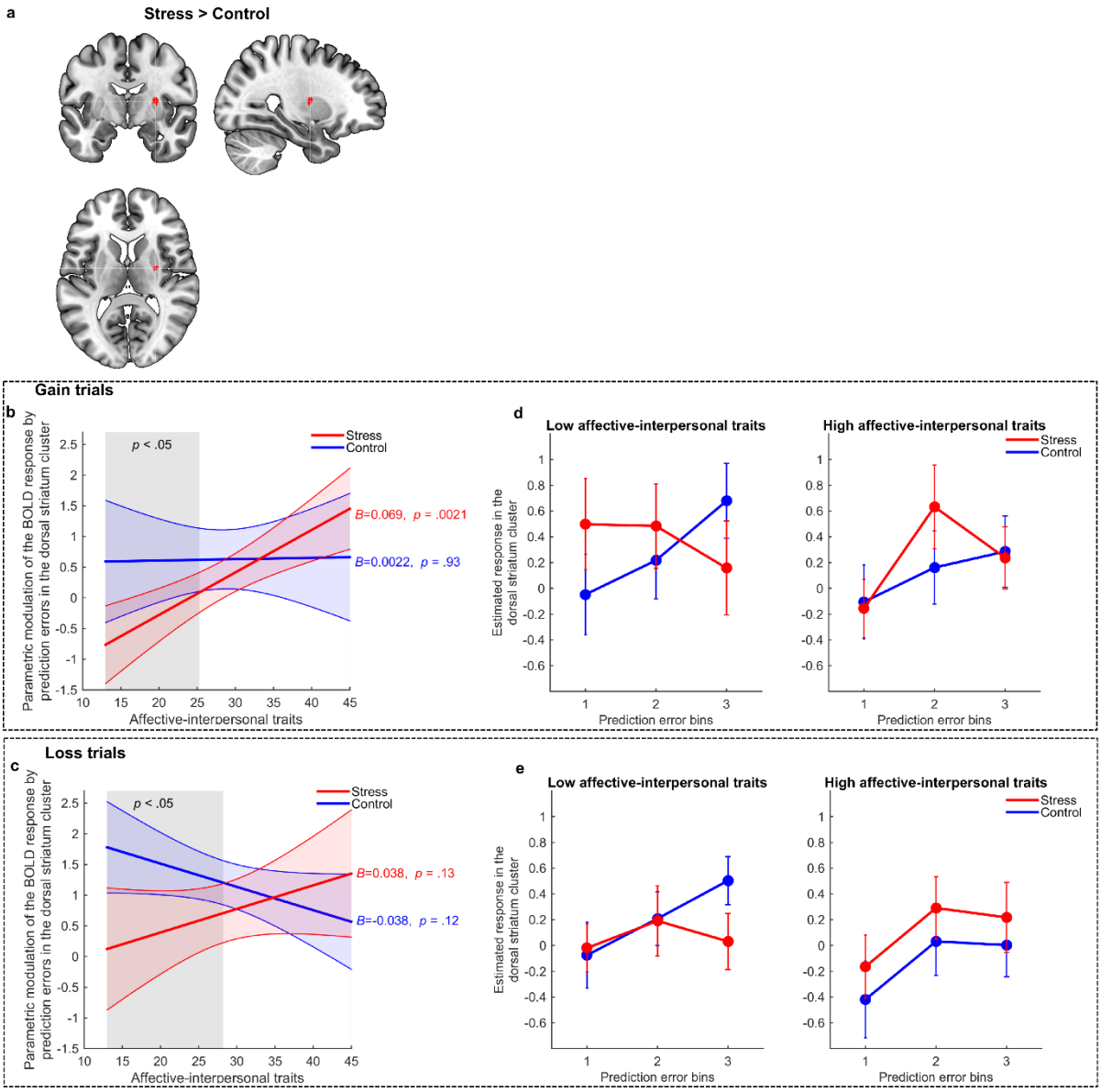


Figure 5.2 Effects of acute stress on prediction error signalling in the dorsal striatum across affective-interpersonal traits. (a) Cluster in the dorsal striatum (peak voxel: [x = 30, y = -5, z = 12]) where affective-interpersonal traits were associated with increased modulation of prediction errors in the stress condition compared with the control condition, after small volume correction (SVC-FWE, p < .05). **(b, c)** Parameter estimates (i.e., regression slopes) for the BOLD response at the dorsal striatum cluster represented in (a) modulated by trial-by-trial prediction errors in gain **(b)** and loss trials **(c)**, during the stress (red) and control (blue) conditions, across affective-interpersonal traits. The central lines represent the linear fits, and the shading around those central lines represent 95% confidence intervals. The slopes (β) and p-values adjacent to the lines were obtained through simple slope analyses for each condition and valence, and the shaded gray areas correspond to Johnson-Neyman regions of significance (p < .05). Slopes, p-values and Johnson-

Neyman regions of significance were presented for the purposes of visualization and interpretation only. **(d, e)** Graphs illustrate the modulation of the BOLD response by prediction errors, in gain trials **(d)** and loss trials **(e)**, within the dorsal striatum cluster identified in the primary parametric modulation model [from (a)] in the stress (red) and control (blue) conditions. Data for illustrative graphs (d) and (e) were derived from a subsidiary model where trials were divided into tertiles of magnitude of prediction errors (with the lowest and highest magnitudes corresponding to bins 1 and 3, respectively) and BOLD response estimates from the dorsal striatum cluster were extracted for each participant, who were further grouped into low and high affective-interpersonal traits, based on the regions of significance identified by Johnson-Neyman analyses. Error bars indicate the mean ± standard error of the mean.

To visualise and interpret the nature of the associations between affective-interpersonal traits and signalling of prediction errors within the dorsal striatum cluster in the stress and control conditions, we extracted individual parameter estimates (i.e., regression slopes) from the dorsal striatum cluster with the strongest activation (cluster with peak voxels [x = 30, y = -5, z = 12], represented in Figure 5.2a). Note that these parameter estimates correspond to the regression slope linking trial-by-trial BOLD response in the dorsal striatum cluster with trial-by-trial prediction errors for each specific participant. Post-hoc simple slope analyses indicated that during gain trials, as levels of affective-interpersonal traits increased, the modulation of prediction errors in the dorsal striatum increased significantly in the stress condition, β = 0.069, p = .0021, but not in the control condition, β = 0.0022, p = .93 (Figure 5.2b); during loss trials, we did not find a significant association between levels of affective-interpersonal traits and the parametric modulation of prediction errors in the stress condition, β = 0.038, p = .13, nor in the control condition, β = -0.038, p = .12 (Figure 5.2c). Further Johnson-Neyman analyses detected that the effect of stress was significant in gain trials when affective-interpersonal scores were lower than 25.33 (representing 45.71 % of the sample) and during loss trials when affective-interpersonal scores were lower than 28.21 (representing 57.14 % of the sample). This suggests that acute stress decreases signalling of prediction errors during gain and loss trials in the dorsal striatum of individuals with lower levels of affective-interpersonal traits, but this effect is absent in individuals with higher levels of these traits.

To further illustrate how affective-interpersonal traits modulated the impact of acute stress on prediction error signals in the dorsal striatum, we conducted a subsidiary fMRI model to extract the BOLD response in the dorsal striatum cluster identified in the parametric modulation model (cluster

with the strongest activation, represented in Figure 5.2a) across different categories of prediction error magnitudes. In this subsidiary model, trials from each condition (i.e., stress and control) and valence (i.e., gain and loss) were further divided into three bins corresponding to tertiles of magnitude of prediction errors. BOLD estimates at the outcome onset were extracted from the dorsal striatum cluster for each subject. Participants were split into low and high affective-interpersonal traits, based on the regions of significance identified above by Johnson-Neyman analyses for gain and loss trials, to illustrate the variation in BOLD response within the dorsal striatum cluster along the magnitude of prediction errors in participants with different levels of psychopathic traits (Figure 5.2d and 5.2e). Inspection of the illustrative plots suggests that in individuals with lower levels of psychopathic traits, the BOLD response within the dorsal striatum cluster increased as the magnitude of prediction errors increased to a larger extent in the control condition than in the stress condition, during both gain (Figure 5.2d) and loss (Figure 5.2e) trials, whereas in individuals with higher levels of psychopathic traits such difference between conditions was attenuated. Additionally, these illustrative data also suggest that acute stress blunted prediction errors in bin 3, mostly constituted by positive prediction errors (see Table S5.2 in Appendix 5 for a description of each bin), in individuals with lower levels of affective-interpersonal traits to a larger extent than in individuals with higher levels these traits (in Figures 5.2d and 5.2e, compare the estimated response in bin 3 between the stress and control conditions in individuals with low and high affective-interpersonal traits).

Overall, our fMRI results indicate that affective-interpersonal traits modulate the impact of acute stress on signalling of prediction errors in the dorsal striatum during reinforcement learning, such that the blunting effect of acute stress on striatal prediction error signals decreases as affective-interpersonal traits increase.

5.4. Discussion

In Chapters 2 and 3 we showed that acute stress impairs reward learning, and in the previous chapter, Chapter 4, we demonstrated that individual differences in psychopathic traits modulate the extent to which acute stress impairs reward learning. However, the neural mechanisms by which psychopathic traits modulate the impact of acute stress on reward learning remain unclear. In this study, we set out to test whether and how psychopathic traits modulated the impact of acute stress on the key neurocomputational mechanism of reinforcement learning — the prediction error — during reward learning, and also during punishment learning. We showed that the impairing effect of acute stress on behavioural performance towards obtaining rewards decreased as levels of affective-

interpersonal traits increased, which was accompanied by a reduction in the extent to which acute stress blunted signalling of prediction errors in the dorsal striatum. Additionally, although our behavioural data did not provide evidence that psychopathic traits modulate the impact of acute stress on punishment learning, signalling of prediction errors in the dorsal striatum during punishment learning was also less affected by acute stress as affective-interpersonal traits increased.

One influential model of psychopathy, the response-modulation hypothesis, postulates that individuals with psychopathy present difficulties in shifting their attention to peripheral stimuli, such as threatful stimuli or stressors, once they are focused on a goal, particularly when the goal involves seeking rewards (Baskin–Sommers et al., 2012; Hamilton & Newman, 2018; Newman, 1998; Newman & Baskin-Sommers, 2012). Our results are in line with this hypothesis and extend the findings from the previous chapter by suggesting a putative protective role of psychopathic traits in the negative impact of acute stress on reward learning.

Importantly, we found that psychopathic traits, and in particular affective-interpersonal traits, accounted for variations in the impact of acute stress on signalling of prediction error within the dorsal striatum during reward learning. Prediction errors are thought to be critical for reward learning by signalling the difference between obtained and expected outcomes. For example, Schönberg et al. (2007) reported decreased signalling of prediction errors in the dorsal striatum — a brain region which has been robustly associated with reward-based action selection (Lohrenz et al., 2007; O'Doherty et al., 2004) — of individuals who did not learn to make optimal choices during a reward-learning task, compared with individuals who did; moreover, they showed that the magnitude of prediction errors in the dorsal striatum correlated positively with behavioural performance during the reward-learning task. Our data is broadly consistent with the idea that stress-induced variability in prediction error signalling within the dorsal striatum underpins individual differences in the extent to which acute stress affects performance towards rewards, and that psychopathic traits may contribute for such variability. More concretely, our finding of stress-induced blunted prediction error signalling in the dorsal striatum of individuals with lower levels of affective-interpersonal traits, but not of individuals with higher levels of these traits, may explain why individuals with lower levels of these traits perform worse when seeking rewards under acute stress.

Our behavioural results further suggest an enhancing effect of acute stress on reward learning in individuals with higher levels of psychopathic traits. This finding is not easily reconciled with our findings from the previous chapter, Chapter 4. Additionally, after exclusion of participants who performed below chance levels, this enhancing effect of stress was only observed in one individual with

high affective-interpersonal traits. Thus, whether individuals with higher levels of psychopathic traits learn better from rewards under stress remains equivocal. Furthermore, our behavioural and neural data suggest that variations in the effects of acute stress on reward learning are more robustly associated with affective-interpersonal traits – which are considered to be the core characteristic of psychopathy. Still, we should remain cautious in interpreting the results for the lifestyle-antisocial dimension and refrain from making differential claims for the dimensions of psychopathy, as lack of variability and right skewed distributions in antisocial traits, which are not uncommon in community samples (Gordts et al., 2017; Seara-Cardoso et al., 2020), may limit the interpretability of the data.

Our findings did not provide evidence for a modulatory role of psychopathic traits on the impact of acute stress on behavioural performance towards avoiding punishments. Instead, our behavioural data, consistent with data from Chapter 4, suggest that psychopathic traits account for variations in the impact of acute stress on reward learning, but not on punishment learning. However, at the neural level, affective-interpersonal psychopathic traits were associated with decreased stress-induced blunted signalling of prediction errors during punishment learning. Although at first sight this may seem contradictory with the behavioural data, this neural finding was not totally surprising.

Neurobiological evidence indicates that positive and negative prediction errors are respectively encoded by dopamine burst firing and dips during both reward and punishment learning (H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009; Oleson et al., 2012; Salinas-Hernández et al., 2018; Ungless et al., 2004; Watabe-Uchida et al., 2017). That is, when an unexpected reward is obtained or an expected punishment is avoided, dopamine neurons burst fire encoding a positive prediction error; when an expected reward is omitted or an unexpected punishment is obtained, dopamine neurons supress their activity, encoding a negative prediction error. Thus, if psychopathic traits modulate the impact of acute stress preferentially on positive prediction errors, given that the corresponding mechanisms are thought to overlap in reward and punishment learning, then we would expect to find a similar association between psychopathic traits and prediction error signalling in reward and punishment learning. Still, this would not necessarily translate into the same behavioural outcomes, given that positive and negative prediction errors seem to have different weights in these types of learning. Specifically, in reinforcement-learning tasks such as ours, reward learning is mostly driven by positive prediction errors, and punishment learning might rely more on negative prediction errors (Palminteri & Pessiglione, 2017). Subsidiary analyses of our data suggest that affective-interpersonal traits might be mostly associated with an attenuated negative effect of acute stress on positive prediction errors, which could result in less impaired reward learning, but not necessarily in altered

punishment learning. However, it should be noted though that our study was not designed to test specific effects on positive and/or negative prediction errors and can only provide tentative evidence for this association. Given the putative impact of acute stress on positive and/or negative prediction errors, further studies are needed to assess the extent to which psychopathic traits modulate that impact.

Our behavioural data did not provide evidence for an association between psychopathic traits and impaired punishment learning in the control condition. This might seem at odds with prior literature, which suggests that psychopathy is characterised by impaired reinforcement learning (K. S. Blair et al., 2006; R. J. R. Blair et al., 2005), particularly when indexed by passive avoidance of punishments (R. J. R. Blair et al., 2004; De Brito et al., 2013; Finger et al., 2011; Newman & Kosson, 1986; Newman & Schmitt, 1998) or by reversal learning (Brazil, Maes, et al., 2013; Budhani et al., 2006; Dargis et al., 2017; De Brito et al., 2013; Gregory et al., 2015; Mitchell et al., 2002). Passive avoidance learning implicates inhibiting responses to stimuli that lead to punishments, whilst reversal learning occurs when previously learned reward/punishment contingencies are reversed. There is less evidence that psychopathy is associated with deficits in early, gradual learning of actions that result in rewards/punishments, as required in the task we used (R. J. R. Blair et al., 2005). Indeed, it has been suggested that psychopathy might be characterised by deficits in specific types of reinforcement learning rather than by a general deficit in reinforcement learning (R. J. R. Blair, 2019; R. J. R. Blair et al., 2005). Further studies in large clinical and large community samples are needed to better understand the extension of such deficits and how they may translate into real-life behaviour.

Some limitations of this work should be noted. Previous literature has associated psychopathic traits — and in particular affective-interpersonal traits — with decreased stress responsiveness (Thompson et al., 2014). These studies typically assess stress responsiveness via cortisol measurements (Holi et al., 2006; Johnson et al., 2015; Loney et al., 2006) or other physiological measures, such as skin conductance or heart rate (Arnett, 1997; Dindo & Fowles, 2011; Fung et al., 2005; Thompson et al., 2014). Although we previously assessed in Chapter 2 the efficacy of our stress manipulation using self-report and skin conductance response measures, in the present work we did not analyse physiological responses. Our findings tentatively suggest that self-reported stress responsivity decreases as affective-interpersonal traits increase, but inclusion of physiological measures in future studies could be important to better understand the interplay between individual differences in psychopathic traits, stress responsivity, and the mechanisms of reinforcement learning.

In this study, we included only male participants as it is known that females' menstrual hormonal-dependent variations are associated with alterations in stress responsivity (Ossewaarde et al.,

2011) as well as in reward (Dreher et al., 2007) and punishment learning (Diekhof & Ratnayake, 2016). Further studies are required to examine whether women with high levels of psychopathic traits are also characterised by a diminished impact of acute stress on reward learning and underlying signalling of prediction errors.

5.5. Conclusions

This chapter described a neuroimaging study where we examined whether and how individual differences in psychopathic traits modulated the impact of acute stress on striatal signalling of prediction errors during reinforcement learning. In Chapter 3, we demonstrated that acute stress blunted signalling of prediction errors in the dorsal striatum. Using the same data set of Chapter 3, we found that psychopathic traits — particularly affective-interpersonal traits — modulated the impact of acute stress on prediction error signals in the dorsal striatum; the blunting effect of acute stress on signalling of prediction errors in the dorsal striatum decreased as the levels of affective-interpersonal traits increased. This neural data may explain why individuals with higher levels of psychopathic traits are less negatively affected by acute stress when seeking rewards than individuals with lower levels of psychopathic traits. Overall, this work extends the findings of the previous chapter, Chapter 4, and highlights the substantial variability in the neurocomputational mechanisms that underly the impact of acute stress on reward learning and the importance of considering individual differences that may contribute for such variability.

In the next and final chapter, I will summarise the findings of the preceding empirical chapters and discuss the implications and future directions of this work.

CHAPTER 6: GENERAL DISCUSSION

6.1. Overview

Reinforcement learning implicates that individuals learn to choose options that lead to rewards and avoid those that result in punishments and is thus critical for adaptive behaviour. Situational factors, such as stress, can have deleterious effects on this ability and have a great impact in how individuals make the best choices and learn from them. Extant behavioural evidence suggests that acute stress impairs reward learning and, possibly, punishment learning, but the computational and neural mechanisms behind such impairing effect are unclear. Furthermore, little is known about whether and how individual differences account for variations in the impact of acute stress on reinforcement learning and underlying neural mechanisms.

Given the pervasiveness of acute stress in modern life, the investigation of the computational and neural mechanisms by which acute stress affects our ability to learn from rewards and punishments is relevant not only to better understand how acute stress may affect our daily functioning but also to develop adequate individual strategies that target the debilitating effects of acute stress.

In this thesis, we have employed behavioural, computational and neuroimaging methods to investigate how acute stress impacts reinforcement learning and how individual differences modulate such impact.

6.2. Research questions

In the current thesis, we set out to investigate how acute stress impacts reinforcement learning at the behavioural, computational and neural levels, and how individual differences in psychopathic traits modulate that impact. The research presented in this thesis aimed to answer three outstanding questions: 1) How does acute stress impact reinforcement learning at the behavioural and computational levels? 2) Does acute stress affect neural signalling of prediction errors during reinforcement learning? and, finally, 3) Do individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning and neural signalling of prediction errors? We will address each of these sequentially in the sections below. Overall, we will argue that acute stress impairs reward learning and, relatedly, decreases the learning rate for positive prediction errors by blunting striatal signalling of prediction errors; we will also argue that differences in psychopathic traits modulate such behavioural and neural effects of acute stress on reward learning.

6.2.1. How does acute stress impact reinforcement learning at the behavioural and computational levels?

In **Chapter 2** of this thesis, we described a study where we employed a behavioural paradigm and computational modelling to investigate how acute stress impacts reinforcement learning and the computational mechanisms behind such impact. Overall, our findings indicate that acute stress impaired behavioural performance towards obtaining rewards, but not towards avoiding punishments, and that this impaired performance may be explained by a decreased learning rate for positive prediction errors.

These findings are consistent with the neurobiological framework of stress-induced disruptions that we propose in the Introduction of this thesis. Acute stress is thought to increase spontaneous dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011), which in turn (and simultaneously) may blunt adaptative phasic dopamine responses that signal positive prediction errors (Daberkow et al., 2013; Maia & Frank, 2017; Werlen et al., 2020). Thus, it is possible that acute stress reduces the integration of positive prediction errors, which can be captured by a reinforcement-learning model as a decreased learning rate for positive prediction errors. By reducing how quickly the unexpected rewarding outcomes of choices are integrated over time, acute stress may impair behavioural performance towards rewards.

One major advantage of our stress manipulation is that it can be implemented during the learning task, so that the stress-induced emotional state is concurrent with the cognitive processes that operate during the task. The threat-of-shock paradigm offers the same advantage and has been widely used to induce robust acute stress responses by instructing participants that they will receive shocks anytime during the learning task (Robinson et al., 2013). Overall, threat-of-shock studies have indicated that acute stress impairs reward learning (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Paret & Bublatzky, 2020), and possibly to a larger extent than punishment learning (Berghorst et al., 2013). Thus, our behavioural findings seem consistent with those from threat-of-shock studies, which suggests that our novel acute stress manipulation is able to recapitulate the results from another well-established stress manipulation.

We found no evidence for an effect of acute stress on punishment learning. Instead, our behavioural analyses suggest a selective effect of acute stress on reward learning. In agreement with this, the literature suggests that the impact of acute stress on punishment learning is more equivocal than on reward learning (Porcelli & Delgado, 2017). Moreover, our computational-modelling analyses did not provide evidence that acute stress affected the learning rate for negative prediction errors —

which are thought to mostly drive punishment learning (Palminteri & Pessiglione, 2017) — to the same extent that it affected the learning rate for positive prediction errors. Whereas the involvement of dopamine release in signalling prediction errors during reward learning is well-established, the functional role of dopamine in punishment learning remains more controversial (Glimcher, 2011). Indeed, it has been proposed that mechanisms other than dopaminergic might be involved in punishment learning (Boureau & Dayan, 2011; Daw et al., 2002; Palminteri & Pessiglione, 2017). Thus, it is possible that stress-induced dopamine release does not affect punishment learning to the same extent as it affects reward learning.

To probe the computational mechanisms underlying stress-induced impaired reward learning, we used a computational reinforcement-learning model (Frank et al., 2007) that, although conceivably more simplistic than the processes engaged by the human brain, has been widely used in the literature and is considered to capture several behavioural and neural processes engaged during learning tasks (e.g., Diederen et al., 2017; Horga et al., 2015; Lefebvre et al., 2017; Niv et al., 2012). Specifically, this well-established, biologically inspired variant of the standard reinforcement-learning model takes into account the differential firing of dopaminergic neurons for positive (burst firing) and negative (dips) prediction errors and has been successfully used to capture variations in dopaminergic functioning (Doll et al., 2011; Frank et al., 2007; Frank & Fossella, 2011; Grogan et al., 2017; Rutledge et al., 2009). In our study, we used this reinforcement-learning model to estimate learning rates for positive and negative prediction errors in order to capture the putative impact of acute stress on adaptive phasic dopamine responses that encode positive and negative prediction errors (note that in the model, the learning rate controls the extent to which prediction errors are used to update current expectations). Although we found that acute stress decreased the learning rate for positive prediction errors, which suggests decreased phasic dopamine responses for positive prediction errors, we can only provide indirect evidence for the impact of acute stress on dopaminergic functioning. The combination of computational approaches with pharmacological and/or neuroimaging studies would be valuable to further support this hypothesis. This motivated us to conduct a study where we employed a model-based fMRI approach to inspect the impact of acute stress on the neural mechanisms of reinforcement learning — this study will be discussed in the next subsection.

In sum, Chapter 2 extends behavioural evidence that acute stress impairs reward learning, possibly to a larger extent than punishment learning. Importantly, computational modelling indicated that acute stress decreased the learning rate for positive prediction errors — consistent with stress-

induced dopamine disruptions. These findings shed new light on the computational mechanisms that underlie the deleterious impact of acute stress on reward learning.

6.2.2. Does acute stress affect neural signalling of prediction errors during reinforcement learning?

In **Chapter 3** of this thesis, we described a study where we employed model-based fMRI to investigate the neural mechanisms by which acute stress impacts reinforcement learning. Specifically, we were interested in examining whether and how acute stress would blunt striatal prediction error signals — the key neurocomputational mechanism of reinforcement learning. Overall, our findings indicate that acute stress blunts signalling of prediction errors in the dorsal striatum during reinforcement learning. Moreover, subsidiary analyses suggest that acute stress preferentially blunts positive prediction error signals — which occur when outcomes are better than expected. These data are consistent with our hypothesis that acute stress disrupts phasic dopamine responses that signal prediction errors, and mostly adaptive dopaminergic phasic burst firing that signal positive prediction errors. By preferentially decreasing positive prediction error signals in the dorsal striatum, acute stress may drive impaired behavioural performance towards rewards but not necessarily away from punishments — which is thought to rely at a larger extent on negative prediction errors than on positive prediction errors (Palminteri & Pessiglione, 2017). Indeed, our task-performance analyses indicate that acute stress selectively impairs reward learning, extending the behavioural findings from Chapter 2 for an independent sample.

It has been proposed that prediction error signals can be dissociated in the dorsal and ventral (nucleus accumbens) regions of the striatum depending on whether an active behavioural choice is required to obtain a reward or not (O'Doherty et al., 2004). O'Doherty et al. (2004) showed that in an active choice task (i.e, instrumental learning) prediction errors were detected both in the dorsal and ventral striatum, whereas in a passive learning task (i.e., pavlovian conditioning) prediction errors were detected only in the ventral striatum. Indeed, extant evidence supports a functional role of the dorsal striatum in implementing reward-based action selection (Lohrenz et al., 2007; Schönberg et al., 2007; Valentin & O'Doherty, 2009; Wang et al., 2016). As expected, during our active choice task, we detected prediction error signals both in the dorsal striatum and nucleus accumbens, but we found a blunting effect of acute stress on prediction error signalling only in the dorsal striatum. Our data thus suggest that acute stress impairs learning of stimulus-action-reward associations, which may be reflected in worse choice performance towards seeking rewards.

Our neural finding of stress-induced blunting signalling of prediction errors — and mostly positive prediction errors — in the dorsal striatum are in striking agreement with our previous computational findings in Chapter 2, which indicate that acute stress decreases the learning rate for positive prediction errors. However, in this study we were not able to replicate the effect of acute stress on parameter estimates, specifically on the learning rate for positive prediction errors. It is possible that due to methodological differences in the version of the fMRI reinforcement-learning task, such as number of trials and/or outcome probabilities, and a smaller sample size we were not able to obtain parameter estimates as stable as in our previous study (Katahira & Toyama, 2021; Wilson & Niv, 2015; Zhang et al., 2020). Still, as in this study we were focused on assessing how acute stress impacted neural signalling of prediction errors, and prediction errors themselves were estimated by the reinforcement-learning model (and then regressed against neural activity in the striatum), we conducted additional checks to validate the used reinforcement-learning model (these were presented in the Supplementary Material of Chapter 3, which can be found in Appendix 3). By confirming that the probability of choices estimated by the reinforcement-learning model followed the same pattern of the actual observed choices, and that parameter recovery was successful, the used reinforcement-learning model proved to be useful for making better sense of our neural data.

Rodent studies suggest that acute stress increases spontaneous phasic dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011), which in turn may blunt striatal dopamine prediction error signals that drive reinforcement learning (Daberkow et al., 2013; Werlen et al., 2020). Although our data cannot provide direct evidence for stress-induced spontaneous phasic dopamine release, given that prediction errors are encoded in the phasic adaptive activity of dopamine neurons (Daw & Tobler, 2014; Glimcher, 2011; Schultz et al., 1997), our findings from Chapters 2 and 3 are broadly consistent with the idea that acute stress impairs reward learning via putative striatal dopaminergic disruptions.

Finally, across Chapters 2 and 3, we observed that the impact of acute stress on reinforcement learning and underpinning prediction error signalling seems to vary substantially between individuals. In the next section we discuss the findings of two studies, using the datasets from Chapters 2 and 3, where we investigated how individual differences modulate the impact of acute stress on reinforcement learning and on striatal prediction error signalling.

In sum, Chapter 3 suggests that acute stress blunts signalling of prediction errors, in particular positive prediction errors, in the dorsal striatum and extends previous behavioural findings from Chapter 2 that acute stress impairs reward learning. Our data point to a neurocomputational mechanism —

consistent with a neurobiological framework of stress-induced dopaminergic disruptions — by which acute stress may impair reward learning.

6.2.3. Do individual differences in psychopathic traits modulate the impact of acute stress on reinforcement learning and neural signalling of prediction errors?

In **Chapters 4 and 5** of this thesis, we described two studies where we employed behavioural and model-based fMRI analyses, respectively, to investigate whether and how individual differences in psychopathic traits — which are constituted by affective-interpersonal features, such as lack of empathy and manipulativeness, and by deviant lifestyle-antisocial characteristics, such as impulsiveness and antisocial behaviour — modulate the impact of acute stress on behavioural performance towards obtaining rewards and avoiding punishments, and whether these modulatory effects could be explained by variations in striatal prediction error signalling. Overall, we found that acute stress impaired behavioural performance towards obtaining rewards in individuals with lower levels of psychopathic traits, but such stress-induced impairment was not observed in individuals with higher levels of psychopathic traits; moreover, we did not find evidence that psychopathic traits modulated the impact of acute stress on punishment learning (Chapter 4). These behavioural findings were replicated in an independent dataset (Chapter 5). Relatedly, at the neural level, we found that, during reward and punishment learning, acute stress blunted signalling of prediction in the dorsal striatum of individuals with lower levels of psychopathic traits — and particularly affective-interpersonal traits — but this stress-induced blunted signalling of prediction errors was attenuated in individuals with higher levels of these traits (Chapter 5). Together, our behavioural and neural findings indicate that the impairing effect of acute stress on reward learning decreases as psychopathic traits increase, which may be driven by diminished blunted prediction errors signals in the dorsal striatum. These results suggest that higher levels of psychopathic traits might have a protective role against the deleterious impact of acute stress on reward learning and underlying neurocomputational mechanisms.

Our neural finding that affective-interpersonal traits account for variations in how acute stress impacts signalling of prediction errors in the dorsal striatum, suggests that affective-interpersonal traits might be associated with altered dopaminergic functioning. Acute stress is thought to increase striatal dopamine release, which in turn may blunt adaptive phasic dopamine responses that signal prediction errors — and preferentially positive prediction errors — in the dorsal striatum (as we showed in Chapter 3). Our data suggest that acute stress blunts prediction error signals in individuals with lower levels of affective-interpersonal traits, but such stress-induced blunting effect seems attenuated in individuals

with higher levels of these traits. Therefore, it is possible that stress-induced dopamine release does not affect phasic dopamine responses that encode prediction errors to the same extent in individuals with different levels of psychopathic traits. Indeed, it has been proposed that psychopathic traits are associated with hyper-reactive striatal dopaminergic functioning, although the evidence is far more robust for the lifestyle-antisocial dimension than for the affective-interpersonal dimension of psychopathy (Buckholtz et al., 2010; Yildirim & Derksen, 2015). Whether and how affective-interpersonal traits — which constitute the core features of psychopathy — are associated with dopaminergic disturbances, remains poorly understood. Consequently, the precise neurobiological mechanisms by which affective-interpersonal traits modulate the impact of acute stress on signalling of prediction errors remain elusive. Further studies, combining computational modelling with pharmacological manipulations and/or positron emission tomography both in community and clinical samples could advance our knowledge about putative dopaminergic disturbances, and how acute stress interacts with those, across the two dimensions of psychopathy.

Our behavioural data across Chapters 4 and 5 suggest that psychopathic traits selectively account for variations in the impact of acute stress on reward learning. However, at the neural level, affective-interpersonal psychopathic traits were also associated with decreased stress-induced blunted signalling of prediction errors during punishment learning. Subsidiary analyses of our neural data suggest, albeit speculatively, that affective-interpersonal traits might be mostly associated with a diminished blunting effect of acute stress on positive prediction error signals across reward and punishment learning. Curiously, whereas reward learning is thought to be mainly driven by positive prediction errors, punishment learning is thought to be mostly driven by negative prediction errors (Palminteri & Pessiglione, 2017). It is thus possible that putative variations in signalling of positive prediction errors result in altered reward learning, but not necessarily in altered punishment learning. Still, further studies are required to directly test this hypothesis.

In sum, our data suggest that individual differences in psychopathic traits modulate the impact of acute stress on reward learning and signalling of prediction errors in the dorsal striatum. Whereas acute stress impairs performance towards obtaining rewards and blunts striatal signalling of prediction errors in individuals with lower levels of psychopathic traits such negative effects of acute stress are not observed in individuals with higher levels of these traits. Additionally, this modulatory effect seems to be mostly driven by stress-induced disruptions in positive prediction errors, which could conceivably explain why psychopathic traits modulate the impact of acute stress on behavioural performance during reward learning, but not during punishment learning. Our data highlights the substantial variability in

how acute stress impacts reward learning and underlying neural mechanisms, and the importance of considering individual differences that may contribute for such variability.

6.3. Limitations

In addition to specific limitations approached in the sections above, there are some general limitations in this thesis that should be considered.

We did not include women in our studies due to hormonal-dependent variations on stress responsivity as well as on reinforcement learning (Diekhof & Ratnayake, 2016; Dreher et al., 2007; Ossewaarde et al., 2010). Although our finding that acute stress impairs reward learning in men seems in line with the findings from other studies in women, future studies are required to assess the extent to which our behavioural and neurocomputational findings in men indeed generalise to women. This is particularly important, given that stress-related disorders, such as depression or generalised anxiety, are more prevalent in women (World Health Organization, 2017).

Our stress manipulation consisted of exposing participants to the stressor during the task to ensure that acute stress was contingent to the learning processes. However, one potential drawback of using acute stressors during a cognitive task, is that it can make it difficult to disentangle whether cognitive impairments occur as a result of acute stress or because the stressor acted as a distractor (Robinson et al., 2013). Although acute stress responses seem to be inherently characterized by a hypervigilant state and redirection of attention to stressors, conceivably causing distraction in tasks that require selective attention (Hermans et al., 2014), we took experimental precautions, such as reducing the unpredictability of the stressor, to minimise a potential distractor effect of the stressor. Our data suggest that is unlikely that our results were due to a distractor effect of the stressor, or else we would expect to observe general impairment instead of a selective impairment on reward learning. Still, additional control conditions, such as a non-stressful sound condition, could be added in future studies to disentangle distraction from stress.

In the fMRI study presented in Chapter 3, the stress induction increased self-reported stress levels, but to a lesser extent than in the study presented in Chapter 2, as the stressor could conceivably be perceived as less salient when administered inside the MRI scanner. Due to technical limitations, we did not analyse physiological responses, but inclusion of such variables in future research could contribute for our understanding of individual variations in stress responsiveness.

Finally, we used a university sample to assess whether and how psychopathic traits modulated the impact of acute stress on reinforcement learning and striatal signalling of prediction errors. By using

a university sample, we were able to capture broader variability in psychopathic traits, although a different pattern of results could have emerged in a clinical sample with a diagnosis of psychopathy. For instance, prior literature suggests that psychopathy is characterised by impairments in reinforcement learning, and more particularly in passive avoidance and reversal learning (R. J. R. Blair et al., 2005). Although our primary hypothesis focused on the interaction between psychopathic traits and acute stress, our data did not provide evidence for a deficit in reinforcement learning in the control condition in individuals with higher levels of psychopathic traits. Future studies using clinical samples should assess whether psychopathy is also characterised by behavioural and neural alterations in the type of reinforcement learning assessed by our task.

6.4. Implications and future directions

Acute stress is pervasive in everyday life and has major implications for well-being and mental health, and consequently has a high societal impact. It has been estimated that stress-related disorders, such as depression, anxiety disorders and alcohol and drug use disorders, affect more than one in six people across European Union countries, and that the total costs of mental ill-health have exceeded €600 billion – or more than 4% of gross domestic product – across the 28 European Union countries (OECD & European Union, 2018). Given that stress can predispose individuals to a broad range of psychopathology, investigating the mechanisms by which acute stress influences cognition and behaviour is critical not only for understanding the effects of acute stress day-to-day life, but may also offer important insights into the design of prevention and treatment strategies for stress-related clinical disorders.

Acute stress is known to be a major factor in triggering and/or worsening of symptoms in a broad range of psychiatric disorders — in particular those involving altered reinforcement learning and dopaminergic dysfunction — such as depression, anxiety and addiction. One putative mechanism that mediates this action is that acute stress increases dopamine release, which may interact with pre-existing differences in dopaminergic functioning, resulting in atypical reinforcement learning. For example, it is thought that acute stress may decrease reward responsiveness, particularly in individuals with anhedonic behaviour — a core symptom of depression — suggesting that vulnerability to depression might involve a susceptibility to stress-induced dopamine disruptions (Bogdan & Pizzagalli, 2006; Pizzagalli, 2014); anxiety has been linked with augmented avoidance of potential negative outcomes (Bishop & Gagne, 2018), particularly under acute stress (Mkrtchian et al., 2017), which possibly involves stress-induced dopaminergic modulation (Stelly et al., 2019); finally, acute stress plays

a powerful role in addiction processes rooted in reinforcement learning, such as the initiation, escalation, and relapse to drug abuse, via dopaminergic mechanisms (Koob & Volkow, 2016; Saal et al., 2003; Sinha, 2007). Thus, investigating how acute stress affects the computational and neural mechanisms of reinforcement learning — especially prediction errors, which implicate adaptative dopaminergic activity — might be important to better understand the role of acute stress on psychiatric disorders.

Reinforcement-learning models appear to reflect the neural activity of dopamine prediction errors in specific brain regions (Pessiglione et al., 2006; Schultz et al., 1997). Furthermore, the parameters estimated by these models, which are built to capture very specific cognitive and neural mechanisms, may offer a unique description of individual behaviour and can be used to tailor adequate interventions (Nair et al., 2020). Thus, such neurocomputational framework provides a better understanding of the mechanisms that drive learning and behaviour and holds potential to translate to psychological, and/or pharmacological, interventions.

In Chapters 2 and 3 we employed a well-established reinforcement-learning model (Frank et al., 2007) to investigate the computational and neural mechanisms behind the impact of acute stress on behavioural performance towards obtaining rewards and avoiding punishments. We found that acute stress decreased the learning rate for positive prediction errors (i.e., decreased how quickly unexpected rewarding outcomes were integrated over time) and blunted neural signalling of prediction errors — preferentially positive prediction errors — in the dorsal striatum, which may explain impaired behavioural performance towards obtaining rewards under acute stress. Thus, our findings suggest that, when under acute stress, the value of actions that resulted in past rewarding outcomes are reduced — due to decreased signalling/integration of positive prediction errors — and it is less likely that such action is chosen in the future. It is not well-known yet how this might translate into real-life settings, but we speculate that under acute stress people might not engage so much in novel pleasant actions, such as starting a new hobby, because the unexpected positive outcomes that result from those actions are poorly integrated, decreasing the value of the actions and the likelihood of selecting them again in the future. This can conceivably explain why acute stress has been robustly associated with habitual behaviours (Schwabe & Wolf, 2013), which are characterised by a loss of sensitivity to the rewarding outcomes of actions and do not require expectations (and respective update) of rewards (Lingawi et al., 2016). In other words, it is possible that acute stress interferes with selection of actions that implicate update of expectations by learning from the unexpected rewarding outcomes of those actions, i.e., positive prediction errors. Interestingly, it has been argued that effective psychotherapies

work by challenging and updating such expectations. Next, we advance some tentative ideas of how our findings could have implications for psychological interventions focused on targeting the negative impact of stress on reward learning and signalling of prediction errors.

Cognitive behavioural therapy (CBT) can be viewed as a form of prediction error-based learning because it challenges erroneous expectations and behavioural patterns in order to induce changes in cognition and behaviour (Nair et al., 2020; Papalini et al., 2020; Queirazza et al., 2019). Considering our findings, this might have two important implications. First, if CBT results in changes in expectations, then computational parameters such as the learning rate, which controls the extent to which the current expectation is updated by prediction errors, should determine the rate of this change. Thus, if acute stress blunts prediction error signals and decreases the learning rate for positive prediction errors, CBT can presumably be less effective when people are under acute stress and struggle to update the value of actions based on new positive information. Indeed, there is recent evidence that prediction error signals track and predict responses to CBT (Queirazza et al., 2019). Second, more targeted CBT approaches that seek to modify how rewarding information is used may increase the effectiveness of interventions by boosting reward learning under acute stress. These approaches may involve instructing the individual to explicitly quantify how rewarding an action will be when under acute stress. For example, to clearly operationalise expectations about what will happen and about how pleasant, or unpleasant, an activity will be and re-adjusting that based on the actual level of pleasure experienced, increases the possibility that a prediction error is computed and used to update reward expectations. Relatedly, approaches focused on promoting attention to the generated prediction errors (Niv et al., 2015) when individuals are under acute stress, may increase learning efficiency, as it increases the chances that prediction errors are optimally attended and used to update future expectations.

The computational characterisation of the effects of acute stress on brain and behaviour may offer additional granularity in understanding how psychological interventions, such as CBT, work when individuals experience acute stress, and therefore may help in tailoring more adequate interventions for individuals to overcome everyday challenges and for those suffering from stress-related disorders. Additionally, further investigation of stress-induced dopamine disruptions using computational modelling combined with pharmacological challenges, might provide insight into pathways through which dopamine-based pharmacological interventions could be used as an adjunct to improve both the response and the long-term effects of psychological therapies in stress-related disorders (Iglesias et al., 2017; Papalini et al., 2020). Finally, we highlight that given the role of individual differences in how

acute stress affects the computational and neural mechanisms of reinforcement learning, both psychological and pharmacological interventions should be specifically tailored to each individual.

In **Chapters 4 and 5** we found that individual differences in psychopathic traits modulated the impact of acute stress on behavioural and neural responses during reinforcement learning. Specifically, whereas in individuals with lower levels of psychopathic traits acute stress impaired reward learning and blunted signalling of prediction errors in the dorsal striatum, in individuals with higher levels of these traits such negative effects of acute stress were absent. Indeed, it has been proposed that psychopathic traits are associated with an ability to remain cool under pressure, and relative immunity from anxiety and stress (Sandvik et al., 2015). These features might even be advantageous in certain professions, such as law enforcement, the military, business and politics. Although it is thought that the behaviour of individuals with psychopathy is relatively unaffected by stressors, and that these individuals retain the ability to perform well under stress (Anestis et al., 2017; Sandvik et al., 2015), to our knowledge, no studies to date have empirically tested how psychopathic traits modulate the impact of acute stress on reinforcement learning and underpinning neural mechanisms.

Psychopathy imposes substantial societal costs and economic burden (Reidy et al., 2015). Investigating how individual differences in levels of psychopathic traits interact with acute stress within the general population can be valuable for furthering our understanding of maladaptive behaviours in psychopathy under acute stress, to prevent it and minimise their consequences. Nevertheless, it would be important to conduct further studies using clinical samples to elucidate how acute stress impacts the mechanisms of reinforcement learning in psychopathy, and, importantly, how this impact is translated into relevant and meaningful real word behaviour (for e.g., as number and types of criminal convictions, incarcerations, etc.).

In addition to specific limitations and future directions approached in the sections above, future work could also extend the data presented in this thesis in several ways. For example, this thesis focused on examining the impact of acute stress on reinforcement learning, but chronic stress, which occurs for longer periods of time, is also known to affect reinforcement learning (Corral-Frías et al., 2015; Novick et al., 2018; Pechtel & Pizzagalli, 2011). Both acute stress and chronic stress seem to be associated with difficulties in reinforcement learning, but through seemingly different underlying mechanisms (Holly & Miczek, 2016). Whereas acute stress has been associated with increased striatal dopamine release (Booij et al., 2016; Cabib & Puglisi-Allegra, 2012; Nagano-Saito et al., 2013; Pruessner et al., 2004), chronic stress, such as chronic psychosocial adversity, has been more strongly associated with dampened striatal dopaminergic functioning (Bloomfield et al., 2019). Thus, it could be

interesting to assess whether chronic stress would be associated with impaired reward learning and/or punishment learning, similarly to what we had found for acute stress, and to examine whether such impairment would also be underpinned by blunted striatal prediction error signals. Additionally, acute stress and chronic stress may interact in their detrimental effect on reinforcement learning (Radenbach et al., 2015). Thus, further studies could examine whether and how chronic stress moderates the effects of acute stress on reinforcement learning and underlying neural mechanisms.

Although our behavioural results obtained across two independent samples (Chapters 2 and 3) seem in agreement with the pattern of results obtained in threat-of shock-studies (Berghorst et al., 2013; Bogdan et al., 2011; Bogdan & Pizzagalli, 2006; Paret & Bublatzky, 2020), given the mixed findings in the stress literature — possibly due to critical methodological differences related to stress and learning operationalization (Porcelli & Delgado, 2017) — it would be important to compare whether and how stress manipulations that induce stress responses before (e.g. cold pressor test of Trier social stress test) and during (e.g. threat of shock or our manipulation) learning tasks result in behavioural and neural differences in the same type of reinforcement learning.

Another potentially interesting future direction for this work would consist of testing whether acute stress promotes aberrant salience of neutral stimuli. It is thought that acute stress increases aberrant spontaneous phasic dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011), which in turn has been linked to aberrant attribution of salience to irrelevant or neutral stimuli and abnormal learning of stimuli that would otherwise be considered irrelevant (Boehme et al., 2015; Maia & Frank, 2017; McCutcheon et al., 2019; Roiser et al., 2013). Thus, it is possible that acute stress promotes aberrant salience (Lopez & Flagel, 2020), driven by stress-induced aberrant spontaneous phasic dopamine release, but this remains unclear. This hypothesis could be tested through a behavioural task specifically designed to measure aberrant learning of neutral stimuli on a trial-by-trial basis, which allows modelling and examination of neural signalling of prediction errors in response to neutral outcomes. Ultimately, this could contribute to understanding whether and how acute stress influences the brain to produce abnormal stimulus-reinforcement associations, which seem to be a hallmark of some psychiatric symptoms, such as psychosis (McCutcheon et al., 2019; Roiser et al., 2013), that are often aggravated by acute stress (Mondelli, 2014; Myin-Germeys & van Os, 2007; van Winkel et al., 2008; Walker & Diforio, 1997).

Likewise, it would be interesting to adapt our fMRI reinforcement-learning task to increase the variance in the magnitude of prediction errors and allow parametric modelling of positive and negative prediction errors (for example, by increasing the number of trials and/or pairs of symbols and changing

the reward and punishment contingencies). This could help to clarify more precisely the impact of acute stress on striatal signalling of positive and negative prediction errors — which are known to be respectively signaled by dopamine burst firing and dips (Schultz et al., 1997) — and further contribute for our understanding of the specific mechanisms of reinforcement learning affected by acute stress.

Moreover, it has been recently proposed that reinforcement learning and working memory do not operate separately, but that working memory may interfere with reinforcement learning computations via prefrontal cortex mechanisms (Collins & Frank, 2012). Future work could dissect the extent to which acute stress affects these two seemingly interconnected processes and related underlying neural mechanisms.

Finally, on a more general note, although parameters from reinforcement-learning models drive individual behaviour, it is also important to develop research that attempts to capture real-life phenomena in ecological settings, to be closer to the environment in which the brain is tuned to and where (mal)adaptative behaviour is generated (Scholl & Klein-Flügge, 2018). This is particularly important in stress studies, to ensure that task-derived parameters affected by a laboratory stress manipulation translate into real-life stress-induced alterations. Recent advances in smartphone-based measurements of mood and decision-making (Eldar et al., 2018; Rutledge et al., 2014) can contribute for the validation of task-derived parameters alongside the real-life phenomena they presumably capture.

6.5. Conclusions

In this thesis we have contributed to the characterisation of the impact of acute stress on the computational and neural mechanisms of reinforcement learning (see Figure 6.1 for a summary of the main findings). Our behavioural and computational findings suggest that acute stress impairs reward learning and decreases the learning rate for positive prediction errors. At the neural level, our results suggest that acute stress blunts signalling of prediction errors, preferentially positive prediction errors, in the dorsal striatum during reinforcement learning. These findings suggest a neurocomputational mechanism by which acute stress may impair reward learning — consistent with a neurobiological framework of stress-induced dopamine disruptions. Furthermore, acute stress seems to impair reward learning to a larger extent than punishment learning. Finally, our results indicate that individual differences in psychopathic traits modulate the impact of acute stress on reward learning and signalling

of prediction errors in the dorsal striatum, suggesting substantial inter-individual variability in how acute stress affects reward learning.

In sum, we consider that the findings of this thesis add important evidence for a more complete and coherent understanding of the impact of acute stress on reinforcement learning.

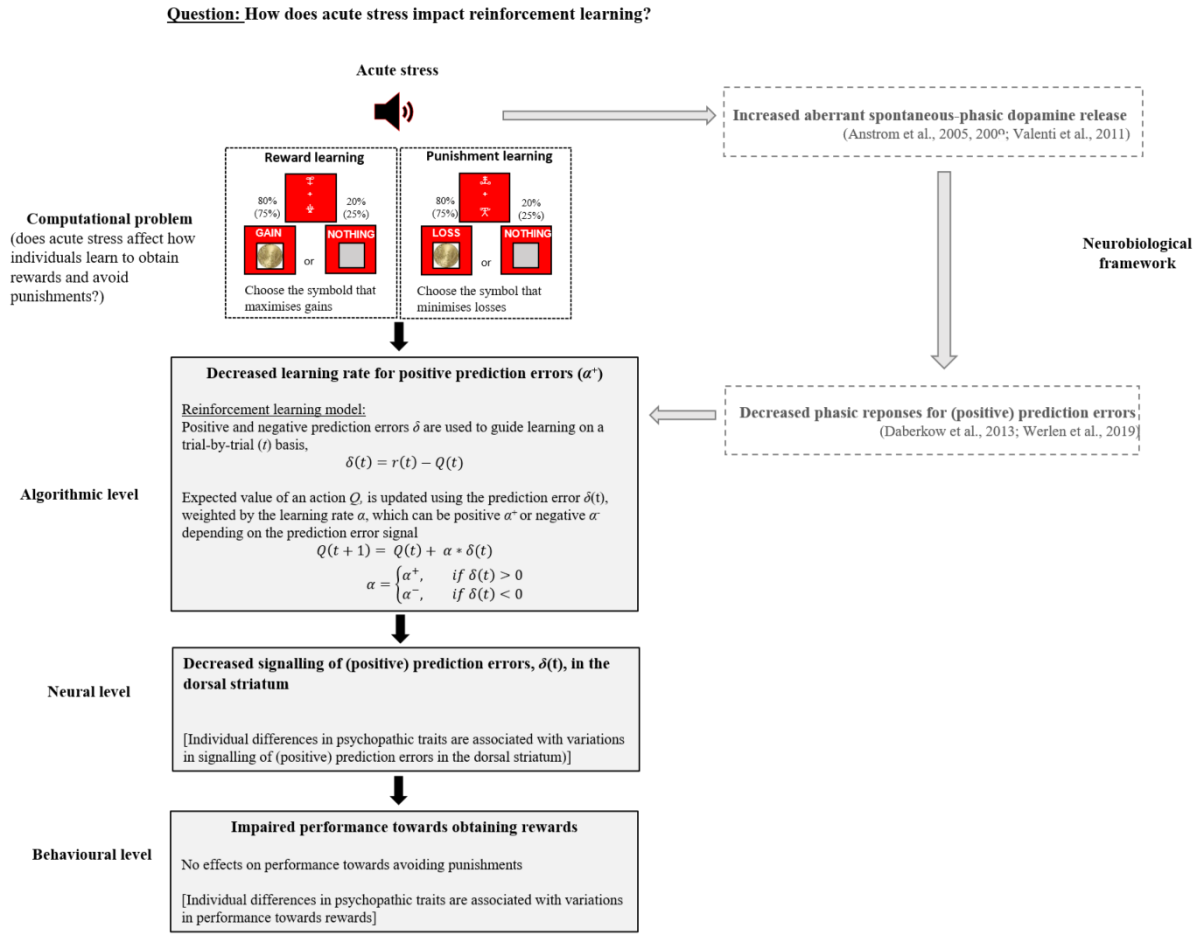


Figure 6.1. Integrated summary of the main findings. In this thesis we aimed to investigate the impact of acute stress on reinforcement learning. Participants were required to solve a computational problem — learning which actions were most and least rewarding/punishing — under acute stress and control conditions. Participants chose between stimuli which differed in terms of reward and punishment probabilities (80% / 20% in Chapter 2 task, and 75%/25% in Chapter 3 task), and through trial and error they learned to choose and to avoid the best and worst options, respectively. To capture the impact of acute stress on their behaviour, a computational reinforcement-learning model (grey box on the top) was fitted to the behavioural data. Model fitting indicated that acute stress decreased the learning rate for positive prediction errors (Chapter 2). In an independent sample, the prediction errors estimated by the computational model were correlated

with striatal BOLD activity. Those model-based fMRI analyses indicated that acute stress decreased signalling of prediction errors — and preferentially positive prediction errors — in the dorsal striatum (Chapter 3). In both studies, behavioural analyses showed that acute stress impaired performance towards obtaining rewards, but no effects were found for performance towards avoiding punishments (Chapters 2 and 3). Additionally, the described behavioural and neural effects of acute stress were modulated by individual differences in psychopathic traits (Chapters 4 and 5). This work suggests a neurocomputational mechanism by which acute stress may impair reward learning, consistent with a neurobiological framework of stress-induced dopaminergic disruptions (dashed boxes on the right).

REFERENCES

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S., & Zigmond, M. J. (1989). Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex. *Journal of Neurochemistry*, *52*(5), 1655–1658. https://doi.org/10.1111/j.1471-4159.1989.tb09224.x

Adler, C. M., Elman, I., Weisenfeld, N., Kestler, L., Pickar, D., & Breier, A. (2000). Effects of acute metabolic stress on striatal dopamine release in healthy volunteers. *Neuropsychopharmacology*, *22*(5), 545–550. https://doi.org/10.1016/S0893-133X(99)00153-0

Ali, N., Nitschke, J. P., Cooperman, C., & Pruessner, J. C. (2017). Suppressing the endocrine and autonomic stress systems does not impact the emotional stress experience after psychosocial stress. *Psychoneuroendocrinology*, *78*, 125–130. https://doi.org/10.1016/j.psyneuen.2017.01.015

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier social stress test: Principles and practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Anestis, J. C., Harrop, T. M., Green, B. A., & Anestis, M. D. (2017). Psychopathic personality traits as protective factors against the development of post-traumatic stress disorder symptoms in a sample of national guard combat veterans. *Journal of Psychopathology and Behavioral Assessment*, *39*(2), 220–229. https://doi.org/10.1007/s10862-017-9588-8

Anstrom, K. K., Miczek, K. A., & Budygin, E. A. (2009). Increased phasic dopamine signaling in the mesolimbic pathway during social defeat in rats. *Neuroscience*, *161*(1), 3–12. https://doi.org/10.1016/j.neuroscience.2009.03.023

Anstrom, K. K., & Woodward, D. J. (2005). Restraint increases dopaminergic burst firing in awake rats. *Neuropsychopharmacology*, *30*(10), 1832–1840. https://doi.org/10.1038/sj.npp.1300730

Arguelles, A. E., Ibeas, D., Ottone, J. P., & Chekherdemian, M. (1962). Pituitary-adrenal stimulation by sound of different frequencies. *The Journal of Clinical Endocrinology and Metabolism*, *22*(8), 846–852. https://doi.org/10.1210/jcem-22-8-846

Arnett, P. A. (1997). Autonomic responsivity in psychopaths: A critical review and theoretical proposal. *Clinical Psychology Review*, *17*(8), 903–936. https://doi.org/10.1016/S0272-7358(97)00045-7

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422. https://doi.org/10.1038/nrn2648

Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive function in monkeys: Evidence for a hyperdopaminergic mechanism. *Archives of General Psychiatry*, *55*(4). https://doi.org/10.1001/archpsyc.55.4.362

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

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. https://doi.org/10.1038/s41562-019-0628-0

Baskin–Sommers, A., Curtin, J. J., Li, W., & Newman, J. P. (2012). Psychopathy-related differences in selective attention are captured by an early event-related potential. *Personality disorders*, *3*(4), 370–378. https://doi.org/10.1037/a0025593

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020

Belujon, P., & Grace, A. A. (2015). Regulation of dopamine system responsivity and its adaptive and pathological response to stress. *Proceedings. Biological Sciences*, *282*(1805), 20142516. PubMed. https://doi.org/10.1098/rspb.2014.2516

Berghorst, L. H., Bogdan, R., Frank, M. J., & Pizzagalli, D. A. (2013). Acute stress selectively reduces reward sensitivity. *Frontiers in Human Neuroscience*, *7*(133). https://doi.org/10.3389/fnhum.2013.00133

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology*, *191*(3), 391–431. https://doi.org/10.1007/s00213-006-0578-x

Binti Affandi, A. H., Pike, A. C., & Robinson, O. J. (2021). Threat of shock promotes passive avoidance, but not active avoidance. *European Journal of Neuroscience*, *00*(n/a), 1–10. https://doi.org/10.1111/ejn.15184

Birk, R. H. (2020). On stress and subjectivity. *Theory & Psychology*, 0959354320953904. https://doi.org/10.1177/0959354320953904

Bishop, S. J., & Gagne, C. (2018). Anxiety, depression, and decision making: A computational perspective. *Annual Review of Neuroscience*, *41*, 371–388. https://doi.org/10.1146/annurev-neuro-080317-062007

Blair, K. S., Morton, J., Leonard, A., & Blair, R. J. R. (2006). Impaired decision-making on the basis of both reward and punishment information in individuals with psychopathy. *Personality and Individual Differences*, *41*(1), 155–165. https://doi.org/10.1016/j.paid.2005.11.031

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387–392. https://doi.org/10.1016/j.tics.2007.07.003

Blair, R. J. R. (2013a). Psychopathy: Cognitive and neural dysfunction. *Dialogues in Clinical Neuroscience*, *15*(2), 181–190. PubMed. https://doi.org/10.31887/DCNS.2013.15.2/rblair

Blair, R. J. R. (2013b). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, *14*(11), 786–799. https://doi.org/10.1038/nrn3577

Blair, R. J. R. (2017). Emotion-based learning systems and the development of morality. *Cognition*, *167*, 38–45. https://doi.org/10.1016/j.cognition.2017.03.013

Blair, R. J. R. (2019). Dysfunctional neurocognition in individuals with clinically significant psychopathic traits. *Dialogues in Clinical Neuroscience*, *21*(3), 291–299. https://doi.org/10.31887/DCNS.2019.21.3/rblair

Blair, R. J. R., Mitchell, D., & Blair, K. S. (2005). *The Psychopath: Emotion and the Brain*. Oxford: Blackwell Publishing.

Blair, R. J. R., & Mitchell, D. G. V. (2009). Psychopathy, attention and emotion. *Psychological Medicine*, *39*(4), 543–555. https://doi.org/10.1017/S0033291708003991

Blair, R. J. R., Mitchell, D. G. V., Leonard, A., Budhani, S., Peschardt, K. S., & Newman, C. (2004). Passive avoidance learning in individuals with psychopathy: Modulation by reward but not by punishment. *Personality and Individual Differences*, *37*(6), 1179–1192. https://doi.org/10.1016/j.paid.2003.12.001

Bloomfield, M. A., McCutcheon, R. A., Kempton, M., Freeman, T. P., & Howes, O. (2019). The effects of psychosocial stress on dopaminergic function and the acute stress response. *eLife*, *8*, e46797. https://doi.org/10.7554/eLife.46797

Boehme, R., Deserno, L., Gleich, T., Katthagen, T., Pankow, A., Behr, J., Buchert, R., Roiser, J. P., Heinz, A., & Schlagenhauf, F. (2015). Aberrant salience is related to reduced reinforcement learning signals and elevated dopamine synthesis capacity in healthy adults. *The Journal of Neuroscience*, *35*(28), 10103–10111. PubMed. https://doi.org/10.1523/JNEUROSCI.0805-15.2015

Bogdan, R., & Pizzagalli, D. A. (2006). Acute stress reduces reward responsiveness: Implications for depression. *Biological Psychiatry*, *60*(10), 1147–1154. https://doi.org/10.1016/j.biopsych.2006.03.037

Bogdan, R., Santesso, D. L., Fagerness, J., Perlis, R. H., & Pizzagalli, D. A. (2011). Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. *The Journal of Neuroscience*, *31*(37), 13246–13254. https://doi.org/10.1523/JNEUROSCI.2661-11.2011

Booij, L., Welfeld, K., Leyton, M., Dagher, A., Boileau, I., Sibon, I., Baker, G. B., Diksic, M., Soucy, J.-P., Pruessner, J. C., Cawley-Fiset, E., Casey, K. F., & Benkelfat, C. (2016). Dopamine cross-sensitization between psychostimulant drugs and stress in healthy male volunteers. *Translational Psychiatry*, *6*(2), e740. https://doi.org/10.1038/tp.2016.6

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-1126-0

Boureau, Y.-L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36*(1), 74–97. https://doi.org/10.1038/npp.2010.151

Boyle, C. C., Kuhlman, K. R., Dooley, L. N., Haydon, M. D., Robles, T. F., Ang, Y.-S., Pizzagalli, D. A., & Bower, J. E. (2019). Inflammation and dimensions of reward processing following exposure to the influenza vaccine. *Psychoneuroendocrinology*, *102*, 16–23. https://doi.org/10.1016/j.psyneuen.2018.11.024

Brazil, I. A., de Bruijn, E. R. A., Bulten, B. H., von Borries, A. K. L., van Lankveld, J. J. D. M., Buitelaar, J. K., & Verkes, R. J. (2009). Early and late components of error monitoring in violent offenders with psychopathy. *Biological Psychiatry*, *65*(2), 137–143. https://doi.org/10.1016/j.biopsych.2008.08.011

Brazil, I. A., Hunt, L. T., Bulten, B. H., Kessels, R. P. C., de Bruijn, E. R. A., & Mars, R. B. (2013). Psychopathy-related traits and the use of reward and social information: A computational approach. *Frontiers in Psychology*, *4*(952). https://doi.org/10.3389/fpsyg.2013.00952

Brazil, I. A., Maes, J. H. R., Scheper, I., Bulten, B. H., Kessels, R. P. C., Verkes, R. J., & de Bruijn, E. R. A. (2013). Reversal deficits in individuals with psychopathy in explicit but not implicit learning conditions. *Journal of Psychiatry and Neuroscience*, *38*(4), E13–E20. https://doi.org/10.1503/jpn.120152

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & De Bruijn, E. R. A. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biological Psychiatry*, *69*(7), 693–699. https://doi.org/10.1016/j.biopsych.2010.11.013

Breier, A., Albus, M., Pickar, D., Zahn, T. P., Wolkowitz, O. M., & Paul, S. M. (1987). Controllable and uncontrollable stress in humans: Alterations in mood and neuroendocrine and psychophysiological function. *American Journal of Psychiatry*, *144*(11), 1419–1425. https://doi.org/10.1176/ajp.144.11.1419

Brett M., Anton J.L., Valabregue R., & Poline J.B. (2002). Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain. Available on CD-ROM. *NeuroImage*.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., Ansari, M. S., Baldwin, R. M., Schwartzman, A. N., Shelby, E. S., Smith, C. E., Cole, D., Kessler, R. M., & Zald, D. H. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nature Neuroscience*, *13*(4), 419–421. https://doi.org/10.1038/nn.2510

Budhani, S., Richell, R. A., & Blair, R. J. R. (2006). Impaired reversal but intact acquisition: Probabilistic response reversal deficits in adult individuals with psychopathy. *Journal of Abnormal Psychology*, *115*(3), 552–558. https://doi.org/10.1037/0021-843X.115.3.552

Bullinger, M., Naber, D., Pickar, D., Cohen, R. M., Kalin, N. H., Pert, A., & Bunney, W. E. J. (1984). Endocrine effects of the cold pressor test: Relationships to subjective pain appraisal and coping. *Psychiatry Research*, *12*(3), 227–233. https://doi.org/10.1016/0165-1781(84)90028-3

Byrne, K. A., Cornwall, A. C., & Worthy, D. A. (2019). Acute stress improves long-term reward maximization in decision-making under uncertainty. *Brain and Cognition*, *133*, 84–93. https://doi.org/10.1016/j.bandc.2019.02.005

Cabib, S., & Puglisi-Allegra, S. (2012). The mesoaccumbens dopamine in coping with stress. *Neuroscience and Biobehavioral Reviews*, *36*(1), 79–89. https://doi.org/10.1016/j.neubiorev.2011.04.012

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*(8), 1111–1134. https://doi.org/10.1016/j.psyneuen.2011.12.010

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H.-J., Bonci, A., & Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature Neuroscience*, *19*(1), 111–116. https://doi.org/10.1038/nn.4191

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition*, *113*(3), 350–364. https://doi.org/10.1016/j.cognition.2008.06.014

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q. J. M., Düzel, E., & Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nature Neuroscience*, *16*(5), 648–653. https://doi.org/10.1038/nn.3364

Cima, M., Smeets, T., & Jelicic, M. (2008). Self-reported trauma, cortisol levels, and aggression in psychopathic and non-psychopathic prison inmates. *Biological Psychology*, *78*(1), 75–86. https://doi.org/10.1016/j.biopsycho.2007.12.011

Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: A review. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(2), 113–125. https://doi.org/10.3758/CABN.8.2.113

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European Journal of Neuroscience*, *35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*(3), 337–366. https://doi.org/10.1037/a0037015

Corral-Frías, N. S., Nikolova, Y. S., Michalski, L. J., Baranger, D. A. A., Hariri, A. R., & Bogdan, R. (2015). Stress-related anhedonia is associated with ventral striatum reactivity to reward and transdiagnostic psychiatric symptomatology. *Psychological Medicine*, *45*(12), 2605–2617. https://doi.org/10.1017/S0033291715000525

Covey, D. P., Roitman, M. F., & Garris, P. A. (2014). Illicit dopamine transients: Reconciling actions of abused drugs. *Trends in Neurosciences*, *37*(4), 200–210. https://doi.org/10.1016/j.tins.2014.02.002

Cremer, A., Kalbe, F., Gläscher, J., & Schwabe, L. (2021). Stress reduces both model-based and model-free neural computations during flexible learning. *NeuroImage*, 117747. https://doi.org/10.1016/j.neuroimage.2021.117747

Daberkow, D. P., Brown, H. D., Bunner, K. D., Kraniotis, S. A., Doellman, M. A., Ragozzino, M. E., Garris, P. A., & Roitman, M. F. (2013). Amphetamine paradoxically augments exocytotic

dopamine release and phasic dopamine signals. *The Journal of Neuroscience*, *33*(2), 452–463. https://doi.org/10.1523/JNEUROSCI.2136-12.2013

Danjo, T., Yoshimi, K., Funabiki, K., Yawata, S., & Nakanishi, S. (2014). Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. *Proceedings of the National Academy of Sciences*, *111*(17), 6455. https://doi.org/10.1073/pnas.1404323111

Dargis, M., Wolf, R. C., & Koenigs, M. (2017). Reversal learning deficits in criminal offenders: Effects of psychopathy, substance use, and childhood maltreatment history. *Journal of Psychopathology and Behavioral Assessment*, *39*(2), 189–197. https://doi.org/10.1007/s10862-016-9574-6

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. 1–26). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Daw, N. D., & Frank, M. J. (2009). Reinforcement learning and higher level cognition: Introduction to special issue. *Cognition*, *113*(3), 259–261. https://doi.org/10.1016/j.cognition.2009.09.005

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4–6), 603–616. https://doi.org/10.1016/s0893-6080(02)00052-7

Daw, N. D., & Tobler, P. N. (2014). Value learning through reinforcement: The basics of dopamine and reinforcement learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain* (2nd ed., pp. 283–298). Elsevier Inc. http://www.sciencedirect.com/science/article/pii/B9780124160088000152

Day, J. J., Roitman, M. F., Wightman, R. M., & Carelli, R. M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neuroscience*, *10*(8), 1020–1028. https://doi.org/10.1038/nn1923

de Berker, A. O., Tirole, M., Rutledge, R. B., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Acute stress selectively impairs learning to act. *Scientific Reports*, *6*, 29816. https://doi.org/10.1038/srep29816

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0065566

de Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience*, *6*(6), 463–475. https://doi.org/10.1038/nrn1683

Diederen, K. M. J., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. *The Journal of Neuroscience*, *37*(7), 1708–1720. https://doi.org/10.1523/JNEUROSCI.1979-16.2016

Diekhof, E. K., & Ratnayake, M. (2016). Menstrual cycle phase modulates reward sensitivity and performance monitoring in young women: Preliminary fMRI evidence. *Neuropsychologia*, *84*, 70–80. https://doi.org/10.1016/j.neuropsychologia.2015.10.016

Dindo, L., & Fowles, D. (2011). Dual temperamental risk factors for psychopathic personality: Evidence from self-report and skin conductance. *Journal of Personality and Social Psychology*, *100*(3), 557–566. https://doi.org/10.1037/a0021848

Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *The Journal of Neuroscience*, *31*(16), 6188–6198. https://doi.org/10.1523/JNEUROSCI.6486-10.2011

Dreher, J.-C., Schmidt, P. J., Kohn, P., Furman, D., Rubinow, D., & Berman, K. F. (2007). Menstrual cycle phase modulates reward-related neural function in women. *Proceedings of the National Academy of Sciences*, *104*(7), 2465–2470. https://doi.org/10.1073/pnas.0605569104

Edelstein, R. S., Yim, I. S., & Quas, J. A. (2010). Narcissism predicts heightened cortisol reactivity to a psychosocial stressor in men. *Journal of Research in Personality*, *44*(5), 565–572. https://doi.org/10.1016/j.jrp.2010.06.008

Ehlers, M. R., & Todd, R. M. (2017). Acute psychophysiological stress impairs human associative learning. *Neurobiology of Learning and Memory*, *145*, 84–93. https://doi.org/10.1016/j.nlm.2017.09.003

Eisenegger, C., Naef, M., Linssen, A., Clark, L., Gandamaneni, P. K., Müller, U., & Robbins, T. W. (2014). Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology*, *39*(10), 2366–2375. https://doi.org/10.1038/npp.2014.84

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of reward learning signals predicts mood fluctuations. *Current Biology*, *28*(9), 1433-1439.e7. https://doi.org/10.1016/j.cub.2018.03.038

Finger, E. C., Marsh, A. A., Blair, K. S., Reid, M. E., Sims, C., Ng, P., Pine, D. S., & Blair, R. J. R. (2011). Disrupted reinforcement signaling in the orbitofrontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. *American Journal of Psychiatry*, *168*(2), 152–162. https://doi.org/10.1176/appi.ajp.2010.10010129

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902. https://doi.org/10.1126/science.1077349

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72. https://doi.org/10.1162/0898929052880093

Frank, M. J. (2015). Linking across levels of computation in model-based cognitive neuroscience. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (p. 159–177). Springer Science + Business Media.

Frank, M. J., & Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology*, *36*(1), 133–152. https://doi.org/10.1038/npp.2010.96

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

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, *120*(3), 497–517. https://doi.org/10.1037/0735-7044.120.3.497

Fung, M. T., Raine, A., Loeber, R., Lynam, D. R., Steinhauer, S. R., Venables, P. H., & Stouthamer-Loeber, M. (2005). Reduced electrodermal activity in psychopathy-prone adolescents. *Journal of Abnormal Psychology*, *114*(2), 187–196. https://doi.org/10.1037/0021-843X.114.2.187

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

Gläscher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(4), 501–510. https://doi.org/10.1002/wcs.57

Glass, D. C., Reim, B., & Singer, J. E. (1971). Behavioral consequences of adaptation to controllable and uncontrollable noise. *Journal of Experimental Social Psychology*, *7*(2), 244–257. https://doi.org/10.1016/0022-1031(71)90070-9

Glienke, K., Wolf, O. T., & Bellebaum, C. (2015). The impact of stress on feedback and error processing during behavioral adaptation. *Neuropsychologia*, *71*, 181–190. https://doi.org/10.1016/j.neuropsychologia.2015.04.004

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

Gordts, S., Uzieblo, K., Neumann, C., Bussche, E. V. D., & Rossi, G. (2017). Validity of the Self-Report Psychopathy Scales (SRP-III full and short versions) in a community sample. *Assessment*, *24*(3), 308–325. https://doi.org/10.1177/1073191115606205

Grace, A. A. (1988). In vivo and in vitro intracellular recordings from rat midbrain dopamine neurons. *Annals of the New York Academy of Sciences*, *537*, 51–76. https://doi.org/10.1111/j.1749-6632.1988.tb42096.x

Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience*, *17*(8), 524–532. https://doi.org/10.1038/nrn.2016.57

Grace, A. A., & Bunney, B. S. (1984). The control of firing pattern in nigral dopamine neurons: Burst firing. *The Journal of Neuroscience*, *4*(11), 2877–2890. https://doi.org/10.1523/JNEUROSCI.04-11-02877.1984

Gregory, S., Blair, R. J. R., Ffytche, D., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: A case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. *The Lancet Psychiatry*, *2*(2), 153–160. https://doi.org/10.1016/S2215-0366(14)00071-6

Grogan, J. P., Tsivos, D., Smith, L., Knight, B. E., Bogacz, R., Whone, A., & Coulthard, E. J. (2017). Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease. *ELife*, *6*. https://doi.org/10.7554/eLife.26801

Guay, J.-P., Ruscio, J., Knight, R. A., & Hare, R. D. (2007). A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology*, *116*(4), 701–716. https://doi.org/10.1037/0021-843X.116.4.701

Hall, S. J., Aisbett, B., Tait, J. L., Turner, A. I., Ferguson, S. A., & Main, L. C. (2016). The acute physiological stress response to an emergency alarm and mobilization during the day and at night. *Noise & Health*, *18*(82), 150–156. https://doi.org/10.4103/1463-1741.181998

Hamilton, R. B., & Newman, J. P. (2018). The response modulation hypothesis: Formulation, development, and implications for psychopathy. In *Handbook of psychopathy* (2nd ed., pp. 80–93). The Guilford Press.

Hare, R. D. (2003). *The Hare Psychopathy Checklist Revised* (2nd ed.). Multi-Health Systems.

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology*, *4*(1), 217–246. https://doi.org/10.1146/annurev.clinpsy.3.022806.091452

Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, *37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian averaging models. *Statistical Science*, *14*(4), 382–417.

Holi, M., Auvinen-Lintunen, L., Lindberg, N., Tani, P., & Virkkunen, M. (2006). Inverse correlation between severity of psychopathic traits and serum cortisol levels in young adult violent male offenders. *Psychopathology*, *39*(2), 102–104. https://doi.org/10.1159/000091021

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*(4), 304–309. https://doi.org/10.1038/1124

Holly, E. N., & Miczek, K. A. (2016). Ventral tegmental area dopamine revisited: Effects of acute and repeated stress. *Psychopharmacology*, *233*(2), 163–186. https://doi.org/10.1007/s00213-015-4151-3

Hoppenbrouwers, S. S., Bulten, Berend H., & Brazil, Inti A. (2016). Parsing fear: A reassessment of the evidence for fear deficits in psychopathy. *Psychological Bulletin*, *142*(6), 573–600.

Horga, G., Maia, T. V., Marsh, R., Hao, X., Xu, D., Duan, Y., Tau, G. Z., Graniello, B., Wang, Z., Kangarlu, A., Martinez, D., Packard, M. G., & Peterson, B. S. (2015). Changes in corticostriatal connectivity during reinforcement learning in humans. *Human Brain Mapping*, *36*(2), 793–803. https://doi.org/10.1002/hbm.22665

House, T. H., & Milligan, W. L. (1976). Autonomic responses to modeled distress in prison psychopaths. *Journal of Personality and Social Psychology*, *34*(4), 556–560. https://doi.org/10.1037//0022-3514.34.4.556

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30–41. https://doi.org/10.1002/pchj.44

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238

Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 12. https://doi.org/10.1186/2045-5380-3-12

Iglesias, S., Tomiello, S., Schneebeli, M., Stephan, K.E. (2017). Models of neuromodulation for computational psychiatry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(3). https://doi.org/10.1002/wcs.1420

Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience*, *10*(6), 459–466. https://doi.org/10.1038/nrn2632

Johnson, M. M., Mikolajewski, A., Shirtcliff, E. A., Eckel, L. A., & Taylor, J. (2015). The association between affective psychopathic traits, time incarcerated, and cortisol response to psychosocial stress. *Hormones and Behavior*, *72*, 20–27. https://doi.org/10.1016/j.yhbeh.2015.04.010

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Katahira, K., & Toyama, A. (2021). Revisiting the importance of model fitting for model-based fMRI: It does matter in computational psychiatry. *PLOS Computational Biology*, *17*(2), e1008738. https://doi.org/10.1371/journal.pcbi.1008738

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & biological engineering & computing*, *42*(3), 419–427. https://doi.org/10.1007/BF02344719

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1–2), 76–81. https://doi.org/10.1159/000119004

Knutson, B., Delgado, M. R., & Phillips, P. E. M. (2009). Chapter 25—Representation of Subjective Value in the Striatum. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 389–406). Academic Press. https://doi.org/10.1016/B978-0-12-374176-9.00025-7

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: A neurocircuitry analysis. *The Lancet. Psychiatry*, *3*(8), 760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Kruse, O., Tapia León, I., Stalder, T., Stark, R., & Klucken, T. (2018). Altered reward learning and hippocampal connectivity following psychosocial stress. *NeuroImage*, *171*, 15–25. https://doi.org/10.1016/j.neuroimage.2017.12.076

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., Brooks, N. H., & Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*, *43*(7), 1581–1588. https://doi.org/10.1038/s41386-018-0032-x

Kuorinka, I., & Korhonen, O. (1981). Firefighters' reaction to alarm, an ECG and heart rate study. *Journal of Occupational Medicine*, *23*(11), 762–766. https://doi.org/10.1097/00043764-198111000-00010

Lebreton, M., Bavard, S., Daunizeau, J., & Palminteri, S. (2019). Assessing inter-individual differences with task-related functional neuroimaging. *Nature Human Behaviour*, *3*(9), 897–905. https://doi.org/10.1038/s41562-019-0681-8

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 0067. https://doi.org/10.1038/s41562-017-0067

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*(3), 334–342. https://doi.org/10.1016/j.cognition.2012.06.002

Lidberg, L., Levander, S., Schalling, D., & Lidberg, Y. (1978). Urinary catecholamines, stress, and psychopathy: A study of arrested men awaiting trial. *Psychosomatic Medicine*, *40*(2), 116–125. https://doi.org/10.1097/00006842-197803000-00004

Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J., & Mather, M. (2013). Stress modulates reinforcement learning in younger and older adults. *Psychology and Aging*, *28*(1), 35–46. https://doi.org/10.1037/a0029823

Lin, Z., Cabrera-Haro, L. E., & Reuter-Lorenz, P. A. (2020). Asymmetrical learning and memory for acquired gain versus loss associations. *Cognition*, *202*, 104318. https://doi.org/10.1016/j.cognition.2020.104318

Lingawi, N.W., Dezfouli, A. and Balleine, B.W. (2016). The psychological and physiological mechanisms of habit formation. In R.A. Murphy and R.C. Honey (eds.), *The Wiley Handbook on the Cognitive Neuroscience of Learning* (pp. 411-441). Wiley Blackwell. https://doi.org/10.1002/9781118650813.ch16

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences*, *104*(22), 9493. https://doi.org/10.1073/pnas.0608842104

Loney, B. R., Butler, M. A., Lima, E. N., Counts, C. A., & Eckel, L. A. (2006). The relation between salivary cortisol, callous-unemotional traits, and conduct problems in an adolescent non-referred sample. *Journal of Child Psychology and Psychiatry*, *47*(1), 30–36. https://doi.org/10.1111/j.1469-7610.2005.01444.x

Lopez, S. A., & Flagel, S. B. (2020). A proposed role for glucocorticoids in mediating dopamine-dependent cue-reward learning. *Stress*, 1–41. https://doi.org/10.1080/10253890.2020.1768240

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: A meta-analysis. *Psychological Bulletin*, *130*(4), 531–552. https://doi.org/10.1037/0033-2909.130.4.531

Luksys, G., & Sandi, C. (2011). Neural mechanisms and computations underlying stress effects on learning and memory. *Behavioural and cognitive neuroscience*, *21*(3), 502–508. https://doi.org/10.1016/j.conb.2011.03.003

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, *55*(1), 6–10. https://doi.org/10.1037/h0047232

Maia, T. V. (2015). Introduction to the series on computational psychiatry. *Clinical Psychological Science*, *3*(3), 374–377. https://doi.org/10.1177/2167702614567350
Maia, T. V., & Conceição, V. A. (2017). The roles of phasic and tonic dopamine in tic learning and expression. *Biological Psychiatry*, *82*(6), 401–412. https://doi.org/10.1016/j.biopsych.2017.05.025
Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723
Maia, T. V., & Frank, M. J. (2017). An integrative perspective on the role of dopamine in schizophrenia. *Biological Psychiatry*, *81*(1), 52–66. https://doi.org/10.1016/j.biopsych.2016.05.021
Matsumoto, H., Tian, J., Uchida, N., & Watabe-Uchida, M. (2016). Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. *eLife*, *5*, e17328. https://doi.org/10.7554/eLife.17328
Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841. https://doi.org/10.1038/nature08028
McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–346. https://doi.org/10.1016/s0896-6273(03)00154-5
McCutcheon, R. A., Abi-Dargham, A., & Howes, O. D. (2019). Schizophrenia, dopamine and the striatum: From biology to symptoms. *Trends in Neurosciences*, *42*(3), 205–220. https://doi.org/10.1016/j.tins.2018.12.004
McRae, A. L., Saladin, M. E., Brady, K. T., Upadhyaya, H., Back, S. E., & Timmerman, M. A. (2006). Stress reactivity: Biological and subjective responses to the cold pressor and Trier Social stressors. *Human Psychopharmacology*, *21*(6), 377–385. https://doi.org/10.1002/hup.778
Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172
Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, *379*(6564), 449–451. https://doi.org/10.1038/379449a0
Mitchell, D. G. V., Colledge, E., Leonard, A., & Blair, R. J. R. (2002). Risky decisions and response reversal: Is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia*, *40*(12), 2013–2022. https://doi.org/10.1016/S0028-3932(02)00056-8
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. https://doi.org/10.1016/j.biopsych.2017.01.017
Möller, M., & Bogacz, R. (2019). Learning the payoffs and costs of actions. *PLOS Computational Biology*, *15*(2), e1006285. https://doi.org/10.1371/journal.pcbi.1006285
Mondelli, V. (2014). From stress to psychosis: Whom, how, when and why? *Epidemiology and Psychiatric Sciences*, *23*(3), 215–218. https://doi.org/10.1017/S204579601400033X
Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience*, *29*, 417–448. https://doi.org/10.1146/annurev.neuro.29.051605.112903
Moran, R. J., Kishida, K. T., Lohrenz, T., Saez, I., Laxton, A. W., Witcher, M. R., Tatter, S. B., Ellis, T. L., Phillips, P. E., Dayan, P., & Montague, P. R. (2018). The protective action encoding of serotonin transients in the human brain. *Neuropsychopharmacology*, *43*(6), 1425–1435. https://doi.org/10.1038/npp.2017.304

Morris, B. H., & Rottenberg, J. (2015). Heightened reward learning under stress in generalized anxiety disorder: A predictor of depression resistance? *Journal of Abnormal Psychology*, *124*(1), 115–127. https://doi.org/10.1037/a0036934

Myin-Germeys, I., & van Os, J. (2007). Stress-reactivity in psychosis: Evidence for an affective pathway to psychosis. *Clinical Psychology Review*, *27*(4), 409–424. https://doi.org/10.1016/j.cpr.2006.09.005

Nagano-Saito, A., Dagher, A., Booij, L., Gravel, P., Welfeld, K., Casey, K. F., Leyton, M., & Benkelfat, C. (2013). Stress-induced dopamine release in human medial prefrontal cortex—18F-fallypride/PET study in healthy volunteers. *Synapse*, *67*(12), 821–830. https://doi.org/10.1002/syn.21700

Nair, A., Rutledge, R. B., & Mason, L. (2020). Under the hood: Using computational psychiatry to make psychological therapies more mechanism-focused. *Frontiers in Psychiatry*, *11*, 140. https://doi.org/10.3389/fpsyt.2020.00140

Neumann, C. S., & Hare, R. D. (2008). Psychopathic traits in a large community sample: Links to violence, alcohol use, and intelligence. *Journal of Consulting and Clinical Psychology*, *76*(5), 893–899. https://doi.org/10.1037/0022-006X.76.5.893

Newman, J. P. (1998). Psychopathic behavior: An information processing perspective. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 81–104). Springer. https://doi.org/10.1007/978-94-011-3965-6_5

Newman, J. P., & Baskin-Sommers, A. R. (2012). Early selective attention abnormalities in psychopathy: Implications for self-regulation. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (2nd ed., pp. 421–440). Guilford Press.

Newman, J. P., Curtin, J. J., Bertsch, J. D., & Baskin-Sommers, A. R. (2010). Attention moderates the fearlessness of psychopathic offenders. *Biological Psychiatry*, *67*(1), 66–70. https://doi.org/10.1016/j.biopsych.2009.07.035

Newman, J. P., & Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *Journal of Abnormal Psychology*, *95*(3), 252–256. https://doi.org/10.1037/0021-843X.95.3.252

Newman, J. P., & Schmitt, W. A. (1998). Passive avoidance in psychopathic offenders: A replication and extension. *Journal of Abnormal Psychology*, *107*(3), 527–532. https://doi.org/10.1037//0021-843x.107.3.527

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, *35*(21), 8145. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Niv, Y., Edlund, J. A., Dayan, P., & Doherty, J. P. O. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *The Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Novick, A. M., Levandowski, M. L., Laumann, L. E., Philip, N. S., Price, L. H., & Tyrka, A. R. (2018). The effects of early life stress on reward processing. *Journal of Psychiatric Research*, *101*, 80–103. PubMed. https://doi.org/10.1016/j.jpsychires.2018.02.002

Oba, T., Katahira, K., & Ohira, H. (2019). The effect of reduced learning ability on avoidance in psychopathy: A computational approach. *Frontiers in Psychology*, *10*(2432), 1–15. https://doi.org/10.3389/fpsyg.2019.02432

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337. https://doi.org/10.1016/s0896-6273(03)00169-7

O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454. https://doi.org/10.1126/science.1094285

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*, 35–53. https://doi.org/10.1196/annals.1390.022

OECD, & European Union. (2018). *Health at a Glance: Europe 2018*. https://www.oecd-ilibrary.org/content/publication/health_glance_eur-2018-en

Oleson, E. B., Gentry, R. N., Chioma, V. C., & Cheer, J. F. (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *The Journal of Neuroscience*, *32*(42), 14804. https://doi.org/10.1523/JNEUROSCI.3087-12.2012

Ossewaarde, L., Hermans, E. J., van Wingen, G. A., Kooijman, S. C., Johansson, I.-M., Bäckström, T., & Fernández, G. (2010). Neural mechanisms underlying changes in stress-sensitivity across the menstrual cycle. *Psychoneuroendocrinology*, *35*(1), 47–55. https://doi.org/10.1016/j.psyneuen.2009.08.011

Ossewaarde, L., Qin, S., Van Marle, H. J. F., van Wingen, G. A., Fernández, G., & Hermans, E. J. (2011). Stress-induced reduction in reward-related prefrontal cortex function. *NeuroImage*, *55*(1), 345–352. https://doi.org/10.1016/j.neuroimage.2010.11.068

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Oyama, K., Hernádi, I., Iijima, T., & Tsutsui, K.-I. (2010). Reward prediction error coding in dorsal striatal neurons. *The Journal of Neuroscience*, *30*(34), 11447. https://doi.org/10.1523/JNEUROSCI.1719-10.2010

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Palminteri, S., & Pessiglione, M. (2013). Reinforcement learning and Tourette syndrome. Em *International Review of Neurobiology* (Vol. 112, pp. 131–153). Academic Press Inc. https://doi.org/10.1016/B978-0-12-411546-0.00005-6

Palminteri, S., & Pessiglione, M. (2017). Opponent brain systems for reward and punishment learning: Causal evidence from drug and lesion studies in humans. In J-C. Dreher & L. Tremblay (Eds.), *Decision Neuroscience: An Integrative Perspective* (pp. 291–303). Elsevier Inc. https://doi.org/10.1016/B978-0-12-805308-9.00023-3

Papalini, S., Beckers, T., & Vervliet, B. (2020). Dopamine: From prediction error to psychotherapy. *Translational Psychiatry*, *10*(1), 164. https://doi.org/10.1038/s41398-020-0814-x

Paret, C., & Bublatzky, F. (2020). Threat rapidly disrupts reward reversal learning. *Behaviour Research and Therapy*, *131*, 103636. https://doi.org/10.1016/j.brat.2020.103636

Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, *78*(3), 321–338. https://doi.org/10.1007/s00426-013-0534-4

Parmentier, F. B. R., Elford, G., Escera, C., Andrés, P., & Miguel, I. S. (2008). The cognitive locus of distraction by acoustic novelty in the cross-modal oddball task. *Cognition*, *106*(1), 408–432. https://doi.org/10.1016/j.cognition.2007.03.008

Parmentier, F. B. R., & Hebrero, M. (2013). Cognitive control of involuntary distraction by deviant sounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1635–1641. https://doi.org/10.1037/a0032421

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology*, *31*(4), 319–330. https://doi.org/10.1111/j.1469-8986.1994.tb02440.x

Patzelt, E. H., Hartley, C. A., & Gershman, S. J. (2018). Computational phenotyping: Using models to understand individual differences in personality, development, and mental illness. *Personality Neuroscience*, *1*, e18. https://doi.org/10.1017/pen.2018.14

Paul, M., Bellebaum, C., Ghio, M., Suchan, B., & Wolf, O. T. (2019). Stress effects on learning and feedback-related neural activity depend on feedback delay. *Psychophysiology*, *57*(2). https://doi.org/10.1111/psyp.13471

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2016). *The SRP-4: Self-report psychopathy scale* (4.ª ed.). Multi-Health Systems.

Pechtel, P., & Pizzagalli, D. A. (2011). Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology*, *214*(1), 55–70. https://doi.org/10.1007/s00213-010-2009-2

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Petzold, A., Plessow, F., Goschke, T., & Kirschbaum, C. (2010). Stress reduces use of negative feedback in a feedback-based learning task. *Behavioral Neuroscience*, *124*(2), 248–255. https://doi.org/10.1037/a0018930

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology*, *10*, 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Porcelli, A. J., & Delgado, M. R. (2017). Stress and decision making: Effects on valuation, learning, and risk-taking. *Current Opinion in Behavioral Sciences*, *14*, 33–39. https://doi.org/10.1016/j.cobeha.2016.11.015

Pruessner, J. C., Champagne, F., Meaney, M. J., & Dagher, A. (2004). Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: A positron emission tomography study using [11C]raclopride. *The Journal of Neuroscience*, *24*(11), 2825–2831. https://doi.org/10.1523/JNEUROSCI.3422-03.2004

Pruessner, J. C., Gaab, J., Hellhammer, D. H., Lintz, D., Schommer, N., & Kirschbaum, C. (1997). Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. *Psychoneuroendocrinology*, *22*(8), 615–625. https://doi.org/10.1016/S0306-4530(97)00072-3

Puglisi-Allegra, S., Imperato, A., Angelucci, L., & Cabib, S. (1991). Acute stress induces time-dependent responses in dopamine mesolimbic system. *Brain Research*, *554*(1–2), 217–222. https://doi.org/10.1016/0006-8993(91)90192-x

Pujara, M., Motzkin, J. C., Newman, J. P., Kiehl, K. A., & Koenigs, M. (2014). Neural correlates of reward and loss sensitivity in psychopathy. *Social Cognitive and Affective Neuroscience*, *9*(6), 794–801. https://doi.org/10.1093/scan/nst054

Queirazza, F., Fouragnan, E., Steele, J. D., Cavanagh, J., & Philiastides, M. G. (2019). Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. *Science Advances*, *5*(7), eaav4962. https://doi.org/10.1126/sciadv.aav4962

Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., Deserno, L., & Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based

behavioral control. *Psychoneuroendocrinology*, *53*, 268–280. https://doi.org/10.1016/j.psyneuen.2014.12.017

Redgrave, P., & Gurney, K. (2006). The short-latency dopamine signal: A role in discovering novel actions? *Nature Reviews Neuroscience*, *7*(12), 967–975. https://doi.org/10.1038/nrn2022

Reidy, D. E., Kearns, M. C., DeGue, S., Lilienfeld, S. O., Massetti, G., & Kiehl, K. A. (2015). Why psychopathy matters: Implications for public health and violence prevention. *Aggression and Violent Behavior*, *24*, 214–225. PubMed. https://doi.org/10.1016/j.avb.2015.05.018

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). Appleton-Century-Crofts

Robinson, O. J., Overstreet, C., Charney, D. R., Vytal, K., & Grillon, C. (2013). Stress increases aversive prediction error signal in the ventral striatum. *Proceedings of the National Academy of Sciences*, *110*(10), 4129–4133. https://doi.org/10.1073/pnas.1213923110

Robinson, O. J., Vytal, K., Cornwell, B. R., & Grillon, C. (2013). The impact of anxiety upon cognition: Perspectives from human threat of shock studies. *Frontiers in Human Neuroscience*, *7*, 203. https://doi.org/10.3389/fnhum.2013.00203

Roiser, J. P., Howes, O. D., Chaddock, C. A., Joyce, E. M., & McGuire, P. (2013). Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. *Schizophrenia Bulletin*, *39*(6), 1328–1336. https://doi.org/10.1093/schbul/sbs147

Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *The Journal of Neuroscience*, *29*(48), 15104–15114. https://doi.org/10.1523/JNEUROSCI.3524-09.2009

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252. https://doi.org/10.1073/pnas.1407535111

Rylander, R. (2004). Physiological aspects of noise-induced stress and annoyance. *Journal of Sound and Vibration*, *277*(3), 471–478. https://doi.org/10.1016/j.jsv.2004.03.008

Saal, D., Dong, Y., Bonci, A., & Malenka, R. C. (2003). Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron*, *37*(4), 577–582. https://doi.org/10.1016/s0896-6273(03)00021-7

Salas, R., Baldwin, P., de Biasi, M., & Montague, P. R. (2010). BOLD responses to negative reward prediction errors in human habenula. *Frontiers in Human Neuroscience*, *4*, 36. https://doi.org/10.3389/fnhum.2010.00036

Salinas-Hernández, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., & Duvarci, S. (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. *eLife*, *7*, e38818. https://doi.org/10.7554/eLife.38818

Sandvik, A. M., Hansen, A. L., Hystad, S. W., Johnsen, B. H., & Bartone, P. T. (2015). Psychopathy, anxiety, and resiliency—Psychological hardiness as a mediator of the psychopathy-anxiety relationship in a prison setting. *Personality and Individual Differences*, *72*, 30–34. https://doi.org/10.1016/j.paid.2014.08.009

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/edrv.21.1.0389

Schad, D. J., Rapp, M. A., Garbusow, M., Nebe, S., Sebold, M., Obst, E., Sommer, C., Deserno, L., Rabovsky, M., Friedel, E., Romanczuk-Seiferth, N., Wittchen, H.-U., Zimmermann, U. S., Walter, H., Sterzer, P., Smolka, M. N., Schlagenhauf, F., Heinz, A., Dayan, P., & Huys, Q. J. M. (2020).

Dissociating neural learning signals in human sign- and goal-trackers. *Nature Human Behaviour*, *4*(2), 201–214. https://doi.org/10.1038/s41562-019-0765-5

Scholl, J., & Klein-Flügge, M. (2018). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behavioural Brain Research*, *355*, 56–75. https://doi.org/10.1016/j.bbr.2017.09.050

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *The Journal of Neuroscience*, *27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schönberg, T., O'Doherty, J. P., Joel, D., Inzelberg, R., Segev, Y., & Daw, N. D. (2010). Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: Evidence from a model-based fMRI study. *NeuroImage*, *49*(1), 772–781. https://doi.org/10.1016/j.neuroimage.2009.08.011

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological Reviews*, *95*(3), 853–951. https://doi.org/10.1152/physrev.00023.2014

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, *33*(6), 890–895. https://doi.org/10.1016/j.psyneuen.2008.03.001

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience and Biobehavioral Reviews*, *36*(7), 1740–1749. https://doi.org/10.1016/j.neubiorev.2011.07.002

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *The Journal of Neuroscience*, *29*(22), 7191. https://doi.org/10.1523/JNEUROSCI.0979-09.2009

Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: From «thinking» to «doing». *Trends in Cognitive Sciences*, *17*(2), 60–68. https://doi.org/10.1016/j.tics.2012.12.001

Seara-Cardoso, A., Queirós, A., Fernandes, E., Coutinho, J., & Neumann, C. (2020). Psychometric properties and construct validity of the short version of the Self-Report Psychopathy Scale in a southern European sample. *Journal of Personality Assessment*, *102*(4), 457–468. https://doi.org/10.1080/00223891.2019.1617297

Seara-Cardoso, A., & Viding, E. (2015). Functional neuroscience of psychopathic personality in adults. *Journal of Personality*, *83*(6), 723–737. https://doi.org/10.1111/jopy.12113

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *The Journal of Neuroscience*, *27*(18), 4826. https://doi.org/10.1523/JNEUROSCI.0400-07.2007

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Sinha, R. (2007). The role of stress in addiction relapse. *Current Psychiatry Reports*, *9*(5), 388–395. https://doi.org/10.1007/s11920-007-0050-6

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudo-type. *Journal of General Psychology*, *12*, 66–77. https://doi.org/10.1080/00221309.1935.9920088

Smith, S. F., & Lilienfeld, S. O. (2015). The response modulation hypothesis of psychopathy: A meta-analytic and narrative analysis. *Psychological Bulletin*, *141*(6), 1145–1177. https://doi.org/10.1037/bul0000024

Soliemanifar, O., Soleymanifar, A., & Afrisham, R. (2018). Relationship between personality and biological reactivity to stress: A review. *Psychiatry Investigation*, *15*(12), 1100–1114. https://doi.org/10.30773/pi.2018.10.14.2

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966–973. https://doi.org/10.1038/nn.3413

Stelly, C. E., Haug, G. C., Fonzi, K. M., Garcia, M. A., Tritley, S. C., Magnon, A. P., Ramos, M. A. P., & Wanat, M. J. (2019). Pattern of dopamine signaling during aversive events predicts active avoidance learning. *Proceedings of the National Academy of Sciences*, *116*(27), 13641–13650. https://doi.org/10.1073/pnas.1904249116

Sulzer, D., Cragg, S. J., & Rice, M. E. (2016). Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia*, *6*(3), 123–148. https://doi.org/10.1016/j.baga.2016.02.001

Suridjan, I., Boileau, I., Bagby, M., Rusjan, P. M., Wilson, A. A., Houle, S., & Mizrahi, R. (2012). Dopamine response to psychosocial stress in humans and its relationship to individual differences in personality traits. *Journal of Psychiatric Research*, *46*(7), 890–897. https://doi.org/10.1016/j.jpsychires.2012.03.009

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637. https://doi.org/10.3758/BF03196525

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

Tan, K. R., Yvon, C., Turiault, M., Mirzabekov, J. J., Doehner, J., Labouèbe, G., Deisseroth, K., Tye, K. M., & Lüscher, C. (2012). GABA neurons of the VTA drive conditioned place aversion. *Neuron*, *73*(6), 1173–1183. https://doi.org/10.1016/j.neuron.2012.02.015

Thompson, D. F., Ramos, C. L., & Willett, J. K. (2014). Psychopathy: Clinical features, developmental basis and therapeutic challenges. *Journal of Clinical Pharmacy and Therapeutics*, *39*(5), 485–495. https://doi.org/10.1111/jcpt.12182

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. Macmillan.

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *The Journal of Neuroscience*, *23*(32), 10402–10410. https://doi.org/10.1523/JNEUROSCI.23-32-10402.2003

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, *307*(5715), 1642–1645. https://doi.org/10.1126/science.1105370

Treadway, M. T., Admon, R., Arulpragasam, A. R., Mehta, M., Douglas, S., Vitaliano, G., Olson, D. P., Cooper, J. A., & Pizzagalli, D. A. (2017). Association between interleukin-6 and striatal prediction-error signals following acute stress in healthy female participants. *Biological Psychiatry*, *82*(8), 570–577. https://doi.org/10.1016/j.biopsych.2017.02.1183

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., de Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, *324*(5930), 1080–1084. PubMed. https://doi.org/10.1126/science.1168878

Tsushima, Y., & Nakayama, K. (2010). Does an auditory distractor allow humans to behave more randomly? *Journal of Vision*, *10*(7), 1028–1028. https://doi.org/10.1167/10.7.1028

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science*, *303*(5666), 2040–2042. https://doi.org/10.1126/science.1093360

Vaessen, T., Hernaus, D., Myin-Germeys, I., & van Amelsvoort, T. (2015). The dopaminergic response to acute stress in health and psychopathology: A systematic review. *Neuroscience and Biobehavioral Reviews*, *56*, 241–251. https://doi.org/10.1016/j.neubiorev.2015.07.008

Valenti, O., Lodge, D. J., & Grace, A. A. (2011). Aversive stimuli alter ventral tegmental area dopamine neuron activity via a common action in the ventral hippocampus. *The Journal of Neuroscience*, *31*(11), 4280–4289. https://doi.org/10.1523/JNEUROSCI.5310-10.2011

Valentin, V. V., & O'Doherty, J. P. (2009). Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *Journal of Neurophysiology*, *102*(6), 3384–3391. https://doi.org/10.1152/jn.91195.2008

van Winkel, R., Stefanis, N. C., & Myin-Germeys, I. (2008). Psychosocial stress and psychosis. A review of the neurobiological mechanisms and the evidence for gene-stress interaction. *Schizophrenia Bulletin*, *34*(6), 1095–1105. https://doi.org/10.1093/schbul/sbn101

Viding, E., McCrory, E., & Seara-Cardoso, A. (2014). Psychopathy. *Current Biology*, *24*(18), R871–R874. https://doi.org/10.1016/j.cub.2014.06.055

Von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & De Bruijn, E. R. A. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychological Medicine*, *40*(9), 1559–1568. https://doi.org/10.1017/S0033291709992017

Walker, E. F., & Diforio, D. (1997). Schizophrenia: A neural diathesis-stress model. *Psychological Review*, *104*(4), 667–685. https://doi.org/10.1037/0033-295x.104.4.667

Wang, K. S., Smith, D. V., & Delgado, M. R. (2016). Using fMRI to study reward processing in humans: Past, present, and future. *Journal of Neurophysiology*, *115*(3), 1664–1678. https://doi.org/10.1152/jn.00333.2015

Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, *40*, 373–394. https://doi.org/10.1146/annurev-neuro-072116-031109

Werlen, E., Shin, S.-L., Gastambide, F., Francois, J., Tricklebank, M. D., Marston, H. M., Huxter, J. R., Gilmour, G., & Walton, M. E. (2020). Amphetamine disrupts haemodynamic correlates of prediction errors in nucleus accumbens and orbitofrontal cortex. *Neuropsychopharmacology*, *45*(5), 793–803. https://doi.org/10.1038/s41386-019-0564-8

Westman, J. C., & Walters, J. R. (1981). Noise and stress: A comprehensive approach. *Environmental Health Perspectives*, *41*, 291–309. https://doi.org/10.1289/ehp.8141291

White, S. F., Pope, K., Sinclair, S., Fowler, K. A., Brislin, S. J., Williams, W. C., Pine, D. S., & Blair, R. J. R. (2013). Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. *American Journal of Psychiatry*, *170*(3), 315–323. https://doi.org/10.1176/appi.ajp.2012.12060840

Wightman, R. M., Heien, M. L. A. V., Wassum, K. M., Sombers, L. A., Aragona, B. J., Khan, A. S., Ariansen, J. L., Cheer, J. F., Phillips, P. E. M., & Carelli, R. M. (2007). Dopamine release is heterogeneous within microenvironments of the rat nucleus accumbens. *The European Journal of Neuroscience*, *26*(7), 2046–2054. https://doi.org/10.1111/j.1460-9568.2007.05772.x

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wilson, R. C., & Niv, Y. (2015). Is model fitting necessary for model-based fMRI? *PLOS Computational Biology*, *11*(6), e1004237. https://doi.org/10.1371/journal.pcbi.1004237

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., Yizhar, O., Cho, S. L., Gong, S., Ramakrishnan, C., Stuber, G. D., Tye, K. M., Janak, P. H., & Deisseroth, K. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*, *72*(5), 721–733. PubMed. https://doi.org/10.1016/j.neuron.2011.10.028

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. World Health Organization; WHO IRIS. https://apps.who.int/iris/handle/10665/254610

Wuensch, L., Pool, E. R., & Sander, D. (2021). Individual differences in learning positive affective value. *Current Opinion in Behavioral Sciences*, *39*, 19–26. https://doi.org/10.1016/j.cobeha.2020.11.001

Yechiam, E. (2019). Acceptable losses: The debatable origins of loss aversion. *Psychological Research*, *83*(7), 1327–1339. https://doi.org/10.1007/s00426-018-1013-8

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*(2), 497–518. https://doi.org/10.1037/a0029383

Yildirim, B. O., & Derksen, J. J. L. (2015). Mesocorticolimbic dopamine functioning in primary psychopathy: A source of within-group heterogeneity. *Psychiatry Research*, *229*(3), 633–677. https://doi.org/10.1016/j.psychres.2015.07.005

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, *15*(6), 695–707. https://doi.org/10.1093/scan/nsaa089

APPENDICES

APPENDIX 1: Ethical approval

03 de Setembro de 2019

Refª 72_2019

Relator: Paulo Gouveia

Parecer emitido em reunião plenária de 15 de Maio de 2019

Nos termos dos Nº 1 e 6 do Artigo 16º da Lei Nº 21/2014, de 16 de Abril, a Comissão de Ética para a Saúde do Hospital de Braga (CESHB) em relação ao estudo ***"Unveiling the neurocomputational mechanisms of acute stress on reinforcement learning: A model-based and neuroimaging approach."***, de que é investigadora principal Joana Rita da Silva Carvalheiro, Alunas de Doutoramento no Centro de Investigação em Psicologia, Universidade do Minho, sob a orientação de Ana Bernardo Seara-Cardoso, Postdocotoral Research Fellow no Centro de Investigação em Psicologia, Universidade do Minho, e Ana Raquel Mesquita, Investigadora da Escola de Psicologia da Universidade do Minho e que decorrerá no Centro Clínico Académico do Hospital de Braga, emite o seguinte parecer:

- O estudo revela-se pertinente face aos objetivos.

Objetivo primário:

Este projeto tem como objetivo compreender os efeitos do stress agudo nos mecanismos neurocomputacionais da aprendizagem por reforço.

Objetivos secundários: 1) Identificar quais as dimensões comportamentais e mecanismos computacionais da aprendizagem por reforço que são afetados pelo stress agudo; 2) caracterizar os mecanismos neurais que estão na base dos efeitos do stress agudo na aprendizagem por reforço; 3) explorar o papel de diferenças individuais na reatividade ao stress agudo na modulação dos mecanismos neurocomputacionais da aprendizagem por reforço.

- O estudo será experimental, combinando métodos de investigação experimental comportamental e de neuroimagem (fMRI).

- Os participantes serão recrutados na Universidade do Minho, correspondendo a 30 indivíduos do sexo masculino pois sabe-se que os mecanismos de recompensa e de resposta a stress aguda podem diferir entre homens e mulheres e a introdução da variável género irá diminuir o poder estatístico. Todos os participantes serão recompensados pela participação com um vale de compras no valor de 15 euros. Os custos associados à compensação dos participantes serão suportados por projetos financiados pela Fundação para a Ciência e Tecnologia.

- O protocolo do estudo é claro e objetivo.

- As Investigadoras Principais possuem aptidões para a realização do mesmo.

- Estão asseguradas as condições materiais e humanas necessárias à realização do estudo.

- A razão risco-benefício é considerada favorável.

- A realização do estudo não implica custos para o Hospital de Braga.

- Não são evidentes conflitos de interesse face aos objetivos do estudo.

Em resumo, o estudo cumpre os princípios da Bioética, pelo que a CE HB nada tem a opor à sua realização.

O Presidente da CESHB

(Dr. Juan R. Garcia)

Braga, 01.12.09

HOSPITAL DE BRAGA, EPE
DIRETOR CLÍNICO
(JORGE MARQUES)

Universidade do Minho

SECSH

Subcomissão de Ética para as Ciências Sociais e Humanas

Identificação do documento: SECSH 004/2018

Título do projeto: Unveiling the neurocomputational mechanisms of acute stress on reinforcement learning: a model-based and neuroimaging approach

Investigador(a) Responsável: Ana Bernardo Seara Cardoso (orientadora), Escola de Psicologia, Centro de Investigação em Psicologia

Outros Investigadores: Joana Rita da Silva Carvalheiro (aluna de Doutoramento), Escola de Psicologia, Centro de Investigação em Psicologia, Universidade do Minho e Ana Raquel Marcelino Mesquita (orientadora), Escola de Psicologia, Centro de Investigação em Psicologia, Universidade do Minho

PARECER

A Subcomissão de Ética para as Ciências Sociais e Humanas (SECSH) analisou o processo relativo ao projeto intitulado *"Unveiling the neurocomputational mechanisms of acute stress on reinforcement learning: a model-based and neuroimaging approach".*

Os documentos apresentados revelam que o projeto obedece aos requisitos exigidos para as boas práticas na investigação com humanos, em conformidade com as normas nacionais e internacionais que regulam a investigação em Ciências Sociais e Humanas.

Face ao exposto, a SECSH nada tem a opor à realização do projeto.

Braga, 05 de março de 2018.

O Presidente

Paulo Manuel Pinto Pereira Almeida Machado

APPENDIX 2: Supplementary material for Chapter 2

This Appendix includes:

Supplementary Analyses:

- Analysis of neutral trials
- Model comparison and Bayesian model averaging

Supplementary Figures (Figure S2.1 to S2.5)

Supplementary Tables (Tables S2.1 to S2.3)

Supplementary References

Supplementary Analyses

Analysis of neutral trials

Evidence from non-human animal studies suggests that acute stress increases aberrant, spontaneous phasic-dopamine release (Anstrom, Miczek, & Budygin, 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Valenti, Lodge, & Grace, 2011). In turn, excessive spontaneous phasic-dopamine release is thought to drive increased behavioural responding and aberrant learning for neutral stimuli (Boehme et al., 2015; Maia & Frank, 2017; McCutcheon, Abi-Dargham, & Howes, 2019). With this in mind, we conducted exploratory analyses to test whether acute stress biased participants' responses during the neutral trials of the reinforcement-learning task (Murray et al., 2008). Each neutral trial consisted of a choice between two different stimuli, which were either associated with a high (0.8), or with a low (0.2), probability of observing as outcome: 1) look at a 0.5€ coin (but the participants did not win nor lose money), or 2) "nothing" and a gray square (participants also did not win nor lose money; Figure S2.1a). Note that the "look" outcome was only present in neutral trials, whereas the "nothing" outcome was also present in gain and loss trials (see Figure 2.1 in the main manuscript for a depiction of gain and loss trials). Although none of the neutral stimuli led to reinforcement (i.e., none of the stimuli was associated with financial outcomes), one of the stimuli was associated, at a high probability, with the observation of a coin ("look"). Aberrant, spontaneous phasic release of dopamine during coin observation ("look") could conceivably act similarly to the phasic release of dopamine that signals prediction errors (Hyland, Reynolds, Hay, Perk, & Miller, 2002) in gain trials following the observation of the 0.5€ coin. Thus, we asked whether stress-induced spontaneous release of dopamine would lead to a behavioural bias during neutral trials towards the selection of the stimulus that led more often to the coin ("look").

To explore the impact of acute stress on behavioural responding during neutral trials, we conducted a generalized linear-mixed effects model with a binomial distribution of the response variable (0 or 1, for the selection of the stimuli with low and high probabilities of obtaining the "look" outcome, respectively; "logit" link function). Predictor variables included condition (stress or control), block (1 or 2), and trial number (from 1 to 40). The model was fitted using Matlab's *fitglme* function. Importantly, we observed a significant main effect of condition on the selection of the high-probability "look" stimuli, $\beta = 0.18$, $p < .001$, 95% confidence interval = [0.096, 0.26] (Figure S2.1b and Table S2.2). Specifically, participants exhibited a bias towards selecting the high-probability "look" stimuli during the stress condition (% choices: $M = 50.0$, $SEM = 2.8$) compared with the control condition (% choices: $M = 46.0$, $SEM = 2.7$). This result suggests that

acute stress modulates behavioural responding for neutral stimuli, possibly due to stress-induced spontaneous phasic-dopamine release. We cannot discard, however, other alternative interpretations for this tentative finding.

Model comparison and Bayesian model averaging

To confirm that the reinforcement-learning model that we used (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007) provided a robust account of the data, we also tested a nested variant of that model. Non-human animal and human studies suggest that dopaminergic neurons are differentially implicated in the coding of positive and negative prediction errors (Frank et al., 2007; Hikida, Kimura, Wada, Funabiki, & Nakanishi, 2010; Schultz, Dayan, & Montague, 1997) and that learning from positive and negative prediction errors likely occurs via the differential effects of dopamine on the plasticity of corticostriatal synapses (Collins & Frank, 2013; Maia & Conceição, 2017; Möller & Bogacz, 2019); such neurobiological evidence motivated us to use a model with separate learning rates for positive (α^+) and negative (α^-) prediction errors (double-α model) [for examples of studies using this model, see Diederen et al., (2016); Frank et al., (2007); Horga et al., (2015); Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde, & Palminteri, (2017); Niv, Edlund, Dayan, & Doherty, (2012)]. In the classical reinforcement-learning framework, however, such distinction between positive and learning rates is not common (Sutton & Barto, 1998). Therefore, the second model that we used was the standard Q-learning model with a single learning rate (α) for both positive and negative prediction errors (single-α model) (Sutton & Barto, 1998). Both models included the inverse temperature parameter (β), which controls for the amount of exploration/exploitation in choice selection (Daw, 2011; Sutton & Barto, 1998). Given that these two models could show a divergent quality of fits for the stress and control conditions, we fitted both models (double-α and single-α) to the stress and control conditions independently. Then, we compared the two models using random-effects Bayesian model comparison (Rigoux, Stephan, Friston, & Daunizeau, 2014; Stephan et al., 2009).

Bayesian model comparison primarily relies on the model evidence, which quantifies how likely the observed data is under a given model (Daw, 2011). For standard reinforcement-learning models, however, the calculation of the corresponding model evidences is not tractable (Daw, 2011), so approximations must be used. Two frequently used approximations are the Bayesian information criterion (BIC) and the Akaike's information criterion (AIC). The BIC and AIC are relatively poor approximations of the model evidence and their use commonly leads to the selection of different models, given that the BIC and AIC typically favor

simpler and more complex models, respectively (Penny, Stephan, Mechelli, & Friston, 2004), especially when there is a considerable number of observations (or trials), n:

$$\mathrm{BIC} = \log\big(P(D|M,\theta_M)\big) - \frac{p}{2}\log(n),$$

$$\mathrm{AIC} = \log\big(P(D|M,\theta_M)\big) - p,$$

where $P(D|M,\theta_M)$ is the likelihood of the data D (in our case, the observed responses from each subject in each condition) given the model M and parameter values θ_M, and p is the number of free parameters of the model. Note that, in the equations above, the BIC and AIC are approximations to the log model evidence, not to the model evidence.

We compared the two models between conditions and at the single condition level, using random-effects Bayesian model comparison (Rigoux et al., 2014; Stephan et al., 2009) as implemented in the Variational Bayesian Analysis (VBA) toolbox (Daunizeau, Adam, & Rigoux, 2014). In the VBA toolbox, the input data to the Bayesian model comparison procedure was composed of the log-model-evidence approximations (BIC or AIC) obtained for each subject, in each condition, for each model (double-α and single-α). This procedure allowed estimating the model frequencies, and the exceedance probabilities, of each model within the tested set of candidate models, given the data gathered from all subjects. The estimated model frequencies quantify the prevalence of a model in the population. The exceedance probability measures how likely it is that any given model is more frequent than all other models in the comparison set (Stephan et al., 2009). Here, we report protected exceedance probabilities (PEPs), which correct exceedance probabilities for the possibility that observed differences in model evidences (across groups or conditions) are due to chance (Rigoux et al., 2014). A PEP above 0.95 provides strong evidence for the selection of a given candidate model (Rigoux et al., 2014).

Bayesian model comparison revealed divergent results when using the different approximations to the log model evidence (i.e., the BIC and AIC; Table S2.3). Model comparison using the BIC did not lead to confident selection of the same model across the stress and control conditions (PEP = 0.37), nor to confident selection of different models in each condition (stress condition: PEP double-α = 0.04, PEP single-α = 0.96; control condition: PEP double-α = 0.29, PEP single-α = 0.71). Similarly, model comparison using the AIC was inconclusive, as it also did not lead to selection of the same model across the stress and control conditions (PEP = 0.92), nor to the selection of different models in each condition (stress condition: PEP double-α = 0.57; PEP single-α = 0.43; control condition: PEP double-α = 0.995, PEP single-α =

0.005). Thus, we turned to Bayesian model averaging (BMA) (Fragoso, Bertoli, & Louzada, 2018; Hoeting, Madigan, Raftery, & Volinsky, 1999), which can be used to combine parameter estimates from multiple nested models, when no single model clearly is best. Specifically, for each approximation to the log model evidence, we calculated weighted point parameter estimates for each subject and condition, by multiplying the subject-condition-specific parameter estimates obtained through each reinforcement-learning model (double-α and single-α) by the posterior model probabilities obtained with each approximation (BIC and AIC) for each subject in each condition.

To confirm that the effects of acute stress on the estimated parameters could be successfully recovered independently of the used approximations to the log model evidence, we analysed the parameters estimated through BMA when using either the BIC or the AIC. The ensuing analyses demonstrated that, regardless of the used approximation (BIC and AIC; Figure S2.5), we were able to recapitulate the effects of acute stress on the estimated parameters reported in the main text, when using only the model with separate learning rates for positive and negative prediction errors (compare Figure S2.5 with Figure 2.3a-c in the manuscript), except for the interaction condition × valence in $\alpha^{\pm}*\beta$ when the BIC was used as an approximation to log model evidence (Figure S2.5e). Still, when we applied post-hoc tests to the BMA results, we were always able to replicate the differences observed in the original parameters between conditions (again, compare Figure S2.5 with Figure 2.3a-c in the manuscript). Additionally, all parameters estimated through BMA were highly correlated with those used in the main paper analyses (Pearson's $r \geq 0.62$, $p < .001$, Spearman's $r > 0.62$, $p < .001$). In sum, although model comparison using the BIC or AIC was not conclusive *per se*, BMA results were always in strikingly agreement with those obtained when using only the model with separate learning rates for positive and negative prediction errors.

Supplementary Figures

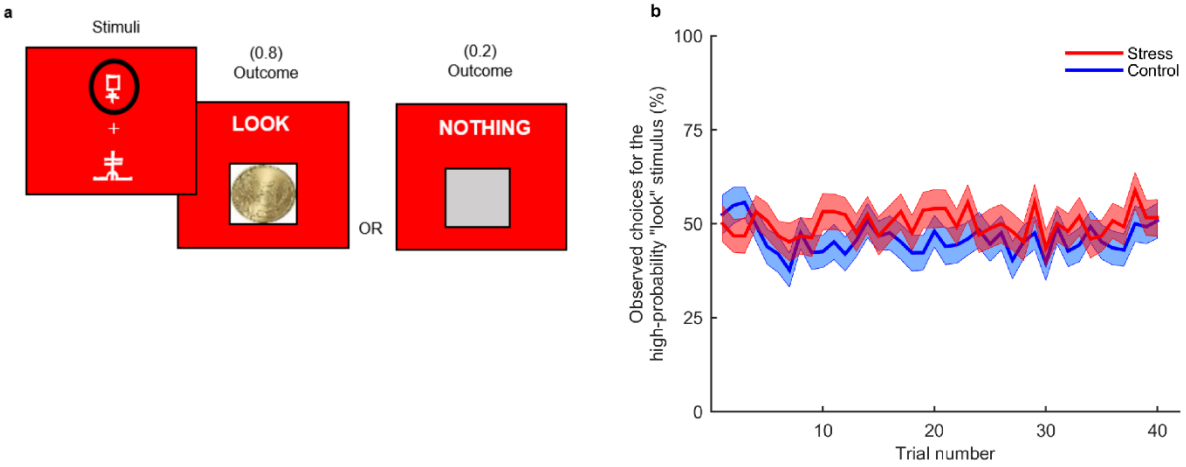


Figure S2.1. Neutral trials and behavioural performance in neutral trials. **(a)** Each neutral stimulus was associated with a certain probability (0.8 or 0.2) of obtaining the outcome "look" and a 0.5€ coin—the obtained monetary outcome was zero, meaning that participants did not win or lose money—and with the reciprocal probability (0.2 or 0.8) of obtaining "nothing" and a gray square. Here, we depict an example for the stress condition in which the participant selected the stimulus with 0.8 probability of obtaining the outcome look at a 0.5€ coin (i.e., the high-probability "look" stimulus). **(b)** Participants chose the stimulus associated at a higher probability with the "look" at the coin outcome significantly more often in the stress condition (red) than in the control condition (blue). The curves represent the trial-by-trial percentage of participants (n = 62) who chose the high-probability "look" stimulus in the stress and control conditions, averaged across the two blocks in each condition. The central lines represent the means and the filled areas represent ± standard errors of the mean.

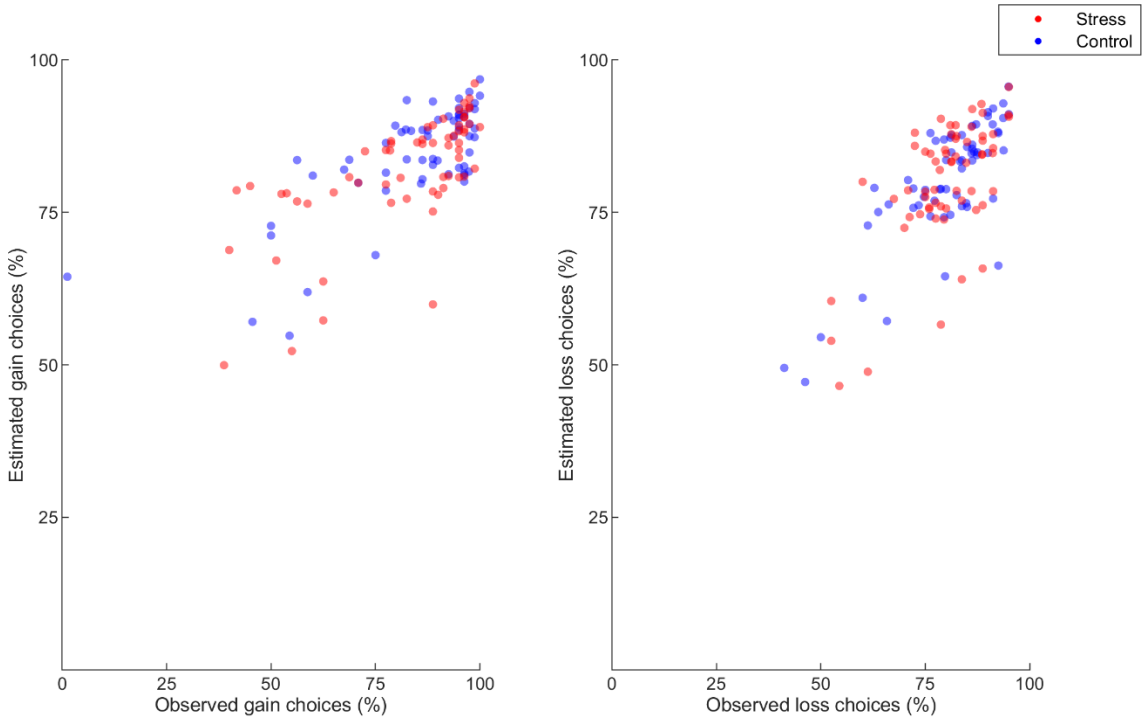


Figure S2.2 Scatterplots of the percentage of observed choices and percentage of estimated choices by the generalized linear mixed effect model, during gain (left) and loss (right) trials, in the stress (red) and control (blue) conditions. More transparent (opaque) dots represent less (more) overlapping data points.

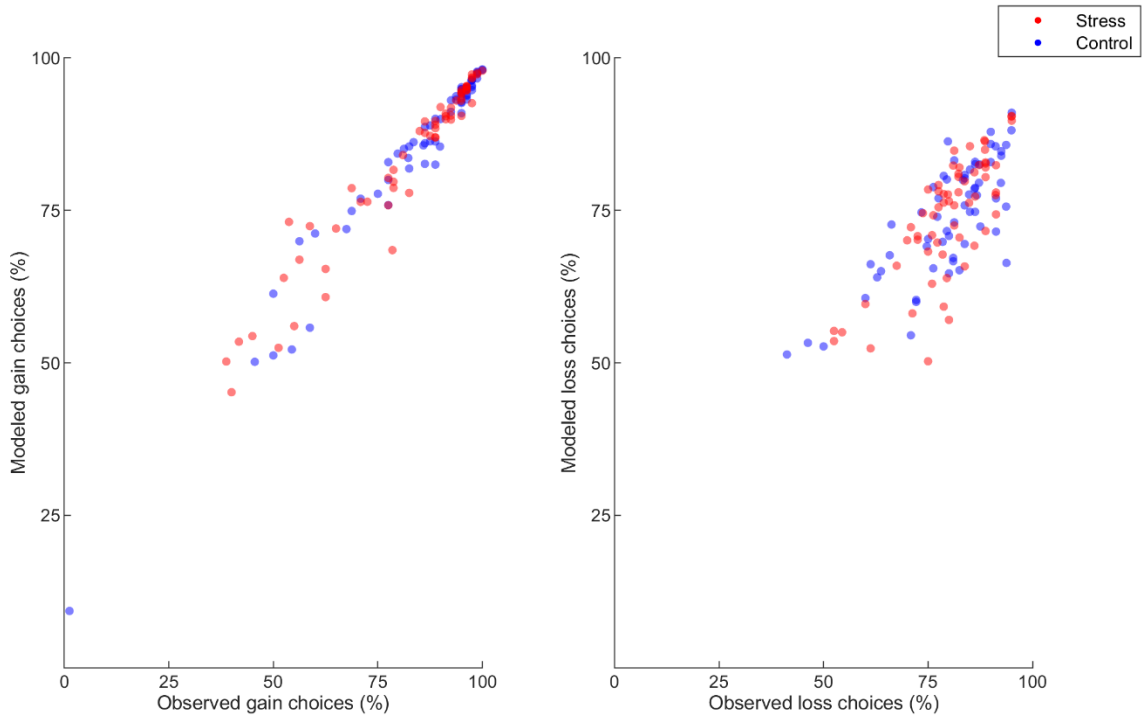


Figure S2.3. Scatterplots of the percentage of observed choices and percentage of estimated choices by the reinforcement-learning model, during gain (left) and loss (right) trials, in the stress (red) and control (blue) conditions. More transparent (opaque) dots represent less (more) overlapping data points.

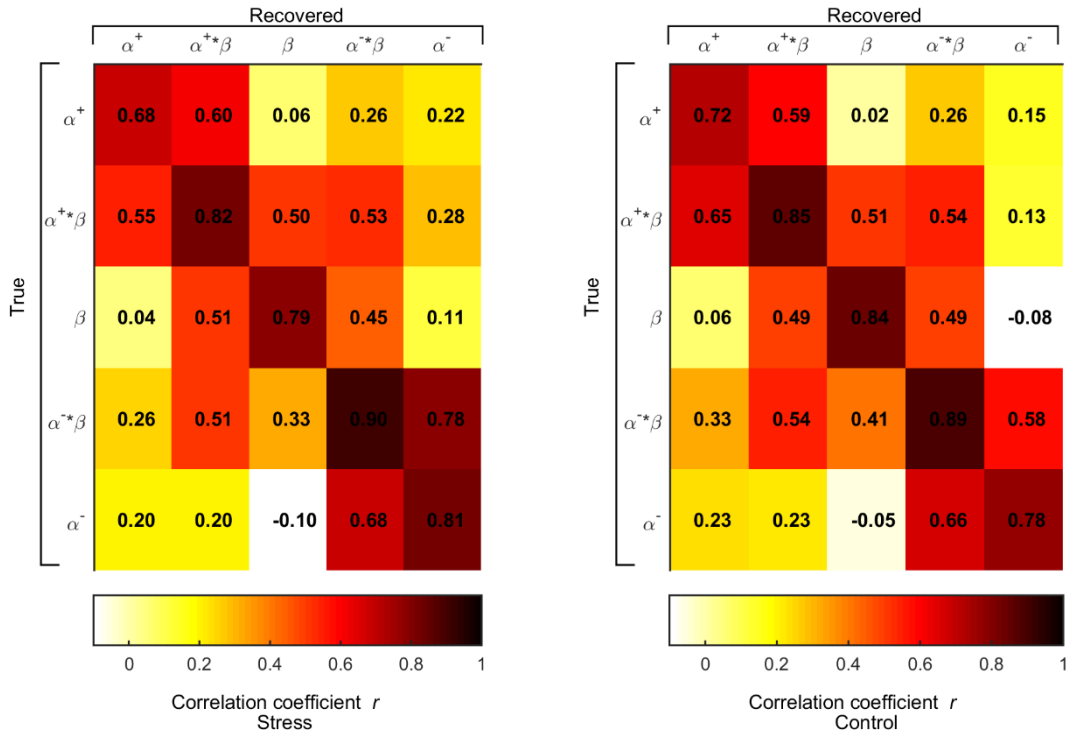


Figure S2.4. Correlation matrices between the estimated and recovered parameters averaged across 100 simulations in the stress (left) and control (right) conditions. The matrices include the values of Pearson's *r* for the correlations between the parameters that were used to generate the simulated data ("True") and those that were obtained by applying the parameter-estimation procedure to the simulated data ("Recovered"). Dark red values of *r* indicate a strong correlation; dark-red diagonal values indicate therefore that parameter recovery was "successful".

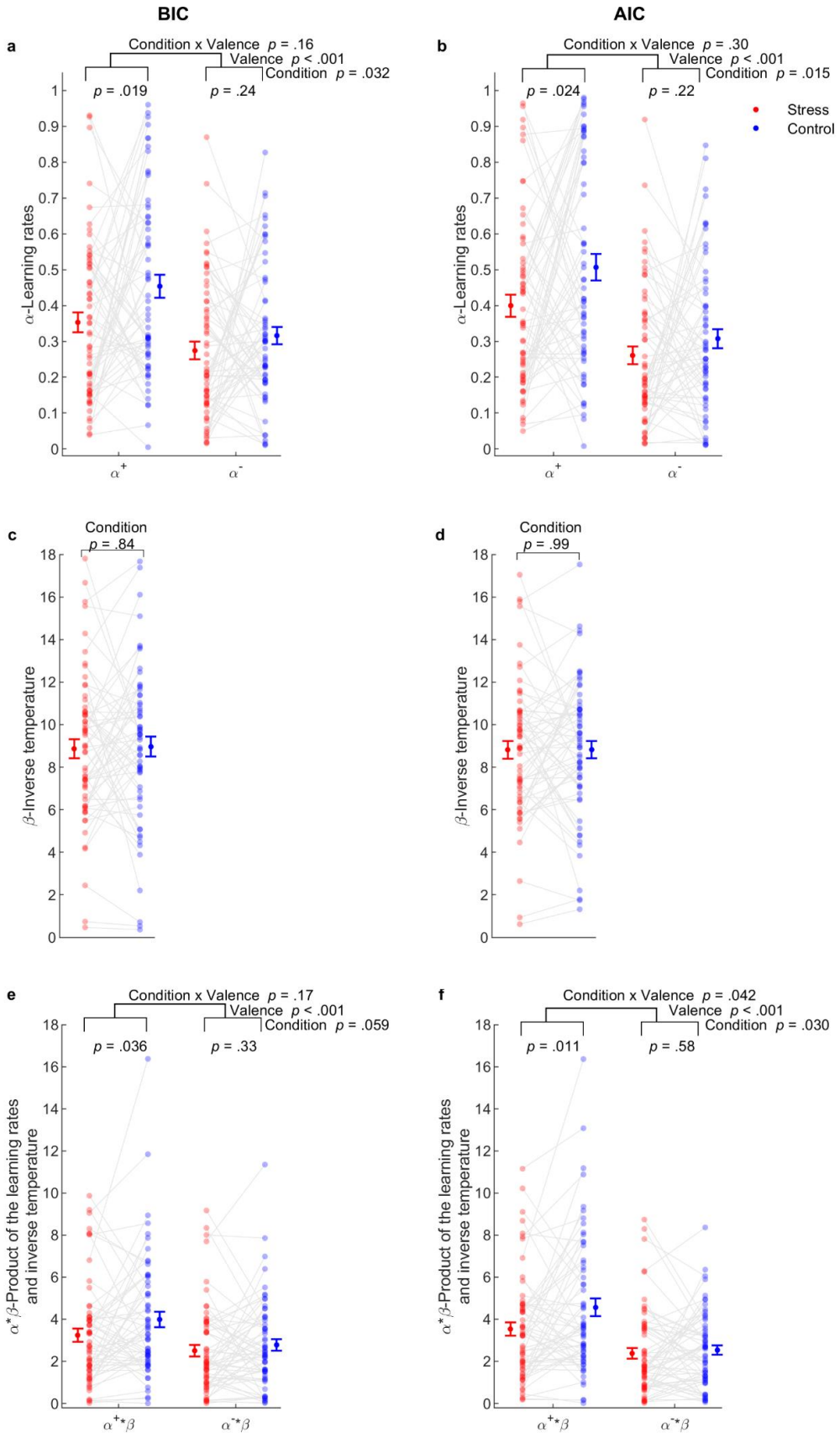

BIC
AIC
a
Condition x Valence p = .16
Valence p < .001
Condition p = .032
p = .019
p = .24
α-Learning rates
α+
α-
b
Condition x Valence p = .30
Valence p < .001
Condition p = .015
p = .024
p = .22
Stress
Control
c
Condition
p = .84
β-Inverse temperature
d
Condition
p = .99
e
Condition x Valence p = .17
Valence p < .001
Condition p = .059
p = .036
p = .33
α*β-Product of the learning rates and inverse temperature
α+*β
α-*β
f
Condition x Valence p = .042
Valence p < .001
Condition p = .030
p = .011
p = .58

Figure S2.5. Parameters estimated using Bayesian model averaging (BMA) across the double-α and single-α models, when using either the Bayesian information criterion (BIC; left panels) or the Akaike information criterion (AIC; right panels) as approximations to the log model evidence, in the stress (red) and control (blue) conditions. **(a, b)** Learning rates for positive, α^{+}, and negative, α^{-}, prediction errors. **(c, d)** Inverse temperature parameter, β. **(e, f)** Products between each learning rate and inverse temperature parameter, $\alpha^{+} * \beta$ and $\alpha^{-} * \beta$. Connected dots represent individual data points in the within-subject design (n = 62). More transparent (opaque) dots represent less (more) overlapping data points. The error bars displayed on the sides of the scatter plots indicate the sample means ± standard errors of the means.

Supplementary Tables

Table S2.1. **Analyses of behavioural performance during the reinforcement-learning task.** Estimated fixed effects coefficients (± standard errors) from generalized linear mixed-effects models. The results reported in the main text refer to a model that included all main effects and the interaction of interest (condition × valence). The significance of the condition × valence interaction did not change when including the other two-factor interactions in the model. As an additional check, we included the significant two-factor interactions ($p < .05$) in a three-factor interaction model. Subject identity was modeled via random intercepts. The reference levels were the control condition, first block, and loss valence.

Fixed effects	Coefficient	Standard error	*t*	*p*	95% Confidence interval
Intercept	0.32	0.11	2.96	.003	[0.11, 0.58]
Trial	0.048	0.0025	18.9	< .001	[0.043, 0.053]
Block	0.65	0.057	11.4	< .001	[0.54, 0.76]
Condition	0.048	0.089	0.54	.59	[-0.12, 0.22]
Valence	0.33	0.056	5.94	< .001	[0.22, 0.45]
Condition × Trial	0.0027	0.0035	0.77	.44	[-0.0042, 0.0097]
Condition × Block	-0.26	0.078	2.96	< .001	[-0.42, -0.11]
Condition × Valence	-0.19	0.078	-2.37	.018	[-0.34, -0.032]
Condition × Valence × Block	-0.088	0.16	-0.55	.58	[-0.40, 0.22]
Condition × Trial × Block	-0.014	0.0072	-1.90	.057	[-0.028, 0.00041]
Condition × Trial × Valence	-0.0042	0.0072	-0.59	.55	[-0.018, 0.0098]

Table S2.2 Analyses of behavioural performance during neutral trials. Estimated fixed-effects coefficients (± standard errors) from generalized linear mixed-effects models. The final model included all main effects. Subject identity was modeled via random intercepts. The reference levels were the control condition, first block, and loss valence.

Fixed effects	Coefficient	Standard error	*t*	*p*	95% Confidence interval
Intercept	0.075	0.11	0.65	.51	[-0.15, 0.30]
Trial	0.000055	0.0018	0.030	.98	[-0.0036, 0.0037]
Block	-0.48	0.043	-11.33	< .001	[-0.57, -0.40]
Condition	0.18	0.043	4.21	< .001	[0.096, 0.26]
Condition × Trial	0.0034	0.0037	0.93	.35	[-0.0038, 0.011]
Condition × Block	0.11	0.085	1.25	.21	[-0.061, 0.27]

Table S2.3. Results from Bayesian model comparison between the reinforcement-learning model used in the main text (double-α) and its nested model with a single learning rate (single- α). NP: number of parameters; M: mean; SEM: standard error of the mean; BIC: Bayesian information criterion; AIC: Akaike information criterion; EF: estimated model frequencies; PEP: protected exceedance probability.

Condition	Model	NP	BIC			AIC		
			M ± SEM	EF	PEP	*M ± SEM*	EF	PEP
Stress	Double-α	3	-67.6 ± 2.8	0.27	0.04	-63.0 ± 2.8	0.61	0.57
	Single-α	2	-67.5 ± 2.9	0.73	0.96	-64.4 ± 2.9	0.39	0.43
Control	Double-α	3	-63.8 ± 2.9	0.34	0.29	-59.2 ± 2.9	0.96	0.995
	Single-α	2	-64.2 ± 3.1	0.66	0.71	-61.2 ± 3.1	0.04	0.005

Supplementary References

Anstrom, K. K., Miczek, K. A., & Budygin, E. A. (2009). Increased phasic dopamine signaling in the mesolimbic pathway during social defeat in rats. *Neuroscience*, *161*, 3–12.

Anstrom, K. K., & Woodward, D. J. (2005). Restraint increases dopaminergic burst firing in awake rats. *Neuropsychopharmacology*, *30*, 1832–1840.

Boehme, R., Deserno, L., Gleich, T., Katthagen, T., Pankow, A., Behr, J., ... Schlagenhauf, F. (2015). Aberrant salience is related to reduced reinforcement learning signals and elevated dopamine synthesis capacity in healthy adults. *Journal of Neuroscience*, *35*, 10103–10111.

Cabib, S., & Puglisi-Allegra, S. (2012). The mesoaccumbens dopamine in coping with stress. *Neuroscience and Biobehavioural Reviews*, *36*, 79–89.

Collins, A. G. E., & Frank, M. J. (2013). Opponent actor learning (OpAL): modelling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*, 337–366.

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*, e1003441.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, 1–26.

Diederen, K. M. J., Spencer, T., Vestergaard, M. D., Fletcher, P. C., Diederen, K. M. J., Spencer, T., ... Schultz, W. (2016). Adaptive prediction error coding in the human midbrain and striatum facilitates behavioural adaptation and learning efficiency adaptive prediction error coding in the human midbrain and striatum facilitates behavioural adaptation and learning Efficiency. *Neuron*, *90*, 1127–1138.

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, *86*, 1–28.

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

Hikida, T., Kimura, K., Wada, N., Funabiki, K., & Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behaviour. *Neuron*, *66*, 896–907.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian averaging models. *Statistical*

Science, *14*, 382–417.

Horga, G., Maia, T. V., Marsh, R., Hao, X., Xu, D., Duan, Y., ... Peterson, B. S. (2015). Changes in corticostriatal connectivity during reinforcement learning in humans. *Human Brain Mapping*, *36*, 793–803.

Hyland, B. I., Reynolds, J. N. J., Hay, J., Perk, C. G., & Miller, R. (2002). Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience*, *114*, 475–492.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*, 0067.

Maia, T. V., & Frank, M. J. (2017). An integrative perspective on the role of dopamine in schizophrenia. *Biological Psychiatry*, *81*, 52–66.

Maia, T. V, & Conceição, V. A. (2017). The roles of phasic and tonic dopamine in tic learning and expression. *Biological Psychiatry*, *82*, 401–412.

McCutcheon, R. A., Abi-Dargham, A., & Howes, O. D. (2019). Schizophrenia, dopamine and the striatum: From biology to symptoms. *Trends in Neurosciences*, *42*, 205–220.

Möller, M., & Bogacz, R. (2019). Learning the payoffs and costs of actions. *PLOS Computational Biology*, *15*, e1006285.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, a D., Honey, G., ... Fletcher, P. C. (2008). Substantia nigra / ventral tegmental reward prediction error disruption in psychosis. *Molecular Psychiatry*, *13*, 1–18.

Niv, Y., Edlund, J. A., Dayan, P., & Doherty, J. P. O. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain, *32*, 551–562.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*, 1157–1172.

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies — Revisited. *NeuroImage*, *84*, 971–985.

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

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., Friston, K. J., Stephan, K. E., ... Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*, 1004–1017.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning : an introduction*. MIT Press.

Valenti, O., Lodge, D. J., & Grace, A. A. (2011). Aversive stimuli alter ventral tegmental area dopamine neuron activity via a common action in the ventral hippocampus. *Journal of Neuroscience*, *31*, 4280–4289.

APPENDIX 3: Supplementary material for Chapter 3

This Appendix includes:

Analyses of the total sample

- Manipulation check (Figure S3.1)
- Task-performance analyses (Figure S3.2)
- fMRI analyses (Figure S3.3 and Figure S3.4)

Primary general linear model (Figure S3.5)

Computational modelling

- Reinforcement-learning model
- Reinforcement-learning modelling results (Figure S3.6)

Supplementary tables (Table S3.1 and Table S3.2)

Supplementary references

Analyses of the total sample

To check the success of the acute-stress manipulation, we collected self-reported stress levels at the end of each task block. As we were interested in examining the impact of acute stress on reinforcement learning, participants who reported to be non-responsive to the stress manipulation (i.e., who did not report higher stress levels in the stress condition than in the control condition) were excluded from the behavioural and neuroimaging analyses presented in the main text. For completeness, here we present data on self-reported stress levels (c.f. section 'Manipulation check'), as well as on task-performance (c.f. section 'Task-performance analyses') and neuroimaging (c.f. section 'fMRI analyses') analyses, in the total sample.

Manipulation check

To assess whether the acute-stress manipulation increased stress levels in the total sample (n = 37), we conducted an analysis of variance (ANOVA) with condition (stress and control) and block (1 and 2) as within-subject factors. Self-reported stress levels differed significantly between conditions, $F(1,36) = 4.53$, $p = .04$, $\eta 2 = 0.11$, (Supplementary Figure S3.1), and there was no main effect of block $F(1,36) = 0.25$, $p = .62$, $\eta 2 = 0.007$. The condition × block interaction was also non-significant $F(1,36) = 2.63$, $p = .11$, $\eta 2 = 0.068$, suggesting that self-reported stress levels remained stable across blocks within each condition.

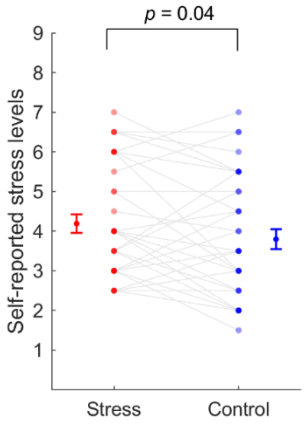


Figure S3.1. Self-reported stress levels in the stress (red) and control (blue) conditions in the total sample (n = 37). Connected dots represent data points from the same participant, and more transparent (opaque)

dots represent less (more) overlapping data points; error bars displayed on the sides of those scatter plots indicate the mean ± standard error of the mean.

Task-performance analyses

We conducted task-performance analyses on the trial-by-trial choice data from the total sample, using a generalized linear mixed-effects model, as described in the main manuscript. We found a significant condition × valence interaction, β = -0.40, p < .001, 95% CI = [-0.62, -0.18], (Supplementary Figure S3.2a). Planned post-hoc analyses revealed that under stress, comparatively to the control condition, participants performed worse when learning to obtain gains F(1, 7078) = 7.47, p = .0063, but better when learning to avoid losses F(1, 7078) = 6.02, p = .014 (Supplementary Figure S3.2b). These somewhat unexpected results concerning punishment learning were driven by the inclusion (in the total sample) of participants who reported lower levels of stress in the stress condition than in the control condition (n = 11) and who performed better, both in gain, F(1, 2097) = 4.63, p = .031, and loss, F(1, 2097) = 25.97, p < 0.001, trials, in the stress condition than in the control condition — note that, like in our subsample of interest, those participants had a worse performance in the condition in which they reported higher stress levels.

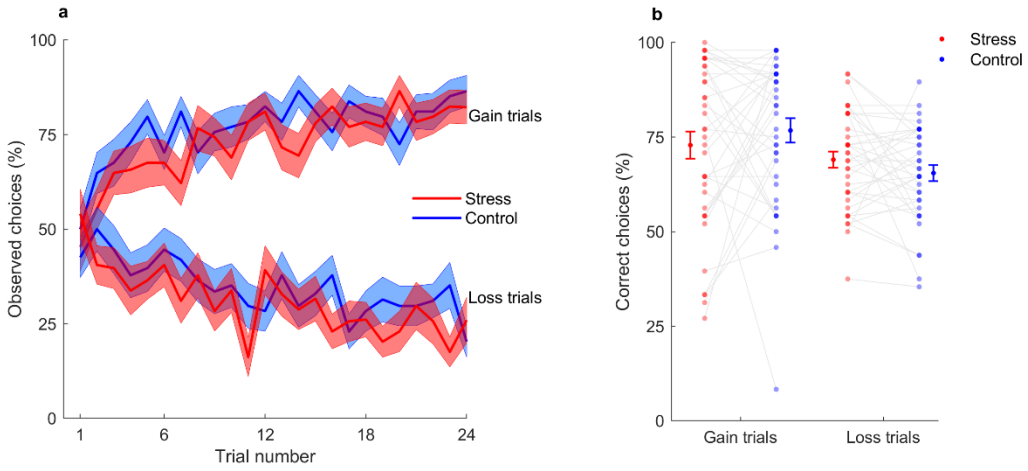


Figure S3.2. Task performance in the total sample. **(a)** Learning curves represent the trial-by-trial percentage of participants (n = 37) who chose the correct gain stimulus (associated with a probability of 0.75 of winning 0.5€; upper part of the graph) and the incorrect loss stimulus (associated with a probability

of 0.75 of losing 0.5€; lower part of the graph), in the stress (red) and control (blue) conditions. Each central line represents the mean and the filled area the ± standard error of the mean. **(b)** Percentage of correct gain and loss choices per participant (n = 37) in stress and control conditions (averaged across blocks). Connected dots represent data points from the same participant, and more transparent (opaque) dots represent less (more) overlapping data points; error bars displayed on the sides of those scatter plots indicate the mean ± standard error of the mean.

fMRI analyses

We repeated the neuroimaging analyses, as described in the main manuscript, for the total sample. Specifically, we took the control > stress contrast to second-level analyses but now including data from the total sample (n = 37). We replicated our main finding that acute stress decreased the parametric modulation of prediction errors in the dorsal striatum, [x = 32, y = 0, z = 12], z = 4.83, k =45, p = .002, SVC-FEW, (Supplementary Figure S3.3a). No significant responses were found in the nucleus accumbens for the control > stress contrast. The fMRI contrast that tested the condition (control and stress) × trial valence (gain and loss) interaction did not show any significant valence-dependent differences in the dorsal striatum nor in the nucleus accumbens.

To explore whether acute stress preferentially blunted positive or negative prediction errors, we conducted subsidiary analyses in the total sample, as described in the main text. Subsidiary data indicated that acute stress differentially affected positive and negative prediction error signals (condition × prediction error valence interaction, F (1,36) = 4.90, p = .033, η^2= 0.12, such that acute stress decreased positive prediction error signals, but not negative prediction error signals (Supplementary Figure S3.3b).

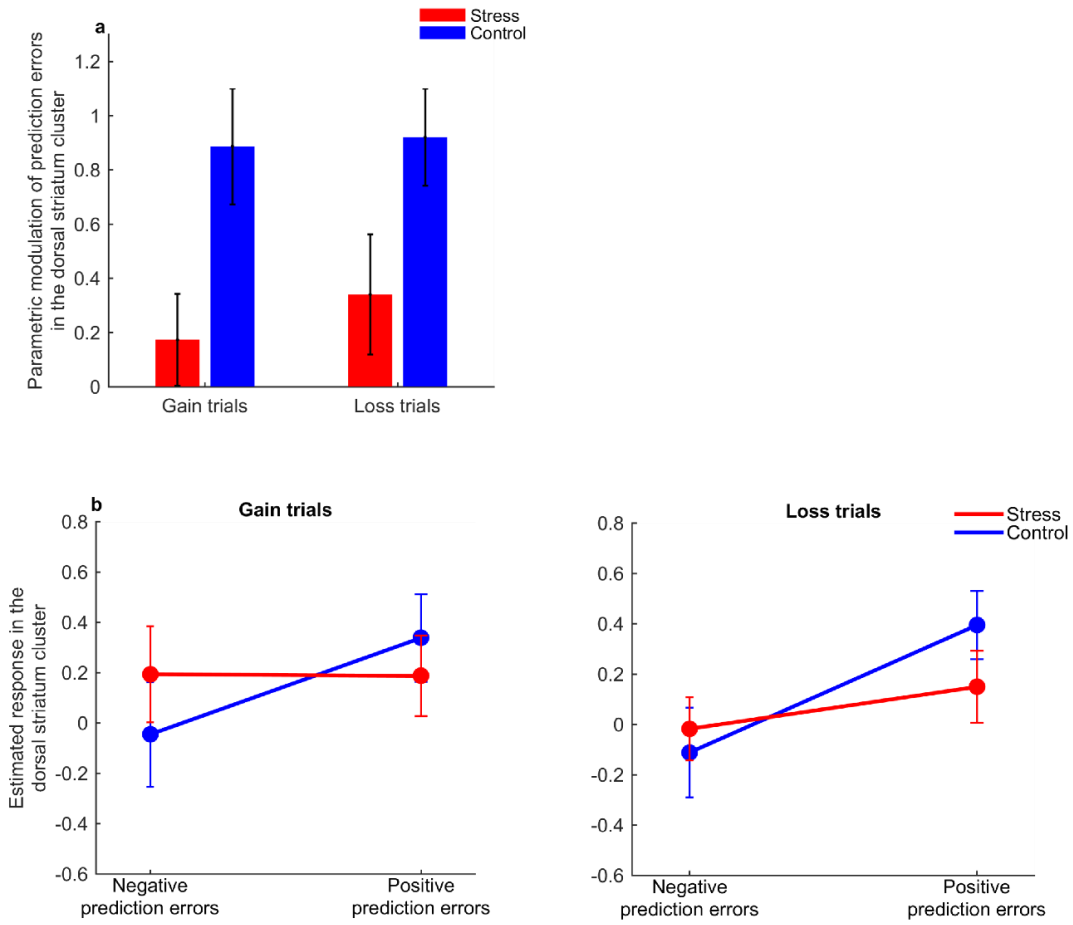


Figure S3.3 Effects of acute stress on prediction error signalling in the dorsal striatum in the total sample**.** **(a)** Bars depict the parameter estimates (i.e., regression slopes) for the BOLD response at the dorsal striatum cluster modulated by trial-by-trial prediction errors, in gain and loss trials, across the stress (red) and control (blue) conditions (n = 37). **(b)** Graphs illustrate the modulation of the BOLD response by prediction errors, during gain (left) and loss (right) trials, at the dorsal striatum cluster in the stress (red) and control (blue) conditions in the total sample (n = 37). Prediction errors were divided into negative and positive, and the BOLD response estimates at the outcome onset were extracted from the dorsal striatum cluster identified in the primary parametric modulation model for each participant. Error bars indicate the mean ± standard error of the mean.

Additionally, as the total sample of participants was characterized by variability in stress responses, with participants reporting to be less or more stressed in the stress condition relative to the control

condition, we conducted exploratory analyses to probe whether the impact of stress on prediction errors signalling was associated with individual differences in stress levels across the total sample. For this, we extracted the parameter estimates (i.e., regression slopes) for the modulation of the BOLD response by prediction errors, in the dorsal striatal cluster identified in the above parametric modulation analysis, and correlated those parameter estimates with the difference in self-reported stress levels between the stress and control conditions. We found that prediction error signals during gain trials in the stress condition correlated negatively with the difference between self-reported stress levels in the stress and control condition (Pearson's correlation: $r = -0.55$, $p < .001$; Spearman's correlation: $rs = -0.52$, $p = .001$) (Supplementary Figure S3.4).

This tentative finding suggests that participants who reported the greatest increase in stress levels in response to the acute stressor, relative to control condition, also showed the greatest reduction in signalling of prediction errors in the dorsal striatum during reward learning when under acute stress.

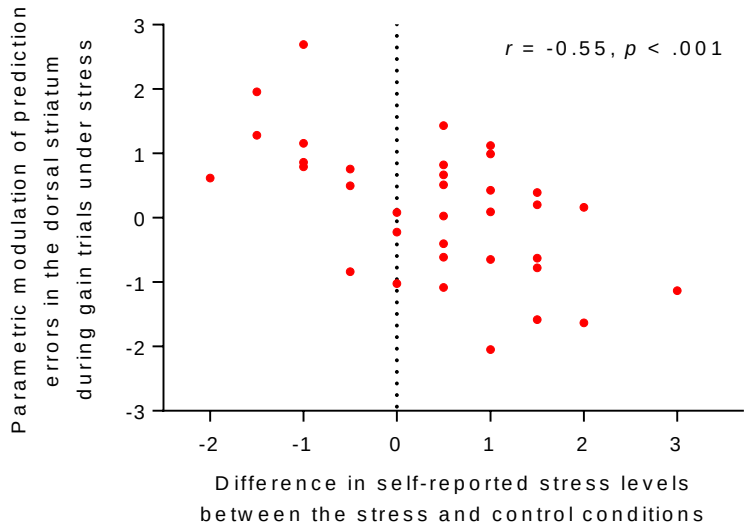


Figure S3.4. Scatterplot of the difference between self-reported stress levels in the stress and control condition and parameter estimates (i.e., regression slopes) of the BOLD response in the dorsal striatum cluster modulated by prediction errors during gain trials in the stress condition (n = 37). The dashed vertical line divides the plot into participants who reported to be responsive to the stress manipulation and participants who did not report to be responsive to the stress manipulation.

Primary general linear model

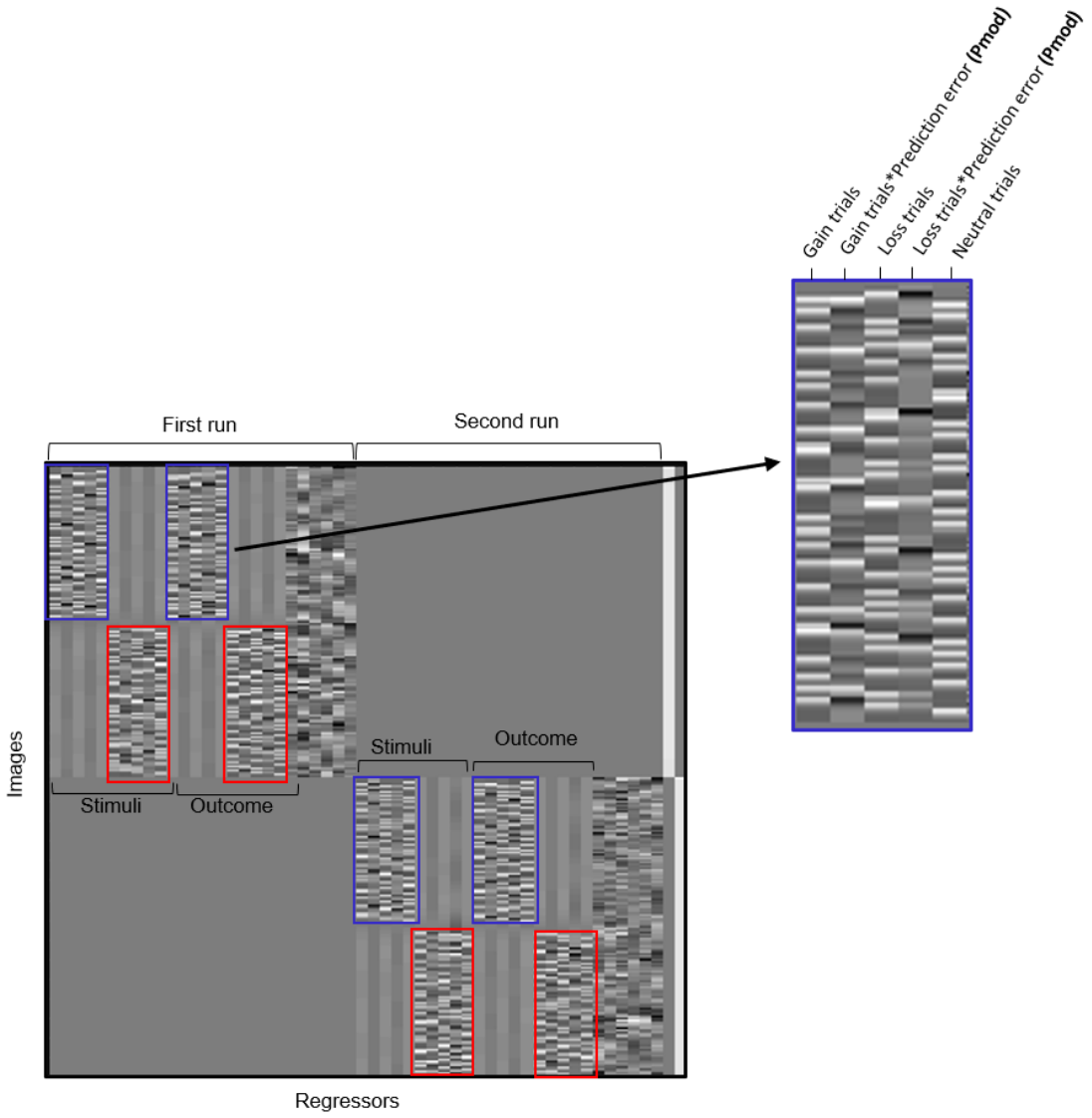


Figure S3.5. Example of a first-level SPM design matrix showing the regressors across the two runs of the task for an exemplary subject: columns (left to right) represent : 1) onsets of stimuli in gain and loss trials — which were parametrically modulated by chosen option values — and neutral trials, in the control (blue) and stress (red) conditions; 2) onsets of outcomes in gain and loss trials — which were parametrically modulated by prediction errors — and neutral trials, in the control (blue) and stress (red) conditions; 3) headmotion regressors (additional regressors for missing trials and/or visible headmotion were also included when needed; not depicted for this subject). The zoomed in regressors on the right are related to

the onset of the outcomes in the control condition during the first run; the regressors of the parametric modulation (Pmod) by the prediction errors in gain and loss trials are depicted in the second and fourth columns, respectively.

Computational modelling

Reinforcement-learning model

Prediction errors were calculated via a reinforcement-learning model that has been extensively used to investigate the behavioural and neural impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans(Diederen et al., 2017; Doll et al., 2011; Frank et al., 2007; Frank & Fossella, 2011; Grogan et al., 2017; Rutledge et al., 2009). The used reinforcement-learning model included separate learning rates for positive (α^{+}) and negative (α) prediction errors, to account both for the differential firing of dopaminergic neurons for positive and negative prediction errors and the differential effects of dopamine onto the plasticity of the corticostriatal synapses implicated in action-value learning(Frank & O'Reilly, 2006; Maia & Frank, 2011). This model also included the inverse temperature parameter, β, which controls randomness in choice selection (Daw, 2011; Sutton & Barto, 1998).

In the context of our task, the used reinforcement-learning model(Frank et al., 2007) assumes that each participant gradually learns the value of choosing a given stimulus (say A or B) from a given pair of stimuli (here, "gain" or "loss" stimuli pairs, as the "neutral" pair of stimuli always yielded null monetary outcomes) as a function of both the outcome that was obtained on that trial following stimulus selection and the participant's prior outcome expectation, as quantified by a Q-value. Specifically, each pair-stimulus value, or Q-value, was initialized to zero, and for each trial, t, within that pair of stimuli, the value of the chosen stimulus (say A was chosen) was updated according to:

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

where δ was the prediction error:

$$\delta(t) = r(t) - Q_A\,(t),$$

where $r(t)$ was 0.5 for winning, 0 for obtaining nothing, and -0.5 for losing. The learning rate, α , was given by:

$$\alpha = \begin{cases} \alpha^{+}, & if\ \delta(t) > 0 \\ \alpha^{-}, & if\ \delta(t) < 0 \end{cases},$$

where α^{+} and α^{-} were the learning rates for positive and negative prediction errors, respectively (Frank et al., 2007).

The probability of choosing one stimulus over another (say A over B) was given by the *softmax* equation:

$$P_A(t) = \frac{e^{[Q_A(t)*\beta]}}{e^{[Q_A(t)*\beta]} + e^{[Q_B(t)*\beta]}},$$

where the β parameter, or inverse temperature, reflects the randomness of choices. The lower the β, the higher the randomness of choices.

We modelled participants' trial-by-trial behaviour in the stress and control conditions using this well-established, biologically inspired reinforcement-learning model (Frank et al., 2007). Model fitting involved estimating the values of the parameters (α^{+}, $\alpha^{-}$and β) that best accounted for the respective trial-by-trial choices in each condition. We estimated the best-fitting model parameters (α^{+}, $\alpha^{-}$and β) for each subject in each condition using maximum *a posteriori* estimation (Daw, 2011). Specifically, to optimize model parameters, we drew the learning rates from Beta distributions [*Beta* (1.1, 1.1)] and the inverse temperature from a Gamma distribution [*Gamma* (1.2, 5)](Palminteri et al., 2015), as in Chapter 2. We then used MATLAB's *fmincon* function, initialized at 100 random starting points of the parameter space, to search for the parameter values that minimized the negative log posterior of the observed sequence of choices, given the previously observed outcomes, with respect to different settings of the model parameters (Daw, 2011).

Reinforcement-learning modelling results

In this study, we were interested in examining the impact of acute stress on the neural coding of prediction errors. For completeness, we also assessed whether acute stress affected the parameters estimated by the reinforcement-learning model, as we showed in the previous chapter, Chapter 2, that acute stress decreased the learning rate for positive prediction errors using a similar version of the reinforcement-learning task. Specifically, we conducted repeated-measures ANOVAs with condition (stress and control) and valence of the prediction error (positive and negative) as within-subject factors. We did not find significant differences in the learning rates between the stress and control conditions (main effect of

condition: F (1,22) = 0.095, p = .76, η^2 = 0.004. There was a main effect of valence, such that the learning rate for positive prediction errors was significantly higher than the learning rate for negative predictions errors in both conditions, F (1,22) = 12.26, p = .002, η^2 = 0.36. The condition × valence interaction was non-significant, F (1,22) = 0.90, p = 0.35, η^2 = 0.039). The inverse temperature parameter also did not differ between the stress and control conditions (F (1,22) = 2.24, p = .15, η^2 = 0.093. We further analysed the products between each learning rate ($\alpha^{\pm}$) and the inverse temperature (β), because in reinforcement-learning models $\alpha^{\pm}$and β tend to be inversely coupled (Daw, 2011), and, as a result, the parameters viewed separately can have larger estimation errors, while their product tends to be more reliably estimated, and thus better recovered (Daw, 2011; Schönberg et al., 2007; Zhang et al., 2020). We did not find evidence for significant differences in the product between learning rates and the inverse temperature parameter ($\alpha^{\pm} * \beta$) between the stress and control conditions (main effect of condition: F (1,22) = 3.02, p = .096, η^2 = 0.12) and the condition × valence interaction was also non-significant (F (1,22) = 0.30, p = .59, η^2 = 0.014). Importantly, the version of the reinforcement-learning task used in Chapter 2, where we found a decreased learning rate for positive prediction errors under acute stress, differed from the version of the task applied in the present study in some important methodological aspects, which contributes to explain the absence of significant results. Specifically, as in this study we were particularly interested in the early phase of learning, and we aimed at avoiding that the task being solved inside the scanner was too repetitive and long, we reduced the number of times we presented each pair of stimuli during each block from 40 to 24 (Lefebvre et al., 2017). We also changed the outcome probabilities from 0.80/0.20 to 0.75/0.25 (Lefebvre et al., 2017) to elicit more non-negligible prediction errors, as in the present study we were specifically interested in examining the neural correlates of prediction errors. It is thus possible that by decreasing the number of trials and/or changing the outcome probabilities, we were not able to obtain parameter estimates as stable (Zhang et al., 2020) as in our previous study of Chapter 2. In addition, our sample size (n = 23) was smaller than in our previous behavioural study (n = 62), which may have led to reduced power to detect significant effects of the stressor (which was itself perceived less intensely on this study; see Discussion on the main manuscript) on parameter estimates.

Although our study did not detect an effect of acute stress on the parameter estimates, we conducted two additional checks to validate the used reinforcement-learning model. First, we validated the used reinforcement-learning model by confirming that the probability of choices estimated under the

reinforcement-learning model followed the same pattern of the actual observed choices. Specifically, we computed trial-by-trial choice probabilities for all participants using the best-fitting set of parameters in each condition (the actual observed choices and outcomes were used to update the choice probabilities). To assess whether the probabilities of choosing the correct answer estimated by the reinforcement-learning model (for each subject, the choice probabilities were averaged across gains or loss trials in each condition) followed the same pattern as the actual correct observed choices, we conducted Pearson's correlations for gains and losses in both conditions, using the respective mean percentages. Results confirmed that the choice probabilities estimated under the reinforcement-learning model showed a close correspondence with the actual observed choices across gain and loss trials in both conditions (gains in the stress condition: $r = 0.99$, $p < .0001$; gains in the control condition: $r = 0.97$, $p < .0001$; losses in the stress condition: $r = 0.53$, $p = .0087$; losses in the control condition: $r = 0.64$, $p = .0094$).

Then, we conducted parameter-recovery analyses. Specifically, we simulated the task-choice behaviour of 23 virtual participants using the parameter values that we had estimated for each of the 23 participants in each condition. We ran 100 simulations. Then, for each simulation, we fitted the model to the virtual participants' data to estimate new (recovered) parameters. Finally, we tested the correlations between the original and recovered parameters using Pearson's correlations. Parameter-recovery analyses demonstrated that the results of our model-fitting procedure were robust both in the stress and control conditions (all $r > 0.55$, all $p < .047$; Supplementary Figure S3.6).

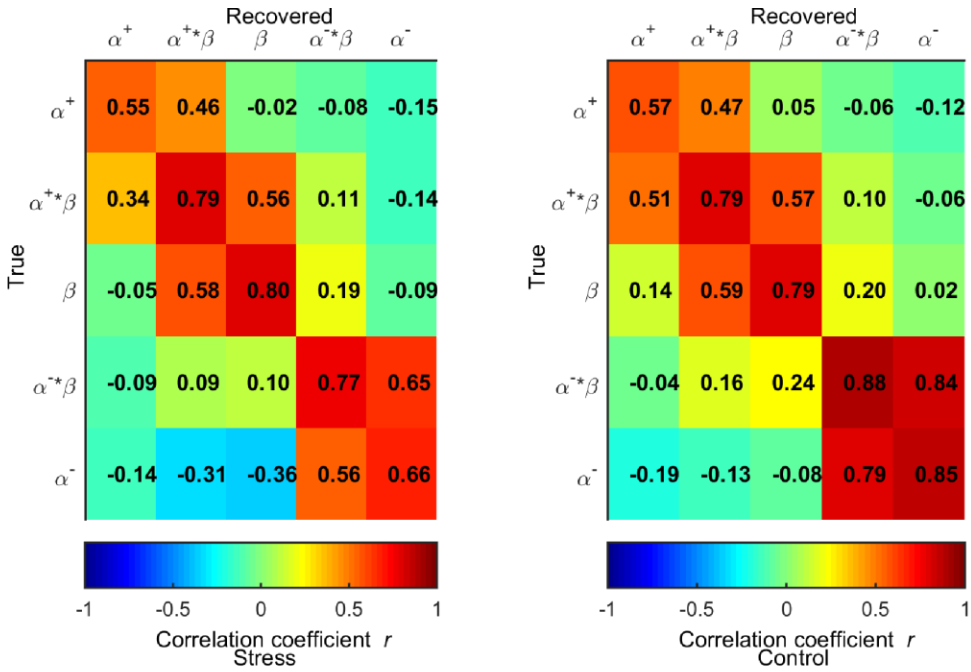


Figure S3.6. Correlation matrices between the estimated and recovered parameters averaged across 100 simulations in the stress (left) and control (right) conditions. The matrices include the values of Pearson's *r* for the correlations between the parameters that were used to generate the simulated data ("True") and those that were obtained by applying the parameter-estimation procedure to the simulated data ("Recovered").

Supplementary tables

Table S3.1. Positive parametric modulation of prediction errors. Regions are reported at a family-wise error (FWE) corrected $p < .05$ with a voxel-level whole-brain correction and a minimum of 10 contiguous voxels, unless otherwise stated. Spatial coordinates (x, y, z) are in the Montreal Neurological Institute (MNI) space, and regions were identified using the automated anatomical labelling (AAL) atlas. R = Right; L = Left. We also tested negative parametric modulations, but as we found no significant activations in our regions of interest, those results are not reported here. [a] Regions reported at $p < .05$, using small-volume FWE voxel-level correction within bilateral masks of the referred regions (dorsal striatum or nucleus accumbens), following an initial threshold of $p < .001$ (uncorrected).

Region	L/R	x	y	z	k	p_{FWE}
Gain trials in the control condition						
Occipital middle	R	26	-90	5	2809	<.001
Occipital inferior	L	-30	-89	8	2419	<.001
Ventral striatum	L	-12	8	-14	35	.001
Ventral striatum	R	14	14	-5	15	.023
Frontal superior	L	-20	32	53	36	.001
Fusiform	L	-32	-36	-23	11	.046
Dorsal striatum [a]	L	-11	9	-11	1355	<.001
Dorsal striatum [a]	R	17	12	-6	791	<.001
Nucleus accumbens [a]	L	-12	8	-14	114	<.001
Nucleus accumbens [a]	R	14	12	-8	135	<.001
Gain trials in the stress condition						
Calcarine	R	21	-92	3	1092	<.001
Occipital middle	L	-15	-98	0	736	<.001
Occipital inferior	R	36	-81	-11	67	<.001
Occipital inferior	L	-32	-84	-12	14	.015
Fusiform	R	30	-42	-20	23	.003
Caudate	L	-8	8	-3	84	<.001
Dorsal striatum [a]	L	-8	8	3	850	<.001
Dorsal striatum [a]	R	11	11	-3	509	.001
Nucleus accumbens [a]	L	-11	6	-8	95	<.001
Nucleus accumbens [a]	R	8	8	-8	119	.001
Loss trials in the control condition						
Parietal Inferior	L	-44	-32	42	21	.005
Dorsal striatum [a]	R	27	-8	12	629	.007
Dorsal striatum [a]	L	-18	-12	-11	199	.014
Dorsal striatum [a]	L	-27	-11	12	297	.022
Dorsal striatum [a]	L	-18	-2	24	65	.035
Nucleus accumbens [a]	L	-15	11	-12	47	<.001
Nucleus accumbens [a]	R	17	9	-11	77	.001
Loss trials in the stress condition						
Dorsal striatum [a]	L	-27	-15	9	207	.028

Table S3.2. Median and boundaries of the magnitudes of the prediction errors for each of the four bins for each condition (stress or control) and each trial valence (gain or loss).

		Median [lower bound; higher bound]			
Condition	**Trial valence**	**Bin 1**	**Bin 2**	**Bin 3**	**Bin 4**
Stress	Gain	-0.39 [-0.43; -0.26]	-0.055 [-0.85; -0.018]	0.080 [0.0037; 0.15]	0.23 [0.17; 0.50]
	Loss	-0.46 [-0.50; -0.43]	-0.23 [-0.38; -0.12]	0.01 [-0.062; 0.055]	0.12 [0.088; 0.17]
Control	Gain	-0.40 [-0.44; -0.24]	-0.0078 [-0.036; 0.014]	0.087 [0.049; 0.15]	0.26 [0.19; 0.5]
	Loss	-0.56 [-0.5; -0.40]	-0.18 [-0.32; -0.12]	0.017 [-0.059; 0.08]	0.15 [0.097; 0.22]

Supplementary references

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. Em *Decision Making, Affect, and Learning: Attention and Performance XXIII* (pp. 1–26).

Diederen, K. M. J., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. *The Journal of Neuroscience*, *37*(7), 1708–1720. https://doi.org/10.1523/JNEUROSCI.1979-16.2016

Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *The Journal of Neuroscience*, *31*(16), 6188–6198. https://doi.org/10.1523/JNEUROSCI.6486-10.2011

Frank, M. J., & Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology*, *36*(1), 133–152. https://doi.org/10.1038/npp.2010.96

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

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, *120*(3), 497–517. https://doi.org/10.1037/0735-7044.120.3.497

Grogan, J. P., Tsivos, D., Smith, L., Knight, B. E., Bogacz, R., Whone, A., & Coulthard, E. J. (2017). Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease. *ELife*, *6*. https://doi.org/10.7554/eLife.26801

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 0067. https://doi.org/10.1038/s41562-017-0067

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *The Journal of Neuroscience*, *29*(48), 15104–15114. https://doi.org/10.1523/JNEUROSCI.3524-09.2009

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *The Journal of Neuroscience*, *27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

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

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, *15*(6), 695–707. https://doi.org/10.1093/scan/nsaa089

APPENDIX 4: Supplementary material for Chapter 4

This Appendix includes:

Supplementary Figures (Figure S4.1 to S4.4)

Supplementary Figures

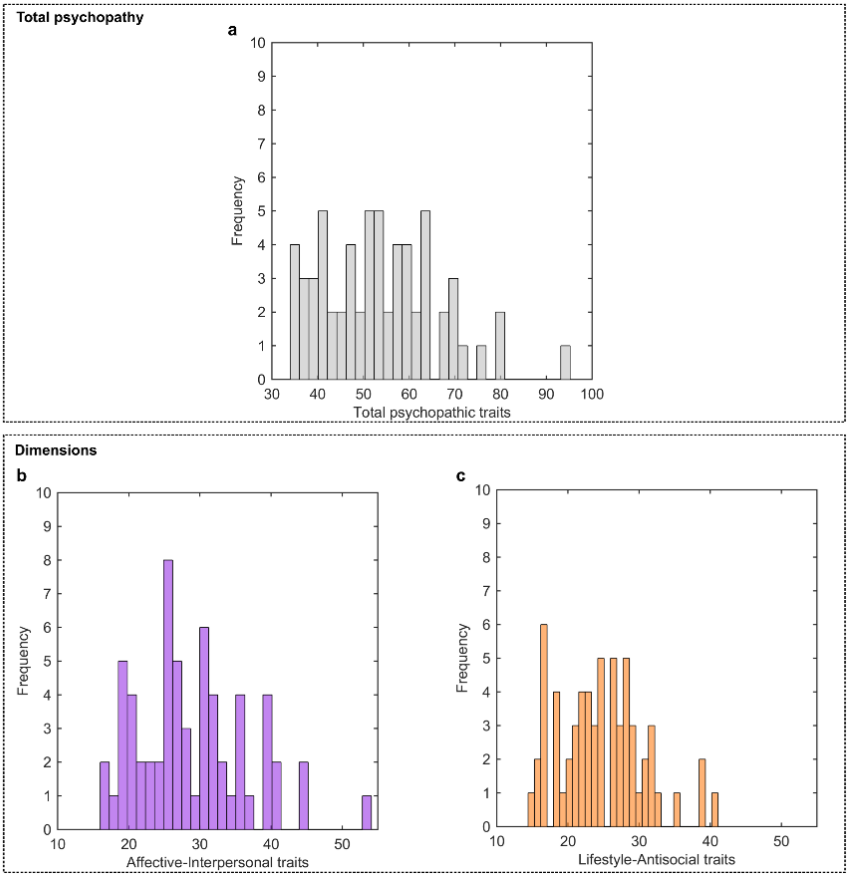


Figure S4.1. Histograms of Self-Report Psychopathy-Short Form (SRP-SF) scores. **(a)** Total SRP-SF scores. **(b, c)** Scores in each SRP-SF dimension: affective-interpersonal **(b)** and lifestyle-antisocial **(c)**.

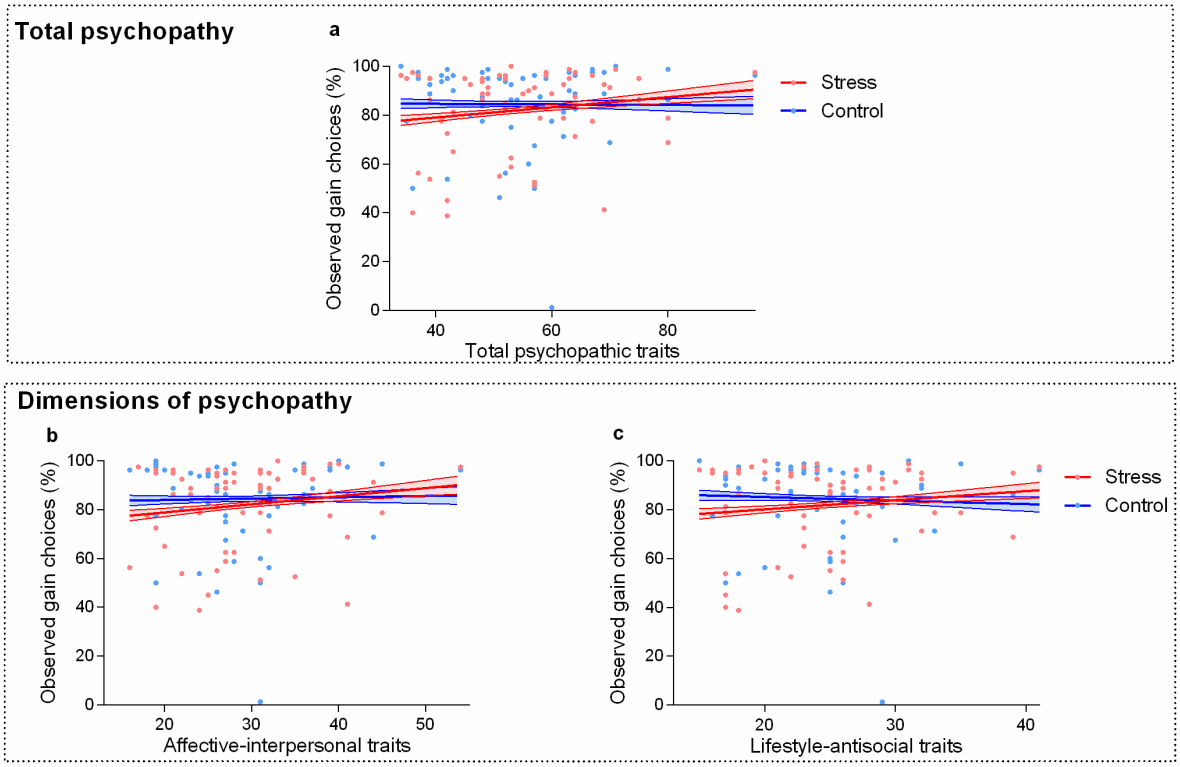


Figure S4.2. Behavioral performance in gain trials across Self-Report Psychopathy-Short Form (SRP-SF) scores in the stress (red) and control (blue) conditions. The plots depict the percentage of correct choices across **(a)** total psychopathy scores, and across **(b, c)** scores on each dimension of psychopathy. The dots represent individual data. The central lines represent the linear fits, and the shading around those central lines represents 95% confidence intervals.

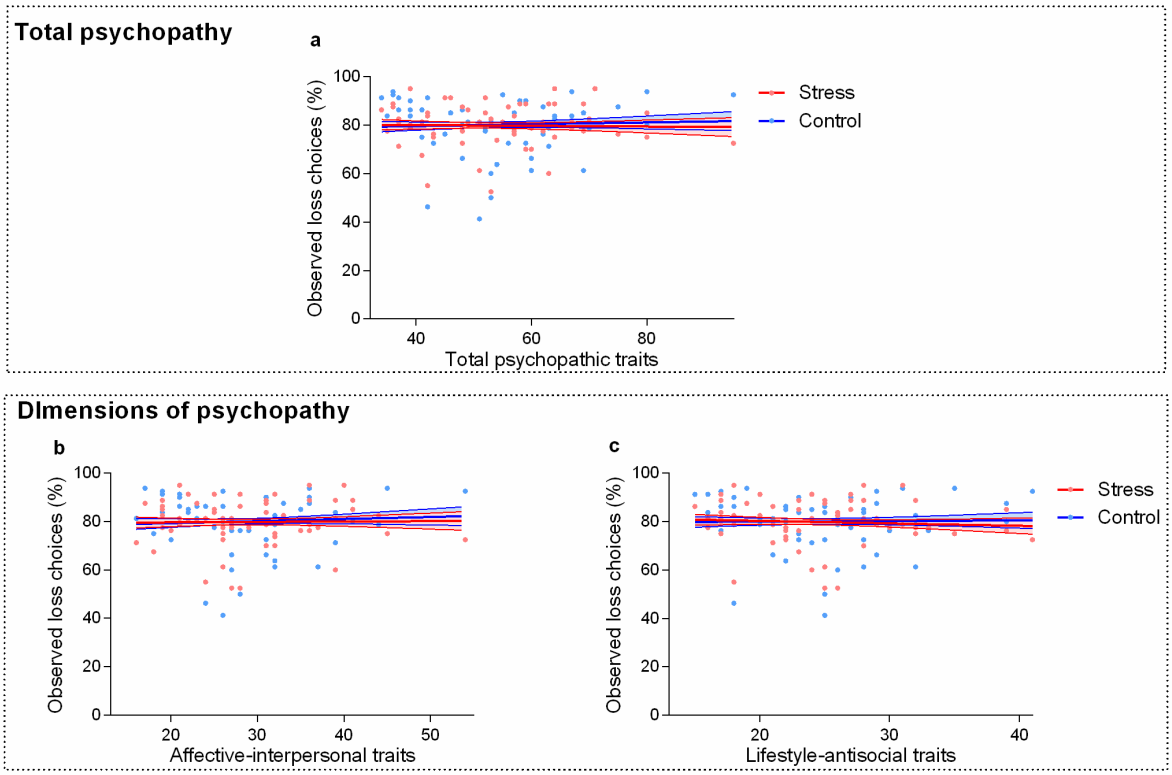


Figure S4.3. Behavioral performance in loss trials across the Self-Report Psychopathy-Short Form (SRP-SF) scores in the stress (red) and control (blue) conditions. The plots depict the percentage of correct choices across **(a)** total psychopathy scores, and across **(b, c)** scores on each dimension of psychopathy. The dots represent individual data. The central lines represent the linear fits, and the shading around those central lines represents 95% confidence intervals.

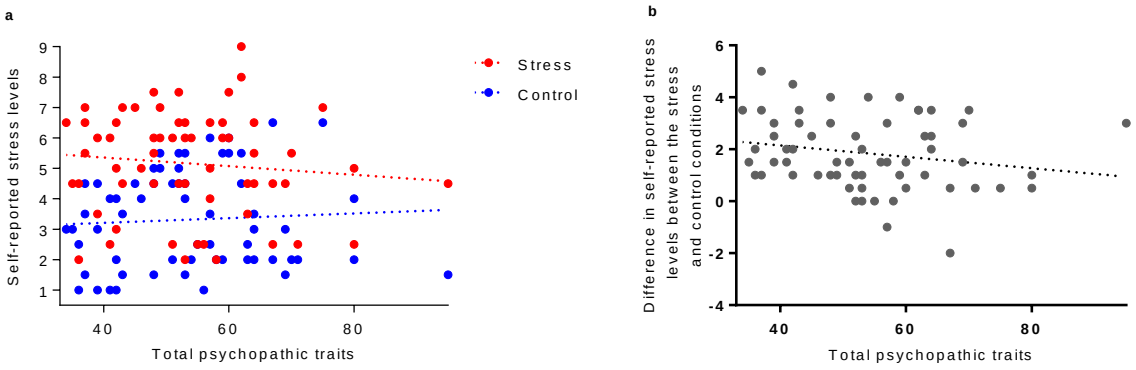


Figure S4.4. Scatterplots of **(a)** self-reported stress levels in the stress (red) and control (blue) conditions as a function of Self-Report Psychopathy-Short Form (SRP-SF) total scores, and **(b)** difference in self-reported stress levels between the stress (red) and control (blue) conditions as a function of SRP-SF total scores. The dots represent individual data, and the dashed lines represent the linear fits.

APPENDIX 5: Supplementary material for Chapter 5

This Appendix includes:

Supplementary Figures (Figure S5.1 to S5.4)

Supplementary Tables (Tables S5.1 and S5.2)

Supplementary Figures

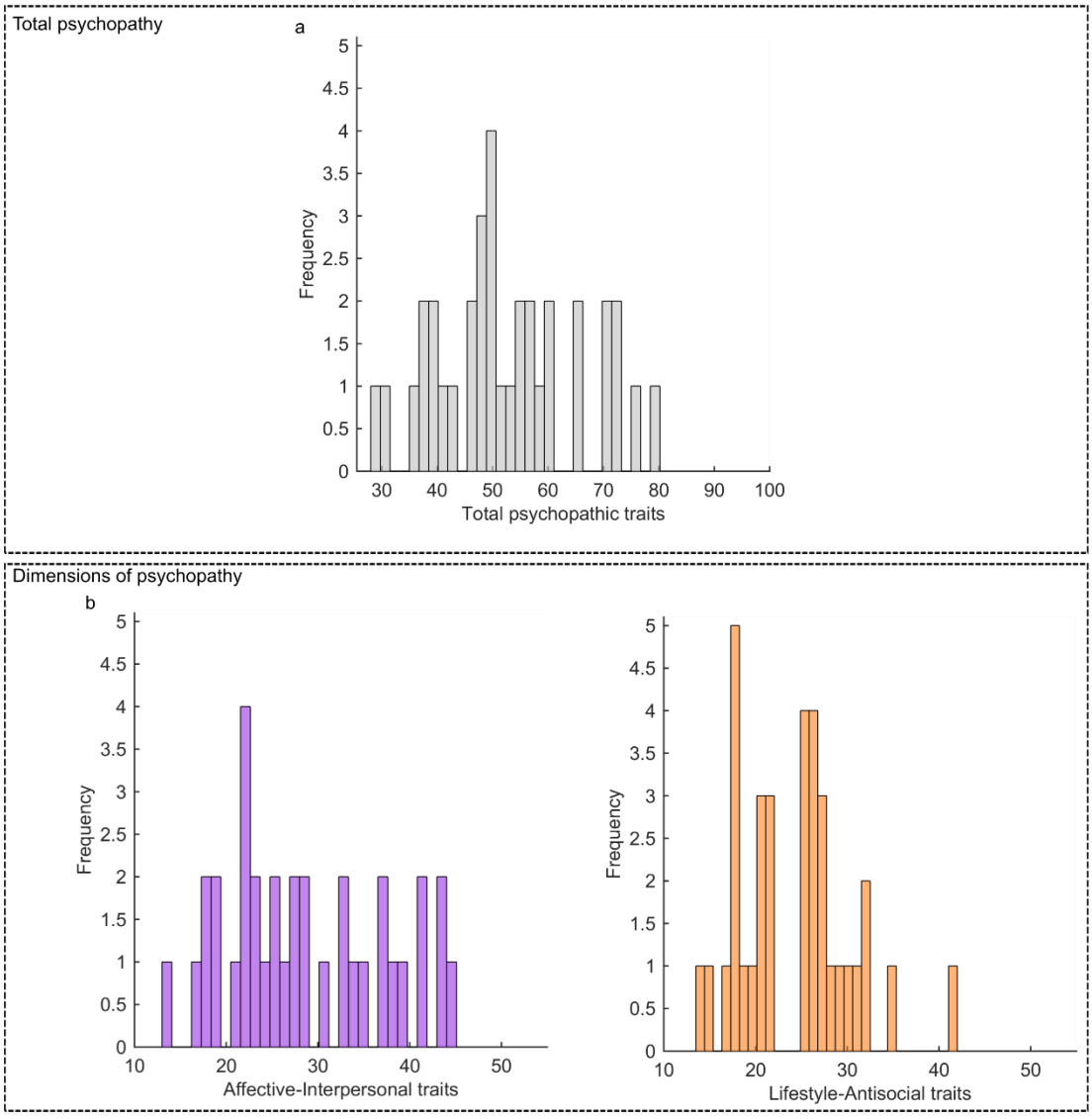


Figure S5.1. Histograms of Self-Report Psychopathy-Short Form (SRP-SF) scores. **(a)** Total SRP-SF scores. **(b)** Scores in each SRP-SF dimension: affective-interpersonal (purple) and lifestyle-antisocial (orange).

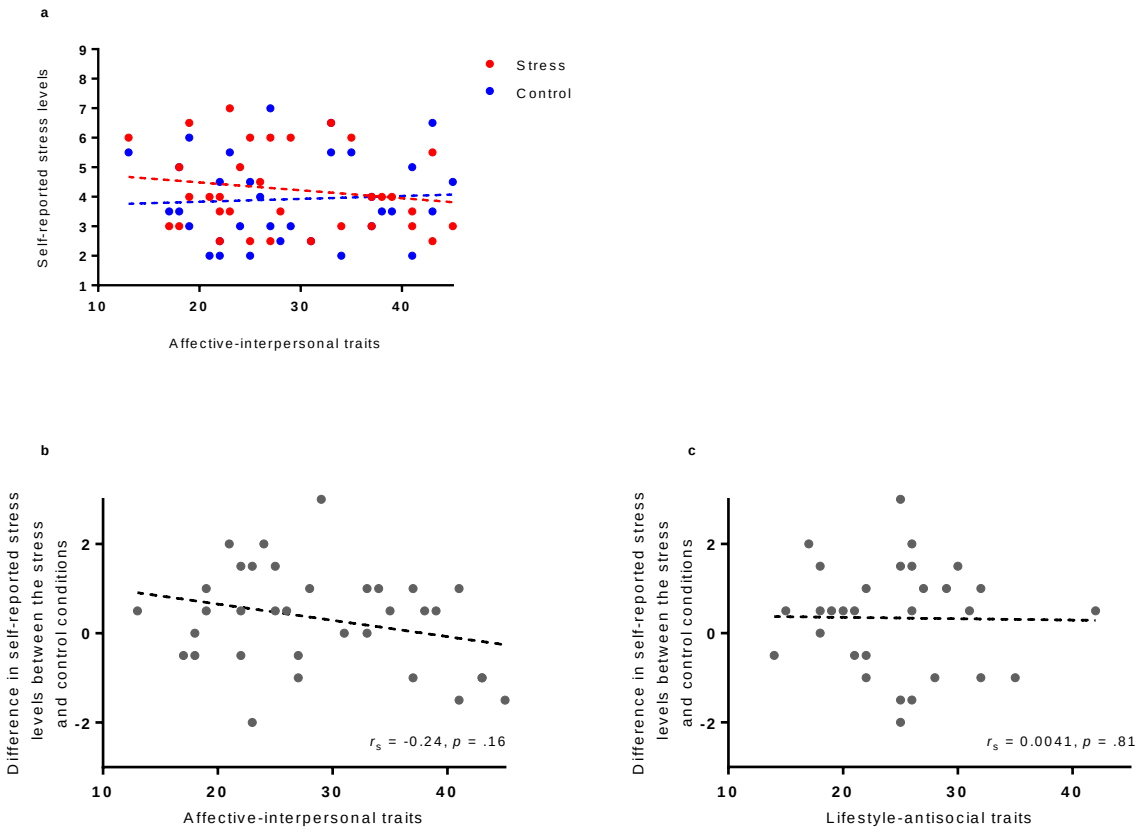


Figure S5.2. Self-reported stress levels as a function of psychopathic traits. **(a)** Self-reported stress levels in the stress (red) and control (blue) condition across affective-interpersonal traits. **(b, c)** Difference in self-reported stress levels between the stress and control condition across the two dimensions of Self-Report Psychopathy-Short Form (SRP-SF): affective-interpersonal dimension and **(b)** lifestyle-antisocial dimension **(c)**. Each dot represents a participant and the dashed lines represent the linear fits.

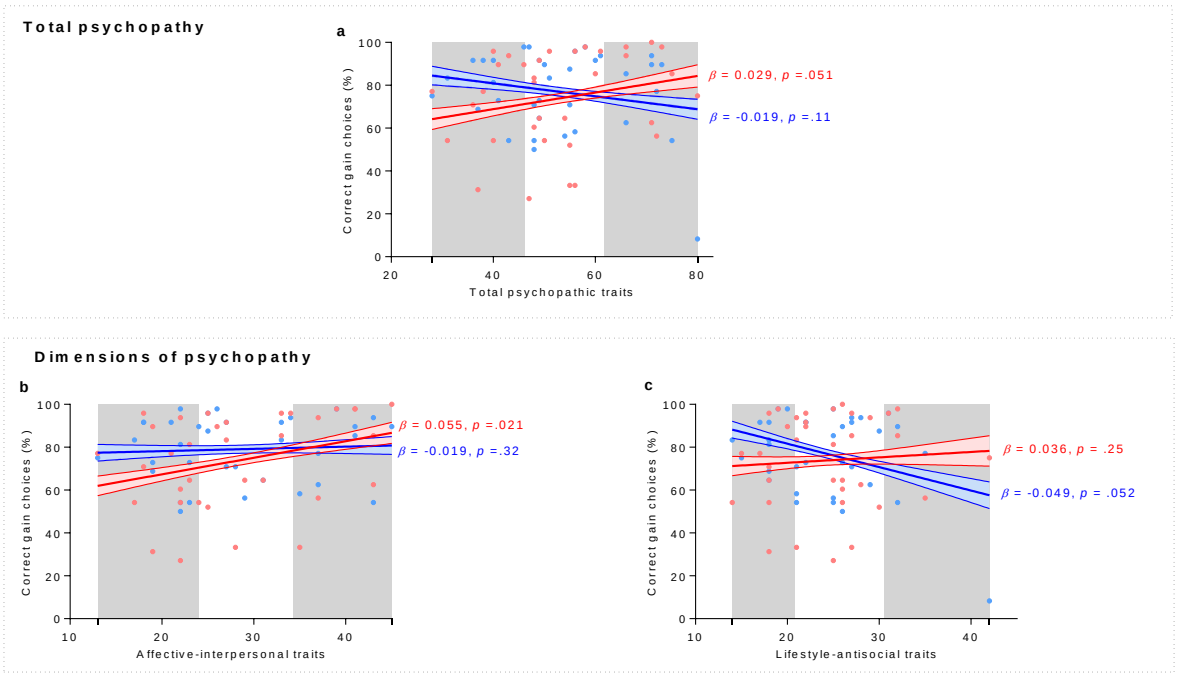


Figure S5.3. Percentage of correct gain choices in the stress (red) and control (blue) conditions across Self-Report Psychopathy-Short Form (SRP-SF) **(a)** total scores, **(b, c)** dimension scores. The interaction and both regions of significance depicted in (c) became non-significant after removing participants that performed below chance levels. Each dot represents a participant. The central lines represent the linear fits of the percentage of correct answers from each subject, and the shading around those central lines represent 95% confidence intervals. The slopes (β) and p-values adjacent to the lines were obtained through simple slope analyses for each condition and valence. The shaded gray areas correspond to Johnson-Neyman regions of significance ($p < .05$).

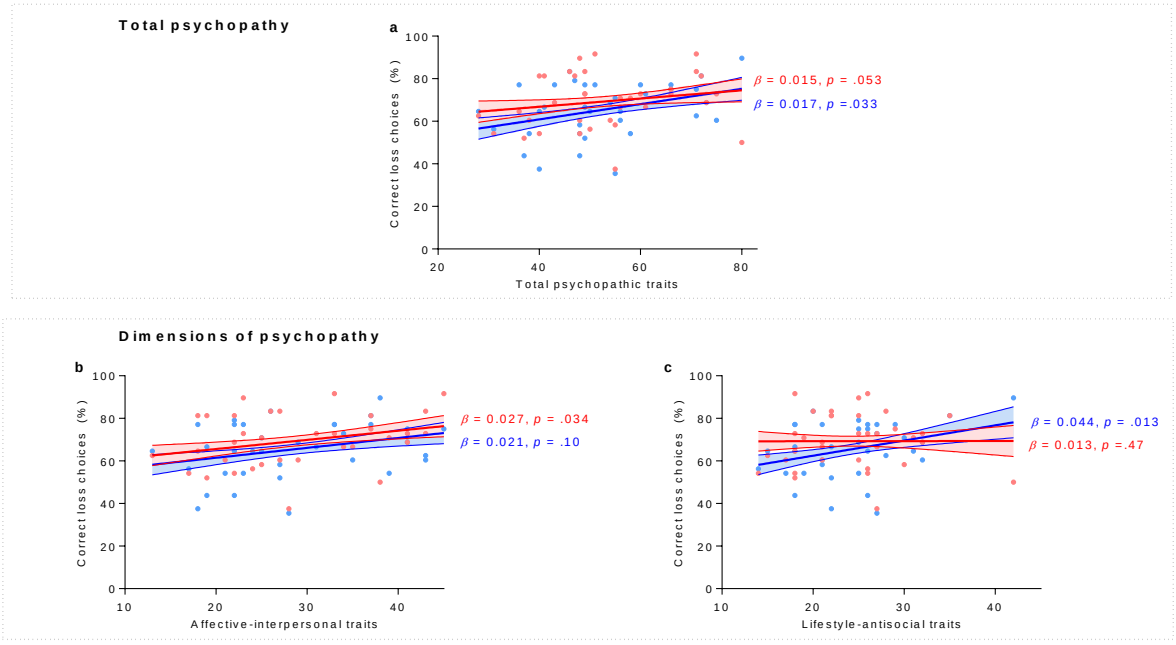


Figure S5.4. Percentage of correct loss choices in the stress (red) and control (blue) conditions across Self-Report Psychopathy-Short Form (SRP-SF) **(a)** total scores, **(b, c)** dimension scores. The interaction depicted in (c) became non-significant after removing participants that performed below chance levels. Each dot represents a participant. The central lines represent the linear fits of the percentage of correct answers from each subject, and the shading around those central lines represent 95% confidence intervals. The slopes (β) and p-values adjacent to the lines were obtained through simple slope analyses for each condition and valence.

Supplementary Tables

Table S5.1. Associations between BOLD response modulation by prediction errors and Self-Report Psychopathy-Short Form (SRP-SF) total and dimension scores. Results are reported at $p < 0.05$, using small-volume FWE voxel-level correction within a bilateral mask of the dorsal striatum (no significant activations were found using a bilateral mask the nucleus accumbens), following an initial threshold of $p < .001$ (uncorrected). Spatial coordinates (x, y, z) are in the Montreal Neurological Institute (MNI) space. The significant results are discriminated for positive/or negative associations. No significant associations are identified by NA. [a]the association became non-significant after removing four participants who performed below chance levels.

Region	x	y	z	k	*Z*	$p\text{-}_{FWE}$
Gain trials in the stress condition						
Total SRP-SF scores (positive association)	30	3	11	464	5.18	<.001
Affective-interpersonal scores (positive association)	32	3	9	728	4.85	<.001
	-20	17	-3	257	4.10	.036
Lifestyle-antisocial scores (positive association)[a]	30	3	11	29	4.00	.048
Gain trials in the control condition						
Total SRP-SF scores				NA		
Affective-interpersonal scores				NA		
Lifestyle-antisocial scores				NA		
Loss trials in the stress condition						
Total SRP-SF scores				NA		
Affective-interpersonal scores				NA		
Lifestyle-antisocial scores				NA		
Loss trials in the control condition						
Total SRP-SF scores				NA		
Affective-interpersonal scores (negative association)	33	2	-9	83	4.21	0.022
Lifestyle-antisocial scores				NA		

Table S5.2. Mean and boundaries of the prediction errors bins for each condition (stress and control) and trial valence (gain and loss) in participants with low and high affective-interpersonal traits.

		Low affective-interpersonal traits			High affective-interpersonal traits		
Condition	**Trial valence**	**Bin 1**	**Bin 2**	**Bin 3**	**Bin 1**	**Bin 2**	**Bin 3**
Stress	**Gain**	-0.29 [-0.42; -0.091]	-0.0091 [-0.068; 0.073]	0.26 [0.11; 0.50]	-0.29 [-0.46;-0.045]	0.037 [-0.0071; 0.096]	0.25 [0.12; 0.50]
	Loss	-0.46 [-0.50; -0.40]	-0.099 [-0.28; 0.022]	0.075 [0.029; 0.14]	-0.41 [-0.50; -0.27]	-0.027 [-0.17; 0.049]	0.16 [0.079; 0.23]
Control	**Gain**	-0.32 [-0.47; -0.084]	-0.0017 [-0.042; 0.044]	0.22 [0.071; 0.50]	-0.26 [-0.43;-0.018]	0.051 [0.0062; 0.10]	0.27 [0.12; 0.50]
	Loss	-0.47 [-0.50; -0.40]	-0.15 [-0.30;-0.053]	0.0055 [-0.0053; 0.13]	-0.40 [-0.50; -0.28]	-0.055 [-0.22; 0.058]	0.17 [0.10; 0.24]

Appendix 6: Published paper

Brain and Cognition 147 (2021) 105657

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Acute stress impairs reward learning in men

Joana Carvalheiro [a,*], Vasco A. Conceição [b], Ana Mesquita [a], Ana Seara-Cardoso [a,*]

[a] Escola de Psicologia, CIPsi, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

ARTICLE INFO

Keywords:
Computational modeling
Dopamine
Prediction error
Reinforcement learning
Stress

ABSTRACT

Acute stress is ubiquitous in everyday life, but the extent to which acute stress affects how people learn from the outcomes of their choices is still poorly understood. Here, we investigate how acute stress impacts reward and punishment learning in men using a reinforcement-learning task. Sixty-two male participants performed the task whilst under stress and control conditions. We observed that acute stress impaired participants' choice performance towards monetary gains, but not losses. To unravel the mechanism(s) underlying such impairment, we fitted a reinforcement-learning model to participants' trial-by-trial choices. Computational modeling indicated that under acute stress participants learned more slowly from positive prediction errors — when the outcomes were better than expected — consistent with stress-induced dopamine disruptions. Such mechanistic understanding of how acute stress impairs reward learning is particularly important given the pervasiveness of stress in our daily life and the impact that stress can have on our wellbeing and mental health.

1. Introduction

Learning to choose options that lead to rewards and to avoid those that result in punishments is crucial for adaptive behavior. Situational factors, such as stress, can have deleterious effects on the ability to make the best choices and learn from them (Porcelli & Delgado, 2017). Stress is present in our day-to-day life, but, notably, how acute stress affects reward and punishment learning remains largely unknown. A growing body of evidence suggests that acute stress impairs reward-seeking behavior (Berghorst, Bogdan, Frank, & Pizzagalli, 2013; Bogdan & Pizzagalli, 2006; Bogdan, Santesso, Fagerness, Perlis, & Pizzagalli, 2011; Ehlers & Todd, 2017; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020; but see Lighthall, Gorlick, Schoeke, Frank, & Mather, 2013), but less is known about the impact of acute stress on punishment-avoidance behavior (Aylward et al., 2019; Petzold, Plessow, Goschke, & Kirschbaum, 2010). More critically, there is even less evidence on the mechanisms that underlie the behavioral effects of acute stress on reward and punishment learning. Here, we use a computational reinforcement-learning framework to investigate the impact of acute stress on reward and punishment learning in men.

In the past decades, the use of computational modeling approaches to describe behavior-brain relationships in healthy humans has played an influential role on cognitive science (Daw & Frank, 2009; Frank, 2015; Huys, Maia, & Frank, 2016; Maia & Frank, 2011; Maia, 2015). Computational models, such as reinforcement-learning models, are built and implemented to capture very specific cognitive and neural mechanisms, thus linking different levels of analysis, from cognitive and behavioral phenomena to neurobiological mechanisms (Chater, 2009; Daw & Frank, 2009; Nair, Rutledge, & Mason, 2020). Reinforcement-learning models are considered extremely useful tools to investigate the neural computations underpinning cognition and behavior (Collins & Frank, 2013; Daw, 2011; Daw & Frank, 2009; Huys et al., 2016; Maia & Frank, 2011; Nair et al., 2020), and can thus provide a mechanistic framework to disentangle the effects of acute stress on reward and punishment learning (Aylward et al., 2019; Huys, Pizzagalli, Bogdan, & Dayan, 2013; Luksys & Sandi, 2011; Otto, Raio, Chiang, Phelps, & Daw, 2013; Radenbach et al., 2015; Robinson, Overstreet, Charney, Vytal, & Grillon, 2013).

According to reinforcement-learning theory, individuals learn to gradually select more and more often the actions that maximize rewards and those that minimize punishments by learning the values of the executed actions, and such learning is driven by prediction errors (Maia & Frank, 2011; Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998). Specifically, prediction errors — which signal the difference between obtained and expected outcomes — are used to progressively update the values of the executed actions (Collins & Frank, 2013; Maia &

* Corresponding authors.
E-mail addresses: joana.rita.carvalheiro@gmail.com (J. Carvalheiro), vasco.conceicao7@gmail.com (V.A. Conceição), ana.mesquita@psi.uminho.pt (A. Mesquita), ana.searacardoso@psi.uminho.pt (A. Seara-Cardoso).

https://doi.org/10.1016/j.bandc.2020.105657
Received 30 July 2020; Received in revised form 20 October 2020; Accepted 23 November 2020
Available online 17 December 2020
0278-2626/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Frank, 2011; Schultz et al., 1997; Sutton & Barto, 1998). Prediction errors can be positive or negative. Positive prediction errors occur when outcomes are better than expected, and can thus signal unexpected rewards or avoidance of expected punishments (Daw & Tobler, 2014; Schultz et al., 1997; Watabe-Uchida, Eshel, & Uchida, 2017). Negative prediction errors occur when outcomes are worse than expected, thus signaling unexpected punishing outcomes or omission of expected rewards (Daw & Tobler, 2014; Schultz et al., 1997; Watabe-Uchida et al., 2017). This means that positive and negative prediction errors can be both present in reward and punishment learning. Interestingly though, recent evidence indicates that in simple reinforcement-learning tasks, reward learning should be mostly driven by positive prediction errors (Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde, & Palminteri, 2017; Palminteri & Pessiglione, 2013), while punishment learning seems to be largely driven by negative prediction errors (Palminteri & Pessiglione, 2013). Studies that do not consider the positive and negative valence of prediction errors when modeling reward and/or punishment learning often assume that, during reward learning, obtaining unexpected rewards (positive prediction errors) is coded similarly to the omission of expected rewards (negative prediction errors), and that, during punishment learning, obtaining unexpected punishments (negative prediction errors) is coded similarly to the omission of expected punishments (positive prediction errors). However, extant evidence suggests that the brain encodes positive and negative prediction errors differently (Frank, 2005; Maia & Frank, 2011; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Schultz et al., 1997). Thus, to better understand the mechanisms underlying reward and punishment learning, computational models could benefit from the integration of such neurobiological evidence and model positive and negative prediction errors separately. Importantly, reinforcement-learning models can capture how quickly unexpected positive and negative outcomes are integrated over time through distinct learning rates for positive and negative prediction errors, respectively (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007). Relatedly, blunted signaling of positive and negative prediction errors can be captured by reinforcement-learning models as reduced learning rates.

Dopaminergic functioning plays a key role in prediction-error-based learning (Frank, Seeberger, & O'Reilly, 2004; Glimcher, 2011; Maia & Frank, 2011; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). Prediction-error signals are known to be encoded in the phasic activity of dopamine neurons (Daw & Tobler, 2014; Maia & Frank, 2011; Schultz et al., 1997). Specifically, phasic bursts of dopaminergic neurons are thought to adaptively encode positive prediction errors, whereas dopamine dips have been associated with the adaptive encoding of negative prediction errors (Daw & Tobler, 2014; Maia & Frank, 2011; Schultz et al., 1997). However, phasic-dopamine responses do not seem to be always adaptive, and there is evidence that dopamine can be phasically released in an aberrant spontaneous manner (Belujon, Grace, & Grace, 2015; Maia & Frank, 2017; Sulzer, Cragg, & Rice, 2016). Studies with non-human male animals suggest that acute stress increases aberrant spontaneous phasic-dopamine release in the mesolimbic pathway, which originates in the ventral tegmental area and projects to terminal regions such as the striatum (Anstrom, Miczek, & Budygin, 2009; Anstrom & Woodward, 2005; Valenti, Lodge, & Grace, 2011). Crucially, exaggerated, aberrant spontaneous dopamine release seems to reduce adaptive phasic bursts that signal positive prediction errors in the striatum (Daberkow et al., 2013; Grace, 2016; Maia & Frank, 2017; Werlen et al., 2020). Additionally, though more speculative (Maia & Frank, 2017), aberrant spontaneous dopamine release might also block the effects of dopamine dips (Frank, 2005; Frank & O'Reilly, 2006) needed to signal negative prediction errors. Still, the extent to which stress-induced aberrant spontaneous dopamine release blunts signaling of positive and/or negative prediction errors remains poorly understood.

Stress can be defined as a subjective state of sensing potentially adverse changes in the environment (Joëls & Baram, 2009). Stressful stimuli (i.e., stressors) can elicit many physiological and functional changes in the brain in response to perceived adverse changes in the environment (de Kloet, Joëls, & Holsboer, 2005; Hermans, Henckens, Joëls, & Fernández, 2014; Joëls & Baram, 2009). The first wave of the stress response occurs rapidly and involves enhanced release of catecholamines, such as noradrenaline and dopamine. The catecholamines release generally initiates almost instantly after the onset of the stressor and the rapid raise in catecholamines levels is quickly translated into changes in neural functioning (Hermans et al., 2014; Joëls & Baram, 2009; Sapolsky, Romero, & Munck, 2000). For instance, increased release of noradrenaline seems to contribute to the reorientation of attention to a more general scanning of the environment (Hermans et al., 2014) and dopamine, which is released during stress exposure in the prefrontal cortex and striatum (Abercrombie, Keefe, DiFrischia, & Zigmond, 1989; Anstrom et al., 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Nagano-Saito et al., 2013; Pruessner, Champagne, Meaney, & Dagher, 2004), is thought to alter reward learning rein (Porcelli & Delgado, 2017). A second, and slower, wave of the stress response involves secretion of glucocorticoids. Cortisol usually has slow, prolonged actions on neural functioning (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009; Sapolsky et al., 2000). Peak concentrations of cortisol are typically not reached before 15–30 min after the stressor onset (de Kloet et al., 2005; Hermans et al., 2014), which points to a limited role of cortisol in the immediate stress response.

Stress exerts various effects on cognitive functioning (de Kloet et al., 2005), namely via fast-initial release of catecholamines and slow-delayed release of cortisol, as mentioned above. Thus, one important factor when studying the impact of acute stress on learning is to consider whether learning is tested immediately after the onset of the stressor when catecholamines levels are higher or later when cortisol levels peak (Hermans et al., 2014; Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; Porcelli & Delgado, 2017; Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012). A significant number of stress and learning studies have considered only slower effects of cortisol in the stress response, adopting designs that may involve relatively longer latencies between stress induction and task solving (Byrne, Cornwall, & Worthy, 2019; de Berker et al., 2016; Ehlers & Todd, 2017; Kruse, Tapia León, Stalder, Stark, & Klucken, 2018; Lighthall et al., 2013; Otto et al., 2013; Paul, Bellebaum, Ghio, Suchan, & Wolf, 2019; Petzold et al., 2010). In those studies, participants are often exposed to a stress manipulation, such as the Trier social stress test or the (socially evaluated) cold pressor test, and the impact of stress on learning is assessed around 10–20 min after the stress manipulation has occurred, so that cortisol levels peak during the learning task. Such stress inductions can combine a social evaluation component, unpredictability, and uncontrollability, which are known to produce robust increases in cortisol levels (McRae et al., 2006; Schwabe, Haddad, & Schachinger, 2008). However, in those paradigms, it is less clear if individuals are actually perceiving stress at the time of learning, as the stressor is experienced outside of the learning context (McRae et al., 2006; Schwabe et al., 2012). Moreover, it is not unlikely that stress recovery occurs during the learning task (see de Berker et al., 2016; Paul et al., 2019; Radenbach et al., 2015; Schwabe & Wolf, 2009, for examples of studies that show decreasing cortisol levels during task solving) due to suppressive actions of glucocorticoids (Hermans et al., 2014; Sapolsky et al., 2000).

Fast-acting stress-induced catecholaminergic effects, such as increased release of dopamine (Abercrombie et al., 1989; Anstrom et al., 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Nagano-Saito et al., 2013; Pruessner et al., 2004), can influence brain function in regions known to be critically associated with learning processes. Thus, by exposing participants to an acute stressor whilst they perform a learning task, one can conceivably capture the impact of stress-induced catecholamine (e.g., dopamine) release on learning. However, one potential drawback of using acute stressors, such as threat of shock, time pressure or noise, during a cognitive task, is that it can make it difficult to disentangle whether cognitive impairments occur as

a result of stress or because the stressor acted as a distractor (Robinson, Vytal, Cornwell, & Grillon, 2013). Indeed, the first wave of the stress response can trigger a sensory hypervigilant state and the redirection of attention to stressors, conceivably causing distraction in tasks that require selective attention (Hermans et al., 2014). Importantly though, distractibility seems to be more robustly induced by unpredictable changes (Parmentier, 2014; Parmentier & Hebrero, 2013; Sussman, Winkler, & Schröger, 2003). Therefore, one possible alternative to minimize the potential distractor effect of an acute stressor during a cognitive task could be to minimize the unpredictability of that stressor.

In this study, we chose to use an acute stressor (with reduced unpredictability) during a reinforcement-learning task, to better understand whether and how acute stress affects reward and/or punishment learning, very likely via putative dopaminergic mechanisms, which can be captured by computational models such as the one we used (Frank et al., 2007). The aim of the present work was twofold. First, to investigate how acute stress impacted behavioral performance during reward and punishment learning. Second, to inspect the computational mechanisms behind the effects of acute stress on reward and punishment learning. Given the putative roles of phasic-dopamine responses on prediction errors signaling and of acute stress on aberrant spontaneous phasic-dopamine release, we hypothesized that acute stress would impair reward learning and, more tentatively punishment learning. Relatedly, we hypothesized that acute stress would decrease the learning rate for positive prediction errors and, more tentatively, the learning rate for negative prediction errors.

To test those hypotheses, we used a well-established reinforcement-learning task involving monetary gains and losses (Pessiglione et al., 2006) combined with a novel acute stress manipulation. Using this task combined with pharmacological manipulations of the dopaminergic system, Pessiglione et al. (2006) showed that dopamine-related drugs modulate prediction errors expressed in the striatum during reward learning. The stress manipulation consisted of exposing participants to an uncontrollable sound (i.e., participants could not put an end to it) whilst they performed the reinforcement-learning task. We chose to use a repetitive alarm sound as stressor, because alarms are known to activate the nervous sympathetic system (e.g., increase heart rate and skin conductance), involved in the earlier phase of the stress response (Hall et al., 2016; Korhonen, 1981; Westman & Walters, 1981). Repetitive, uncontrollable sounds for long periods of time seem to be particularly effective in increasing dopamine release (Arnsten & Goldman-Rakic, 1998) and levels of adrenocorticotropic hormone (Arguelles, Ibeas, Ottone, & Chekherdemian, 1962; Breier et al., 1987). To minimize the potential distractor effect of the stressor, we used a sound that was always constant and repetitive, and therefore predictable, as unexpected changes in sound sequences robustly induce distraction and negatively affect performance in an unrelated task (Hughes, 2014; Parmentier, 2014; Parmentier, Elford, Escera, Andrés, & Miguel, 2008). To further minimize unpredictability, we also instructed participants, before they started the task, that they would hear an alarm, not associated with the task itself or with their performance, during blocks cued with a warning signal and a red background (Hughes, 2014; Parmentier & Hebrero, 2013; Sussman et al., 2003). Although the stressor was not unpredictable, it was uncontrollable (i.e., participants could not put an end to it). Uncontrollable auditory stimuli can elevate stress responses (Arguelles et al., 1962; Breier et al., 1987; Rylander, 2004; Westman & Walters, 1981) and may disrupt dopaminergic mechanisms (Arnsten & Goldman-Rakic, 1998) and cognitive functioning (Glass, Reim, & Singer, 1971). To check the success of the acute-stress manipulation, we collected self-report stress levels at the end of each block of the task and measured skin conductance response (SCR) rate throughout the task. Then, we inspected how acute stress altered reward and punishment learning during the reinforcement-learning task using both classical statistical analyses and computational-model-based analyses of participants' behavioral data. For the latter analyses, we fitted participants' choices with a previously established, biologically inspired, reinforcement-learning model (Frank et al., 2007), which has been extensively used to investigate the cognitive and behavioral impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans (Diederen et al., 2017; Doll, Hutchison, & Frank, 2011; Frank & Fossella, 2011; Frank et al., 2007; Grogan et al., 2017; Rutledge et al., 2009). The fitted reinforcement-learning model allowed us to examine the effects of acute stress on learning rates for positive and negative prediction errors. In this study, we included only male participants due to females' hormonal-dependent variations on stress responsivity, as well as on reward and punishment learning (Diekhof, Korf, Ott, Schädlich, & Holtfrerich, 2020; Dreher et al., 2007; Ossewaarde et al., 2010).

2. Material and Methods

2.1. Participants

Sixty-two healthy male participants (age range = 18 – 35 years; M = 21.9 years, SD = 3.7 years) were recruited at University of Minho, Portugal. Due to similarities in study design, we used the study of Pezold et al. (2010) as a reference to determine our minimum sample size. To allow for possible data exclusions due to technical problems and/or nonresponders in skin conductance data, participant enrollment in our study was stopped once we reached a sample size two-fold the sample size analyzed in Petzold et. al (2010). Data collection continued until all participants enrolled up to that point were tested. Data collection was stopped before data analyses, independently of the statistical significance of the data. Four participants were excluded from skin-conductance analyses due to poor signal quality. No participants were *a priori* excluded from any other data analyses, although we conducted confirmatory analyses excluding potential outliers to ensure that our results were not driven by extreme values.

All participants provided their informed consent before the experimental session. All experimental procedures were approved by the Ethics Committee of University of Minho.

2.2. Reinforcement-learning task

After a short practice (12 trials), to familiarize participants with the task timings and response keys, participants completed four blocks of an adapted version of a well-established reinforcement-learning task (Pessiglione et al., 2006) (Fig. 1). Each block included three pairs of abstract stimuli, and each pair of stimuli was presented 40 times, making 120 trials in total per block. Different abstract stimuli were used in each block. Each pair of stimuli was associated with a valence: one pair of stimuli was associated with gains (gain 0.5€ or no change), a second pair associated with losses (loss 0.5€ or no change), and a third pair associated with neutral, or non-financial outcomes (look at a 0.5€ coin or no change; for a depiction of neutral trials see Fig. S1a in the Supplementary Material). The outcome probabilities were reciprocally 0.8 and 0.2 for the stimuli in each of the three pairs. On each trial, one pair was randomly presented on the screen, with one stimulus above and the other below a central fixation cross (the stimuli position was counterbalanced across trials). Participants were instructed to choose between the two visual stimuli displayed on the computer screen to maximize payoffs. Missing choices occurred when participants did not press the response keys within 2000 ms (0.38% missing choices: 40 in the stress condition and 72 in the control condition, in a total of 29,760 trials across all participants) and were signaled with a "Missed" message (no other outcome was provided). Missing choices were excluded from data analyses. Before starting the task, participants were informed that they would be paid the amount of money obtained during a randomly selected block, although they all left with the same fixed compensation (15€) for their participation. The experiment was programmed and presented with Cogent 2000 (http://www.vislab.ucl.ac.uk/cogent.php) implemented in MATLAB R2015a (MathWorks).

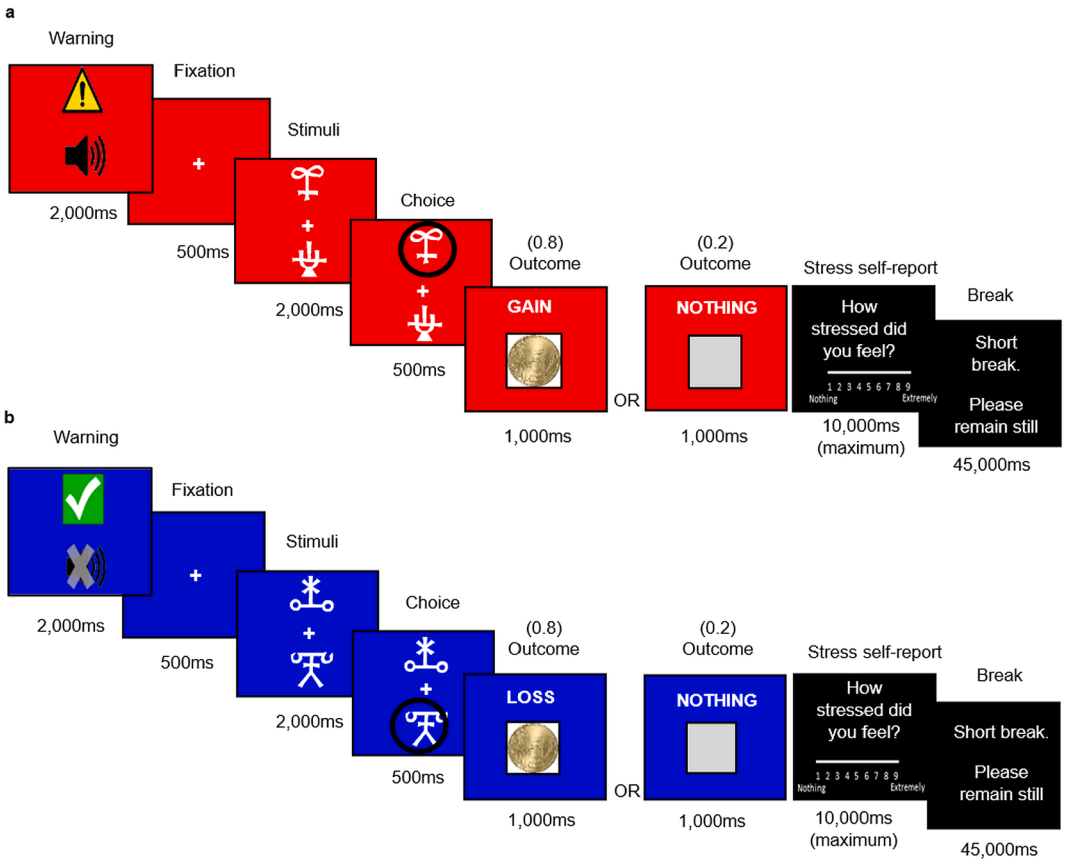


Fig. 1. Reinforcement-learning task. On each trial of the task, participants had to choose either the upper or the lower of two abstract visual stimuli presented on a computer screen and to subsequently observe the obtained outcome, whilst under acute stress and under control conditions. **(a)** In the depicted stress-condition example, the chosen stimulus was associated with a probability of 0.8 of winning 0.5€ and with a probability of 0.2 of obtaining nothing. The other (not chosen) stimulus was associated with a probability of 0.8 of obtaining nothing and with a probability of 0.2 of winning 0.5€. **(b)** In the depicted control-condition example, the chosen stimulus was associated with a probability of 0.8 of losing 0.5€ and with a probability of 0.2 of obtaining nothing. The other (not chosen) stimulus was associated with a probability of 0.8 of obtaining nothing and with a probability of 0.2 of losing 0.5€. Participants completed a total of four blocks, consisting of an alternation between two stress and two control blocks. To assess stress responses, self-reported stress levels were collected at the end of each block and skin conductance was measured throughout the task.

2.3. Acute stress manipulation

During the experimental session, participants performed two blocks of the reinforcement-learning task whilst exposed to a stressor (i.e., stress condition; Fig. 1a) and two blocks without the stressor (i.e., control condition; Fig. 1b). To elicit stress responses we exposed participants to a predictable, but uncontrollable auditory stimulus, a constant alarm ("Annoying modern office building alarm.wav", retrieved from freesound.org, and programmed to loop uninterruptedly), played through the same set of over-ear headphones (GOODIS, model GWH4093, with the volume set to the maximum), with the sound volume adjusted in the laptop to level 27 (in a scale from 0 to 100) for all participants. We conducted a brief pilot to qualitatively confirm that this volume was tolerable, yet stressful. Stress blocks were signaled by a warning sign and a red background (Fig. 1a), and control blocks were signaled by a safe sign and blue background (Fig. 1b). Stress and control blocks were administered alternately and in counterbalanced order. Each block of the task lasted 8 min, and the whole task took around 36 min (including warnings, breaks between blocks, and stress ratings with variable durations). The experiment was conducted in a soundproof room to avoid interference due to environmental noise, and between 12 pm and 6 pm to minimize diurnal variability in stress responses.

2.4. Manipulation check

Stress levels were assessed by asking participants at the end of each block to rate how stressed they felt during that block on a scale of 1 (nothing) to 9 (extremely). To further assess the impact of the acute stressor on autonomic responses, we acquired skin conductance responses (SCRs) using BIOPAC MP150 and a pair of finger electrodes. Electrodes were attached to participants' left index and ring fingers; the gain was set to 5, the low pass filter to 10 Hz, and the high pass filters to DC. Recordings were performed using Acknowledge 4.4. Data were acquired at 200 Hz, downsampled to 62 Hz, and smoothed with a median filter in order to remove outliers. Each participant's SCRs were detected using a threshold of 0 μs and a rejection rate of 10% (Kim, Bang, & Kim, 2004), meaning that we excluded SCR with an amplitude smaller that 10% of the maximum SCR amplitude in each subject, rather than using an arbitrary threshold that would not account for inter-individual differences in SCRs. The SCR rate was calculated by dividing the number of SCRs detected in each block by the duration of that block (in minutes).

Self-reported stress levels and SCR rates were analyzed using repeated-measures analyses of variance (ANOVAs), with condition (stress and control) and block (1 and 2) as within-subject factors, and post-hoc paired *t*-tests. ANOVAs effect sizes are reported as eta-squared, η^2, and post-hoc paired *t*-tests' effect sizes are reported as Cohen's *d* and 95% confidence intervals. We further conducted non-parametric Wilcoxon signed-rank tests, which are more robust to outliers, to confirm

the results from post-hoc paired *t*-tests. Additionally, to confirm that our findings were robust to extreme values, we repeated the SCR rate analyses excluding participants with abnormally large SCR rates. Statistical analyses were conducted using JASP 0.9.

2.5. Task performance analyses

To examine the impact of acute stress on choice performance during the reinforcement-learning task, we applied a generalized linear mixed-effects (glme) model to participants' choice data (with correct and incorrect choices coded as 1 and 0, respectively). We used a "logit" link function to account for the binomial distribution of the data. As predictor variables in the glme model we included condition (stress or control), valence (gains or losses), block number (1 or 2), and trial number (1 to 40), and the interaction of interest (condition × valence; see Table S1 in the Supplementary Material for a full description of the glme model). The glme included a fixed intercept, as well as random intercepts for each participant. We fitted the glme model to the behavioral data using MATLAB's *fitglme* function and conducted post-hoc analyses via contrast matrices using the MATLAB's *coefTest* function. To assess the robustness of our findings, we also tested the significance of the interaction of interest (condition × valence) using confirmatory likelihood ratio tests (Daw, 2011) between the aforementioned full glme model and two nested models, which assumed equal performance in stress and control conditions during either gain or loss trials, through the MATLAB's *lratiotest* function. Additionally, we repeated the analyses excluding participants that performed below chance levels, which is indicative that participants did not learn to perform the task correctly and may additionally reflect non-compliance with the experimental setting.

To confirm that the choice probabilities estimated by the glme model showed a close correspondence with the actual observed choices, we used MATLAB's *predict* function. Then, we assessed the Pearson's correlation between the percentage of actual "correct" choices (i.e., choice of the stimuli associated with a probability of 0.8 of winning or a probability of 0.2 of losing) for each participant in each condition (averaged across blocks) and the percentage of "correct" choices as estimated by the glme model. We also performed confirmatory Spearman's correlations, which are more robust to outliers.

2.6. Computational modeling

2.6.1. Reinforcement-learning model

We modeled participants' trial-by-trial behavior in the stress and control conditions using a reinforcement-learning framework (Sutton & Barto, 1998) that has been extensively used to investigate the behavioral and neural impact of pharmacological manipulations and genetic variations in the dopaminergic system in humans (Diederen et al., 2017; Doll et al., 2011; Frank & Fossella, 2011; Frank et al., 2007; Grogan et al., 2017; Rutledge et al., 2009). Importantly, the fitted model included separate learning rates for positive (α^+) and negative (α^-) prediction errors, to account both for the differential firing of dopaminergic neurons for positive and negative prediction errors (Daw & Tobler, 2014; Maia & Frank, 2011; Maia & Conceição, 2017) and the differential effects of dopamine onto the plasticity of the corticostriatal synapses implicated in action-value learning (Frank & O'Reilly, 2006; Maia & Frank, 2011; Maia & Conceição, 2017; Möller & Bogacz, 2019). This model also included the inverse temperature parameter, β, which controls the stochasticity of choice selection, or the decision noise (Daw, 2011; Sutton & Barto, 1998), as detailed below.

In the context of our experimental study, the specific reinforcement-learning model used (Frank et al., 2007) assumes that each participant gradually learns the value of choosing a given stimulus (say A or B) from a given pair of stimuli (here, "gain" or "loss" stimuli pairs, as the "neutral" pair of stimuli always yielded null monetary outcomes) as a function of the outcome that was obtained on that trial following stimulus selection. Specifically, each expected pair-stimulus value, or *Q*-value, was initialized to zero, and for each trial, *t*, within that pair of stimuli, the value of the chosen stimulus (say A was chosen) was updated according to:

$$Q_A(t+1) = Q_A(t) + \alpha^* \delta(t)$$

where δ was the prediction error:

$$\delta(t) = r(t) - Q_A(t)$$

where $r(t)$ was 0.5 for gains, 0 for neutral outcomes, and −0.5 for losses. The learning rate, α, was given by:

$$\alpha = \begin{cases} \alpha^+, if \delta(t) > 0 \\ \alpha^-, if \delta(t) < 0 \end{cases},$$

where α^+ and α^- were the learning rates for positive and negative prediction errors, respectively (Frank et al., 2007).

The probability of choosing one stimulus over another (say A over B) was given by the *softmax* equation:

$$P_A(t) = \frac{e^{[Q_A(t)*\beta]}}{e^{[Q_A(t)*\beta]} + e^{[Q_B(t)*\beta]}}$$

where the β parameter, or inverse temperature, controlled the amount of decision noise. The lower the β, the higher the decision noise (i.e., decisions were made more at random).

2.6.2. Model fitting, parameter analyses, and model validation

We fitted the reinforcement-learning model to the trial-by-trial choice data from each participant in each condition. Model fitting involved estimating the values of the parameters (α^+, α^-, and β) that best accounted for the respective trial-by-trial choices in each condition.

We estimated the best-fitting model parameters (α^+, α^-, and β) for each subject in each condition using maximum *a posteriori* estimation (Daw, 2011). Specifically, to optimize model parameters, we drew the learning rates from Beta distributions [*Beta* (1.1, 1.1)] and the inverse temperature from a Gamma distribution [*Gamma* (1.2, 5)] (Palminteri, Khamassi, Joffily, & Coricelli, 2015). We then used the MATLAB's *fmincon* function, initialized at 100 random starting points of the parameter space, to search for the parameter values that minimized the negative log posterior of the observed sequence of choices, given the previously observed outcomes, with respect to different settings of the model parameters (Daw, 2011).

To assess the effects of acute stress on the parameters estimated by the reinforcement-learning model, we conducted repeated-measures ANOVAs with condition (stress and control) and prediction error valence (positive and negative) as within-subject factors, and post-hoc paired *t*-tests. ANOVAs' effect sizes are reported as eta-squared, η^2, and post-hoc paired *t*-tests effect sizes are reported as Cohen's *d* and 95% confidence intervals. We also confirmed all results from post-hoc tests using non-parametric tests (Wilcoxon signed-rank tests), which are more robust to outliers. Additionally, we repeated all analyses excluding participants that performed below chance levels to test whether the significance of the results remained unchanged. These statistical analyses were conducted using JASP 0.9.

To validate the used reinforcement-learning model, we computed trial-by-trial choice probabilities for all participants using the best-fitting set of parameters in each condition. The actual observed choices and outcomes were used to update the choice probabilities. To assess whether the choice probabilities estimated by the reinforcement-learning model (for each subject, the choice probabilities were averaged across gains or loss trials in each condition) followed the same pattern as the actual observed choices, we conducted Pearson's correlations and confirmatory Spearman's correlations for both conditions, using the respective mean percentages.

To further validate the robustness of our model-fitting procedure, we examined the capacity of recovering subject-condition-specific

parameters using simulated datasets. Specifically, we simulated the task-choice behavior of 62 virtual participants using the parameter values that we had estimated for each of the 62 participants in each condition. We ran 100 simulations. Then, for each simulation, we fitted the model to the virtual participants' data to estimate new (recovered) parameters. Finally, we tested the correlations between the original and these recovered parameters using Pearson's and confirmatory Spearman's correlations.

3. Results

3.1. Manipulation check

First, we confirmed that the acute stress manipulation successfully elicited stress responses in the participants. Self-reported stress levels differed significantly between conditions, $F(1, 61) = 107.67$, $p < .001$, $\eta^2 = 0.64$, as participants reported higher levels of stress in the stress condition ($M = 5.16$, $SEM = 0.21$) than in the control condition ($M = 3.31$, $SEM = 0.20$), $t(61) = 10.38$, $p < .001$, $d = 1.32$, 95% confidence interval (CI) = [1.49, 2.20] (Fig. 2a). SCR rate also differed significantly between conditions, $F(1, 57) = 20.61$, $p < .001$, $\eta^2 = 0.27$, being higher in the stress condition ($M = 2.89$, $SEM = 0.27$) than in the control condition ($M = 2.46$, $SEM = 0.22$), $t(57) = 4.54$, $p < .001$, $d = 0.57$, 95% CI = [0.24, 0.62] (Fig. 2b). The condition × block interactions were non-significant for the self-reported stress levels, $F(1, 61) = 0.004$, $p = .95$, $\eta^2 = 0$, and for the SCR rate, $F(1, 57) = 0.46$, $p = .50$, $\eta^2 = 0.0080$, suggesting that both the self-reported stress levels and the SCR rate remained stable across blocks within conditions. Furthermore, confirmatory non-parametric Wilcoxon signed-rank tests replicated the effects of acute stress on self-reported stress levels, when comparing the stress condition with the control condition, $Z = 6.29$, $p < .001$, and SCR rate, $Z = 4.08$, $p < .001$. Additionally, to check whether our results were robust to extreme values, we identified two participants with abnormally large SCR rate and reanalyzed the data without those participants. The significance of the results remained unchanged (main effect of condition: $F(1, 55) = 19.47$, $p < .001$, $\eta^2 = 0.26$), with higher SCR rate in the stress condition than in the control condition, $t(55) = 4.41$, $p < .001$, $d = 0.59$, 95% CI = [0.21, 0.56], even after excluding the two participants.

In sum, these results suggest that the acute stress manipulation successfully elicited stress responses.

3.2. Task performance

After confirming that self-reported stress levels and SCR rate were augmented in the stress condition, we examined the impact of acute stress on choice performance during the reinforcement-learning task (Fig. 2c) using a glme model, which accounted for the binomial distribution of the data (correct or incorrect responses) and included as predictor variables condition (stress or control), valence (gains or losses), block number (1 or 2), trial number (1 to 40), and the interaction of interest (condition × valence). We found a significant condition × valence interaction, β = -0.19, $p = .018$, 95% CI = [-0.34, −0.031] (Fig. 2d), and post-hoc analyses revealed that under stress, relative to the control condition, participants performed significantly worse when seeking monetary gains, $F(1, 19755) = 12.87$, $p < .001$, but not when avoiding losses, $F(1, 19755) = 0.14$, $p = .71$. As performance below chance levels might be indicative of non-compliance with the experimental setting, we also inspected whether each participant's behavioral performance across gains and losses was below chance levels (i.e., less than 50% of correct choices) in the stress or control conditions. One participant performed below chance levels in the stress condition, and two participants performed below chance levels in the control condition. Thus, we repeated the aforementioned analyses excluding these three participants. We found that the condition × valence interaction was still significant after excluding participants that did not learn how to perform the task, β = -0.28, $p < .001$, 95% CI = [-0.44, −0.12]. We also confirmed the robustness of our findings using likelihood ratio tests. We found that the full glme model (which assumed different performance towards gains between the stress and control conditions) had a significantly better fit than a model that assumed no differences in performance towards gains between conditions, $\chi^2(1) = 12.80$, $p < .001$, and that the same full model did not have a significantly better fit than a model that assumed no differences in performance towards losses between the stress and control conditions $\chi^2(1) = 0.20$, $p = .66$. These analyses indicate that behavioral performance towards gains, but not losses, significantly differed between the stress and control conditions.

In sum, acute stress selectively impaired choice performance towards monetary gains during the reinforcement-learning task. As an additional check, we confirmed that the choices estimated by the glme model showed close correspondence with the observed choices across trials in both conditions, Pearson's $r > 0.65$, $p < .001$, Spearman's $r > 0.52$, $p < .001$ (Fig. 2e, see Fig. S2 in the Supplementary Material for scatterplots).

Additionally, given that acute stress is thought to increase aberrant spontaneous phasic-dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Valenti et al., 2011), which, in turn, may lead to augmented learning and behavioral responding for neutral stimuli (Maia & Frank, 2017; Roiser, Howes, Chaddock, Joyce, & McGuire, 2013), we performed an exploratory analysis of participants' choices for the neutral stimuli pairs (the pairs not associated with financial outcomes; for further details see Table S2 and "Analyses of neutral trials" in the "Supplementary Analyses" section of the Supplementary Material). We found that, within the pairs of neutral stimuli, acute stress increased behavioral responding towards the stimuli that more often yielded as outcome a coin with no financial value (the high-probability "look" stimuli) relative to the stimuli that yielded no outcome at all (Fig. S1b in the Supplementary Material). This tentative finding seems consistent with the idea that acute stress might bias behavioral responding for neutral stimuli due to augmented aberrant spontaneous phasic-dopamine release.

3.3. Computational modeling

To further probe the nature of the effects of acute stress on reward and punishment learning, we fitted a biologically inspired reinforcement-learning model (Frank et al., 2007) to participants' trial-by-trial choices (see subsection 2.6.1. in "Materials and Methods" for a full description of the model). The fitted model included separate learning rates for positive (α^+) and negative (α^-) prediction errors, to account for the differential firing of dopaminergic neurons for positive and negative prediction errors (Schultz et al., 1997). The model also included the inverse temperature parameter, β, which controls the decision noise (Daw, 2011; Sutton & Barto, 1998). We then analyzed the best-fitting model parameters using ANOVAs.

These analyses revealed a main effect of condition on learning rates, $F(1, 61) = 4.69$, $p = .034$, $\eta^2 = 0.071$, but a non-significant condition × valence interaction, $F(1, 61) = 2.13$, $p = .15$, $\eta^2 = 0.034$ (Fig. 3a). For completeness, we performed paired t-tests. Despite not formally well-justified (due to the lack of a significant interaction), such paired t-tests yielded results consistent with those from task-performance analyses and those from ensuing computational modeling analyses (cf. analyses that excluded participants who performed below chance levels and analyses of α±*β). Paired t-tests revealed that α^+ was significantly lower in the stress condition ($M = 0.40$, $SEM = 0.033$) than in the control condition ($M = 0.51$, $SEM = 0.038$), $t(61) = -2.25$, $p = .028$, $d = -0.29$, 95% CI = [-0.21, −0.013], while α^- was not significantly different between the stress ($M = 0.25$, $SEM = 0.025$) and control conditions ($M = 0.27$, $SEM = 0.027$), $t(61) = -0.72$, $p = .47$, $d = -0.092$, 95% CI = [-0.098, 0.046] (Fig. 3a). Additionally, as for task-performance analyses, we repeated all statistical analyses excluding the three participants that performed below chance levels. Exclusion of those three participants revealed that the condition × valence interaction reached significance, $F(1, 58) = 4.61$, $p = .036$, $\eta^2 = 0.074$), such that α^+ was significantly

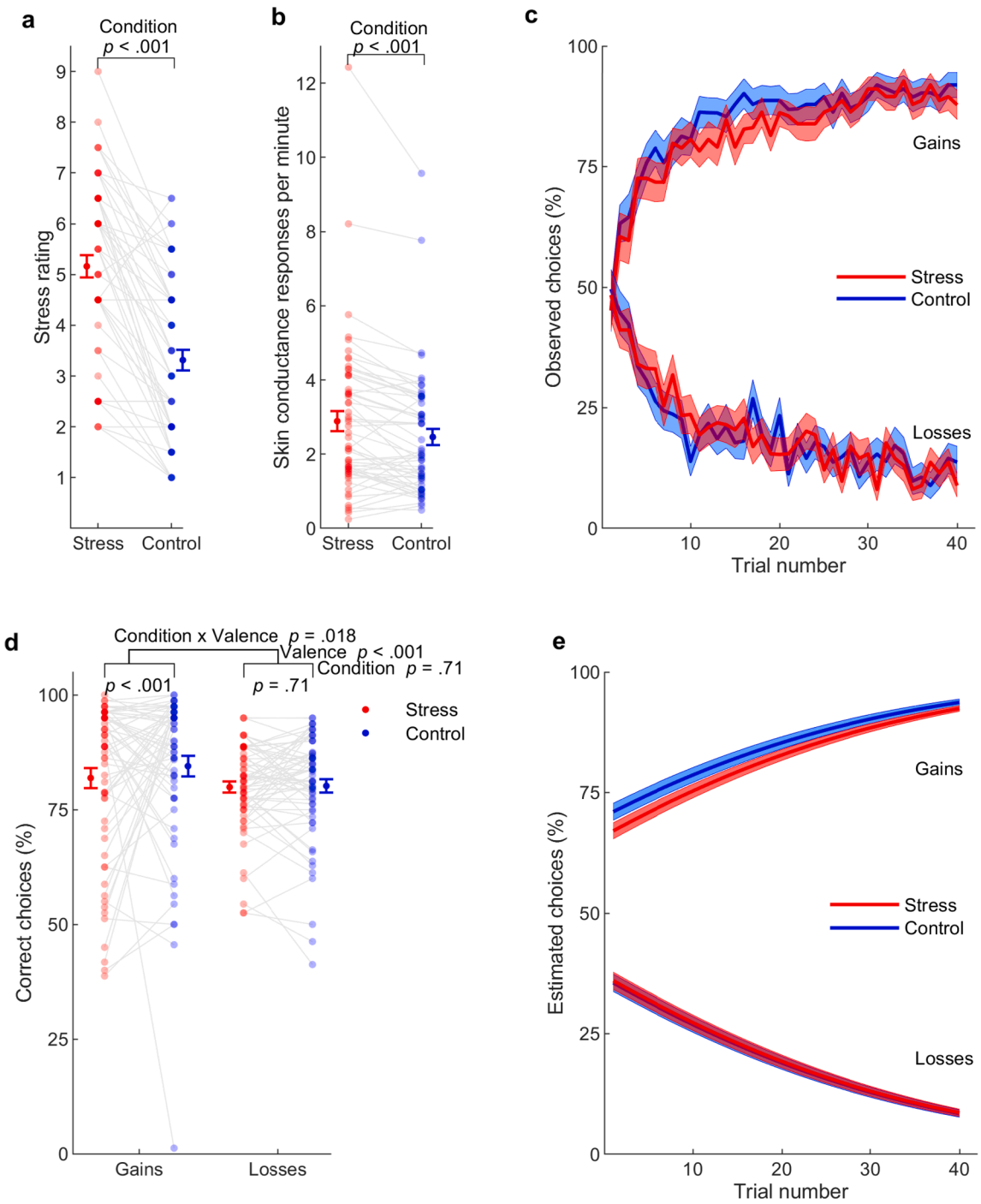


Fig. 2. Manipulation check and task performance. **(a)** Participants (n = 62) reported higher stress levels in the stress condition (red) compared with the control condition (blue). **(b)** Skin conductance response rate (n = 58) was significantly higher in the stress condition than in the control condition. **(c)** Learning curves represent the trial-by-trial percentage of participants (n = 62) who chose the "correct" gain stimulus (associated with a probability of 0.8 of winning 0.5€; upper part of the graph) and the "incorrect" loss stimulus (associated with a probability of 0.8 of losing 0.5€; lower part of the graph), in the stress and control conditions. **(d)** Participants performed significantly worse when seeking monetary gains, but not when avoiding monetary losses, in the stress condition relative to the control condition. The reported *p*-values are from the generalized linear mixed-effects model that included each participant's trial-by-trial choices and respective post-hoc tests. **(e)** The choices estimated by the generalized linear mixed-effects model captured the evolution of the actual observed choices during the reinforcement-learning task (to compare the choices estimated by the model with the actual observed choices, compare the overlap of the curves between the stress and control conditions depicted here with the overlap of the curves depicted in Fig. 2c). In panels a, b, and d, connected dots represent data points from the same participant, and more transparent (opaque) dots represent less (more) overlapping data points; the error bar displayed on the side of the scatter plots indicate the sample mean ± standard error of the mean. In panels c and e, each central line represents the mean and the filled area represents the ± standard error of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

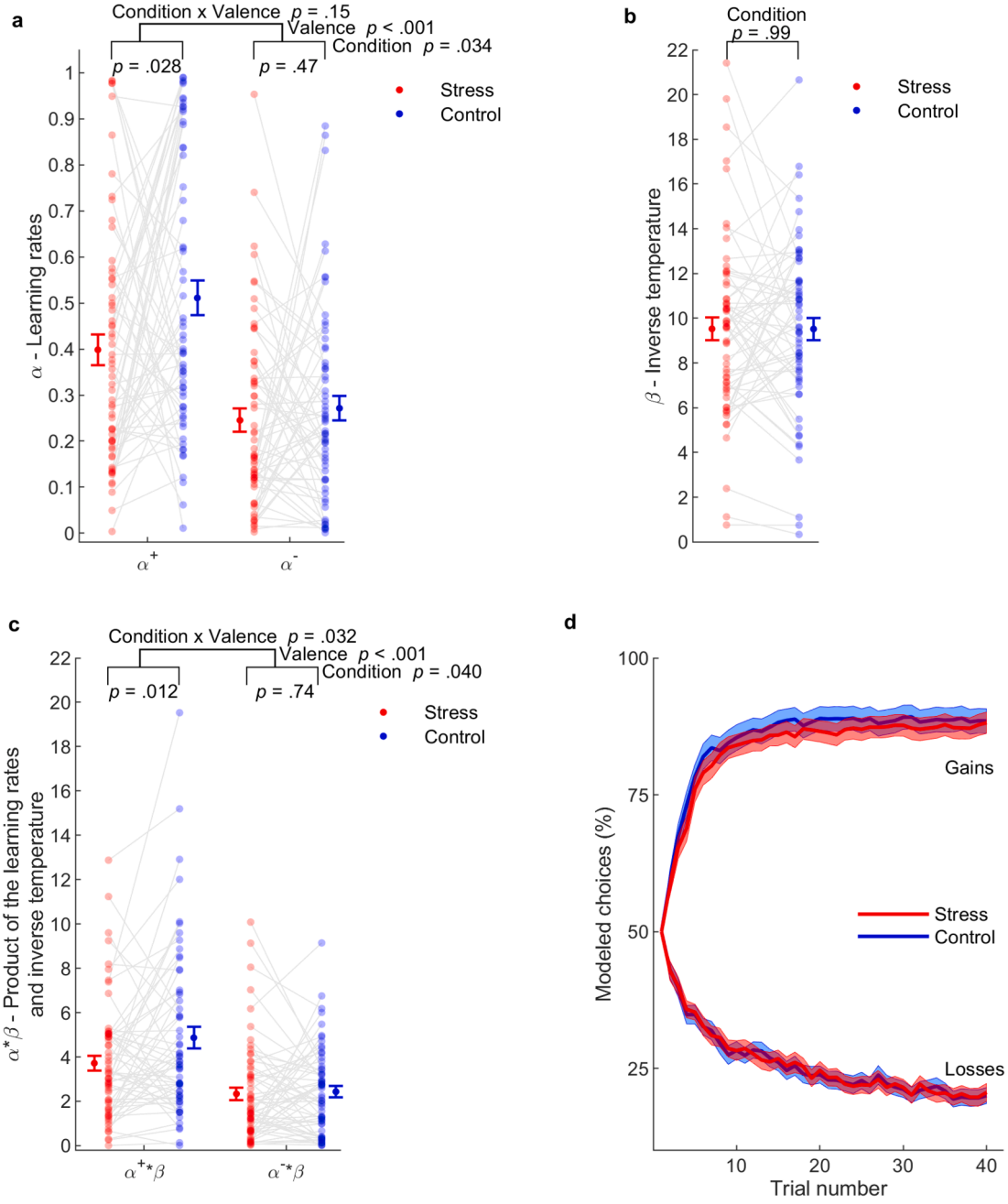


Fig. 3. Model-fitting results. **(a)** Participants (n = 62) had lower learning rates (α^+ and α^-) in the stress condition (red) compared with the control condition (blue). Paired t-tests revealed a significantly lower learning rate for positive prediction errors (α^+) in the stress condition comparatively to the control condition. **(b)** The inverse temperature (β) did not differ between conditions. **(c)** $\alpha^+*\beta$, but not $\alpha^-*\beta$, was significantly reduced in the stress condition compared with the control condition. In panels a, b and c, connected dots represent data points from the same participant, and more transparent (opaque) dots represent less (more) overlapping data points. The error bar displayed on the side of the scatter plots indicate the sample mean ± standard error of the mean. **(d)** The probabilities of choosing the "correct" gain (upper part of the graph) and the "incorrect" loss (lower part of the graph) stimuli estimated by the reinforcement-learning model followed the same pattern of the actual observed choices (compare the overlap of the curves between the stress and control conditions depicted here with the overlap of the curves that represent the actual observed choices depicted in Fig. 2c). Each central line represents the mean and the filled area represents the ± standard error of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lower in the stress condition than in the control condition, $t(58) = -2.48$, $p = .016$, $d = -0.32$, 95% CI = [-0.22, −0.024], and α^- did not differ significantly between conditions $t(58) = -0.16$, $p = .88$, $d = -0.021$, 95% CI = [-0.073, 0.063]. Thus, our findings suggest that acute stress decreases α^+.

The parameter β, which controls the amount of noise in choice selection, also did not differ significantly between the stress ($M = 9.52$, $SEM = 0.49$) and control ($M = 9.53$, $SEM = 0.51$) conditions, $t(61) = -0.016$, $p = .99$, $d = 0.0020$, 95% CI = [-0.98, 0.99] (Fig. 3b).

We further tested whether stress had a differential effect on reward compared with punishment learning by analysing the products between each learning rate ($\alpha^\pm$) and the inverse temperature (β). In reinforcement-learning models, $\alpha^\pm$ and β tend to be inversely coupled (Daw, 2011; Supplementary Material of Maia & Conceição, 2017) because $\alpha^\pm$ multiply by state-action values and the state-action values themselves are multiplied by β to compute choice probabilities (see

equations in subsection 2.6.1). As a result, the parameters viewed separately can have larger estimation errors, while their product tends to be more reliably estimated (Daw, 2011; Schonberg et al., 2007; Waltz et al., 2018; Zhang, Lengersdorff, Mikus, Gläscher, & Lamm, 2020), and thus better recovered. Note that the products $\alpha^{\pm}*\beta$ control how strongly the outcomes impact subsequent choice preferences (Schonberg et al., 2007). Statistical analyses revealed a significant condition × valence interaction, $F(1, 61) = 4.85$, $p = .032$, $\eta^2 = 0.074$, meaning that acute stress significantly decreased $\alpha^{+}*\beta$ (stress: $M = 3.72$, $SEM = 0.34$, control: $M = 4.87$, $SEM = 0.48$), $t(61) = -2.58$, $p = .012$, $d = -0.33$, 95% CI = [-2.04, −0.26], while not significantly affecting $\alpha^{-}*\beta$ (stress: $M = 2.33$, $SEM = 0.28$, control: $M = 2.44$, $SEM = 0.25$), $t(61) = -0.33$, $p = .74$, $d = -0.042$, 95% CI = [-0.71, 0.51] (Fig. 3c). Exclusion of the three participants that performed below chance levels did not change the significance of the results for the parameter β, $t(58) = 0.15$, $p = .88$, $d = 0.020$, 95% CI = [-0.94, 1.10], nor for the $\alpha^{\pm}*\beta$ condition × valence interaction, $F(1, 58) = 4.45$, $p = .039$, $\eta^2 = 0.071$. Notably, the previous results were fully replicated when comparing the estimated parameters between the stress and control conditions using non-parametric Wilcoxon signed-rank, rather than parametric, post-hoc tests (α^{+}: $Z = -2.22$, $p = .026$; α^{-}: $Z = -0.17$, $p = .86$; β: $Z = -0.11$, $p = .91$; $\alpha^{+}*\beta$: $Z = -2.19$, $p = .028$; $\alpha^{-}*\beta$: $Z = -0.35$, $p = .73$).

To validate the used reinforcement-learning model, we confirmed that the probability of choices estimated under the reinforcement-learning model had a close correspondence with the actual observed choices across trials in both conditions, Pearson's $r > 0.76$, $p < .001$, Spearman's $r \geq 0.70$, $p < .001$ (Fig. 3d, see Fig. S3 in the Supplementary Material for scatterplots). Next, to further validate that our previous model-fitting results were reliable, we conducted parameter-recovery analyses. Those parameter-recovery analyses demonstrated that the results of our model-fitting procedure were robust both in the stress and control conditions, Pearson's $r > 0.67$, $p < .001$, Spearman's $r > 0.69$, $p < .001$ (see Fig. S4 in the Supplementary Material). Additionally, Bayesian model averaging analyses (Hoeting, Madigan, Raftery, & Volinsky, 1999) using both the aforementioned, neurobiologically inspired reinforcement-learning model (with separate learning rates for positive and negative prediction errors) and a nested, alternative candidate model (with a single learning rate) provided further support for the robustness of the parameters estimated by the model with separate learning rates (see Fig. S5, Table S3 and "Model comparison and Bayesian model averaging" in the "Supplementary Analyses" section of the Supplementary Material).

Taken together, these computational findings indicate that under acute stress participants incorporated positive prediction errors at a lower rate, which seems to explain why acute stress impaired behavioral performance towards monetary gains during the reinforcement-learning task.

4. Discussion

Acute stress is present in day-to-day life, and people recurrently need to make choices and learn from the rewarding or punishing outcomes of those choices whilst under stress. In this study we investigated whether and how acute stress impacted reward and punishment learning in men using a reinforcement-learning framework. Given the putative roles of phasic-dopamine responses on prediction errors signaling and of acute stress on aberrant spontaneous phasic-dopamine release, we hypothesized that acute stress would impair reward learning and, more tentatively punishment learning. Relatedly, we hypothesized that acute stress would decrease the learning rate for positive prediction errors and, more tentatively, the learning rate for negative prediction errors. We found that acute stress impaired behavioral performance towards monetary gains, but not losses, and that this impaired performance could be explained by a decreased learning rate for positive prediction errors.

4.1. Effect of acute stress on reward learning

Our finding that acute stress impaired reward-seeking performance is consistent with several previous studies which found reduced reward responsiveness under acute stress (Bogdan & Pizzagalli, 2006; Bogdan et al., 2011; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), particularly in high stress-reactive individuals (Berghorst et al., 2013). At a first sight, however, some of the extant literature may seem equivocal, possibly due to critical methodological differences related to stress operationalization (Porcelli & Delgado, 2017). A significant number of studies have investigated the effects of stress on learning using different paradigms, such as the Trier social stress test (Boyle, Stanton, Eisenberger, Seeman, & Bower, 2019; Kruse et al., 2018; Petzold et al., 2010; Radenbach et al., 2015), or the (socially evaluated) cold pressor test (Byrne et al., 2019; Ehlers & Todd, 2017; Glienke, Wolf, & Bellebaum, 2015; Lighthall et al., 2013; Otto et al., 2013; Paul et al., 2019), in which acute stress is induced before the learning task. In these paradigms, stress induction precedes any learning processes, thus the stress-induced emotional state may be less concurrent with the cognitive processes that operate during the task. As such, it is unclear whether these paradigms probe the effects of acute stress or of recovery from stress on learning (Hermans et al., 2014). In contrast, in our study, we induced stress during the learning task, and we used relatively short task blocks to avoid the slow cortisol response from peaking and subsequent recovery within blocks. The timing of the stress induction relative to the task is an important factor to take into account when comparing studies, as early-fast catecholamine release and delayed-slow glucocorticoids actions might account for different effects on brain function (de Kloet et al., 2005; Hermans et al., 2014; Joëls & Baram, 2009). Critically, studies that induce stress before the learning task may suggest divergent behavioral results from ours (Byrne et al., 2019; Lighthall et al., 2013), whereas studies that have induced stress during the learning task (as ours) seem to be in agreement with our findings of impaired reward learning during acute-stress exposure (Berghorst et al., 2013; Bogdan & Pizzagalli, 2006; Bogdan et al., 2011; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020). Therefore, cross-study comparisons suggest that individuals may perform more poorly when learning to maximize rewards whilst exposed to an acute stressor.

In this study we further inspected the computational mechanisms behind impaired reward learning under acute stress. Using a reinforcement-learning model with separate learning rates for positive and negative prediction errors (Frank et al., 2007), we found that acute stress reduced the rate at which participants learned from positive prediction errors. A reduced learning rate for positive predictions errors means that under acute stress individuals learned more slowly about unexpected rewards and therefore took longer to adapt their behavior on the basis of the more recent rewarding outcomes of their choices. Other studies have successfully applied this reinforcement-learning model to analyze data from distinct tasks to describe the genetic (Doll et al., 2011; Frank et al., 2007) and neural (Diederen et al., 2016; Lefebvre et al., 2017; Niv, Edlund, Dayan, & Doherty, 2012) correlates of cognition and behavior. However, to the best of our knowledge, no studies had attempted to model how acute stress affects cognition, particularly reward and punishment learning, using such model. Another reinforcement-learning model commonly used in the stress literature includes distinct learning rates for gain and loss trials regardless of the valence of the prediction error (e.g., Aylward et al., 2019; Robinson et al., 2013; Treadway et al., 2017), meaning that learning from rewards in gain trials (or punishments in loss trials) is modeled the same way as learning from omission of rewards in gain trials (or omission of punishments in loss trials, respectively), which may be misleading and not be so well supported by neurobiological evidence as a model that assumes distinct learning rates based on the valence of the prediction error (as the one we used) (Maia & Frank, 2011; O'Doherty et al., 2003; Schultz et al., 1997). In simple reinforcement-learning tasks, such as ours, reward learning seems to mostly occur via positive prediction

errors (Lefebvre et al., 2017). But, given the structure of our reinforcement-learning model, stress-induced impaired reward learning could have conceivably been underpinned by a higher learning rate for negative prediction errors — which could have led to unlearning of the correct stimulus — and/or by lower inverse temperature parameter (β) — which could have been reflected in choices more at random. The fact that our data indicated that the learning rate for negative prediction errors was not higher in the stress condition and that the inverse temperature parameter did not differ between conditions, but that the learning rate for positive prediction errors was decreased in the stress condition, suggests that acute stress specifically impaired reward learning via disruption of positive prediction errors. Thus, our findings contribute to a better mechanistic understanding of how acute stress may impact reward learning.

Given that the (quantifiable) parameters from the reinforcement-learning model that we used seem to reflect specific dopaminergic-related neural mechanisms (e.g., Frank et al., 2007; Frank & O'Reilly, 2006) our computational findings may shed light on putative neural mechanisms underlying the impact of acute stress on cognition and behavior. Specifically, our computational findings seem broadly consistent with the proposed neurobiological account of dopaminergic neurons functioning under acute stress. Acute stress is thought to induce aberrant spontaneous dopamine release (Anstrom et al., 2009; Anstrom & Woodward, 2005; Cabib & Puglisi-Allegra, 2012; Valenti et al., 2011); stress-induced spontaneous dopamine release, in turn, may disrupt the adaptive striatal phasic-burst dopamine responses that signal positive prediction errors (Daberkow et al., 2013; Grace, 2016; Maia & Frank, 2017; Werlen et al., 2020), which would explain the stress-induced impairment of reward learning. While our computational results suggest that acute stress disrupts striatal responses to prediction errors during reward learning (see also Huys et al., 2013), we cannot disregard, however, that acute stress may also affect reward learning via prefrontal cortex disturbances (Arnsten, 2009; Lighthall et al., 2013; Otto et al., 2013).

4.2. Effect of acute stress on punishment learning

Acute stress impaired reward learning, but we found no evidence for an effect of acute stress on punishment learning. Computational-modeling analyses did not provide evidence that acute stress affected the learning rate for negative prediction errors to the same extent that it affected the learning rate for positive prediction errors, which might explain why behavioral performance during loss trials — which might rely more on negative prediction errors compared with performance during gain trials (Palminteri & Pessiglione, 2013) — was not impaired under acute stress.

According to a long standing influential loss aversion framework (Kahneman & Tversky, 1979), losses can have more debilitative potential than gains; therefore, as an adaptive strategy, it is possible that individuals may be more attuned to losses than to gains (Lejarraga, Hertwig, & Gonzalez, 2012; Yechiam, 2019; Yechiam & Hochman, 2013), explaining why punishment learning was spared under acute stress. However, recent evidence suggests that individuals learn gain associations better than loss associations in reinforcement-learning tasks despite symmetrical task structure and symmetrical outcome probabilities (Lin, Cabrera-Haro, & Reuter-Lorenz, 2020) (as in our task). Indeed, our behavioral data indicated that participants performed better when learning to seek gains than when learning to avoid losses in both the stress and control conditions. In addition, our computational data indicated that the learning rate for positive prediction errors was significantly higher than the learning rate for negative prediction errors, which might suggest underestimation of negative outcomes relative to positive outcomes (see also Lefebvre et al., 2017). Thus, our data suggest that it is unlikely that acute stress spared the mechanisms involved in punishment learning, but not in reward learning, due to heightened attention towards losses relative to gains.

One potential neurocognitive explanation for such lack of effect of acute stress on loss avoidance is that D2 dopamine receptors, which mediate punishment learning (Frank & O'Reilly, 2006; Maia & Frank, 2011), are already mostly activated at baseline dopamine levels, so their activation might be affected by decreases, but less so by increases, in dopamine levels (Maia & Conceição, 2017; Möller & Bogacz, 2019). Finally, non-dopaminergic mechanisms may also be involved in punishment learning (Boureau & Dayan, 2011; Moran et al., 2018), which may partially explain why previous studies using the same reinforcement-learning task also did not find significant effects of pharmacological manipulations of the dopaminergic system on punishment learning (Eisenegger et al., 2014; Pessiglione et al., 2006).

4.3. Acute stress and behavioral performance in neutral trials

We found preliminary evidence that acute stress biased behavioral responding to neutral stimuli, as the choice of the high-probability "look" stimuli — compared to the stimuli with a high probability of yielding "nothing" — was augmented under stress compared to the control condition (for a detailed discussion about this finding, see the Supplementary Material). Although very tentative, this finding might be of relevance as excessive, aberrant spontaneous dopamine release is thought to underlie increased behavioral responding and aberrant learning for neutral stimuli (Maia & Frank, 2017; Roiser et al., 2013). Still, whether acute stress promotes aberrant valuation of neutral stimuli via dopaminergic disturbances remains unknown and should be further investigated.

4.4. Limitations

In this study, we induced acute stress in participants, using a repetitive and uncontrollable sound, whilst they completed a reinforcement-learning task. We exposed participants to the sound during the task to ensure that acute stress was contingent on the learning processes, but at the expense of possibly confounding the induction of stress with distraction. To minimize this potential confound, we used a sound that was always constant and repetitive, as unpredictable changes in sound sequences induce distraction more robustly (Hughes, 2014; Parmentier, 2014; Parmentier et al., 2008; Sussman et al., 2003). Moreover, and also aiming to reduce the unpredictability of the sound, before starting the task, participants were instructed that they would hear a sound during blocks cued with a warning signal and a red background (Hughes, 2014; Parmentier & Hebrero, 2013; Sussman et al., 2003). Besides the experimental precautions we took to avoid that the stressor acted as a distractor, our data indicated a selective effect of the stress manipulation on behavioral performance during gain trials. If the sound was acting as a distractor, rather than as a stressor, it would be more likely to find instead a general behavioral impairment; that is, worse performance towards both gains and losses, rather than a selective effect on gains. Relatedly, if the manipulation had acted mostly as a distractor, it would be expected that participants behaved more at random. For example, previous data indicate that when participants are asked to generate novel combinations of numbers whilst exposed to an auditory distractor, they generate combinations more at random than participants who are not exposed to the auditory distractor, suggesting that enhanced distraction during a cognitive task may enhance the ability to generate randomness (Tsushima & Nakayama, 2010). Thus, if participants were distracted by our auditory stressor, we would have expected them to make choices more at random, which would likely be captured by a reduction of the inverse temperature parameter — reflecting increased random behavior — in the stress condition. Still, although our data and experimental design make it unlikely that the observed effect on reward learning was due to distractibility rather than stress, additional control conditions should be added in future studies to experimentally disentangle distraction from stress.

Finally, only men were included in this study, to avoid the potential

confounding effects of menstrual-cycle-dependent variation on stress responsivity (Ossewaarde et al., 2010), as well as on reward and punishment learning (Diekhof et al., 2020; Dreher et al., 2007). Our finding that acute stress disrupts reward learning in men seems to be in line with previous reports showing that acute stress disrupts reward-seeking behavior in women (Berghorst, Bogdan, Frank, & Pizzagalli, 2013; Bogdan, Perlis, Fagerness, & Pizzagalli, 2010; Bogdan, Santesso, Fagerness, Perlis, & Pizzagalli, 2011; Bogdan & Pizzagalli, 2006; Morris & Rottenberg, 2015; Paret & Bublatzky, 2020), but further studies are needed to assess whether acute stress has the same computational effects on reward and punishment learning in men and women.

5. Conclusions

We present evidence that acute stress reduces how quickly male adults integrate the unexpected rewarding outcomes of their choices over time. Our results are consistent with a neurobiological framework of stress-induced dopaminergic disturbances and can thus contribute to a better understanding of the computational mechanisms that underlie the deleterious impact of acute stress on reward learning. Ultimately, this study might offer key mechanistic insights into the impact of acute stress in everyday life.

Author contributions

J. Carvalheiro, A. Mesquita and A. Seara-Cardoso developed the study concept with critical input from V. Conceição. Testing and data collection were performed by J. Carvalheiro under the supervision of A. Mesquita and A. Seara-Cardoso. Data analyses and interpretation was carried out by J. Carvalheiro and V. Conceição. J. Carvalheiro drafted the manuscript. All authors revised and edited the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Maël Lebreton for helpful discussions. This work was supported by grants from the Portuguese Foundation for Science and Technology (FCT) to A. Seara-Cardoso [PTDC/MHC-PCN/2296/2014, co-financed by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-016747)] and to A. Mesquita (IF/00750/2015). J. Carvalheiro was supported by a FCT PhD fellowship (PD/BD/128467/2017). This study was conducted at the Psychology Research Centre (PSI/01662), School of Psychology, University of Minho, supported by FCT and the Portuguese Ministry of Science, Technology and Higher Education (UID/PSI/01662/2019), through national funds (PIDDAC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2020.105657.

References

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S., & Zigmond, M. J. (1989). Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex. *Journal of Neurochemistry, 52*, 1655–1658.

Anstrom, K. K., Miczek, K. A., & Budygin, E. A. (2009). Increased phasic dopamine signaling in the mesolimbic pathway during social defeat in rats. *Neuroscience, 161*, 3–12.

Anstrom, K. K., & Woodward, D. J. (2005). Restraint increases dopaminergic burst firing in awake rats. *Neuropsychopharmacology, 30*, 1832–1840.

Arguelles, A. E., Ibeas, D., Ottone, J. P., & Chekherdemian, M. (1962). Pituitary-adrenal stimulation by sound of different frequencies. *The Journal of Clinical Endocrinology and Metabolism, 22*, 846–852.

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience, 10*, 410–422.

Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive function in monkeys: evidence for a hyperdopaminergic mechanism. *Archives of General Psychiatry, 55*, 362–368.

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*, 1116–1123.

Belujon, P., Grace, A. A., & Grace, A. A. (2015). Regulation of dopamine system responsivity and its adaptive and pathological response to stress. *Proceedings of the Royal Society B: Biological Sciences, 282*, 20142516.

Bogdan, R., Perlis, R. H., Fagerness, J., & Pizzagalli, D. A. (2010). The impact of mineralocorticoid receptor ISO/VAL genotype (rs5522) and stress on reward learning. *Genes, Brain and Behavior, 9*, 658–667.

Berghorst, L. H., Bogdan, R., Frank, M. J., & Pizzagalli, D. A. (2013). Acute stress selectively reduces reward sensitivity. *Frontiers in Human Neuroscience, 7*, 1–15.

Bogdan, R., & Pizzagalli, D. A. (2006). Acute stress reduces reward responsiveness: Implications for depression. *Biological Psychiatry, 60*, 1147–1154.

Bogdan, R., Santesso, D. L., Fagerness, J., Perlis, R. H., & Pizzagalli, D. A. (2011). Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. *Journal of Neuroscience, 31*, 13246–13254.

Boureau, Y.-L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology, 36*, 74–97.

Boyle, C. C., Stanton, A. L., Eisenberger, N. I., Seeman, T. E., & Bower, J. E. (2019). Effects of stress-induced inflammation on reward processing in healthy young women. *Brain Behavior and Immunity, 83*, 126–134.

Breier, A., Albus, M., Pickar, D., Zahn, T. P., Wolkowitz, O. M., & Paul, S. M. (1987). Controllable and uncontrollable stress in humans: Alterations in mood and neuroendocrine and psychophysiological function. *The American Journal of Psychiatry, 144*, 1419–1425.

Byrne, K. A., Cornwall, A. C., & Worthy, D. A. (2019). Acute stress improves long-term reward maximization in decision-making under uncertainty. *Brain and Cognition, 133*, 84–93.

Cabib, S., & Puglisi-Allegra, S. (2012). The mesoaccumbens dopamine in coping with stress. *Neuroscience and Biobehavioral Reviews, 36*, 79–89.

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition, 113*, 350–364.

Collins, A. G. E., & Frank, M. J. (2013). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review, 121*, 337–366.

Daberkow, D. P., Brown, H. D., Bunner, K. D., Kraniotis, S. A., Doellman, M. A., Ragozzino, M. E., … Roitman, M. F. (2013). Amphetamine paradoxically augments exocytotic dopamine release and phasic dopamine signals. *Journal of Neuroscience, 33*, 452–463.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*, 1–26.

Daw, N. D., & Frank, M. J. (2009). Reinforcement learning and higher level cognition: Introduction to special issue. *Cognition, 113*, 259–261.

Daw, N. D., & Tobler, P. N. (2014). Value learning through reinforcement: The basics of dopamine and reinforcement learning. In P. W. Glimcher, & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 283–298). Amsterdam: Elsevier.

de Berker, A. O., Tirole, M., Rutledge, R. B., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Acute stress selectively impairs learning to act. *Scientific Reports, 6*, 29816.

de Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience, 6*, 463–475.

Diederen, K. M. J., Spencer, T., Vestergaard, M. D., Fletcher, P. C., Diederen, K. M. J., Spencer, T., … Schultz, W. (2016). Adaptive prediction error coding in the human midbrain and striatum facilitates behavioral adaptation and learning efficiency. *Neuron, 90*, 1127–1138.

Diederen, K. M. J., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. *Journal of Neuroscience, 37*, 1708–1720.

Diekhof, E. K., Korf, S., Ott, F., Schädlich, C., & Holtfrerich, S. K. C. (2020). Avoidance learning across the menstrual cycle: A conceptual replication. *Frontiers in Endocrinology, 11*, 1–12.

Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *Journal of Neuroscience, 31*, 6188–6198.

Dreher, J., Schmidt, P. J., Kohn, P., Furman, D., Rubinow, D., & Berman, K. F. (2007). Menstrual cycle phase modulates reward-related neural function in women. *Proceedings of the National Academy of Sciences, 104*, 2465–2470.

Ehlers, M. R., & Todd, R. M. (2017). Acute psychophysiological stress impairs human associative learning. *Neurobiology of Learning and Memory, 145*, 84–93.

Eisenegger, C., Naef, M., Linssen, A., Clark, L., Gandamaneni, P. K., Müller, U., & Robbins, T. W. (2014). Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology, 39*, 2366–2375.

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

Frank, M. J. (2015). Linking across levels of computation in model-based cognitive neuroscience. In *An introduction to model-based cognitive neuroscience* (pp. 159–177). New York, NY, US: Springer.

Frank, M. J., & Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology, 36*, 133–152.
Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, 104*, 16311–16316.
Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience, 120*, 497–517.
Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306*, 1940–1943.
Glass, D. C., Reim, B., & Singer, J. E. (1971). Behavioral consequences of adaptation to controllable and uncontrollable noise. *Journal of Experimental Social Psychology, 7*, 244–257.
Glienke, K., Wolf, O. T., & Bellebaum, C. (2015). The impact of stress on feedback and error processing during behavioral adaptation. *Neuropsychologia, 71*, 181–190.
Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, 108*, 15647–15654.
Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience, 17*, 524–532.
Grogan, J. P., Tsivos, D., Smith, L., Knight, B. E., Bogacz, R., Whone, A., & Coulthard, E. J. (2017). Effects of dopamine on reinforcement learning and consolidation in Parkinson's disease. *ELife, 6*, 1–23.
Hall, S., Aisbett, B., Tait, J., Turner, A., Ferguson, S., & Main, L. (2016). The acute physiological stress response to an emergency alarm and mobilization during the day and at night. *Noise and Health, 18*, 150–156.
Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences, 37*, 304–314.
Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian averaging models. *Statistical Science, 14*, 382–417.
Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal, 3*, 30–41.
Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience, 19*, 404–413.
Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders, 3*, 1–16.
Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience, 10*, 459–466.
Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: How does it work? *Trends in Cognitive Sciences, 10*, 152–158.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.
Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing, 42*, 419–427.
Korhonen, O. (1981). Firefighters' reaction to alarm, an ECG and heart rate study. *Journal of Occupational Medicine, 23*, 762–766.
Kruse, O., Tapia León, I., Stalder, T., Stark, R., & Klucken, T. (2018). Altered reward learning and hippocampal connectivity following psychosocial stress. *NeuroImage, 171*, 15–25.
Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human. Behaviour, 1*.
Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition, 124*, 334–342.
Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J., & Mather, M. (2013). Stress modulates reinforcement learning in younger and older adults. *Psychology and Aging, 28*, 35–46.
Lin, Z., Cabrera-Haro, L. E., & Reuter-Lorenz, P. A. (2020). Asymmetrical learning and memory for acquired gain versus loss associations. Cognition, 202.
Luksys, G., & Sandi, C. (2011). Neural mechanisms and computations underlying stress effects on learning and memory. *Current Opinion in Neurobiology, 21*, 502–508.
Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience, 14*, 154–162.
Maia, T. V., & Frank, M. J. (2017). An integrative perspective on the role of dopamine in schizophrenia. *Biological Psychiatry, 81*, 52–66.
Maia, T. V. (2015). Introduction to the series on computational psychiatry. *Clinical Psychological Science, 3*, 374–377.
Maia, T. V., & Conceição, V. A. (2017). The roles of phasic and tonic dopamine in tic learning and expression. *Biological Psychiatry, 82*, 401–412.
McRae, A. L., Saladin, M. E., Brady, K. T., Upadhyaya, H., Back, S. E., & Timmerman, M. A. (2006). Stress reactivity: Biological and subjective responses to the cold pressor and Trier Social stressors. *Human Psychopharmacology, 21*, 377–385.
Möller, M., & Bogacz, R. (2019). Learning the payoffs and costs of actions. *PLOS Computational Biology, 15*.
Moran, R. J., Kishida, K. T., Lohrenz, T., Saez, I., Laxton, A. W., Witcher, M. R., … Montague, P. R. (2018). The protective action encoding of serotonin transients in the human brain. *Neuropsychopharmacology, 43*, 1425–1435.
Morris, B. H., & Rottenberg, J. (2015). Heightened reward learning under stress in generalized anxiety disorder: A predictor of depression resistance? *Journal of Abnormal Psychology, 124*, 115–127.
Nagano-Saito, A., Dagher, A., Booij, L., Gravel, P., Welfeld, K., Casey, K. F., … Benkelfat, C. (2013). Stress-induced dopamine release in human medial prefrontal cortex-18F-fallypride/PET study in healthy volunteers. *Synapse, 67*, 821–830.
Nair, A., Rutledge, R. B., & Mason, L. (2020). Under the hood: Using computational psychiatry to make psychological therapies more mechanism-focused. *Frontiers in Psychiatry, 11*, 140.
Niv, Y., Edlund, J. A., Dayan, P., & Doherty, J. P. O. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience, 32*, 551–562.
O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*, 329–337.
Ossewaarde, L., Hermans, E. J., van Wingen, G. A., Kooijman, S. C., Johansson, I.-M., Bäckström, T., & Fernández, G. (2010). Neural mechanisms underlying changes in stress-sensitivity across the menstrual cycle. *Psychoneuroendocrinology, 35*, 47–55.
Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences, 110*, 20941–20946.
Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications, 6*, 8096.
Palminteri, S., & Pessiglione, M. (2013). *Reinforcement learning and Tourette syndrome, in International Review of Neurobiology* (pp. 131–153). Elsevier Inc..
Paret, C., & Bublatzky, F. (2020). Threat rapidly disrupts reward reversal learning. *Behaviour Research and Therapy, 131*, Article 103636.
Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research, 78*, 321–338.
Parmentier, F. B. R., Elford, G., Escera, C., Andrés, P., & Miguel, I. S. (2008). The cognitive locus of distraction by acoustic novelty in the cross-modal oddball task. *Cognition, 106*, 408–432.
Parmentier, F. B. R., & Hebrero, M. (2013). Cognitive control of involuntary distraction by deviant sounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1635–1641.
Paul, M., Bellebaum, C., Ghio, M., Suchan, B., & Wolf, O. T. (2019). Stress effects on learning and feedback-related neural activity depend on feedback delay. *Psychophysiology, 57*.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442*, 1042–1045.
Petzold, A., Plessow, F., Goschke, T., & Kirschbaum, C. (2010). Stress reduces use of negative feedback in a feedback-based learning task. *Behavioral Neuroscience, 124*, 248–255.
Porcelli, A. J., & Delgado, M. R. (2017). Stress and decision making: Effects on valuation, learning, and risk-taking. *Current Opinion in Behavioral Sciences, 14*, 33–39.
Pruessner, J. C., Champagne, F., Meaney, M. J., & Dagher, A. (2004). Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: A positron emission tomography study using [11C]raclopride. *Journal of Neuroscience, 24*, 2825–2831.
Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H. J., … Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology, 53*, 268–280.
Robinson, O. J., Overstreet, C., Charney, D. R., Vytal, K., & Grillon, C. (2013). Stress increases aversive prediction error signal in the ventral striatum. *Proceedings of the National Academy of Sciences, 110*, 4129–4133.
Robinson, O. J., Vytal, K., Cornwell, B. R., & Grillon, C. (2013). The impact of anxiety upon cognition: Perspectives from human threat of shock studies. *Frontiers in Human Neuroscience, 7*, 203.
Roiser, J. P., Howes, O. D., Chaddock, C. A., Joyce, E. M., & McGuire, P. (2013). Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. *Schizophrenia Bulletin, 39*, 1328–1336.
Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience, 29*, 15104–15114.
Rylander, R. (2004). Physiological aspects of noise-induced stress and annoyance. *Journal of Sound and Vibration, 277*, 471–478.
Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews, 21*, 55–89.
Schonberg, T., Daw, N. D., Joel, D., O'Doherty, J. P., Doherty, J. P. O., Schonberg, T., … O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27*, 12860–12867.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593–1599.
Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology, 33*, 890–895.
Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience and Biobehavioral Reviews, 36*, 1740–1749.
Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience, 29*, 7191–7198.
Sulzer, D., Cragg, S. J., & Rice, M. E. (2016). Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia, 6*, 123–148.
Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin and Review, 10*, 630–637.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Treadway, M. T., Admon, R., Arulpragasam, A. R., Mehta, M., Douglas, S., Vitaliano, G., … Pizzagalli, D. A. (2017). Association between interleukin-6 and striatal prediction-error signals following acute stress in healthy female participants. *Biological Psychiatry, 82*, 1–8.

Tsushima, Y., & Nakayama, K. (2010). Does an auditory distractor allow humans to behave more randomly? *Journal of Vision, 10*, 1028.

Valenti, O., Lodge, D. J., & Grace, A. A. (2011). Aversive stimuli alter ventral tegmental area dopamine neuron activity via a common action in the ventral hippocampus. *Journal of Neuroscience, 31*, 4280–4289.

Waltz, J. A., Xu, Z., Brown, E. C., Ruiz, R. R., Frank, M. J., & Gold, J. M. (2018). Motivational deficits in schizophrenia are associated with reduced differentiation between gain and loss-avoidance feedback in the striatum. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*, 239–247.

Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience, 40*, 373–394.

Werlen, E., Shin, S. L., Gastambide, F., Francois, J., Tricklebank, M. D., Marston, H. M., … Walton, M. E. (2020). Amphetamine disrupts haemodynamic correlates of prediction errors in nucleus accumbens and orbitofrontal cortex. *Neuropsychopharmacology, 45*, 793–803.

Westman, J. C., & Walters, J. R. (1981). Noise and stress: A comprehensive approach. *Environmental Health Perspectives, 41*, 291–309.

Yechiam, E. (2019). Acceptable losses: The debatable origins of loss aversion. *Psychological Research, 83*, 1327–1339.

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin, 139*, 497–518.

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience, nsaa089*.