Observational fear learning: From mouse to man

Shana Elyse Silverstein

Division of Psychology and Language Sciences
University College London

National Institutes of Health

Prepared under the supervision of Professor Essi Viding, Doctor Andrew Holmes, & Professor Jonathan Roiser

A thesis submitted to University College London for the degree of
Doctor of Philosophy
2021

Declaration

I, Shana Silverstein, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed,
Shana Silverstein

Abstract

Observational fear learning (OFL) is a means of conveying threatening information to another and demonstrates how social organisms learn from environmental interactions to promote safety without exposure to harm. This thesis was conducted with the aim of advancing understanding of the mechanisms underlying OFL in a behaviourally translational manner in mice and humans.

In Chapter 2, development and characterization of a cued-OFL task for mice is described. Mice form a robust and lasting stimulus specific fear memory through observation of a distressed conspecific that does not produce a phenotype of generalized anxiety nor alterations in socialization. In Chapter 3, anatomical interrogation of cortico-limbic-mid-brain regions and pathways of importance for direct fear learning (DFL) and social learning are identified as potential components of the neural network subserving OFL. In Chapter 4, the functional circuitry of OFL is studied using in vivo optogenetics and Ca^{2+} imaging via fibre photometry to reveal a causal role of the ventral hippocampus (vHPC) in constraining prelimbic (PL) projections to the lateral/ventrolateral periaqueductal grey (l/vlPAG) to modulate OFL.

Study 1 of Chapter 5 addresses the development and validation of an OFL task in humans using reinforcement learning framework and computational modelling. OFL was best characterized by a single learning rate for both high- and low-shock associated stimuli (CS). Moreover, model parameters described participant OFL as not heavily influenced by recent prediction errors and valuation of the CSs was updated relatively slowly. Study 2 of Chapter 5 investigates the relationship of dispositional traits commonly associated with fear, namely anxiety and psychopathy, on individual differences in OFL. Besides a modest association between trait anxiety and prediction response time, limitations with the OFL task and deployment on an online platform were suboptimal for engaging these relationships. Together, these findings provide a complementary approach to characterize OFL in mice and humans.

Impact statement

The ability to learn from our environment through social transmission is a highly conserved mechanism found across species. Social animals derive information relating to threat and safety through interactions with others. One example of this is observational fear learning (OFL), a form of social fear transmission occurring through watching another in distress. Current diagnostic criteria for Trauma- and Stressor-Related Disorders state that causal traumatic event(s) can be witnessed (DSM-5, 2013), acknowledging that observation of harrowing occurrences poses a serious health risk when experienced at pathological levels. Moreover, OFL can be processed and experienced in diverse ways that may provide critical information for clinical disorders, such as post-traumatic stress or generalized anxiety, or conversely, personality disorders, like psychopathy.

Despite the ecological relevance and importance of associative learning in a social context, research on fear learning has largely been in isolation from social influences. Current literature on OFL is comparatively recent and relatively sparse, with little known about the mechanisms and conditions that contribute towards this distinct form of learning. The work undertaken in this thesis sought to establish and characterize cued-OFL in mice and humans in order to define the underlying neural circuitry and investigate the impact of dispositional traits on OFL variability.

In mice, a combination of anatomical tracing with in vivo optogenetics and calcium imaging technologies were used to define a novel disynaptic circuit subserving OFL. An interacting network of brain regions including the ventral hippocampus (vHPC), prelimbic cortex (PL), and lateral/ventrolateral periaqueductal grey (l/vlPAG) was found to calibrate OFL. vHPC constraint of PL outputs to PAG was determined to provide a safety signal during OFL which, when compromised, amplifies observed fear. This neural network incorporates circuits previously demonstrated in direct fear or social learning, but never together.

Concurrently, a complementary OFL task for humans was developed using a reinforcement learning framework to model trial-by-trial variation in learning. Computational modelling of learning, supported by Bayesian model comparisons,

characterized OFL as incremental and valuation of stimuli was slowly updated in response to prediction errors. Moreover, the role of dispositional traits on individual differences of OFL revealed that trait anxiety was modestly associated with slower response time; however no other correlations were found between anxious or psychopathic traits and measures of OFL. While the results were largely non-significant, addressing the association of trait factors that might be informative about differences in OFL had not previously been investigated. Moreover, it revealed crucial limitations of the task and testing platform and, in comparison with the mouse research, a need for a more ecologically relevant paradigm.

To date, research on OFL has largely been developed and conducted in a siloed manner between research fields and animal models. As demonstrated within this thesis, however, OFL is a complex behaviour integrating components of fear and social learning with the contribution of individual differences in disposition and learning. A collaborative biopsychosocial approach is required in order to characterize OFL and address the clinical implications for the development of fear related pathologies.

Acknowledgements

This thesis is the result of a cumulative effort by some of the most committed, supportive, and kind individuals on this globe, without whom, none of this could have been possible.

First and foremost, I want to thank my supervisors Andrew, Essi, and Jon for the opportunity to pursue and complete my doctoral work. There are not enough words to express my gratitude to each of you for your constant support and patience through all of the peaks and valleys over the last six years. I lucked out beyond anything I could have imagined with you three – your uniqueness has enriched my scientific-growth, and more importantly, my life, in tremendous ways.

I want to thank Andrew for taking a chance on me. You saw some potential (or maybe madness) in me and held me to that high standard. I would never have known the extent of my capabilities as a scientist or human had it not been for you. Thank you for showing me how to think about science, what questions to ask, and how to continue to persevere in the face of adversity. The grey hairs and years off my life seem much more worthwhile from this side of the journey.

I would like to thank Essi for being my cheerleader and bestowing upon me endless encouragement. Your ability to bring a little perspective to moments where I teetered on the edge of sanity and perfectionism is something I will hold tightly to in challenging times. Thank you for all of your time spent reading endless drafts of my writing and dragging me through to the finish line. I am so fortunate to have such a brilliant and fierce female role-model to look up to in science and in life (and fashion).

I would like to thank Jon for being convinced that my ambitious (verging on unrealistic) research proposal was worth taking a chance on. Your patience and kindness to sit with me for weeks, holding my hand through every line of MATLAB code and data analysis, means more than I can express. You opened up an entirely new world of science for me and supported me through all phases of my understanding and I am forever grateful. תודה רבה

To the Holmes lab, past and present, thank you for being my science home base, no matter where I go. I would like to give particular thanks to Olena, who took me on as a post-bac seven years ago. I would not be a scientist without your patience and training. Ozge, for always lending support and oxytocin. Ayesha, for being my grad school soul sister and forever conference roommate. Katie, Adrina, Abby, and Aaron for making the lab a much better place because of your presence. The success and

advancement of the OFL project in my absence would not have been possible if not for Takayuki, Mio, Julia, and Leo. You have all my gratitude for helping this project to blossom.

I would like to thank the members of the DRRU for welcoming me with open arms and becoming my London family. There is no greater joy than a sunny Thursday trip to the farmers market for Pasta e Basta or a perfectly timed visit to Homeslice. I want to especially thank Ruth for being my spirit animal and soul-person and Christina for being the best work wife. I am extremely grateful for also want to thank the members of the Neuroscience and Mental Health group for coming to my rescue and becoming my friends. Alex, I am forever grateful for your support, patience, and friendship during the scanning study, and to Vincent, for being truly brilliant with all things modelling and so generous with your time.

To all of my mice, thank you for your contribution to science.

To my beloved friends and family, thank you for being there with me through this process, providing unconditional love and support, forgiving my absences, and coming to visit regardless of where I am. To Julie, for holding space for all of me. Thank you to my in-laws, Debbie and Tom, who have been such an unconditional and steady force in my life. My parents, Alan and Leza, for never putting limitations on what I could achieve and always reminding me of my true self.

To Doug, thank you for giving me space to soar and for being the safest place to land. This doctorate has seen us through our engagement, getting married, finishing law school, two years of a transatlantic relationship, a global pandemic, and expanding our family to include Marcel and Poppy – if we could make it through all of that, we can do anything together.

Lastly, to my Poppyseed – you are the most exceptional creature I have ever known. In one unimaginable year you have redefined my understanding of strength and resilience. Thank you for showing me that anything is possible. I am forever grateful to be your mommy.

Table of contents

Declaration 1

Abstract 2

Impact statement 3

Acknowledgements 5

Table of contents 7

List of tables 12

List of figures 13

Key abbreviations 14

Chapter 1 General Introduction 16

1.1 Social fear learning 16

1.1.1 Paradigms for studying social fear learning 17

1.1.2 Introduction to observational fear learning 19

1.2 Experimental OFL in rodents 20

1.2.1 The development of OFL tasks in mice 20

1.2.2 Behavioural and genetic considerations impacting OFL 21

1.2.3 Neurocircuitry underlying OFL 26

1.3 Experimental OFL in humans 32

1.3.1 OFL in humans 33

1.3.2 Physiological and psychological factors impacting OFL in humans 34

1.3.3 Dispositional trait considerations impacting OFL 37

1.3.4 Further development of OFL in humans 38

1.4 Thesis research aims 39

1.5 Dissemination 41
Chapter 2 Observational fear learning behaviour in mice 42
2.1 Chapter introduction 42
2.1.1 Observational fear learning 42
2.1.2 Behavioural considerations of OFL 43
2.1.3 OFL paradigm development, characterization, and validation 45
2.2 Materials and Methods 46
2.2.1 Subjects 46
2.2.2 Cued-OFL paradigm 46
2.2.3 Task validation and characterization 48
2.2.4 Anxiety-like behaviour 50
2.2.5 Sociability 50
2.2.6 Data analyses 51
2.3 Results 51
2.3.1 Mice learn a cued-shock association through observation 51
2.3.2 Cued-OFL does not affect anxiety-like behaviour or sociability 56
2.4 Discussion 58
Chapter 3 Anatomical circuitry underlying observational fear learning 61
3.1 Chapter Introduction 61
3.1.1 Role of the PL in fear learning 61
3.1.2 Inputs to the PL involved in DFL and social learning 63
3.1.3 Outputs from the PL involved in fear expression 65
3.1.4 Role of PL in a broader anatomical circuit underlying OFL 66
3.2 Materials and Methods 66
3.2.1 Subjects 66
3.2.2 Regional activation mapping using c-fos immunohistochemistry 67

3.2.3 Input-specific PL c-fos quantification 68

3.2.4 Tracing vHPC monosynaptic inputs to PL neuronal subtypes via immunohistochemistry 69

3.2.5 Intersectional viral tracing of monosynaptic inputs to PL neuronal subtypes 70

3.2.6 Anterograde tracing of vHPC inputs to PAG-projecting PL neurons 70

3.2.7 Data analyses 71

3.3 Results 71

3.3.1 OFL differentially activates key regions of emotion regulation ... 71

3.3.2 Cued-OFL recruits vHPC→PL and BA→PL pathways similarly to DFL 74

3.3.3 vHPC inputs to PL selectively target glutamatergic neurons and PV INs 75

3.3.4 vHPC→PL outputs preferentially innervate several regions involved in emotion processing 75

3.4 Discussion 78

Chapter 4 Functional circuitry subserving observational fear learning 81

4.1 Chapter Introduction 81

4.1.1 Technologies used to explore an OFL functional circuitry 81

4.1.2 Brief summary of the proposed OFL circuit 83

4.2 Materials and Methods 84

4.2.1 Subjects 84

4.2.2 In vivo PL photosilencing during conditioning 84

4.2.3 In vivo PL photosilencing during retrieval 86

4.2.4 In vivo PL PV IN photosilencing 86

4.2.5 In vivo vHPC→PL photosilencing 86

4.2.6 In vivo BA→PL photosilencing 87

4.2.7 In vivo vHPC→PL Ca^{2+} imaging via fibre photometry..... 87

4.2.8 In vivo PL→l/vlPAG Ca^{2+} imaging via fibre photometry and concurring vHPC→PL photosilencing with in vivo PL→l/vlPAG Ca^{2+} imaging ... 88

4.2.9 Statistical analysis 90

4.3 Results 91

4.3.1 Cued-OFL recruits and requires PL 91

4.3.2 PV INs in PL are involved in the consolidation of OFL 93

4.3.3 vHPC inputs to PL constrain OFL..... 94

4.3.4 Photosilencing BA inputs to PL does not have a lasting impact on OFL 95

4.3.5 vHPC→PL Ca^{2+} signals during retrieval negatively correlate with cued-OFL 96

4.3.6 Ca^{2+} imaging shows OFL is signalled by PL projections to l/vlPAG 98

4.3.7 OFL promotes vHPC inhibition of PL→PAG pathway 99

4.4 Discussion 101

4.4.1 PL recruitment to integrate threat and social information in OFL... 102

4.4.2 PL inputs from vHPC restrict OFL..... 103

4.4.3 vHPC inhibition of PL→l/vlPAG pathway in OFL..... 104

4.4.4 Conclusions 104

Chapter 5 Individual differences contributing to observational fear learning.. 106

5.1 Chapter Introduction 106

5.1.1 OFL in humans 106

5.2 Study 1: Paradigm validation of a new OFL task 108

5.3 Materials and Methods 110
5.3.1 Participants 110
5.3.2 Materials 110
5.3.3 Procedure 112
5.3.4 Analyses 113
5.4 Results 115
5.4.1 Characterization of OFL task to medium shock reaction 115
5.4.2 Characterization of OFL task with high shock expression 116
5.4.3 Behavioural comparison between medium and high intensity reaction stimuli 117
5.4.4 Behavioural data: differences between high and low probability stimuli 117
5.4.5 Cognitive ability is not associated with OFL 118
5.5 Discussion 119
5.6 Study 2: Study of dispositional traits and their association with individual differences in OFL 123
5.6.1 The relationship between dispositional traits and learning 123
5.6.2 Dispositional traits in the context of OFL 126
5.7 Materials and Methods 127
5.7.1 MTurk participants 127
5.7.2 Materials 128
5.7.3 Procedure 129
5.7.4 Computational model 130
5.7.5 Analyses 130
5.8 Results 131
5.8.1 Characterization of OFL 131

5.8.2 Associations between dispositional traits and OFL 132
5.9 Discussion 137
Chapter 6 General Discussion 141
6.1 Summary of findings 141
6.2 Synthesis, limitations, and future directions 144
6.2.1 How the current findings contribute to our understanding of OFL in mice 144
6.2.2 How the current findings contribute to our understanding of OFL in humans 158
6.2.3 Limitations and considerations of OFL as a behaviourally translational research task 163
6.3 Conclusion 164
References 167

List of figures

Figure 2.1. Example of OFL task setup 48
Figure 2.2. Behavioural tests of OFL 52
Figure 2.3. Behavioural ethogram of observers and demonstrators 53
Figure 2.4. Role of obscuring US during OFL 54
Figure 2.5. Fear generalization/sensitization following OFL 55
Figure 2.6. Presence of an observer on DFL 56
Figure 2.7. OFL effects on subsequent anxiety-like behaviour and sociability 57
Figure 3.1. C-fos activity in response to OFL conditioning 73
Figure 3.2. Pathway specific cellular activation of CTb inputs to PL 74
Figure 3.3. Visualization of cell type specific inputs in PL from the vHPC 77
Figure 3.4. PL outputs from vHPC driven pathway 78
Figure 4.1. PL photosilencing during OFL 92
Figure 4.2. PV IN in PL silencing during OFL conditioning 94
Figure 4.3. Silencing vHPC inputs to PL during OFL 95
Figure 4.4. Silencing BA inputs to PL during OFL 96
Figure 4.5. Ca^{2+} imaging of vHPC inputs to PL during retrieval 97
Figure 4.6. Ca^{2+} imaging of PL output to l/vlPAG during OFL conditioning ... 100
Figure 4.7. Concurrent silencing vHPC→PL while recording from PL→l/vlPAG 101
Figure 5.1. Example stimuli 111
Figure 5.2. Comparison of medium versus high intensity shock response ... 116
Figure 5.3. High- versus low-probability stimuli 118
Figure 5.4. OFL behavioural characterization 131
Figure 5.5. Computational model 132
Figure 5.6 Correlations between trait anxiety and OFL 134

List of tables

Table 5.1. Participants' descriptive statistics 134
Table 5.2 Correlations between all measures 136

Key abbreviations

AAV	Adeno-associated virus
ACC	Anterior cingulate cortex
AI	Anterior insula
AMG	Amygdala
BA	Basal amygdala
BLA	Basolateral amygdala
Ca2+	Calcium
CLA	Claustrum
CS	Conditioned stimulus
CTb	Cholera toxin subunit B
DFL	Direct fear learning
dmPFC	Dorsomedial prefrontal cortex
dPAG	Dorsal periaqueductal grey
HPC	Hippocampus
IL	Infralimbic cortex
IN	Interneuron
LA	Lateral amygdala
l/vlPAG	Lateral/ventrolateral periaqueductal grey
mPFC	Medial prefrontal cortex
NAc	Nucleus accumbens
NAs	Nucleus accumbens shell
OFL	Observational fear learning

pACC	Posterior anterior cingulate cortex
PAG	Periaqueductal grey
PB	Phosphate buffer
PBS	Phosphate buffered saline
PFA	Paraformaldehyde
PL	Prelimbic cortex
PN	Pyramidal neuron
PV	Parvalbumin
SCR	Skin conductance response
Sst	Somatostatin
STAI	State-Trait Anxiety Inventory
US	Unconditioned stimulus
Vglut1	Vesicular glutamate transporter 1
vHPC	Ventral hippocampus
Vip	Vasoactive intestinal peptide
vlPAG	Ventrolateral periaqueductal grey

Chapter 1 General Introduction

1.1 Social fear learning

How we learn from our environment is largely a combination of our own direct experiences with the world around us and learning through the interactions and experiences of others (Cook & Mineka, 1990; Kim, Keum, & Shin, 2019; Mineka & Öhman, 2002; Olsson, Knapska, & Lindström, 2020). As a social species, the social transmission of information is vitally important, as we can quickly learn about safety and security without having to put ourselves directly at risk (Debiec & Olsson, 2017; Keum & Shin, 2019; Kim et al., 2019; Olsson et al., 2020; Szczepanik et al., 2020). The responses to dangerous circumstances we are faced with can be innate or learned, driving a behavioural response to adapt, adjust, and accommodate to the constantly changing uncertainty around us.

Fear learning in the presence of a threat is a highly conserved response across species and most commonly studied experimentally through Pavlovian fear conditioning (Debiec & Olsson, 2017; Keum & Shin, 2019; LeDoux, 2000; Olsson et al., 2020). Fear conditioning is a form of associative learning, which pairs a neutral conditioned stimulus (CS; e.g., a white noise tone, a light) with an aversive unconditioned stimulus (US; e.g., an electrical shock, an air puff). This learned association between a CS-US pairing can subsequently trigger a fear response to the CS in the absence of US exposure. Fear learning from this direct experience often persists far beyond the initial event, demonstrating that it is a crucial mechanism for defence and survival by associating environmental cues with potential harm (Fanselow, 1994; Maren, 2001). Moreover, it combines a unique interface for studying memory and emotionality, and has given rise to early studies on neural systems and cognition (LeDoux, 2000).

Exposure to social cues lends itself to learning about environmental outcomes by integrating information from others' behaviour to update understanding of the world. Whilst much of the literature has focused on fear learning in isolation from social influences, more recent research has turned to forms of fear learning that rely

on social interactions to transmit information related to threat and safety. This is in part due to a change in the current diagnostic criteria for Trauma- and Stressor-Related Disorders, which now states that causal traumatic event(s) can either be directly experienced *or* witnessed (American Psychiatric Association, 2013). With this crucial inclusion, the DSM-5 acknowledges that the vicarious transmission of distressing information poses a serious health risk when experienced at pathological levels. Moreover, fear learning through social mechanisms can be processed and experienced in diverse ways that may provide critical information for clinical disorders, such as post-traumatic stress or generalized anxiety, and on the opposite end of the spectrum, callous unemotional traits or psychopathy.

1.1.1 Paradigms for studying social fear learning

While classical fear conditioning requires a subject to directly experience a US, social forms of fear learning use the transmission of another's fearful response to a CS as a means of learning. Rachman (1977) suggested three ways in which fear can be acquired: direct experience, indirectly through exposure, or indirectly through instruction. For the two indirect forms of fear learning, indirect exposure provides the most translational approach to study social fear learning across species. Indirect fear learning relies on a host of sensory cues to convey information, such as through visual observation, olfactory cues, and auditory stimuli. From this, various behavioural tasks across species and throughout the lifespan have been established to assay social fear learning.

An early experimental study on indirect learning showed that a rat trained to lever press for a food reward would stop pressing when observing that this action caused another rat to receive a footshock, thereby illustrating the possible transmission of affective states in rodents (Church, 1959). Since then, a number of paradigms have demonstrated different ways in which rodents acquire and express fear learning by proxy. Kavaliers and colleagues (2003) used a paradigm in which naïve mice observed a conspecific subjected to biting-flies. Without direct experience of the biting-flies, the naïve mice responded to non-biting flies with the same conditioned response that they had observed. Another paradigm reported by

Knapska and colleagues (2010) demonstrated that mice interacting with a fear-conditioned cagemate subsequently showed a conditioned fear response without directly experiencing fear conditioning. Similarly, rats observing a conspecific's conditioned fear response to a CS exhibited fearful behaviour to a cue as well (Jones, Agee, & Monfils, 2018). Moreover, social fear learning can be disrupted by irregularities in social development, such as an impairment in observational fear conditioning due to social isolation following weaning (Yusufishaq & Rosenkranz, 2013).

Not surprisingly, studies in non-human primates demonstrate a similar ability to socially acquire fear learning. For example, Cook and Mineka (1990) demonstrated that captive rhesus monkeys learned to fear snakes (which they had never personally encountered) through watching videos of other monkeys' fearful behaviours in response to snakes. Similarly, ververt monkeys demonstrated food avoidance after seeing other vervet monkeys' averse response to bitter foods (Van De Waal, Borgeaud, & Whiten, 2013).

Finally, social fear learning is a highly conserved mechanism as evidenced by research in humans. From a very early age there are examples of social fear learning, such as the transmission of fear and avoidance through maternal modelling (Gerull & Rapee, 2002), or through early associations between certain animals or objects and others' facial expressions (Askew & Field, 2007). This type of learning persists well beyond childhood, as Olsson and Phelps (2004) demonstrated using a paradigm in which people learn a cued-fear association from observing another in pain. Additionally, the rise of social media has increased people's access to real displays of violence and harm, and exposure to these images and videos is associated with a rise of fear-induced psychological disorders (Hopwood & Schutte, 2017); although the direction of causation between exposure and clinical diagnoses is debatable. These studies, as well as many other recent paradigms assaying social fear learning, demonstrate how exposure to social cues signalling danger are a powerful means of conveying information about threat and safety (Debiec & Olsson, 2017; Olsson et al., 2020). While social fear learning can be conveyed in countless ways, the scope of this dissertation focuses specifically on social fear learning through observation because

of the translational potential to use this approach experimentally in both mice and humans.

1.1.2 Introduction to observational fear learning

Observational fear learning (OFL) is a means of social fear transmission through visual perception of another in distress. This is a robust means of vicarious learning in both rodents and humans. Additionally, the OFL model provides critical insight into how social organisms learn from interactions about the environment around them.

An OFL paradigm was first developed for human participants by Olsson & Phelps (2004) whereby an individual learns an association between a CS (a blue square) and a US (electrical shock given to an actor) on a video screen. Unlike direct fear learning (DFL), which requires the participant to directly experience the aversive US, they instead experience the US vicariously through observing another's discomfort and/or distressed reaction to a US. OFL is commonly measured through the participant's skin conductance response (SCR), a biological measure of electrical conductance driven by sweat secretion in response to arousing stimuli (Lonsdorf et al., 2017); however, more recently, behavioural and neural measurements have been implemented to measure OFL (Haaker, Golkar, Selbing, & Olsson, 2017).

Following the institution of Olsson and Phelps's (2004) OFL task in humans, Jeon and Shin (2011) developed a similar behavioural protocol for mice. In their paradigm, two mice are placed within a conditioning chamber – the 'demonstrator' is placed on an exposed metal grid floor to directly receive footshocks (US), while the 'observer' is separated by a transparent divider and is safe from the aversive stimuli. The 'demonstrator' is repeatedly shocked while the 'observer' looks on, without directly experiencing the shock themself. OFL is tested the following day by returning the observer on their own to the conditioning context and quantifying freezing behaviour, a standard measure of conditioned fear response in rodents (Fanselow, 1980). This original contextual-OFL paradigm has since been modified in various ways including the addition of a cue or pre-training the observer to the US (Allsop et al., 2018).

Drawing upon these well-established paradigms to study OFL from mouse to man, I will now explore in depth what has been established from prior research. I will first discuss OFL in rodents, delving into the behavioural and genetic considerations contributing to learning and memory as well as what is known so far about the underlying neural circuitry. Additionally, I will briefly touch upon what the extensive research on DFL can contribute to directing subsequent interrogation into the mechanisms of OFL. Following this, I will turn towards the human research on OFL by discussing factors that might provide clearer insight into how this form of learning occurs as well as the potential influence of dispositional traits in both non-clinical and pathological circumstances of OFL.

1.2 Experimental OFL in rodents

As already discussed, OFL is a robust behavioural assay established to study how information about threat can be transmitted through vicariously. As introduced earlier, social transmission of fear has been demonstrated in a variety of ways experimentally, mostly through fear-by-proxy paradigms in which a rodent is exposed to a previously conditioned cagemate and learns a conditioned fear response without ever having direct US exposure (Burgos-Robles et al., 2019; Debiec & Olsson, 2017; Keum & Shin, 2016; Kim et al., 2019; Kondrakiewicz et al., 2019; Olsson et al., 2020). These studies provide compelling evidence that mice have the ability to convey socioenvironmental information to conspecifics; however, the tasks used to study this vary tremendously. It is therefore necessary to hone in on a specific aspect of social fear learning in order to effectively study the underlying mechanisms in a standardized and reproducible way so as to build upon each other's findings to advance our collective understanding. For this, I will focus on the specific role of observation in transmitting fear learning.

1.2.1 The development of OFL tasks in mice

Until relatively recently, there had not been a behavioural paradigm designed to study OFL specifically in rodents. Drawing upon the task designed by Olsson and

Phelps (2004) in humans, Jeon and colleagues (2010, 2011) characterized a contextual behavioural assay where fear learning can occur solely through observation. Similar to contextual-DFL, the original OFL paradigm used aversive footshocks (US) paired with a previously neutral context (CS) to form a fearful association in mice. However, unlike DFL, the footshock was always delivered to a 'demonstrator,' whilst a second mouse, the 'observer,' only viewed the shock delivery, developing a vicarious conditioned fear response through observation. On the OFL conditioning day, the 'demonstrator' was placed on one side of the conditioning chamber, exposed to a metal grid floor capable of delivering footshocks, while the 'observer' was placed on the opposite side of the chamber within a clear plastic container in order to keep the mice separate and to protect the observer from shock delivery. Following a brief habituation phase, the demonstrator received repeated 2-s 1-mA footshocks every 10 s for 4 minutes. The next day OFL is tested by returning the observer alone to the conditioning chamber for several minutes to retrieve the fear memory (measured by quantifying freezing behaviour).

Studies using the OFL paradigm typically find that conditioned responses acquired through observation tend to be lesser than those acquired through a similar level of conditioning involving direct experience with a shock (Allsop et al., 2018; Jeon et al., 2010). Lower conditioned fear response in OFL could be explained by lesser intensity of the variable(s) serving as the US in OFL (the reaction of the demonstrator to shock), as compared to direct fear learning (the experience of the pain of a shock). Additionally, neural systems may be recruited to calibrate fear to a level appropriate of a threat that is experienced vicariously but not directly, resulting in a constraint on OFL. Such internal modulations might have an adaptive value by preventing responses to witnessed threats that exceed the level of threat that is present.

1.2.2 Behavioural and genetic considerations impacting OFL

Whilst the study of OFL is still relatively new, some progress has been made on how OFL is transmitted and retained in rodents. However, there are several factors to consider that can influence OFL in rodents. These aspects range from behavioural considerations, such as the degree of familiarity between conditioning pairs to the

genetic considerations between mouse strains that impact the degree of OFL transmission.

1.2.2.1 Familiarity of conditioning pairs

It has been reported that the degree of familiarity between demonstrators and observers plays a significant role on the strength of OFL conditioning and retrieval. For instance, conditioning pairs that are cagemates or mating-pairs enhances observers' conditioned fear responses as compared to demonstrator-observer pairs that are unfamiliar with one another (Jeon et al., 2010; Jeon & Shin, 2011). Whilst eliciting stronger OFL allows for a greater degree of behavioural manipulation, using familiar conditioning pairs is not always possible, such as when using surgical procedures that require mice to be single-housed permanently following surgical operation. In addition to familiarity, the degree of social exposure at developmentally sensitive periods can also impact OFL. Yusufishaq and Rozenkranz (2013) demonstrated that socially-isolated mice pups post-weaning had an impairment in OFL.

1.2.2.2 Sensory information

Another component influencing the degree of OFL is the need for additional sensory information beyond vision. Whilst visual information is essential for vicarious fear learning (Ito, Erisir, & Morozov, 2015; Jeon et al., 2010), both olfactory (Aoued et al., 2020) and auditory inputs (Kim, Kim, Covey, & Kim, 2010; Pereira, Cruz, Lima, & Moita, 2012) enhance the effective transmission of OFL. The inclusion of additional sensory information may prove useful when integrated with visual information to support robust OFL.

1.2.2.3 Contextual- and cued-variations of conditioning

Freezing in response to a context occurs when that environment triggers an association with an aversive stimulus. The original OFL paradigm relies on the recall of OFL solely through a contextual association for the observer (Jeon & Shin, 2011). While the authors demonstrated sufficient conditioned fear response to the context

alone, Yusufishaq & Rosenkranz (2013) showed more robust OFL with an explicit shock-associated cue. Cued-fear conditioning has previously been shown to be a better predictor of an aversive situation than context alone: an increase in the strength of the US or number of US presentations is required to form a reliable and lasting associative fear memory to a context (Rustay, Browman, & Curzon, 2008).

1.2.2.4 Prior shock exposure

An adaptation of the OFL paradigm entails direct exposure of the aversive US to the observer before undergoing observational fear conditioning. Allsop and colleagues (2018) reported a significant increase in freezing behaviour in observers pre-exposed to a footshock as compared to observers naïve to the direct pain of a shock experience. They proposed that a previous understanding of the US helps to further detect and integrate social cues relating to aversive experiences. However, there is long-standing evidence to suggest that exposure to a single footshock produces a strong context-dependent fear response and memory in and of itself (Wiltgen, Sanders, Behne, & Fanselow, 2001). The addition of pre-training an observer to the US may be a potential confound when assessing the mechanisms of OFL, despite the enhanced fear response of observers during OFL conditioning and later during retrieval. Additionally, pre-exposure is less ecologically valid as a test of purely vicarious fear learning, insofar as natural examples of OFL are crucial, because they do *not* directly expose an individual to a harmful stimulus.

1.2.2.5 Behavioural differences across mouse strains

There are a large number of inbred mouse strains commonly used in behavioural neuroscience, each with their own unique genetic characteristics. These individual differences give rise to a variety of behavioural responses in closely controlled paradigms. Chen and colleagues (2009) initially reported the differences of cued-social learning between two well studied mouse strains: C57BL/6J and BALB/cJ. They identified that while the C57BL/6J strain exhibited robust and consistent freezing behaviour in response to vicarious conditioning, the behaviour of the BALB/cJ was much more variable. It is well documented that C57BL/6J typically

are more prone to social exploration and investigation, while BALB/cJ tend to be less responsive or averse to socialization (Sankoorikal, Kaercher, Boon, Lee, & Brodkin, 2006). Despite both strains exhibiting learning from a social context, it may be that the significantly lower freezing behaviour recorded in the BALB/cJ strain is less of a lack of learning *per se*, and might instead be an expression of other fearful behaviours other than freezing that were not quantified in this study.

Building upon this research, Keum and colleagues (2016) designed a study comparing the OFL of 11 inbred mouse strains in order to assess the possible varying genetic contributions underlying contextual-observational fear conditioning and retrieval. The authors reported that of the 11 strains tested, five demonstrated OFL, while the remaining six did not. Importantly, they found a clear relationship between conditioning and retrieval freezing behaviour amongst all mouse lines tested. However, this mouse-strain-specific correlation for OFL behaviour did not extend to additional tasks related to DFL including anxiety, locomotive ability, and sociability. This suggests that there may be innate genetic variations specific to moderating OFL that do not necessarily contribute to other emotional behaviours or learning. This highlights both the requirement to consider carefully the mouse strain choice in studying OFL, as well as the importance of future research on the underlying genetic mechanisms that may give rise to the spectrum of OFL behavioural responses beyond freezing.

1.2.2.6 Genetic contributions to OFL

The work demonstrating variations between mouse strains suggests the potential of specific genes contributing to OFL. Jeon and colleagues (2010) originally posited a role for the highly expressed $Ca_v1.2$ calcium (Ca^{2+}) channels in the anterior cingulate cortex (ACC) as being one of the key modulators of OFL. This type of Ca^{2+} channel is largely involved in synaptic transmission, neuronal excitability, and moderating pain response. By engineering a $Ca_v1.2^{ACC/Cre}$ mouse line, the authors were able to selectively delete this gene in the ACC leading to a significant impairment of OFL. However, this effect was specific to observational learning as the genetic deletion did not impact DFL, anxiety, or object recognition. This finding

suggests a critical role of the ACC $Ca_v1.2$ Ca^{2+} channels in moderating the observational or social aspect of fear learning rather than fear learning or anxiety in general. Additionally, the role $Ca_v1.2$ Ca^{2+} channels in moderating pain responses advocates for the engagement of a larger pain response circuit potentially underlying OFL.

More recently, a unique missense variant in neurexin 3 (Nrxn3) expressed in the inbred strain 129S1/SvlmJ was identified as being critical for this mouse line's particularly strong OFL response (Keum et al., 2018). Whole-genome and DNA sequencing identified the single nucleotide polymorphism on the Nrxn3 gene contributing to the altered phenotype of 129S1/SvlmJ mice. The Nrxn3 gene is involved in encoding an evolutionarily conserved synaptic cell adhesion molecule critical for synapse assembly and synaptic transmission (Reissner, Runkel, & Missler, 2013; Südhof, 2008). In order to target Nrxn3 in vivo, Keum and colleagues used CRISPR/Cas9 technology to knock in the gene in the C57BL/6J mouse line. While the knock-in mice demonstrated a significantly stronger OFL response during conditioning, their retrieval of the observational fear memory did not differ from the wild-type mice. Moreover, the knock-in mice showed typical behavioural conditioning to DFL. This suggests that Nrxn3 may play a role in acquiring a fear memory through observation, but the memory may depend on the presence of a demonstrator for recall. Whilst the identification of these two specific genetic components may certainly contribute to OFL, there are a large variety of neuropeptides and neurotransmitters known to influence social learning and interaction, such as oxytocin and serotonin, that may have varying genetic components adding to the diversity and complexity underlying OFL as well (Keum & Shin, 2019).

In illustrating a few of the behavioural and genetic considerations influencing OFL, it is clear that building a foundational paradigm for studying OFL is crucial in order to control for as many variables as possible. These examples also demonstrate the multitude of directions that research on OFL can take as genetics unquestionably contribute to the effectiveness of observational fear.

1.2.3 Neurocircuitry underlying OFL

One of the major advantages of studying rodent models is the degree of specificity and directionality possible in interrogating brain circuitry and function. Recent technological advances, such as the development of optogenetics, allow for rapid, causal control of cell-type specific activity using a combination of genetic and optical approaches. Another crucial tool advancing current understanding of neuronal functioning is Ca^{2+} imaging, a technique used to record activity dynamics in a pathway-specific manner. Despite the rapid development of new techniques, relatively little is understood of the neural circuits enabling and attenuating OFL. Because of this, I will first discuss what has been discovered thus far about the brain regions and pathways subserving OFL. I will then provide a very brief overview of the wealth of data describing the mechanisms of DFL as it may provide helpful insight into determining where to look next in OFL circuitry.

1.2.3.1 Brain regions and pathways involved in OFL

Despite research on the neurocircuitry underlying OFL being in its early stages, recent studies have revealed a few potential mechanisms involved in observational learning (Allsop et al., 2018; Burgos-Robles et al., 2019; Ito et al., 2015; Ito & Morozov, 2019; Jeon et al., 2010; Keum et al., 2018; Keum & Shin, 2016, 2019; Kim, Mátyás, Lee, Acsády, & Shin, 2012; Liu, Ito, & Morozov, 2017). The focus has largely been on the contribution of the ACC, because of its distinct ability to process social information and its vast interconnectedness with other regions involved in social cognition and emotion processing. The seminal study on OFL by Jeon and colleagues (2010) demonstrated a critical contribution of the posterior ACC (pACC) in acquiring observational fear. The authors focused on the role of pain perception in OFL, thus targeting the ACC and thalamus, which together make up the 'medial pain circuit' by representing pain affect and sensation. Pharmacological inactivation of the pACC and the medial thalamic nuclei (parafascicular and mediodorsal nuclei) prior to conditioning, inhibited acquisition of OFL, while inactivating the lateral sensory thalamic nuclei had no effect on OFL conditioning. This suggests that OFL engages the emotional aspect of pain, but not necessarily the lateral system of sensory pain

perception. Moreover, the authors report the lateral amygdala (LA), a region known for fear learning and memory storage, is responsible for both the conditioning *and* retrieval of OFL as it closely interacts with the ACC through theta-oscillations to modulate OFL.

More recently, Allsop and colleagues (2018) expanded upon these findings to build a proposed causal model of OFL. Through the use of optogenetics and single-unit in vivo electrophysiology, they elaborated on the functional mechanisms supporting a cortico-amygdala transfer of vicariously learned fear. The study demonstrated that neurons in the ACC and the basolateral amygdala (BLA) encode a CS-US association during OFL, with specific inputs from the ACC to the BLA demonstrating enhanced representation of the CS. Moreover, they showed that photoinhibition of the ACC→BLA circuit disrupted the retrieval of OFL, but not acutely during conditioning. Lastly, the authors demonstrated that the ACC→BLA pathway is critical to OFL, but not DFL, suggesting that the ACC inputs to BLA are responsible for the observational US component of cued-fear learning, which are first processed by the ACC and then sent to the BLA to produce appropriate behavioural responses.

In addition to focus on the ACC, Ito and colleagues (2015) identified the dorsomedial prefrontal cortex (dmPFC), a region which comprises both the anterior part of the ACC and the prelimbic cortex (PL), as being involved in OFL. They demonstrated that contextual-OFL engaged neurons in the dmPFC and BLA by altering the synaptic transmission of BLA inputs from the dmPFC. Moreover, AMPAR silent synapses between dmPFC→BLA were generated in response to OFL suggesting that this pathway may be prone to facilitation through additional Ca^{2+} influx following OFL conditioning, in turn allowing for greater plasticity of the circuit.

Taken together, these studies point to a critical role of the ACC specifically (the dmPFC more largely) in regulating the BLA to properly acquire OFL. In contrast to the larger focus on the cortico-limbic circuit, some studies have begun to look into the microcircuitry within the dmPFC to better understand cell specific effects on the neuromodulatory role of OFL processing. Cortical structures, such as the ACC and the dmPFC, are a mixture of excitatory pyramidal neurons (PNs) and GABAergic

interneurons (INs). Whilst the previous circuit level studies focused on the role of PNs, INs have been shown to regulate neural networks through inhibition of PNs. This mechanism is involved in modulating DFL by providing internal control of glutamatergic activity within a region (Markram et al., 2004; Courtin et al., 2014). For example, Zhou and colleagues (2018) demonstrated that parvalbumin (PV) INs, a subtype of INs in the ACC, are involved in OFL, but not DFL. Specifically, upon activating PV INs prior to fear conditioning, the authors reported an attenuation in freezing behaviour. It should be noted that inactivating PNs in the ACC produces the same behavioural effect (Jeon et al., 2010).

In addition to the role of PV INs, Keum and colleagues (2018) assessed the role of somatostatin (Sst) INs, another subtype of INs, in the ACC. They demonstrate that Sst INs, not PV INs, in the pACC are responsible for controlling the degree of OFL response. Specifically, inhibition of Sst INs during conditioning increased freezing behaviour, while activating these same neurons during conditioning inhibited OFL. While the authors reported a lack of effect in manipulating PV INs, it should be noted that a similar directionality of effect was found by activating PV INs by Zhou and colleagues (2018). Xu and colleagues (2019) recently showed that Sst INs supress PV INs, which in turn disinhibits PNs in the dmPFC. The consequence of this cascade is that disinhibition of PNs augments dmPFC mediation of fear responses.

Together these studies begin to describe a circuit moderating OFL conditioning that is processed largely in the dmPFC and sends inputs to the BLA that are necessary for fear expression. Moreover, dmPFC INs play a significant role in modulating PNs ability to acquire OFL. It must be mentioned, however, that each of these publications used different variations of OFL paradigms, which may contribute to some of the inconsistencies between studies.

1.2.3.2 Neuronal circuits subserving DFL

The interrogation of the neural circuitry involved in OFL is relatively recent and notably sparse. However, there is a wealth of research on the neural networks underlying DFL, which may help expand the currently limited understanding of the

observational component of fear learning. From this well-established paradigm, several key regions have been identified for their contribution to the processing and retrieval of an associated fear memory. The network of brain areas most commonly studied include the amygdala (AMG), hippocampus (HPC), PFC, periaqueductal grey (PAG), thalamus, and brainstem.

Very briefly, it is understood that sensory information from fear conditioning are processed in the midbrain, thalamus, and cortex, which then send outputs to the LA where the association between a CS and US is processed (LeDoux, Cicchetti, Xagoraris, & Romanski, 1990). The LA and basal amygdala (BA) simultaneously receive inputs from the HPC and PFC regarding memories and information associated with the stimulus (de Voogd et al., 2020; Maren, Phan, & Liberzon, 2013). This information is processed in the AMG, which then send signals downstream to areas like the midbrain PAG and brainstem where a behavioural response is generated (T. B. Franklin, 2019; Olsson et al., 2020; Wright & McDannald, 2019a).

1.2.3.3 Integrating DFL circuitry to understand OFL

A few regions within the DFL circuit have already been explored within the context of OFL – namely the roles of specific parts of the PFC and the BLA. Additionally, there are several other regions and circuits highlighted in the DFL literature that, while not yet studied in OFL, have either been implicated in social learning or may contribute to components of OFL.

The PFC, broadly, is crucial for processes of cognitive control working to balance sensory and emotional information, as well as managing internal states, and regulating response outputs (Miller & Cohen, 2001). More specifically, the PL, which lies just ventral to the anterior ACC, is necessary for the acquisition, expression, and maintenance of fear memories in both cued- and contextual-learning (Burgos-Robles, Vidal-Gonzalez, & Quirk, 2009; Corcoran & Quirk, 2007; Milad & Quirk, 2012; Shibano et al., 2020; Sierra-Mercado, Padilla-Coreano, & Quirk, 2010; Tovote et al., 2015; Vidal-Gonzalez, Vidal-Gonzalez, Rauch, & Quirk, 2006). Additionally, the PL has been shown to be involved in attentional modulation towards cues based on their predictive association with an aversive outcome (Sharpe & Killcross, 2015a).

Moreover, PL PV INs are critical for driving fear expression (Courtin et al., 2014), whilst more recently PL Sst INs have been shown to be responsible for bidirectional modulation of fear memory expression and for encoding cue-specific memories (Cummings & Clem, 2019). Together, these results suggest a potential role not only in DFL, but in OFL as well.

The AMG, as already discussed, has been shown to play a major part in DFL and OFL. The AMG directly receives a vast amount of sensory and contextual information priming the brain for cues about threat and safety (Sah, Faber, De Armentia, & Power, 2003). In particular, the BA has robust reciprocal projections between the mPFC (McDonald, 1991) allowing for quickly adaptable responses to salient information, which permits a large degree of emotional regulation. Within the context of cued-fear learning, a fear response is initiated by the AMG, but sustained by the PL (Burgos-Robles et al., 2009), such that the bottom-up directionality of the circuit is important for coding fear associated cues that contribute to fear expression (Jimenez & Maren, 2009; Klavir, Prigge, Sarel, Paz, & Yizhar, 2017; Senn et al., 2014). The current implications for the contributions of the AMG during OFL have been described in section 1.2.3.1.

The HPC is understood to be involved in processing contextual fear learning (Jin & Maren, 2015). It has been shown to be critical in the development of mood and anxiety disorders (Jimenez et al., 2018; Padilla-Coreano et al., 2016; Parfitt et al., 2017), while the ventral hippocampus (vHPC) specifically plays an important role in emotional regulation (Fanselow & Dong, 2010; Strange, Witter, Lein, & Moser, 2014). Because of the robust inputs from the vHPC to PL (Hoover & Vertes, 2007), this pathway has been the focus of DFL research as one component of a broader anxiety-modulating network (Padilla-Coreano et al., 2016). Moreover, the vHPC inputs to PL appear to be crucial for contextual encoding of fear (Hallock et al., 2019) and are responsible for gating a conditioned fear response by exciting PL PNs, which further drives fear expression (Burgos-Robles et al., 2009; Sotres-Bayon, Sierra-Mercado, Pardilla-Delgado, & Quirk, 2012). More recently, Abbas and colleagues (2018) reported that Sst INs in the mPFC are necessary for modulating vHPC→mPFC inputs involved in working memory and spatial encoding.

Recent attention to the HPC has suggested a potential role in social learning (Montagrin, Saiote, & Schiller, 2018; Okuyama, 2018) with CA1 neurons in the vHPC being activated in response to familiar social interactions (Okuyama, Kitamura, Roy, Itohara, & Tonegawa, 2016) and necessary for consolidating social olfactory information (Zinn et al., 2016). Moreover, excitation of PV INs in the vHPC are essential for discerning familiar versus novel conspecifics, further substantiating a role for this region in storing and retrieving social memories (Deng, Gu, Sui, Guo, & Liang, 2019). To date, however, there are no social fear learning studies that have addressed the vHPC→PL pathway.

Lastly, the outputs from the PL are critical for the mediation of fear expression. DFL studies have largely explored PL projections to the BA and the microcircuits within the AMG as the primary source of responses elicited by fear learning (Herry et al., 2008). Specifically, the synchronized activity of BA inputs from the PL have been shown to be responsible for driving freezing behaviour acquired during fear conditioning (Karalis et al., 2016).

Focus has recently turned towards exploring additional regions receiving innervation from the PL to further understand the complexity of fear response. The PAG in particular is involved in defensive responses (Amorapanth, Nader, & Ledoux, 1999; De Oca, DeCola, Maren, & Fanselow, 1998; Franklin, 2019). Specifically, the ventrolateral PAG (vlPAG), has been shown to be necessary for organizing fear responses according to threat predictions (Wright & McDannald, 2019). This may in part be driven by local GABAergic neurons controlling freezing behaviour output (Tovote et al., 2016). Moreover, PL projections to the PAG have been shown to be crucial for contextual fear discrimination providing further understanding of how responses to fear and threat are determined (Rozeske et al., 2018). Within the context of social fear learning, the dorsal PAG (dPAG) has been found to be responsible for eliciting defensive behaviours in response to social threats (Faturi, Rangel, Baldo, & Canteras, 2014). Franklin and colleagues (2017) showed that inhibition of mPFC inputs to the PAG mimicked social defeat behaviour. Additionally, a social defeat paradigm weakened mPFC-dPAG connections leading to an increase in social avoidance.

Overall, progress has been made in a relatively brief period of time in the study of OFL in rodents. It is already clear that rodents can not only serve as a reliable model for OFL, but that the advantages of cutting-edge neuroscience technologies allow for rapidly expanding our understanding of this behaviour. Whilst OFL has been shown to engage aspects of pain and fear circuits, the studies to date primarily focus on a cortico-amygdalar pathway. Research on DFL and pain has substantiated the role of this circuit, however, it is not the only pathway within the brain modulating OFL. From the characterization of circuitry on DFL as well as research on social learning, there is a strong case for observational learning and memory regulation utilizing neural regions and pathways including the PL, BA, vHPC, and PAG, which will be interrogated within this thesis. Learning from observation of others has a significant adaptive value, yet the impact of viewing distressing or harrowing events occurring to others can be a significant contributor to trauma-related disorders. Understanding the mechanisms of brain activity involved in processing and consolidating OFL provides potential opportunity to address pathological functioning and disorder stemming from observed information.

1.3 Experimental OFL in humans

Having a behavioural paradigm that can be modelled across different species provides possibilities to better characterize behaviours, emotions, and disorders. In the above section, I explored the contributions of rodent OFL research, reflecting on how the technology available and current understanding of general fear learning can provide a detailed account of the behavioural and neural mechanisms underlying OFL. However, the mechanisms of brain activity and function are only a piece of what occurs during and contributes to OFL. Whilst mice demonstrate the ability to transmit information about threat through observation, the dispositional characteristics contributing to variation in fear expression are much less easy to untangle.

Some groups have been quick to describe OFL as an example of rodent empathy (Keum et al., 2016; Kim et al., 2019; Luo et al., 2020; Ueno et al., 2018); however, it is not possible at this time to know the internal emotional state of a mouse and what

might be driving their behaviour. Is the response a form of mimicry, simply replicating the fearful behaviour of a demonstrator? Is the observed distress of another mouse interpreted as an imminent or potential threat to the self, or is it a transference of emotional contagion? Why do some mice exhibit stronger freezing behaviour whilst others present a more anxious flight response? Because mice cannot communicate to us their thoughts and feelings during behavioural tasks, research using human participants is the best option for gaining a greater understanding of the psychological factors, dispositional traits, and individual differences contributing to OFL.

Working across species provides a more complete view of the biopsychosocial experience of highly conserved responses. With this in mind, it is crucial to develop research studies that rely on complementary approaches across species in order to fill in the gaps when using only one research model versus another. In this thesis, I not only develop a cued-OFL paradigm and characterize a novel neuronal network in mice, but I also create and validate an OFL task in humans based on reinforcement learning theory in order to understand how learning through observation occurs and the influence of dispositional traits contribute towards OFL.

1.3.1 OFL in humans

In contrast to the variety of OFL behavioural assays in rodents, the human study of OFL has largely deployed a paradigm originated by Olsson and Phelps (2004; Haaker et al., 2017). This OFL paradigm is comprised of two stages: a learning stage and a direct-expression stage. Prior to testing, the participant observer is connected to electrodes to monitor SCR, which could also potentially deliver electrical shocks (importantly, the observer never receives electrical shocks, but is also not informed that they will not). During the observational learning stage, the observer watches a video on a screen of a demonstrator being presented with two different CSs (a blue or yellow square). During the learning stage, one CS (CS+) is paired with the demonstrator reacting to an electrical shock delivered to the forearm (US), whilst the other CS (CS-) is never associated with the US. Next, during the direct-expression stage the observer is presented with the CS+ and CS- in the absence of the

demonstrator on screen. The observer's expression of conditioned fear response is measured using SCRs to the CS+ and CS-, which are typically enhanced in response to the CS+ (Haaker et al., 2017).

Critically, this paradigm allows for the interrogation into value-based learning, an innate behavioural prioritization of maximizing rewards and minimizing punishment, modelling how animals must continually learn to predict and respond to environmental cues to ensure survival. Whilst the OFL protocol is essentially Pavlovian in nature, meaning the task elicits behavioural reflexes automatically in response to aversive cues, it can be modified for instrumental learning (Olsson et al., 2020). In general, Pavlovian learning is stimulus driven while instrumental learning is action driven, meaning that the instrumental system gives value to actions based on prior experience (Bach & Dayan, 2017).

Reinforcement learning is a framework describing *how* learning occurs to maximize rewards and minimize punishment by forming expectations about the value of actions and environmental cues to minimize error. Reinforcement learning theory can powerfully predict how learning occurs by studying the difference between expected and actual outcomes (prediction errors), which is thought to drive learning. A reinforcement learning model holds that when an actual outcome of a decision differs from what is expected, the associative value of a stimulus changes due to the influence of this new information (Sutton & Barto, 1998). With respect to OFL, the value of the CS is constantly being updated to determine whether or not it will predict a vicarious aversive shock delivery. Using two separate CSs in the OFL task allows for the interrogation of a learning strategy, such that the cue associated with the US will receive greater attention and will be learned about preferentially (Olsson et al., 2020).

1.3.2 Physiological and psychological factors impacting OFL in humans

1.3.2.1 Physiological measurements

The primary measurement of OFL in the protocol used by Olsson and colleagues (2017) examined SCRs to the different stimuli and cues. SCRs are a measure of sweat

gland activity widely used to determine conditioned fear responses in DFL (Lonsdorf et al., 2017). The authors reported greater transmission of OFL was consistently associated with higher SCR (Haaker et al., 2017). Additionally, they demonstrated that SCRs of demonstrator and observer pairs during the conditioning phase of OFL predicted the strength of the observers' conditioned fear responses during the direct-expression test phase. This suggests that the greater the synchrony of physiological arousal between a demonstrator and observer, the greater the SCR of the observer in the absence of the demonstrator at testing (Pärnamets, Espinosa, & Olsson, 2020).

In addition to SCRs, another psychophysiological measure is the use of eye-tracking during OFL. This can be useful for research questions related to gaze patterns and attentional focus during different phases of OFL (Kleberg, Selbing, Lundqvist, Hofvander, & Olsson, 2015). The authors reported that observers' gaze patterns were dependent upon whether they were presented with a CS+ or a CS-, such that observers spent more time looking at the demonstrator's face when the CS+ was presented than when the CS- was shown. Additionally, they demonstrated that greater fixation time at the CS+ during the conditioning phase predicted a stronger conditioned fear response during the test phase. Similar to eye-tracking, fear-potentiated startle responses are another measure of conditioned fear responses that have been used in a variety of species and provide further psychophysiological information about fear reaction (Lonsdorf et al., 2017). Recently, Selbing and Olsson (2019) used the startle response as an additional measure of conditioned fear response to OFL. When testing observers' responses to demonstrators displaying differing amounts of anticipatory anxiety, they reported that startle responses aligned with participants' expectancy ratings, and that learning was better when viewing a more anxious demonstrator.

Lastly, a more advanced physiological measure of learning is the haemodynamic response indexed by blood-oxygen-level-dependent (BOLD) contrast imaging using functional magnetic resonance imaging (fMRI). By examining changes in BOLD signals in the brain in response to OFL stimuli, it has been determined that regions including the AMG, ACC, and anterior insula cortex (AI) are activated during OFL (Lindström, Haaker, & Olsson, 2018; Olsson, Nearing, & Phelps, 2007).

1.3.2.2 Biases and beliefs about demonstrator model

OFL inherently relies on the effective transmission of information; however, not all information is valued equally. Whom the information comes from plays a part in the degree of OFL transfer, suggesting that the demonstrator model should be considered when designing social experiments. As an example, Golkar and colleagues (2015) showed that a demonstrator's race influenced the degree of OFL transmission. Specifically, the greater the similarity of the demonstrator to the observer, the larger the facilitation of fear learning through observation. A subsequent study expanded on the role of race on OFL to include social group biases (Golkar & Olsson, 2017). Cleverly, the authors used football club membership support as a means of creating experimental social groups. During the OFL conditioning stage, the observer watched a demonstrator who was either in the same racial group as them or not, and whom they believed to either support their football club or not. They reported that observers' SCRs were significantly higher when learning from a demonstrator of the same race who supported the same team. Moreover, the role of social grouping does not diminish the role of racial bias on OFL.

Additionally, beliefs about a demonstrator's ability also contributes to successful OFL (Selbing & Olsson, 2017). Instructions about a demonstrator's learning abilities, regardless of their actual aptitude, significantly influenced observers' OFL acquisition. Participants displayed poorer performance learning from a demonstrator described to be a 'low learner' in their ability to avoid harmful consequences, versus a demonstrator described as a 'high learner' and adept at learning to avoid aversive situations. The authors suggested that the ascribed ability of the demonstrator affected the attention given to observational information. This demonstrates that observers' prior beliefs cause them to attribute greater attentional value to some cues versus others, regardless of whether those beliefs are valid or not.

Lastly, the behaviour of a demonstrator can influence the degree of successful OFL transmission. Selbing and Olsson (2019) reported that the degree of anxious anticipatory behaviour exhibited by the demonstrator was associated with observers' ability to discriminate between cues. Observers learning from a demonstrator

displaying equally anxious behaviour in response to both CS+ and CS- as compared to a demonstrator anxiously anticipating only the CS+ led to more robust discrimination between safe versus unsafe stimuli. Together, these studies indicate several factors surrounding choosing a demonstrator model that can impact on the strength of OFL.

1.3.3 Dispositional trait considerations impacting OFL

Both animal and human research have identified individual differences in reactivity and expression of OFL. These differences may help explain why some individuals, but not others, have a pathological response to witnessing distressing events happen to others (Debiec & Olsson, 2017; Jeon et al., 2010; Keum et al., 2016; Mikosz, Nowak, Werka, & Knapska, 2016; Olsson et al., 2016). Broadly speaking, dispositional traits are the patterns of behaviour, emotion, and thought that constitute an individual's personality; they can contribute to differences in vulnerability to certain situations or events (Zinbarg & Mohlman, 1998). Trait differences, such as anxiety and psychopathy, are known to be characterised by atypical patterns of learning and threat processing (Decety, Skelly, & Kiehl, 2013; Mkrtchian, Aylward, Dayan, Roiser, & Robinson, 2017; Seara-Cardoso, Viding, Lickley, & Sebastian, 2015). Moreover, individuals with mood and anxiety disorders may have a bias for interpreting situations as threatening, potentially making them hyper-vigilant towards observationally acquired fear (Debiec & Olsson, 2017; Helsen, Goubert, Peters, & Vlaeyen, 2011; Ueno et al., 2018). Conversely, attenuated fear responses in individuals with high-psychopathic traits may lead them to pay less attention to and become less aroused by the distress of others (Decety et al., 2013; Seara-Cardoso, Sebastian, Viding, & Roiser, 2016; Seara-Cardoso et al., 2015). This may result in fewer pairings between events that are threatening and observed fear responses, which may potentially contribute to atypical development of fear learning and empathy over time (Bird & Viding, 2014).

A more clinically based contribution to individual differences in OFL is the influence of observers' dispositional traits. With the exception of the study by Selbing and Olsson (2019) discussed in the previous section, which tangentially examined the

role of the demonstrators' anxiety on OFL, the role of trait characteristics impacting OFL has largely been absent from the literature to date.

The one trait that has received attention, however, is empathy (Bernhardt & Singer, 2012; de Waal & Preston, 2017; Keum & Shin, 2016; Kim et al., 2019; Olsson et al., 2016). Much of the focus surrounding empathy and OFL is based on the principle that OFL is an experimental example of empathy (Debiec & Olsson, 2017; Kim et al., 2019). Building on the case for OFL as a model for exploring empathy, Olsson and colleagues (2016) show that empathy can be amplified or reduced depending on the instructions provided during OFL. Specifically, observers were told to either enhance or reduce their empathic responses to the demonstrator, or were not given any instruction. Participants instructed to pay particular attention to the discomfort expressed by the demonstrator exhibited the strongest conditioned fear response, which was even greater in observers with high reported trait empathy.

1.3.4 Further development of OFL in humans

The basis of the OFL task defined by Olsson and colleagues (2017) allows for various manipulations that can provide crucial information into how OFL is acquired, processed, and expressed. One possible modification to the classic paradigm is to alter how predictive the cues in the task are by modifying the reinforcement ratio, such that instead of the CS+ always being associated with the US and the CS- never being associated with the US, one or both of the CSs could be partially reinforced by pairing to the US only some percentage of the time. This approach has been used previously, for example in one study where the CS+ was associated with the US 50% of the time, whilst the CS- was never predictive of the US (Lindström, Golkar, Jangard, Tobler, & Olsson, 2019; Lindström et al., 2018; Olsson et al., 2020; Szczepanik et al., 2020c); however, this could also extend to the predictability of the CS- as well. Another potential modification is making the OFL paradigm an instrumental learning assay, so that learning is measured not by a physiological response, but by participants predicting the association of each CS with the US. Measuring learning through participants' predictions allows for the study of *how* learning occurs through observation, and not just that it occurs. Additionally, by measuring learning through

participants' predictions and not through physiological measures, which requires equipment that might be restricted to certain spaces, this task can be deployed in a greater range of settings, including clinical ones. Additionally, the task could be modified to include different intensities of shock response exhibited by the demonstrator to test whether the degree of pain/distress influenced OFL.

Beyond modifications to the OFL task itself, very little research has been done on observers' individual dispositional traits and how they may contribute to the success or degree of OFL. Given the current DSM-5 inclusion of observation of fearful events contributing to trauma- and stressor-related disorders, there is a significant gap in the literature linking this innate and preserved form of learning to variability in dispositional traits. By addressing this link, greater understanding of how OFL can be individually experienced can provide better knowledge surrounding the development of clinical pathologies. Ultimately, the framework of the OFL paradigm in humans provides a tremendous degree of flexibility in the ways it can be altered to study the various components critical to social learning that I address within this thesis.

1.4 Thesis research aims

The overarching aim of this thesis is to create a behaviourally translational OFL task in mice and humans, and use it to understand the mechanisms underlying this important type of social learning. By utilizing recent technological advances in behavioural neuroscience in mice, coupled with computational modelling of reinforcement learning and dispositional trait assessment in humans, I hope to provide a characterization of OFL across species that integrates neural network investigations with the social aspects of learning. I present three empirical chapters addressing the development and characterization of a cued-OFL task in mice, followed by anatomical and functional investigations into the neural circuitry underlying OFL (conducted at NIAAA). I then present an empirical chapter validating an OFL task for human participants and investigate the role of dispositional traits on the ability to learn about threat through observation (conducted at UCL).

Chapter 2 outlines the development of a novel cued-OFL task in mice. This first relies on developing and validating the behavioural assay to ensure robust and retained fear expression. This is followed by investigating whether the paradigm impacts anxiety and sociability, two well-defined potential comorbid pathologies of social fear transmission.

Chapter 3 presents studies that anatomically define the neural circuitry subserving OFL in mice. Immediate-early gene mapping and neuronal tracing techniques are used independently and in combination to assess the brain network engaged during OFL. The regions and pathways of interest are a combination of well-defined components of DFL as well as areas shown to be involved in social learning.

Chapter 4 presents studies that functionally define a novel neural circuit underlying OFL and retrieval based upon the anatomical findings from the previous chapter. In vivo optogenetics and Ca^{2+} imaging via fibre photometry is used to demonstrate an interacting network of brain structures including the vHPC, PL, and PAG that serve to calibrate OFL.

Study 1 of Chapter 5 outlines the development and validation of an OFL task for human participants assessed in person. This paradigm relies on the reinforcement learning framework, which allows for the behaviour of observers to be characterized using computational models, supported by Bayesian model comparison, to understand differences in basic reinforcement learning parameters: learning rate (representing the impact of feedback on internal value representations) and temperature (representing the subjective value of reinforcers, as well as choice stochasticity).

Study 2 of Chapter 5 reports on data from a separate, larger, online sample in which I investigated the role of dispositional traits on individual differences in OFL. Using the validated paradigm, individual differences in dispositional traits commonly associated with fear, specifically anxiety and psychopathy, were related to individual differences in OFL. Additionally, the role of self-reported empathy was examined to see whether it could explain associations between OFL and anxiety or psychopathic traits.

Chapter 6 summarizes the findings from the four empirical chapters, relates them to the broader literature, outlines future directions of study, and considers potential translational implications.

1.5 Dissemination

The findings from Chapters 2-4 were presented as posters at the NIH Graduate Research Symposium (Bethesda, MD; 2016), the Gordon Research Conference: Optogenetic Approaches to Understanding Neural Circuits & Behaviour (Sunday River, ME; 2016), the Gordon Research Conference: Amygdala in Health and Disease (Easton, MA; 2017), and the Society for Neuroscience Meeting (Washington, DC; 2017). Additionally, these findings were presented as an invited talk delivered at the NIAAA Fellows Research Seminar (Rockville, MD; 2017) and awarded an NIH Fellows Award for Research Excellence (2017). The findings from Chapters 2-5 were presented as posters at the MQ Mental Health Science Meeting (London, UK; 2019), the UCL Doctoral School Research Poster Competition (London, UK; 2019), the NIH/KI/UCL Joint Neuroscience Symposium (Bethesda, MD; 2019), the UCL Neuroscience Symposium (London, UK; 2019 – First Prize Poster Winner), the British Association for Psychopharmacology Summer Meeting (Manchester, UK; 2019 – President's Poster Prize Winner), and the NIH Graduate Research Symposium (Bethesda, MD; 2021). These finding were awarded an NIH Fellows Award for Research Excellence (2020, 2022) and selected by the Neurobiology Interest Group at the NIH for their Neuroscience Award (2020). Additionally, these findings were presented as invited talks during 2019 at the MQ Mental Health Science Meeting, the NIH/KI/UCL Joint Neuroscience Symposium, the UCL NPP Graduate Programme "Works in Progress" meeting, and the Neural Dynamics Forum at the University of Bristol.

Chapter 2 Observational fear learning behaviour in mice

2.1 Chapter introduction

As set out in the introduction of this thesis, OFL is a crucial means of learning about our environment. However, there is a need to institute and define a paradigm by which to study this behaviour reliably in mice. The current chapter describes the establishment and characterization of a robust rodent OFL behavioural assay.

2.1.1 Observational fear learning

Fear learning is a natural, innate response to threating or aversive stimuli necessary to ensuring safety and survival (LeDoux, 2000). Learning about potential threats can occur through direct experience, or indirectly through social interactions, as identified experimentally in a number of animal models (Atsak, Orre, Bakker, Cerliani, & Roozendaal, 2011; Jeon & Shin, 2011; Kim et al., 2019; Olsson et al., 2020; Olsson & Phelps, 2004). While it is evolutionarily essential to learn from others' experiences, especially when the information relates to potential threat or danger, much of the research on fear learning to date has been concentrated on direct forms of fear acquisition through the extensive development of classical Pavlovian fear learning. The current literature on vicarious forms of fear learning is comparatively recent and relatively sparse, with little known about how fear can be socially transferred and the impact that has on learning and processing aversive information.

Various behavioural paradigms have been established examining socially acquired fear learning in rodents (Chen et al., 2009; Jeon et al., 2010; Jones et al., 2018; Knapska et al., 2006; Warren et al., 2013). Social fear transmission has been assayed in a variety of ways, such as exposing naïve mice to those whom have previously been attacked by biting flies (Kavaliers et al., 2003) or having pairs of mice observe pain behaviour in one another (Langford et al., 2006). From these diverse vicarious paradigms, it is now established that rodents undergo fear learning from a social context.

One behaviourally translational and robust means of social fear learning is through observation – i.e., learning from visualizing another's behaviour (Jeon et al., 2010; Olsson & Phelps, 2004). Jeon and Shin (2011) developed a behavioural protocol for mice closely related to classical Pavlovian conditioning, termed OFL. Their behavioural assay is closely comparable to DFL, but includes learning about a shock-associated context solely through observation of a conspecific. In their paradigm, two mice are placed within a conditioning chamber – the 'demonstrator' is placed on an exposed metal grid floor to directly receive footshocks, while the 'observer' is separated by a transparent divider and is safe from the aversive stimuli. The 'demonstrator' is repeatedly shocked while the 'observer' looks on, without directly experiencing the shock. In contrast to DFL, the aversive US takes the form of observing the 'demonstrator' receive footshocks. The 'observer's' learning is tested the following day by returning the mouse to the conditioning context and measuring the amount of time spent exhibiting freezing behaviour. This protocol provided a foundational assay from which further manipulations of OFL could be used to better understand the transmission of fear via observation.

2.1.2 Behavioural considerations of OFL

OFL is a highly conserved process, which has been speculated to contribute to complex behaviours, such as empathy and altruism (Bernhardt & Singer, 2012; de Waal, 2008). The ability to acquire learned fear without direct exposure becomes a powerful research tool with respect to psychological disorders, such as trauma- and stressor-related disorders, depression, and psychopathy (Bird & Viding, 2014; Blair et al., 2016; Bora, Yucel, & Pantelis, 2009; Decety et al., 2013; Foulkes, McCrory, Neumann, & Viding, 2014; Harnett et al., 2018; Lockwood, Apps, Valton, Viding, & Roiser, 2016; Olsson et al., 2016). Previous research has begun to investigate factors that may influence the ability to acquire and retain OFL and the subsequent consequences these features may have on mental health and socialization.

While our understanding of what may contribute to varying degrees of learning and the possible likelihood of developing a fear driven pathology from OFL is incomplete, several factors have been identified as influencing the degree and

strength of OFL. It has been reported that the degree of familiarity between demonstrator and observer plays a significant role on the strength of fear learning and retrieval. Observed pairings of cagemates or mating pairs enhanced the degree of fearful behaviour exhibited by the observer compared with conditioning pairs that were unfamiliar with one another (Jeon et al., 2010; Jeon & Shin, 2011b). Conversely, Yusufishaq and Rozenkranz (2013) found that the social isolation of observer mice post-weaning reduced OFL transmission.

Another component impacting the strength of OFL is the observer's access to olfactory, auditory, and visual information. Work on this question indicates that visual observation remains the critical sensory modality in OFL transmission and memory (Ito et al., 2015; Jeon et al., 2010). Additionally, Keum and colleagues (2016) found a great degree of variability in OFL between different strains of mice driven by essential genetic differences that can impact sociability and fear learning (Chen et al., 2009; Keum et al., 2016; Keum & Shin, 2019). Finally, Allsop and colleagues (2018) found that an observer's prior direct experience of the US (e.g., directly receiving footshocks) enhances OFL, and proposed that a previous experience with the US helps to further detect and integrate social cues relating to aversive experiences.

In addition to these procedural variables that can augment or reduce OFL, there are notable differences in behavioural output as well. While the classic behavioural measure of DFL is freezing behaviour (e.g., cessation of movement apart from necessary respiratory activity) (Fanselow, 1980), there are a variety of more subtle behaviours that may be informative towards understanding a behavioural task, like OFL. For observer mice, it is possible that OFL engages some anxiety-like behaviours in addition to freezing behaviour only. Such strategies include escape-like movements (e.g., darting), defensiveness (e.g., defensive digging, orienting body away), or exploratory behaviours (e.g., rearing, sniffing, orienting towards) (Gruene, Flick, Stefano, Shea, & Shansky, 2015; Holmes & Rodgers, 1998; Rodgers, Cao, Dalvi, & Holmes, 1997). These additional behavioural strategies may be important components of characterizing and understanding the emotionality of OFL (Kondrakiewicz et al., 2019; Olsson et al., 2020).

Lastly, an additional consideration for studying OFL is the comorbid nature of fear learning – specifically potential impacts on stress, anxiety, and sociability that OFL can contribute to and be detected in subsequent behavioural tasks. Various rodent models of social fear learning report an increase in anxiety- and depressive-like symptoms following conditioning (Krishnan et al., 2007; Warren et al., 2013), while Ito and colleagues (Ito et al., 2015) demonstrate passive avoidance behaviour following contextual-OFL. Furthermore, some social fear learning paradigms report pro-social behaviours modelling similarities with human empathy, such as allogrooming, allolicking, increased pain sensitivity, and helping free a trapped conspecific (Bartal, Decety, & Mason, 2011; Bartal, Rodgers, Bernardez Sarria, Decety, & Mason, 2014; Karakilic et al., 2018; Langford et al., 2006; Lu et al., 2018; Luo et al., 2020). Similarly, rodents may express social buffering, or helping one another recover from stress, following fear learning (Kikusui, Winslow, & Mori, 2006; Morozov & Ito, 2018). Hence, while social fear learning can lead to subsequent behaviours, and even pathologies, it is necessary to characterize OFL in terms of potential additional emotional states that may be elicited beyond fear.

2.1.3 OFL paradigm development, characterization, and validation

In the present study, I begin by modifying the contextual-OFL protocol presented by Jeon and Shin (2011) to include a cued component to both conditioning and retrieval tests. Instead of relying solely on the apparatus to serve as a CS, I include a 30 s neutral white-noise tone that co-terminates with a footshock to the demonstrator. By adding a cued component, the paradigm provides additional associative information from which an observer can better attend to the US beyond context alone (Rustay et al., 2008).

In addition to establishing a robust cued-OFL paradigm, I also define a number of behaviours beyond freezing that are unique to observer mice during both conditioning and retrieval. Moreover, I address whether the OFL paradigm impacts subsequent anxiety and sociability – two well-defined and comorbid potential pathologies of social fear transmission. Based on previous research, I predicted observers would form a lasting, cue-specific memory through observation.

Moreover, I hypothesized that there would be discernible behavioural differences between observer and demonstrator mice during conditioning and retrieval such that observers would have less of a response to the CS than demonstrators, but would engage in additional behaviours, like exploration or defensive responses, in addition to freezing. Lastly, I hypothesized that OFL would cause heightened anxiety and decreased socialization relative to mice not exposed to a US.

2.2 Materials and Methods

2.2.1 Subjects

Subjects were adult male C57BL/6J mice obtained from The Jackson Laboratory (Bar Harbor, ME, USA) and housed in a temperature (72±5°F) and humidity (45±15%) controlled vivarium under a 12-hour light/dark cycle (lights on 0630 h). The C57BL/6J strain was chosen because it exhibits robust contextual OFL, as compared to other inbred mouse strains (Keum et al., 2016). For consistency, demonstrators were also adult male C57BL/6J-background mice unfamiliar (i.e., never cohoused) with the observer. All observer and demonstrator mice were at least 8-weeks old at the time of testing. Mice were pair-housed by sex and strain/line until at least 48-hours prior to behavioural testing when all mice were separated into their own homecages. The number of mice used in each experiment is given in the figure legends. No formal power calculations were applied to determine animal numbers for experiments. All experimental procedures were approved by the NIAAA Animal Care and Use Committee and followed the NIH guidelines outlined in 'Using Animals in Intramural Research' and the local Animal Care and Use Committees.

2.2.2 Cued-OFL paradigm

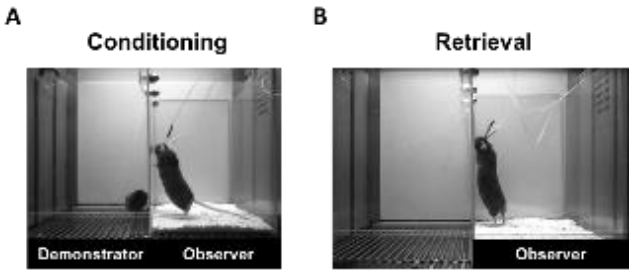
OFL of an auditory cue was tested with modifications from a previously described procedure measuring contextual OFL (Jeon & Shin, 2011).

2.2.2.1 OFL Conditioning

Conditioning was conducted on Day 1 during the light-cycle in a 30 x 25 x 25 cm chamber with transparent front and rear-facing walls, opaque metal-plated side walls, and a metal rod floor, cleaned with a 79% water/20% ethanol/1% vanilla-extract solution to provide a distinctive odour. The chamber was divided by a transparent Plexiglas partition into 2 equally-sized sub-compartments: 1 in which a single demonstrator was placed and the other in which a single observer was placed. Observer-demonstrator pairs were always novel to one another. The floor of the observer compartment was covered with Plexiglas and wood chips to insulate the observer from footshock, whilst the demonstrator was exposed to the electrified metal rod floor (Figure 2.1A). After a 180 s baseline period, the demonstrator received 30 pairings (10 s inter-pairing interval) of a 30 s, 75 dB, white noise (CS), audible to the demonstrator and observer, and 2 s, 1 mA scrambled footshock (US), presented during the last 2 s of the CS. After the final pairing, there was a 120 s no-stimulus period before mice were returned to the homecage. Additional protocols were previously attempted when designing the task, including using only 9 CS-US pairings with a 0.6 mA footshock and a non-cued task similar to what is used by Jeon and colleagues (2010); however, neither protocol was sufficient in producing lasting OFL.

CS and US presentation were controlled by the Med Associates VideoFreeze system (Med Associates, Burlington, VT, USA). Freezing, scored manually every 5 s (as no visible movement except that required for breathing), was measured as an index of fear (Blanchard & Blanchard, 1972), and converted to a percentage [(number of freezing observations/total number of observations) x 100].

Figure 2.1. Example of OFL task setup



A. OFL conditioning setup with naïve pair of mice acting as demonstrator (L) and observer (R). **B.** Retrieval setup 24 h following condition with only observer present.

2.2.2.2 Retrieval

Fear retrieval was tested the day following conditioning. The observer was placed in the same Plexiglas compartment in which conditioning occurred, but without the presence of the demonstrator (Figure 2.1B). After a 180 s baseline period, there were 5 x CS presentations (5 s inter-pairing interval). The observer was then returned to their homecage.

2.2.3 Task validation and characterization

Prior to the start of this study, cued-OFL had not been characterized in mice. I therefore first sought to validate OFL as a robust means of learning.

2.2.3.1 Cued-OFL

C57BL/6J observer-demonstrator pairs underwent the standard OFL conditioning and retrieval procedure described above.

2.2.3.2 CS/no-US OFL

Mice underwent the standard conditioning and retrieval procedure, but with no US delivered during conditioning.

2.2.3.3 Cued-DFL

C57BL/6J observer-demonstrator pairs underwent the standard OFL conditioning and retrieval procedure as above. In this experiment, freezing was measured in the demonstrators.

2.2.3.4 Behavioural ethogram

In addition to freezing behaviour, a battery of additional behaviours was hand-scored every 5 s in order to provide a more complete understanding of OFL on behaviour. In the above behavioural profiling experiments, the following behaviours were measured during conditioning and retrieval: grooming (sitting on the hind legs and rubbing forepaws together or against their face), rearing (placing forepaws on a wall or standing unsupported on the hind legs), digging (pushing woodchips away or towards the body with the snout or forepaws), and other behaviours that fall outside of the aforementioned actions.

2.2.3.5 Opaque partition OFL

Mice underwent the standard conditioning and retrieval procedure, but an opaque partition was turned on 1 s prior to shock delivery until 1 s after shock delivery for each trial of OFL.

2.2.3.6 Fear sensitization/generalization

C57BL/6J observer-demonstrator pairs underwent the standard OFL conditioning and retrieval procedure described above. After the final CS presentation of retrieval, there was a 180 s no-stimulus interval followed by 5 presentations (5 s inter-pairing interval) of a novel 30 s, 75 dB, 7 kHz tone.

2.2.3.7 Impact of observer presence on DFL

Demonstrators underwent the conditioning and retrieval procedure as described above, but in one group the observer was present and in another group the observer was not.

2.2.4 Anxiety-like behaviour

Three groups of mice - OFL, DFL and CS/no-US controls - underwent conditioning and retrieval. The next day mice were tested in the elevated plus-maze (Holmes & Rodgers, 2003) and the following day, tested in a novel open field (Karlsson, Tanaka, Heilig, & Holmes, 2008). The elevated plus-maze was ABS plastic (San Diego Instruments, San Diego, CA, USA), consisting of 2 x 30 x 5 cm open arms (illuminated to 90 lux) and 2 x 30 x 5 x 15 cm closed arms (illuminated to 20 lux) extending from a 5 cm^2 central square and elevated 20 cm from the ground. The mouse was placed in the centre square to begin a 5 min test. Percent time spent in the open arms and total (open + closed) arm entries were measured by the Ethovision video-tracking system (Noldus Information Technology Leesburg, VA, USA).

The open field was a 39 x 39 x 35 cm white Plexiglas square arena (centre of the arena illuminated to 50 lux). The mouse was placed in a corner and allowed to freely explore the apparatus for 10 min. Total distance travelled and percent time spent in the 20 x 20 cm centre square was measured using the Ethovision video-tracking system (Noldus Information Technology).

2.2.5 Sociability

C57BL/6J observer-demonstrator pairs underwent the standard OFL conditioning procedure. In addition, a CS-only group was also tested during the same session. The next day mice were assessed for social interest behaviour, using a modified version of a previously described procedure (Feyder et al., 2010). The mice were placed in the corner of an empty 39 x 39 x 35 cm square arena constructed of white Plexiglas for a 2.5-min baseline period (centre of the arena illuminated to 50

lux). Next, an unfamiliar adult male C57BL/6J stimulus mouse was placed in a corner of the arena within a 10.5-diameter cage (inverted 'pencil holder'), with 1 cm-spaced bars to allow physical contact, for a 5 min interaction period. The stimulus mouse was then removed for a 5 min post-interaction period. The number of whole-body visits, total time spent and average time spent per visit in a 5 cm radius proximal to the stimulus mouse was manually scored from video by an experimenter blind to group.

2.2.6 Data analyses

All analyses were run on a Windows OS. Group differences were analysed using Student's t-tests or analysis of variance (ANOVA) followed by Newman-Keuls *post-hoc* tests. The threshold for statistical significance was set at $p < 0.05$. Mice with freezing scores greater or lesser than 3 standard deviations from the group mean during conditioning were excluded from the analysis.

2.3 Results

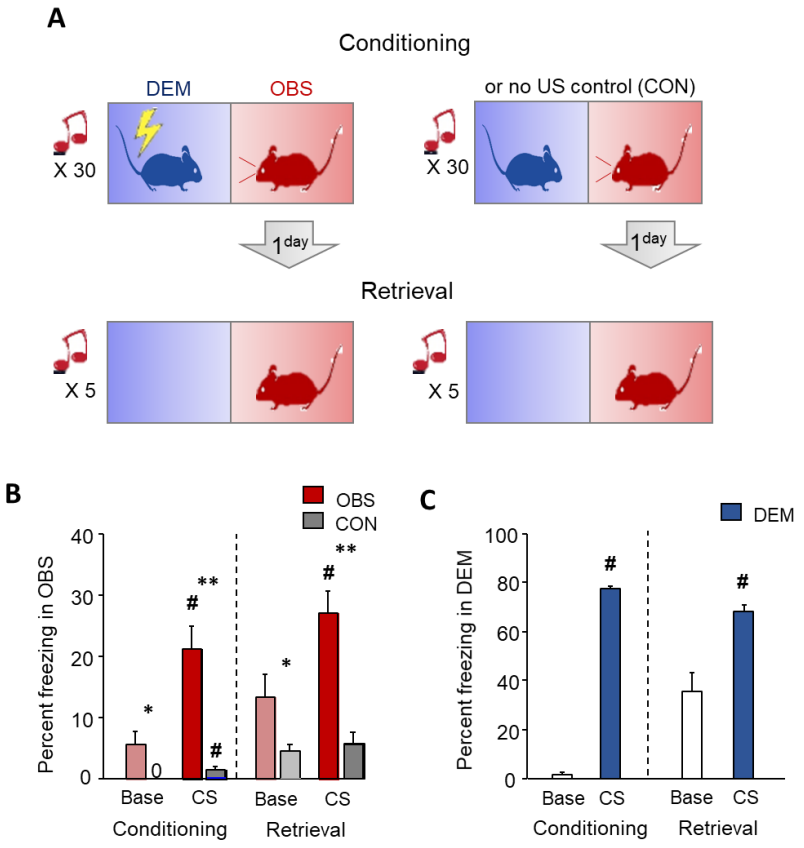
2.3.1 Mice learn a cued-shock association through observation

2.3.1.1 Comparison of OFL behaviour between observers, demonstrators, and no-US controls

First, I tested the OFL paradigm comparing traditional observers witnessing a demonstrator receive footshocks against control observers witnessing a demonstrator in the absence of footshocks (Figure 2.2A). Observers witnessing demonstrators receive repeated CS-US pairings in an adjacent chamber displayed significantly more freezing to the CS during conditioning (unpaired t-test: $t(16) = 4.62$, $p < 0.001$; CS vs base in CS-only CON group, paired t-test: $t(7) = 2.69$, $p = 0.031$; CS vs base in CS-US group, paired t-test: $t(9) = 5.36$, $p < 0.001$) and retrieval the day following conditioning (unpaired t-test: $t(16) = 4.85$, $p < 0.001$; CS vs base in CS-only CON group, paired t-test: $p > 0.05$; CS vs base in CS-US group, paired t-test: $t(9) =$

5.16, $p < 0.001$), as compared to controls observing a demonstrator repeatedly exposed to the CS and no US (Figure 2.2B). It should be noted that freezing levels during both conditioning and retrieval (~20-30%) are less than is typically induced by DFL. By comparison, demonstrators undergoing the same 30-trial DFL procedure showed freezing during conditioning (CS vs base, paired t-test: $t(9) = 73.46$, $p < 0.001$) and retrieval (CS vs base, paired t-test: $t(9) = 4.71$, $p = 0.001$) in the range of 60-80% (Figure 2.2C).

Figure 2.2. Behavioural tests of OFL

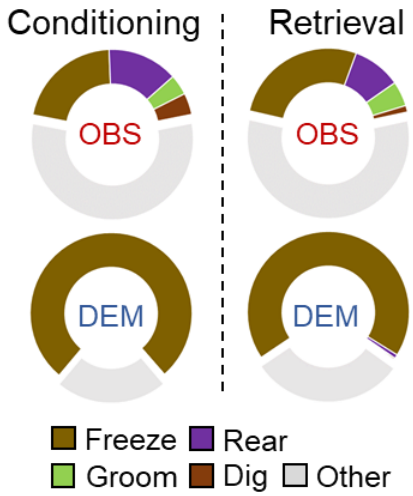


A. Schematic of OFL conditioning and retrieval with US and without US. **B.** OFL conditioning and retrieval data comparing CS-US observers (OBS; n=10) to CS only observers (CON; n=8). **C.** DFL of demonstrators during conditioning and retrieval (DEM; n=10). **p<0.01, *0.01<p<0.05 (2-tailed, between groups); #p<0.05 (2-tailed, within groups).

2.3.1.2 Behavioural ethogram demonstrates differing OFL responses between observers and demonstrators

To assess whether freezing behaviour best characterized fear transmission in the present OFL paradigm, a profile was created by hand-scoring mouse behaviour every 5 s during conditioning and retrieval. In addition to freezing behaviour, I also recorded grooming, rearing, digging, and other behaviours (i.e., any actions that fell beyond these defined criteria). Observers most consistently exhibited freezing behaviour during both conditioning (21.26% ± 3.82) and retrieval (27.00% ± 3.63), suggesting this may be the most reliable behaviour to measure OFL. However, observers did perform other behaviours as well such as rearing (conditioning: 14.06% ± 2.10; retrieval: 9.67% ± 2.14), grooming (conditioning: 4.33% ± 1.23; retrieval: 5.00% ± 2.00), and digging (conditioning: 4.11% ± 1.15; retrieval: 1.33% ± 0.54). Consistent with DFL literature (Fanselow, 1980), demonstrators almost entirely presented freezing in response to CS-US learning (conditioning: 77.50% ± 0.93; retrieval: 68.33% ± 2.45) (Figure 2.3).

Figure 2.3. Behavioural ethogram of observers and demonstrators

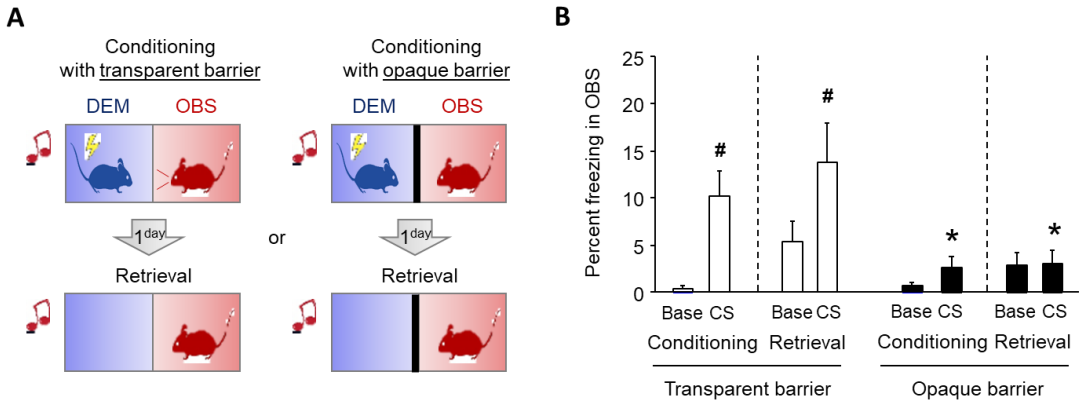


Comparison of behavioural strategies employed during conditioning and retrieval between OBS (n=10) and DEM (n=10).

2.3.1.3 Observation of US is critical for OFL transmission

Whilst Jeon and Shin (2011) demonstrated that observation is critical for the transmission of social fear learning, they made this assessment based on placing a permanent opaque barrier between demonstrators and observers. I wanted to know whether obscuring only the US delivery was sufficient in blocking OFL (Figure 2.4A). Observers conditioned in the transparent-partition version froze to the CS, relative to pre-CS baseline, during conditioning (paired t-test: $t(7) = 3.49$, $p = 0.010$) and retrieval (paired t-test: $t(7) = 3.32$, $p = 0.013$). Observers conditioned in the opaque-partition did not differ in their freezing behaviour than those froze significantly less to the transparent conditioning (paired t-test: $t(8) = 1.50$, $p = 0.172$) and transparent retrieval (paired t-test: $t(8) = 0.07$, $p = 0.943$) (Figure 2.4B).

Figure 2.4. Role of obscuring US during OFL



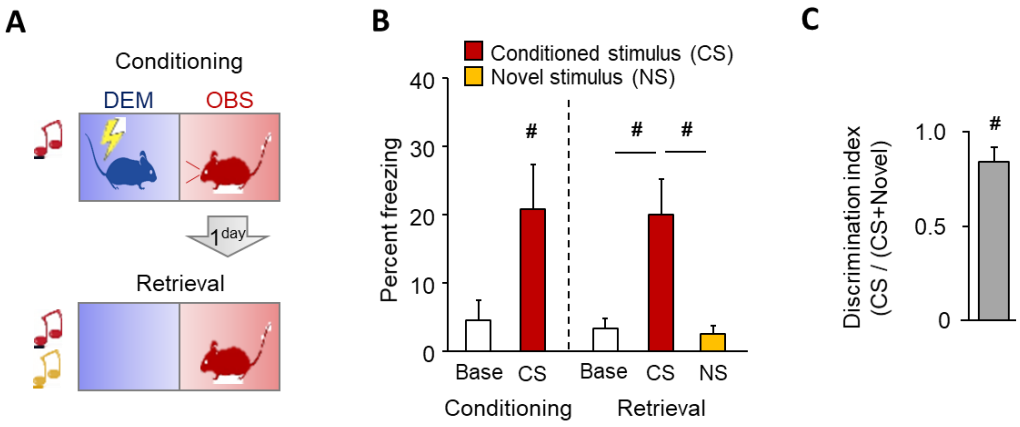
A. Schematic of regular OFL compared to OFL with an opaque barrier during US presentations only followed by regular retrieval. **B.** Comparison of freezing rates during OFL conditioning and retrieval between mice with a transparent partition (n=8) and mice with an opaque partition during the US (n=9).

2.3.1.4 Observers do not generalize fear following OFL

An additional component to test when establishing and validating cued-OFL was the possibility of fear generalization or sensitization to other stimuli never

associated with a US. To test if the transference of OFL was specific to the CS-US presented on conditioning day, a novel tone was introduced following retrieval to measure any fear generalization that might occur or any non-specific fear sensitization (Figure 2.5A). Discounting this possibility, mice that had undergone OFL conditioning CS showed significantly increased freezing to the CS relative to baseline (CS vs base, paired t-test: $t(6) = 3.33$, $p = 0.016$), which did not generalize to a novel tone (tone vs base, paired t-test: $p > 0.05$; CS vs tone, paired t-test: $t(6) = 3.92$, $p = 0.008$) (Figure 2.5B). The specificity of the CS-elicited response was further evidenced by a high CS versus novel cue discrimination index (Figure 2.5C).

Figure 2.5. Fear generalization/sensitization following OFL



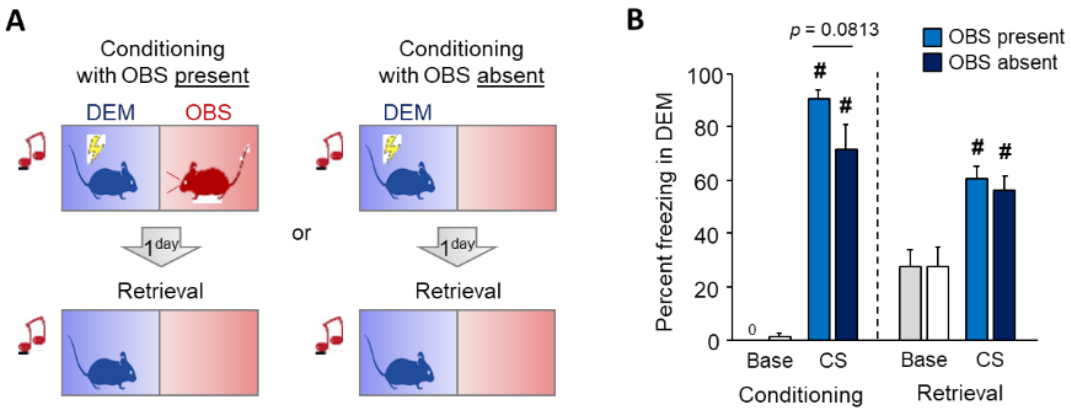
A. Schematic of fear generalization task following OFL conditioning and retrieval. **B.** Percentages of time spent freezing to the CS and the novel stimulus (NS) by observers (n=7). **C.** Discrimination index of observers' ability to associate fear with a CS instead of a NS.

2.3.1.5 Presence of an observer during OFL conditioning does not impact DFL

Additionally, I wanted to test if DFL was affected by the presence of an observer. In one group, demonstrators underwent the OFL task with an observer present as per the standard protocol, while the second group of demonstrators underwent the same number of CS-US trials but in the absence of an observer (Figure

2.6A). During conditioning, demonstrators undergoing DFL in the presence of an observer present froze less to the CS than demonstrators conditioned alone, albeit this result narrowly missed statistical significance. Both groups froze more to the CS, relative to pre-CS baseline (ANOVA effect of group: $F(1,12) = 3.00$, $p = 0.109$; ANOVA effect of CS: $F(1,12) = 266.68$, $p < 0.001$; ANOVA group x CS interaction: $F(1,12) = 4.23$, $p = 0.062$). During retrieval, demonstrators conditioned with an observer present froze to the CS to a similar extent as demonstrators conditioned alone; and both groups froze more to the CS, relative to pre-CS baseline (ANOVA effect of group: $F(1,12) = 0.13$, $p = 0.722$; ANOVA effect of CS: $F(1,12) = 32.10$, $p < 0.001$; ANOVA group x CS interaction: $F(1,12) = 0.13$, $p = 0.728$) (Figure 2.6B).

Figure 2.6. Presence of an observer on DFL



A. Schematic of OFL task determining if presence of an observer impacts DFL behaviour. **B.** Demonstrators freeze marginally more in the presence of an observer than without during conditioning, but retrieval behaviour shows no lasting impact of observed DFL (OBS present n=7; OBS absent n=7).

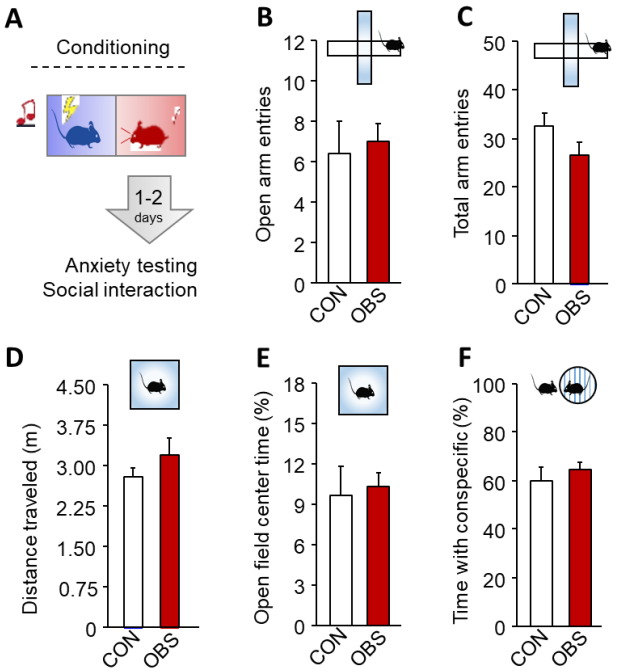
2.3.2 Cued-OFL does not affect anxiety-like behaviour or sociability

Social stressors, such as repeated defeat by a dominant conspecific, produce increases in anxiety-like behaviour and social aversion (Krishnan et al., 2007), while similarly, contextual-OFL conditioning facilitates later passive avoidance learning (Ito

et al., 2015). I therefore assessed whether OFL affected these behavioural dimensions by testing mice within two days of OFL in the elevated plus-maze and novel open field assays as well as on a simple measure of sociability (Figure 2.7A). There were no significant differences in the frequency of open arm (ANOVA: $p > 0.05$) or total arm (ANOVA: $p > 0.05$) entries in the elevated plus-maze between OFL and naïve homecage controls (Figure 2.7B-C). Likewise, neither total distance travelled (ANOVA: $p > 0.05$) nor percentage of time spent in the centre of the open field (ANOVA: $p > 0.05$) differed significantly between groups (Figure 2.7D-E). Lastly, no significant group differences were identified for time spent in proximity to a novel conspecific (ANOVA: $p > 0.05$) (Figure 2.7F).

These data show that in contrast to more extreme and prolonged social stressors, such as repeated social defeat, OFL does not produce a phenotype of heightened anxiety-like behaviour nor social aversion.

Figure 2.7. OFL effects on subsequent anxiety-like behaviour and sociability



A. Schematic of OFL conditioning followed by elevated plus-maze the day following training. Open field and sociability tasks were performed 2 days after OFL conditioning. **B.** Number of open arm entries by home-cage controls (CON; white n=6) compared to observers (OBS; red n=6) **C.** Total number of entries into any arms of the elevated plus-maze. **D.** Total distance travelled in open field task. **E.** Percent of time spent in the centre of the open field box. **F.** Percent of time spent interacting with a novel conspecific during sociability task.

2.4 Discussion

The aim of this chapter was to develop, validate, and characterize a cued-OFL behavioural assay in mice expanding upon the contextual paradigm established by Jeon and Shin (2011). The present study advances this work by demonstrating that mice form a lasting, stimulus-specific memory for a discrete environmental cue paired with a footshock solely through observation of conditioning in an unfamiliar conspecific. The proposed paradigm proves to be specific to the socially-acquired associative memory and is not the result of generalized over-sensitization to fear learning, nor does it produce an enduring phenotype of anxiety-like behaviour or social disturbance.

Prior research has shown mice can reliably and robustly form a fear memory through observation of a distressed conspecific (Allsop et al., 2018; Bruchey, Jones, & Monfils, 2010; Church, 1959; Kavaliers et al., 2003; Keum & Shin, 2016; Knapska et al., 2010; Langford et al., 2006; Sterley et al., 2018). Based upon the original non-cued version of the OFL paradigm put forth by Jeon & Shin (2011), I modified the assay to study observed learning to cued-information. Through this I demonstrated that observers can learn a cued-fear response through observational conditioning that is also able to be retrieved through exposure to the CS. This was evidenced by the significantly higher rates of freezing during CS presentation compared to baseline context exposure. Additionally, this effect is absent in the presence of a novel tone. While the context where the conditioning took place has been shown to be important for OFL in mice in prior studies (Ito et al., 2015; Jeon et al., 2010; Keum & Shin, 2019), the present study shows that a CS-US pairing provides reliably stronger expression and better retrieval of an observed fear memory.

In addition to establishing a cued-OFL paradigm, I wanted to examine how vicarious learning might impact anxiety and sociability. Previous studies using different paradigms of socially-transmitted fear learning have demonstrated increases in anxiety levels and decreases or changes in sociability following their conditioning paradigms (Knapska et al., 2010; Krishnan et al., 2007); however, in the present study there were no increases in anxiety- or sociability-related behaviours following OFL conditioning. This may be in part because other studies have used much more aggressive forms of social learning, like social defeat, or have run their paradigm for multiple days. It is perhaps not surprising, then, that this proposed paradigm does not find changes in anxiety-levels, which may require a direct experiential component that poses a greater immediate threat, or changes in sociability, which is more typically seen when conditioning pairs are closely related (Ito et al., 2015; Jeon et al., 2010) or in aggressor contexts (Krishnan et al., 2007).

Various studies have debated whether observation alone is sufficient in transmitting a fear response in mice. Some argue that giving an observer a prior experience of a footshock before conducting OFL conditioning produces a much more robust effect on acquisition and retrieval (Allsop et al., 2018; Carrillo et al., 2019). However, other studies, the present one included, demonstrate pre-exposure to the US is *not* necessary for the formation of a strong and stable observed fear memory (Ito et al., 2015; Jeon et al., 2010). Moreover, there is long-standing evidence to suggest that exposure to a single footshock produces a strong context-dependent fear response and memory in and of itself (Wiltgen et al., 2001). Whilst the freezing behaviour presented within this chapter illustrated that OFL fear response is substantially lower than that induced by DFL, observers still developed a robust fear memory, which is important for future manipulation studies. Moreover, observers exhibited additional behavioural strategies not seen in DFL, that while less consistent than freezing, may inform individual differences in fear responses to an observed threat and should not be discounted from a learning and memory perspective.

The capacity to learn about environmental threats vicariously and avoid the need for potentially harmful, direct experience with danger, has enormous adaptive value across species, and is a common mode by which learning occurs. Given the vast

literature on directly-acquired forms of cued- and contextual-fear, it is therefore surprising that comparatively little attention has been paid to the study of observational fear. The current establishment of a reliable cued-OFL behavioural assay in addition to the contextual-OFL protocol by Jeon and Shin (2011) is crucial towards beginning to redress this relative lack of attention to social fear learning.

In the following chapter I expand upon this paradigm validation by anatomically mapping OFL neural circuitry using a combination of immediate-early gene activation labelling and adeno-associated viruses (AAVs). With the foundation of a behavioural assay, research into the underlying neural mechanisms of OFL can provide a better understanding of the learning and memory of social fear, how it differs from direct forms of fear learning, and circuits that may contribute to the development of pathologies linked to OFL.

Chapter 3 Anatomical circuitry underlying observational fear learning

3.1 Chapter Introduction

The present chapter begins by looking at the general differences in neural activity between OFL and DFL in key regions known to be involved in direct fear circuitry. I first focus on the contribution of the PL, located within the dmPFC, and its established roll in processing fear learning. I look at regional activation in response to conditioning, then define anatomical inputs and outputs of the PL using a combination of AAVs and immediate-early gene activation labelling using c-fos immunohistochemistry. From this, an anatomically specific circuit emerges involving the engagement of PL through inputs from the vHPC and BA, as well as PL outputs to the PAG.

3.1.1 Role of the PL in fear learning

The bulk of what is known about the underlying neural circuitry of fear learning comes from research on the direct experience of an aversive event. As described in the general introduction, DFL has been a model behavioural task for studying remote fear memory. From this well-established paradigm, several key regions have been identified for their contribution to the processing and retrieval of an associated fear memory. The PL has received much attention in DFL studies as it has been shown to be critically involved in executive control and emotion processing (Miller & Cohen, 2001), acquisition and maintenance of fear (Sierra-Mercado et al., 2010; Vidal-Gonzalez et al., 2006), and the modulation of attention towards cues based on their predictive value of an aversive outcome (Sharpe & Killcross, 2015a). More specifically, the PL has been found to play a role in the acquisition, expression, and sustenance of both cued- and contextual-fear memories (Burgos-Robles et al., 2009; Corcoran & Quirk, 2007; Milad & Quirk, 2012; Rozeske, Valerio, Chaudun, & Herry, 2015; Shibano et al., 2020; Sierra-Mercado et al., 2010; Tovote et al., 2015). Giustino

and colleagues (2016) demonstrated suppression of PL activity in response to cued-footshocks. The authors' reported this was driven by preferential activation of the PL over the slightly more ventral structure, the infralimbic cortex (IL), an area believed to work in direct contrast to the PL, which ultimately may underlie fear-induced freezing behaviour.

In contrast with the studies cited above, which largely focus on excitatory PNs in the PFC, research on DFL has identified specific contributions of the much more sparsely populated inhibitory GABAergic INs that are critical for neuromodulation of learning and memory. Whilst PFC PNs have been shown to encode stimulus-specific associations through modifications to synaptic strength and density (Josselyn, Köhler, & Frankland, 2015), INs are believed to inhibit PNs as a means of moderating the specificity of fear learning (Courtin et al., 2014). Specifically, PV and Sst INs account for the majority of GABAergic inhibitory cells in the PFC (Rudy, Fishell, Lee, & Hjerling-Leffler, 2011). While it has been demonstrated that PL PV INs are critical for driving fear expression (Courtin et al., 2014), more recently PL Sst INs have been shown to be responsible for bidirectional modulation of fear memory expression and encoding cue-specific memories (Cummings & Clem, 2020).

Whilst research into the neural circuitry underlying OFL is still in its early stages, recent studies reveal potential mechanisms involved in vicarious learning (Allsop et al., 2018; Ito et al., 2015; Ito & Morozov, 2019; Jeon et al., 2010; Keum et al., 2016; Keum et al., 2018; Kim et al., 2014; Liu et al., 2017). These studies have provided evidence for the involvement of the dmPFC, more specifically focusing on the pACC, in processing and regulating OFL. These studies highlight this region's contribution in processing and perceiving pain in both oneself and in others (Burgos-Robles et al., 2019; Sivaselvachandran, Acland, Abdallah, & Martin, 2018). Additionally, social fear learning paradigms have suggested a neuromodulatory role of PV and Sst INs in the PFC similar to what has been shown in DFL (Keum et al., 2018; Xu et al., 2019; Zhou et al., 2018). Even though preliminary studies show the PFC to be engaged in observational fear (Ito et al., 2015; Li et al., 2014; L. Liu et al., 2017), the mechanistic and regulatory role the PL plays in cued-OFL remains unknown.

3.1.2 Inputs to the PL involved in DFL and social learning

The role of the PL in DFL is well characterized; however, it is not singularly responsible for representing and sustaining fear. The breadth of research on DFL provides a framework for cued forms of fear learning and understanding of how these brain-wide processes work. The PL is heavily innervated by a number of cortical and subcortical regions, consistent with a role in cognitive processes (Hoover & Vertes, 2007). Moreover, the PL provides top-down regulation of emotional responses during fear learning by integrating diverse neural inputs from various regions, often in a reciprocal manner (Giustino & Maren, 2015; Jackson, Karnani, Zemelman, Burdakov, & Lee, 2018; Marek, Strobel, Bredy, & Sah, 2013; Marek, Xu, Sullivan, & Sah, 2018; Padilla-Coreano et al., 2016; Parfitt et al., 2017b; Senn et al., 2014; Sotres-Bayon & Quirk, 2010; Sotres-Bayon et al., 2012; Ye, Kapeller-Libermann, Travaglia, Inda, & Alberini, 2017; Yizhar & Klavir, 2018). Particular attention has been placed on AMG and HPC inputs to PL.

The AMG directly receives a vast amount of sensory and contextual information, priming the brain for cues about threat and safety (Sah et al., 2003). In particular, the BA has robust reciprocal projections with the mPFC (McDonald, 1991) allowing for quickly adaptable responses to salient information to support emotion regulation. Within the context of cued-DFL, a fear response is initiated by the AMG, but sustained by the PL (Burgos-Robles et al., 2009) such that the bottom-up directionality of the circuit is important for coding fear-associated cues that contribute to fear expression (Jimenez & Maren, 2009; Klavir et al., 2017; Senn et al., 2014).

More recently, research on OFL has indicated a role for the bidirectional pathway between the BA and the ACC (Allsop et al., 2018; Ito et al., 2015; Ito & Morozov, 2019; Jeon et al., 2010). Critically, these studies suggest preferential encoding of observationally acquired fear from the ACC inputs *to* the BA, but do not find significant contributions of the reciprocal pathway. Whilst the anterior portions of the ACC lie proximally to the PL, it is important to remember these are distinct structures with differing roles in learning and memory.

In DFL, the AMG inputs to PL are largely involved in the acquisition and expression of conditioned fear, whilst the HPC processes contextual-fear learning (Jin & Maren, 2015). The HPC has been shown to be critical in the development of mood and anxiety disorders (Jimenez et al., 2018; Padilla-Coreano et al., 2016; Parfitt et al., 2017), while the vHPC specifically plays an important role in emotion regulation (Fanselow & Dong, 2010; Strange et al., 2014). Because of the robust inputs from the vHPC to PL (Hoover & Vertes, 2007) this pathway has been the focus of a broader anxiety modulating network (Padilla-Coreano et al., 2016). Moreover, the vHPC inputs to PL appear to be crucial for contextual encoding of fear (Hallock et al., 2019) and are responsible for gating the conditioned fear response by exciting PL PNs, further driving fear expression (Burgos-Robles et al., 2009; Sierra-Mercado et al., 2010; Sotres-Bayon et al., 2012). More recently, Abbas and colleagues (2018) reported that Sst INs in the mPFC are necessary for modulating vHPC→mPFC inputs involved in working memory and spatial encoding.

To date, there are no social fear learning studies that have addressed the vHPC→PL pathway; however, recent attention to the HPC has suggested a potential role in social learning (Montagrin et al., 2018; Okuyama, 2018). CA1 neurons in the vHPC have been shown to be activated in response to social interaction with a familiar conspecific (Okuyama et al., 2016) and are necessary for consolidating social olfactory information (Zinn et al., 2016). Additionally, Kitamura and colleagues (2017) reported HPC inputs helped generate rapid PFC memory engram cells during contextual learning. One mechanism of vHPC-PL memory storage and retrieval may be through excitation of PV INs in the vHPC, which have been shown to be essential for discerning familiar versus novel conspecifics (Deng et al., 2019).

BA and vHPC inputs to PL have been studied for their contribution in sustaining learned fear through associating aversive events with contexts and cues. The PL then modulates these inputs to respond to fear accordingly. While this framework has been studied in DFL, potential contributions of these circuits to cued-OFL remain unknown.

3.1.3 Outputs from the PL involved in fear expression

As discussed above, inputs to the PL are crucial for fear regulation, yet the outputs from this region are critical for the mediation of fear expression. DFL has largely explored PL projections to the BA and the microcircuits within the AMG as the primary source of response elicited by fear learning (Herry et al., 2008). Specifically, synchronized activity of BA inputs from the dmPFC has been shown to be responsible for driving freezing behaviour acquired during fear conditioning (Karalis et al., 2016).

More recently, focus has turned to additional regions receiving innervation from the PL to further understand the complexity of fear. The PAG, in particular, is involved in defensive responses (Amorapanth et al., 1999; De Oca et al., 1998; Franklin, 2019), but recent work shown the vlPAG to be primed for organizing fear responses according to threat predictions (Wright & McDannald, 2019). This may in part be driven by local GABAergic neurons controlling freezing behaviour output (Tovote et al., 2016). Moreover, PL projections to the l/vlPAG have been shown to be crucial for contextual fear discrimination providing further understanding of how responses to fear and threat are determined (Rozeske et al., 2018). Within the context of social fear learning, the dPAG has been found to be responsible for eliciting defensive behaviours in response to social threats (Faturi et al., 2014). Franklin and colleagues (2017) showed that inhibition of mPFC inputs to the dPAG mimics social defeat behaviour and that a social defeat paradigm weakens mPFC-dPAG connections leading to increase in social avoidance.

Another PL output region involved in fear response is the nucleus accumbens (NAc). The expression of fear motivated behaviours has been found to be driven from BA inputs to the mPFC, resulting in a NAc led conditioned fear response (McGinty & Grace, 2008). In situations of low threat, Moscarello and Maren (2018) proposed a HPC→mPFC→NAc circuit responsible for behavioural responses, such as avoidance, exploration, and risk assessment, as opposed to reactive fear behaviours, such as freezing. With regards to social behaviour, PL inputs to the NAc may contribute to social-spatial learning by activating in response to social investigation, but only in specific locations (Murugan et al., 2017).

In summary, investigating outputs from the PL during OFL conditioning and retrieval may provide a better understanding of the behavioural subtleties seen in the present paradigm and the delicate balance between an adaptive response and one that is pathological.

3.1.4 Role of PL in a broader anatomical circuit underlying OFL

Drawing on our understanding of the circuits involved in DFL, and what little is known about social fear learning, the present chapter focuses on a combination of anatomical and immediate-early gene labelling to begin mapping the anatomical circuitry underlying cued-OFL. Here I compare region- and projection-specific activation of a PL-centred network of OFL to DFL. Through these anatomical mapping studies, a potentially functional vHPC→PL→PAG circuit emerges.

3.2 Materials and Methods

3.2.1 Subjects

Unless otherwise specified, observers were adult male C57BL/6J mice obtained from The Jackson Laboratory (Bar Harbor, ME, USA) and housed in a temperature (72±5°F) and humidity (45±15%) controlled *vivarium* under a 12-hour light/dark cycle (lights on 0630 h). Demonstrators were also adult male C57BL/6J-background mice unfamiliar (i.e., never cohoused) with the observer. Mice were singly-housed until at least 2 days before testing.

Four, C57BL/6J-background, Cre mutant lines were used for tracing experiments: B6.Cg-*Pvalb*$^{tm1.1(cre)Aibs}$/J (PV-Cre) mice expressing Cre recombinase in PV-expressing cells (JAX strain 012358), *Sst*$^{tm2.1(cre)Zjh}$/J (Sst-Cre) mice expressing Cre in Sst-expressing cells (JAX strain 013044), *Vip*$^{tm1(cre)Zjh}$/J (Vip-Cre) mice expressing Cre in vasoactive intestinal peptide-expressing cells (JAX strain 010908), and B6.129S-*Slc17a7*$^{tm1.1(cre)Hze}$/J (Vglut1-Cre) mice expressing Cre in vesicular glutamate transporter 1-expressing cells (JAX strain 023527). Male PV-Cre, Sst-Cre, and Vglut1-

Cre mice were bred in-house by mating Cre+ sires with C57BL/6J dams. Male Vip-Cre mice were purchased directly from The Jackson Laboratory.

The number of mice used in each experiment is given in the figure legends. All experimental procedures were approved by the NIAAA Animal Care and Use Committee and followed the NIH guidelines outlined in 'Using Animals in Intramural Research' and the local Animal Care and Use Committees.

3.2.2 Regional activation mapping using c-fos immunohistochemistry

Two groups of mice – OFL and CS/no-US OFL controls - underwent conditioning as described in Chapter 2 and, 2 hours later (Zhong et al., 2014), were deeply anaesthetized with sodium pentobarbital and transcardially perfused with ice cold phosphate buffered saline (PBS, pH 7.4) followed by ice cold 4% paraformaldehyde (PFA). Another set of test-naïve controls were sacrificed immediately on removal from the home-cage. Given the lack of c-fos differences between the home-cage and CS/no-US OFL controls (all regions $p > 0.05$), these groups were combined for analysis.

Brains were removed and 50-µm coronal sections cut on a vibratome (Leica VT1000 S, Leica Biosystems Inc, Buffalo Grove, IL, USA) and stored free-floating in phosphate buffer (PB) 0.1M at 4° C for no longer than 1 week. Sections were first thoroughly rinsed 3X for 10 min in PBS and then blocked in 10% normal goat serum and 1% bovine serum albumin in PBS-TritonX (0.3%) for 2 hours. Sections were incubated over 2 nights in a mixture of rabbit anti-c-fos (9F6) (cat#: 2250S, 1:1000, Cell Signaling Technology, Danvers, MA, USA) and a mouse monoclonal anti-NeuN antibody (MAB377, 1:1000, MilliporeSigma, Burlington, MA, USA) in a dilution of 1% normal goat serum and 0.1% bovine serum albumin in PBS-TritonX (0.3%) on a platform rocker at 4°C. Sections were then rinsed 3X for 10 min in PBS and incubated in anti-rabbit Alexa 488 secondary antibody (cat#: A-11034, 1:500, Invitrogen, Eugene, OR, USA) and Alexa Fluor 555 anti-mouse antibody (cat#: A-21422, 1:500, Invitrogen) in a dilution of 1% normal goat serum and 0.1% bovine serum albumin in PBS-TritonX (0.3%) at room temperature on a platform rocker for 2 hours. Sections were rinsed in PBS 2X for 10 min and then counterstained with 5 µg/mL Hoechst 33342 (cat#: H1399, Thermo Fisher Scientific, Waltham, MA, USA) in PBS and rinsed

1X 0.1M PB for 10 min. Serial sections were mounted onto slides, air-dried, and cover slipped with DABCO (Sigma-Aldrich, St. Louis, MO, USA), 10mM Tris-HCl (pH 8.0), and Glycerol and sealed with clear nail polish.

Images of all 3 channels (c-fos, NeuN, Hoechst) for all collected sections were acquired using an Olympus VS120 Virtual Slide Microscope system (Olympus, Center Valley, PA, USA, VS_ASW software) with a 20x objective (U Plan S Apo; 20x NA 0.75). The NeuN channel was used as the focus-reference, in autofocus mode. For image analysis, the FIJI (https://imagej.net/Fiji) VSI reader plugin (BIOP, Zurich, Switzerland) (Schindelin et al., 2012) was used and a contour of each brain area was manually drawn and imaged with reference to a mouse brain atlas (Franklin & Paxinos, 2008). The number of c-fos+ neurons were counted in the following brain regions: PL (from AP = +2.10, ML = ± 0.25, DV = 2.00 to AP = +1.78, ML = ±0.25, DV = −2.25), posterior ACC (pACC) (from AP = +1.10, ML = ±0.50, DV = -2.00 to AP = +0.74, ML = ±0.50, DV = −1.75), anterior region of the claustrum (CLA) (from AP = +1.18, ML = ±2.50, DV = -3.50 to AP = +0.86, ML = ±2.75, DV = -3.75), AI (from AP = +1.18, ML = ±3.00, DV = -3.75 to AP = +0.86, ML = ±3.25, DV = -3.50), BA (from AP = -1.46, ML = ±2.75, DV = -4.75 to AP = -1.82, ML = ±2.75, DV = -4.75), and vHPC (from AP = -2.92, ML = ±3.00, DV = −4.50 to AP = -3.28-, ML = ±3.25, DV = -4.50). For each brain region, cell counts were conducted (blind to experimental group) in 3 sections from each hemisphere, for a total of 6 data points per region per mouse. It was unnecessary to correct for double counting because sections were non-consecutive. The mean number of c-fos+ cells per 0.25 mm^2 was quantified in a semi-automated manner using a custom-written Fiji macro. As there were no differences in c-fos counts between the home-cage and CS-only controls (all regions $p > 0.05$), these groups were combined for analysis.

3.2.3 Input-specific PL c-fos quantification

Mice were placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA, USA) to bilaterally infuse 0.20 µl of the retrograde tracer, cholera toxin subunit B (CTb, Alexa Fluor 555 conjugate, cat# C34775, 1% wt/vol dilution) (Thermo Fisher Scientific), into the PL (coordinates relative to bregma: AP +1.95 mm, ML ±1.00

mm, DV -1.90 mm, at a 20°) over 10 min using a Hamilton syringe and 33-gauge needle. The needle was left in place for a further 5 min to ensure diffusion. One week later, mice were randomly assigned to either DFL, OFL, or naïve home-cage control. Depending on their group, mice underwent conditioning as described in the previous chapter and, 2 hours later, were perfused as described above. The naïve control group were sacrificed immediately on removal from the home-cage.

Fifty-µm thick coronal sections were prepared and immunostained for c-fos as above. Sections were imaged using a Zeiss LSM 700 confocal microscope under a Plan-Apochromat 20x/0.8 M27 objective (Carl Zeiss Microscopy, Thornwood, NY, USA). For each brain region, 3x 321 μm^2 images were acquired from each hemisphere, for a total of 6 data points per region per mouse. CTb+ cells and c-fos+ cells were manually counted, blind to experimental group, using the FIJI Time Series Analyzer plugin. The number of c-fos+ cells that were also CTb+ were expressed as a percentage of total CTb+ cells.

3.2.4 Tracing vHPC monosynaptic inputs to PL neuronal subtypes via immunohistochemistry

Mice were placed in a stereotaxic alignment system (Kopf Instruments) under isoflurane anesthesia. Monosynaptic labelling was achieved using a previously described AAV1-containing viral vector (Zingg et al., 2017). C57BL/6J mice had an AAV1-Cre-containing viral vector (AAV1-hSyn-Cre, titre: 3.5 x 10^{13} GC/mL, plasmid #55637, Addgene, Cambridge, MA, USA, and packaged by Vigene Biosciences) bilaterally infused (0.30 µL/hemisphere) into the vHPC (coordinates AP -3.10 mm, ML ±3.30 mm, DL -4.25 mm, relative to bregma) and a Cre-dependent vector (AAV5-Ef1a-DIO-eYFP, titre: 2.1 x 10^{12}, UNC Vector Core) bilaterally infused (0.30 µL/hemisphere) into the PL (same coordinates as above). Six weeks later, mice were perfused and coronal (50 µm thick) sections containing the PL and vHPC were obtained. Sections were immunostained for anti-GFP by first using blocking solution [10% normal goat serum] (Vector Laboratories Inc, Burlingame, CA, USA) and 2% bovine serum albumin (MP Biomedicals, Santa Ana, CA, USA) in 0.05 M PBS with 0.2% Triton X-100] for 2 hours at room temperature (20°C) and then incubated at 4°C overnight with chicken

anti-GFP (1:3000 dilution, cat# 13970, Abcam, Cambridge, UK). Duplicate sections from each mouse were immunostained for either mouse monoclonal anti-PV (cat# 235, 1:1000, Swant) or mouse monoclonal anti-CaMKIIα (cat# 05-532, 1:1000, MilliporeSigma), together with secondary Alexa Fluor 488 goat anti-chicken antibody (cat# ab150169, 1:500, Abcam) and Alexa Fluor 555 goat anti-mouse. Non-consecutive sections (3 per mouse) containing the PL were mounted and coverslipped with Vectashield HardSet mounting medium with 4',6-diamidino-2-phenylindole (Vector Laboratories) and examined with the aid of an Olympus BX41 microscope (Olympus Corporation).

3.2.5 Intersectional viral tracing of monosynaptic inputs to PL neuronal subtypes

PV-Cre, Sst-Cre, Vip-Cre, and Vglut1-Cre mice were placed in a stereotaxic alignment system (Kopf Instruments) under isoflurane anesthesia. AAV1-mediated anterograde trans-synaptic labelling of vHPC inputs to the PL was achieved as described section **Error! Reference source not found.**, with 2 exceptions: 1) 0.20 µL of AAV1-EF1a-Flp-WPRE (titre: 1.69 x 10^{13} GC/mL, Addgene plasmid #55637, packaged by Vigene Biosciences) were infused into the vHPC; 2) 0.15 µL of AAVdj-hSyn-Con/Fon-Arch3.3-eYFP-WPRE (titre: 2 x 10^{12} vg/mL, Stanford University) were infused into the PL.

Six weeks later, mice were perfused and brains were removed and stored in PBS. Coronal sections (50 µm thick) containing the PL were prepared and immunostained as described above, with the additional rat anti-Sst (cat# MAB354, 1:1000, MilliporeSigma) and Alexa Fluor 555 goat anti-rat (cat# A-21434, 1:500, Invitrogen) antibody.

3.2.6 Anterograde tracing of vHPC inputs to PAG-projecting PL neurons

Mice were placed in a stereotaxic alignment system (Kopf Instruments) under isoflurane anesthesia. Anterograde trans-synaptic labelling was achieved as described in section **Error! Reference source not found.**, using the same AAV1-hSyn-

Cre bilaterally infused into the vHPC and AAV5-Ef1a-DIO-eYFP bilaterally infused into the PL. The same immunohistochemistry procedure was used as described above.

3.2.7 Data analyses

Data are expressed as the mean ± standard error of the mean (SEM), unless stated otherwise. Group differences were analysed using paired (in cases of 2 dependent factors) or unpaired (in cases of 2 independent factors) Student's t-tests, or analyses of variance (ANOVA) (in cases of >2 dependent or independent factors) followed by Fisher's LSD *post hoc* tests. Experiments were powered to match sample sizes typical of the technique reported in the field, though no formal power analysis was performed. The threshold for statistical significance was set at $p < 0.05$.

3.3 Results

Previous studies in mice have implicated the BA and pACC in the contextual version of OFL (Burgos-Robles et al., 2019; Carrillo et al., 2019; Jeon et al., 2010; Keum et al., 2018; Kim et al., 2012; Morozov & Ito, 2018; Pisansky, Hanson, Gottesman, & Gewirtz, 2017; Sakaguchi et al., 2018). In addition, another recent study found that pACC inputs to the BA are required for formation of an observationally acquired fear memory in which the observer is given a 'priming' footshock experience prior to OFL conditioning (Allsop et al., 2018). Beyond these findings, however, remarkably little remains known about the activity of neural substrates underlying OFL.

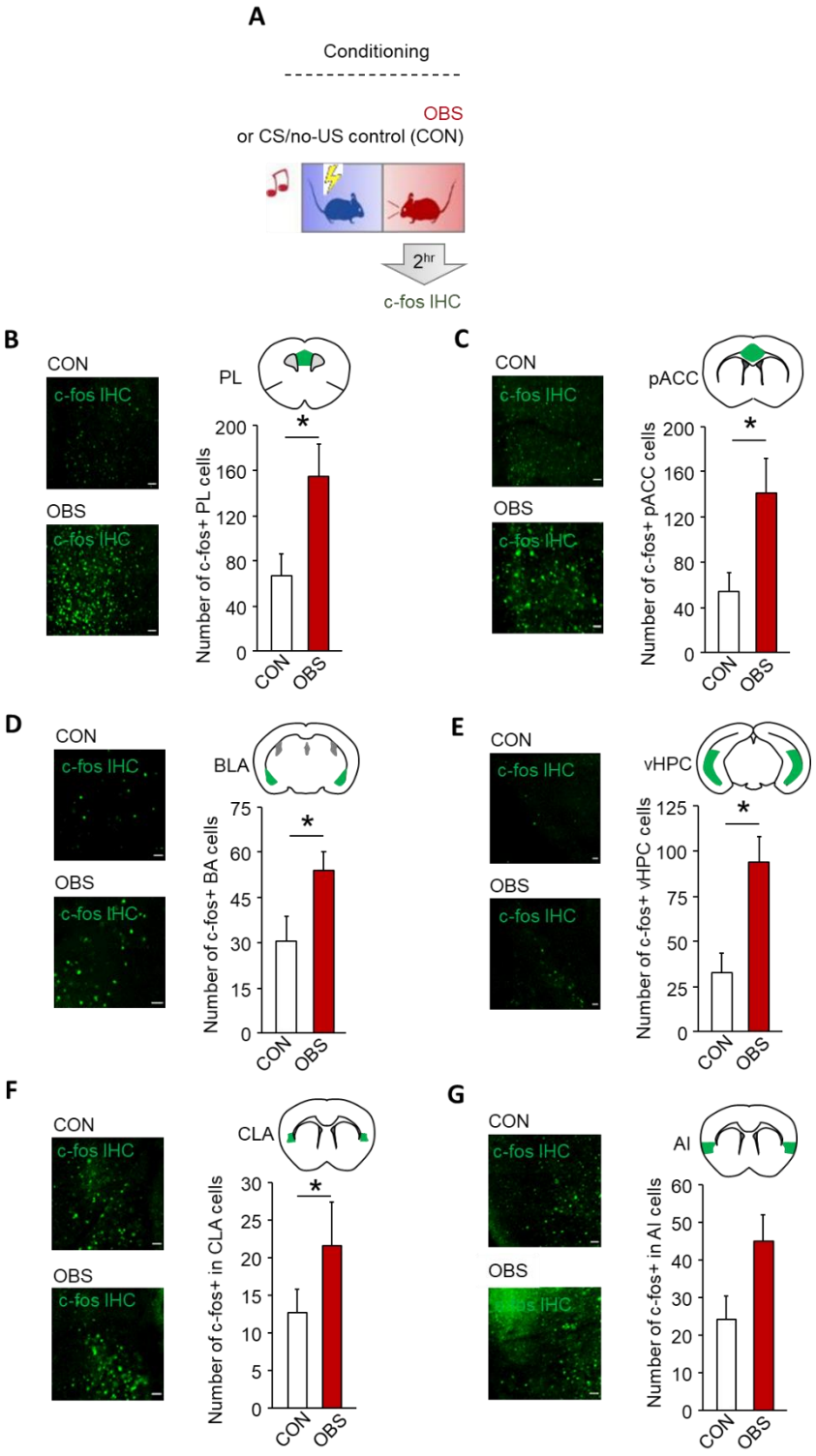
3.3.1 OFL differentially activates key regions of emotion regulation

I began my investigation into the anatomical activity mapping of OFL using c-fos immunohistochemistry, an immediate-early gene indicator of cellular activity, to survey OFL-related neuronal activity in six cortico-limbic regions involved in emotion regulation. Mice were either typical observers or CS/no-US controls. Following conditioning mice were sacrificed and brain sections were immunohistochemically stained for c-fos to quantify the number of c-fos+ cells in a representative area from

each region of interest (Figure 3.1A). Cell counting was conducted blind to treatment. More c-fos+ cells were seen in observers versus CS/no-US controls in the PL (unpaired t-test; $t(26) = 2.67$, $p = 0.013$), pACC ($t(26) = 2.35$, $p = 0.027$), BA ($t(26) = 2.15$, $p = 0.041$), vHPC ($t(23) = 3.36$, $p = 0.003$), and CLA ($t(26) = 2.29$, $p = 0.030$) (Figure 3.1B-F). The only region where there was not a statistical difference between observers and CS/no-US control was the AI ($t(26) = 1.48$, $p > 0.05$) (Figure 3.1G), albeit the difference was in the same direction as the other regions.

Quantification of these brain sections indicate higher c-fos+ counts in the pACC, PL, and BA of observers relative to CS/no-US controls, in line with electrophysiological and optogenetic data establishing a role for these regions in cued-OFL (Ito et al., 2015; Allsop et al., 2018). Additionally, more c-fos+ cells were found in a number of brain regions not previously implicated in OFL, including the vHPC and CLA.

Figure 3.1. C-fos activity in response to OFL conditioning



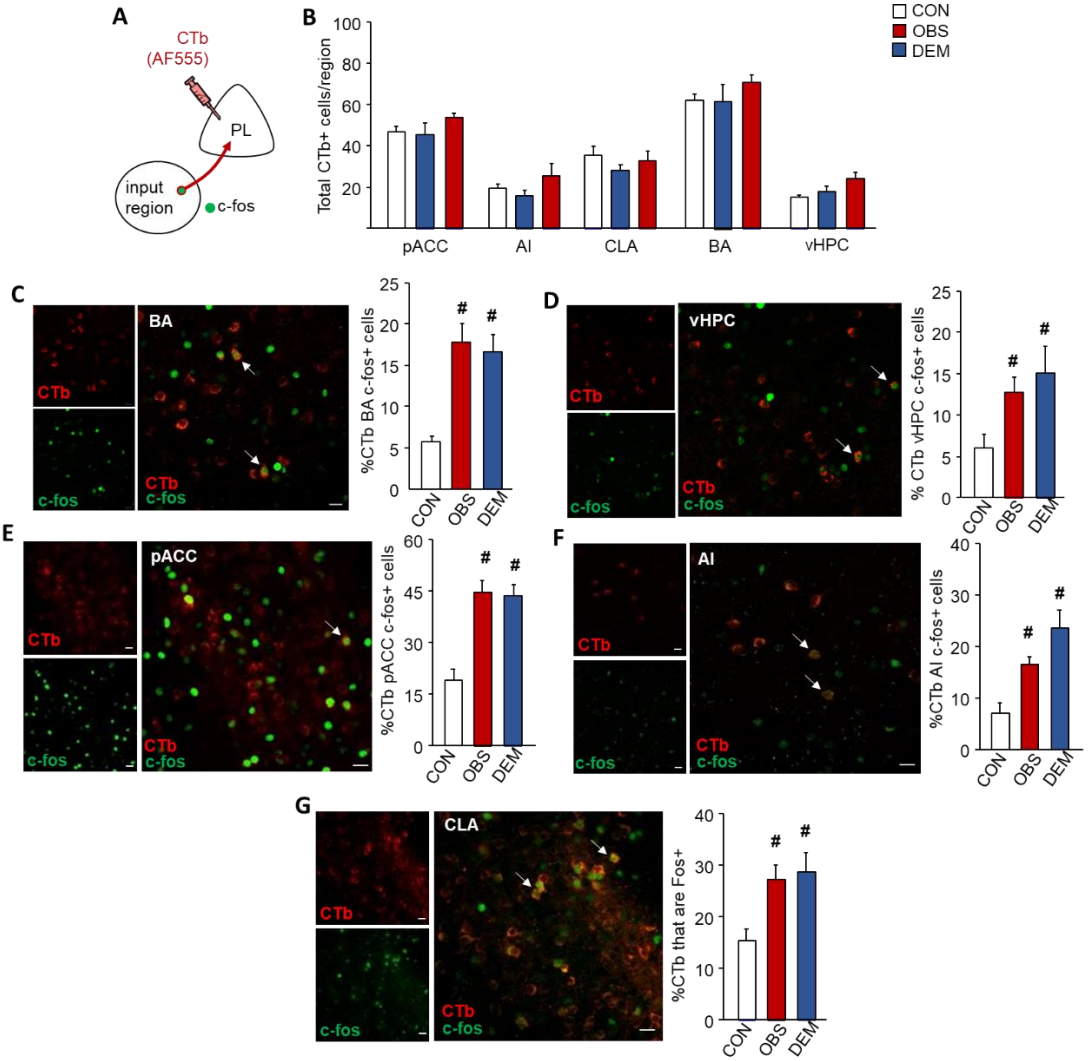
A. Schematic representing observer (OBS) conditioning versus CS/no-US control (CON) followed by c-fos immunohistochemistry (IHC). **B-G.** Representative images of CON and OBS c-fos activity and quantifications of c-fos+ cells in PL, pACC, BA, vHPC, CLA, and AI (n=12-16 mice/group). Scale bars; 500 µm.

3.3.2 Cued-OFL recruits vHPC→PL and BA→PL pathways similarly to DFL

Next, I wanted to visualize and quantify activation of the key input pathways to PL, identified in the previous section as being significantly activated in response to OFL. To this end, I first sought to ascertain the major loci of direct neuronal input to the PL using the fluorescent retrograde tracer, CTb. This revealed projections from multiple distal regions including the BA, CA1 subregion of the vHPC, pACC, AI, and CLA in line with prior tracing studies (Hoover & Vertes, 2007). I then assessed the functional recruitment of these input-pathways during OFL by quantifying the number of CTb-labelled cells expressing c-fos (**Error! Not a valid bookmark self-reference.**A). Mice had either undergone OFL, DFL, or were naïve home-cage controls. As anticipated, there were no differences between groups in the number of counted CTb-labelled cells across regions (ANOVA group-effect: $p > 0.05$ for all 5 regions) (**Error! Not a valid bookmark self-reference.**B).

As compared to naïve home-cage controls, there were a higher percentage of CTb/c-fos+ double-labelled cells after OFL or DFL in the BA (ANOVA group-effect: $F(2,12) = 12.83$, $p = 0.010$), vHPC (ANOVA group-effect: $F(2,12) = 3.87$, $p = 0.050$), ACC (ANOVA group-effect: $F(2,12) = 20.19$, $p < 0.001$), AI (ANOVA group-effect: $F(2,12) = 11.89$, $p = 0.001$), and CLA: (ANOVA group-effect: $F(2,12) = 5.96$, $p = 0.016$) (**Error! Not a valid bookmark self-reference.**C-G). Notably, there were no significant differences in pathway specific activity between demonstrators and observers ($p > 0.05$) suggesting fear learning, regardless of whether it be direct or observed, engages these circuits to a similar degree. Moreover, there were significantly more CTb/c-fos+ cells overall in demonstrators than observers in the pACC, AI, and CLA; however, BA and vHPC inputs to PL were similarly engaged regardless of whether fear conditioning was direct or indirect.

Figure 3.2. Pathway specific cellular activation of CTb inputs to PL



A. Schematic of CTb injection in the PL to retrogradely label inputs to PL followed by c-fos IHC. **B.** Comparison of the number of CTb+ cells across 5 regions projecting to PL between home-cage controls (CON), observers (OBS), and demonstrators (DEM). **C-G.** Representative images of CTb only, c-fos only, and CTb/c-fos+ cells in OFL mice (arrows indicate examples of CTb/c-fos colocalization) as well as graphical comparisons of CTb/c-fos+ cells in BA, vHPC, pACC, AI, and CLA between CON, OBS, and DEM (n=5 mice/group). Scale bars = 20 µm.

3.3.3 vHPC inputs to PL selectively target glutamatergic neurons and PV INs

Drawing upon the pathway specific activation of vHPC inputs to the PL recruited during OFL, I used 4 mouse lines – PV-Cre, Sst-Cre, Vip-Cre, and Vglut1-Cre – to label vHPC monosynaptic inputs to PL. An anterograde virus containing a Flp construct was

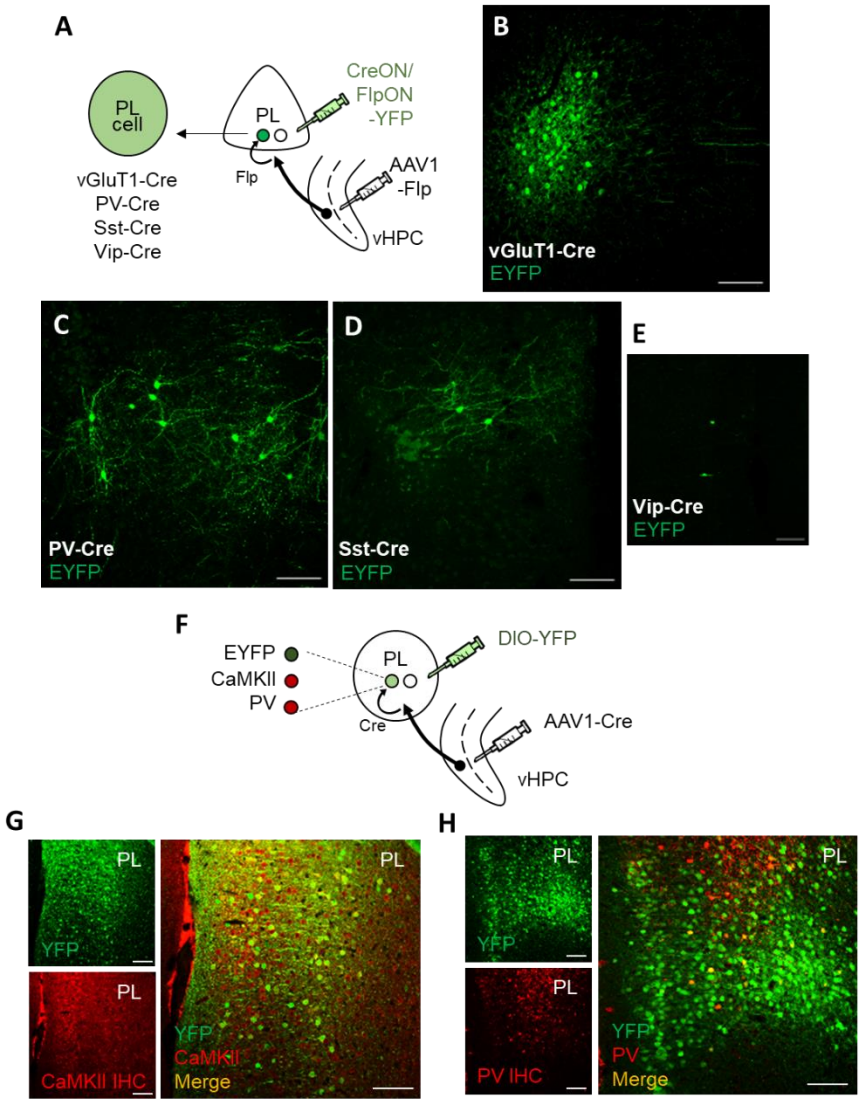
injected into the vHPC and a CreON/FlpON virus fused to YFP into the PL in order to selectively visualize cell-type specific labelling from the vHPC to the PL (Figure 3.3A). Comparatively, vHPC preferentially targeted glutamatergic PNs in the PL as well as a considerable number of PV INs (Figure 3.3B-C). To a much lesser extent, vHPC inputs directly targeted Sst INs and even fewer Vip INs in the PL (Figure 3.3D-E).

To further confirm these findings, a Cre-containing AAV (AAV1) (Zingg et al., 2017) was injected into the vHPC of C57BL/6J mice along with a Cre-dependent synaptophysin-containing virus fused to YFP, into the PL. This viral combination allowed visualization of the terminals of those PL neurons that in turn receive vHPC inputs. The PL neurons were immunostained for either CaMKII or for PV (Figure 3.3F). YFP/AF555 double-labelling can be visualized in CaMKII as well as in PV stained tissue, indicative of vHPC innervation of PV-expressing INs in the PL (Figure 3.3G-H).

3.3.4 vHPC→PL outputs preferentially innervate several regions involved in emotion processing

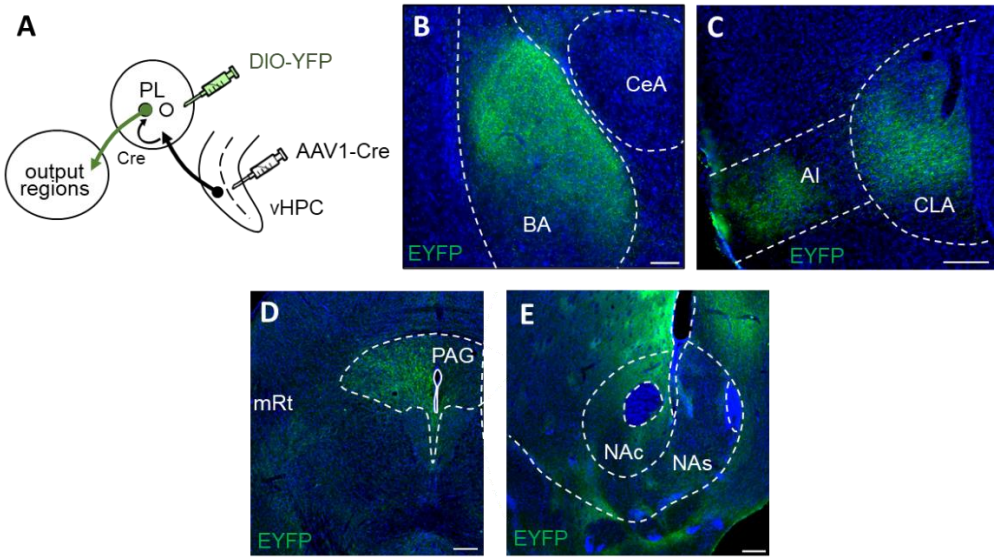
Having established an anatomical circuit by which vHPC selectively targets majority glutamatergic PNs and PV INs in PL, I sought to characterize the output regions of the vHPC→PL pathway. To achieve this, a combination of the previously described labelling with a Cre-dependent vector to visualize regions engaged specifically by vHPC driven outputs from the PL was used (Figure 3.4A). Visualization of strong pathway specific innervation was observed in areas previously found to be engaged in sending outputs to the PL during OFL, such as the BA, CLA, and AI (Figure 3.4B-C) as well as regions known to be involved in the expression of fear, such as the PAG, as well as the NAc (specifically the shell region, NAs) (Figure 3.4D-E). From this an anatomical picture emerges of how OFL might be acquired, processed, and expressed in a functional circuit.

Figure 3.3. Visualization of cell type specific inputs in PL from the vHPC



A. Schematic of combined anterograde virus with 4 different mouse lines (vGlut1-Cre, PV-Cre, Sst-Cre, and Vip-Cre) for identifying PL neuronal subtypes receiving monosynaptic vHPC inputs. **B-E.** vHPC input targets in PL specific cell types (glutamatergic PNs, PV, Sst, and Vip, respectively. **F.** Schematic depiction of combined anterograde virus and immunohistochemistry (IHC) strategy for identifying PL neuronal subtypes receiving monosynaptic vHPC inputs. **G.** YFP-labelled vHPC-innervated (upper left), CaMKII-labelled (lower left), and double-labelled (right) PL projection neurons. **H.** YFP-labelled vHPC-innervated (upper left), PV-labelled (lower left), and double-labelled (right) PL PV+ interneurons. Scale bars = 100 µm.

Figure 3.4. PL outputs from vHPC driven pathway



A. Schematic depiction of combined anterograde virus and immunohistochemistry (IHC) strategy for identifying vHPC→PL outputs. **B.** BA targeted output cells from vHPC→PL projections. **C.** vHPC→PL output to AI and CLA. **D.** PAG output from vHPC→PL pathway. **E.** NAc and NAs inputs from vHPC→PL. Scale bars = 500 µm.

3.4 Discussion

The current chapter is an investigation into the anatomical underpinnings of OFL. Using a combination of tracing and circuit-activity mapping it is possible to determine how OFL may engage certain neural networks. The present study showed that observation of a footshocked conspecific engages several cortico-limbic areas involved in emotion processing including the PL, pACC, AI, CLA, BA, and vHPC. Moreover, PL-projecting neurons in the five latter regions showed activation following OFL. Interestingly, the engagement of vHPC and BA projections to PL did not significantly differ between DFL and OFL, which suggests these pathways may be similarly engaged regardless of direct or indirect exposure. Relying on the relatively recent viral technology of trans-neuronal labelling, I demonstrated a vHPC→PL circuit that preferentially targets glutamatergic neurons as well as PV INs in the PL that requires further functional investigation, as will be discussed in the following chapter. Additionally, this technology allows for the visualization of vHPC→PL output regions,

providing some anatomical direction towards a possible PL output pathway for processing and expressing OFL.

Unsurprisingly, a general increase in cellular activity was found across regions well characterized for their involvement in learning and memory in response to cued-OFL, independent of the effects of context, cue, or novelty of the presence of an unknown mouse. More interesting, however, were the results demonstrating that cued-OFL strongly recruits neurons at various distal loci, including the vHPC and BA, that project directly to the PL. The pattern of PL input activation after OFL overlaps with DFL, suggesting that OFL may tap into many of the same neural circuits as are recruited in direct shock experience. It is important to bear in mind that though these data indicate a degree of equivalency in input recruitment between OFL and DFL, they do not necessarily show that the same projection neurons are engaged or that the same neuronal populations are targeted in the PL. Additionally, the recruitment of inputs from the pACC is noteworthy given this region has been shown to be a critical locus for OFL in prior studies (Burgos-Robles et al., 2019; Carrillo et al., 2019; Jeon et al., 2010; Keum et al., 2018; Kim et al., 2012; Morozov & Ito, 2018; Pisansky et al., 2017; Sakaguchi et al., 2018).

Utilizing the relatively recent technology of trans-synaptic labelling, visualization of cell-type specific innervation in the PL from vHPC was possible. It has long been understood that vHPC provides feedforward inhibitory control of the PL PNs through excitatory signalling to PL INs (Stores-Bayon et al., 2012; Tierney, Degentais, Thierry, Glowinski, & Gioanni, 2004); however, until this technology became available, it was not clear how. Because visualization of monosynaptic connections onto inhibitory PV INs is possible, the inhibitory and excitatory nature of this circuit suggests PV INs may exhibit an internal modulatory inhibition of the PL PNs to affect an inhibitory response. Further investigation into the functional role of the vHPC→PL circuit with regards to OFL is required to understand the pathway's role in a behavioural context, which will be addressed in the following chapter.

Lastly, the circuit mapping of vHPC outputs from the PL identified several downstream targets that may be important for OFL. One possible target is the midbrain PAG. Emerging evidence shows that l/vlPAG-projecting mPFC neurons

regulate a number of processes that are likely important for OFL, including social stress, fear discrimination, and punished conflict (Franklin et al., 2017; Rozeske et al., 2015; Rozeske et al., 2018; Siciliano et al., 2019; Vander Weele et al., 2018). Based on the present anatomical investigation, PL→l/vlPAG circuit could be a pathway to functionally explore as a potential means of processing higher-order environmental information to disambiguate threat from relative safety in OFL.

In summary, by mapping the anatomical circuitry activated during OFL, I have identified several regions and pathways throughout the cortical, limbic, and midbrain areas that may contribute a functional role on OFL. Specifically, pathways previously shown to be involved in fear learning (e.g. vHPC and BLA inputs to the PL) and processing social information (e.g. PL output regions like l/vlPAG and NAc) are of particular interest to explore with functional techniques. The following chapter will expand upon the anatomical framework of OFL by employing optogenetics and fibre photometry to understand the functional circuitry underlying OFL.

Chapter 4 Functional circuitry subserving observational fear learning

4.1 Chapter Introduction

The previous chapter provided an anatomical assessment of the circuitry engaged by OFL. Moreover, the information gleaned from the tracing studies suggested potential pathways and directionality to be explored with functional techniques. Building upon the anatomical investigation, the present chapter uses both optogenetics and Ca^{2+} imaging via fibre photometry technology to study how circuits may be engaged in OFL. The chapter begins by examining the role of the PL in OFL conditioning and retrieval. I then look at input-specific effects of the BA and vHPC on PL during conditioning. Lastly, I use a combination of photometry and optogenetics to understand how the l/vlPAG is engaged when the vHPC→PL circuit is photosilenced. Taken together, these results reveal that OFL is attenuated by vHPC modulation of a cortico-midbrain circuit. Additionally, these data demonstrate that multiple brain circuits are recruited to calibrate OFL, including circuits that serve to actively constrain such learning. Ultimately, these findings could have important implications for understanding how witnessing distressing events can cause trauma-related disorders, and possible avenues for intervention.

4.1.1 Technologies used to explore an OFL functional circuitry

4.1.1.1 Optogenetics

Prior to the rise of optogenetics, interrogation of brain regions and neural circuits were largely studied through lesion, pharmacological, and genetic manipulations. Whilst effective in providing a rudimentary understanding of neural functioning, these strategies are largely insensitive to the cellular specificity and temporal precision present in brain signalling and encoding of information. Conversely, optogenetic technology provides for rapid, causal control of cell-type-

specific activity. Moreover, this can be achieved in freely moving and acting mammals (Aravanis et al., 2007; Deisseroth, 2011; Fenno, Yizhar, & Deisseroth, 2011; Yizhar, Fenno, Davidson, Mogri, & Deisseroth, 2011).

Optogenetic control generally works by using light to modulate neurons that have been infected to express light-sensitive ion channels. There are currently a variety of light-activated molecules in use to excite, inhibit, and record neuronal activity; however, in the present chapter I focus on the use of the inhibitory opsin, archaerhodopsin-3 (eArch), which I have used in all of my optogenetic experiments. eArch is a light-driven outward proton pump that silences expressing neurons with pulses of green light. The proton pump actively transports protons out of the cell, reducing the external pH, which in turn causes membrane potentials to be more negative causing hyperpolarization to silence neurons (Yamanashi et al., 2019). Additionally, eArch has been shown to rapidly recover its light-driven capabilities making it ideal for behavioural tasks requiring repeated inhibition (Chow et al., 2010; Han et al., 2011). It should be acknowledged that there has been speculation that sustained eArch activation produces spontaneous proton release (Mahn, Prigge, Ron, Levy, & Yizhar, 2016); however, this event is only observed after 5 min of continuous inhibition. According to the data collected for this present study, it does not appear that 30 s inhibitory intervals produce this effect.

4.1.1.2 Ca^{2+} imaging via fibre photometry

Around the same time as the development of optogenetics for studying circuit-level contributions to behaviour in behaving animals, also came the approach of recording activity dynamics through Ca^{2+} imaging. Generally, Ca^{2+} fluctuations in neurons correspond to cellular activity including the generation of action potentials, neurotransmitter release, synaptic plasticity, and gene transcription. Ca^{2+} influx occurs in response to electrical activity, making the recording of Ca^{2+} signalling a powerful tool for studying the function of neuronal networking (Chen et al., 2013; Gunaydin et al., 2014; Wang, DeMarco, Witzel, & Keighron, 2021).

In the present study, fibre photometry is specifically used to provide information on the synchronization of activity dynamics within the proposed circuits

underlying OFL. Here, I use genetically encoded Ca^{2+} indicators, GCaMP6 and GCaMP7, which allow for the detection of changes in fluorescent intensity on the order of single action potentials by imaging intracellular free Ca^{2+} (Chen et al., 2013; Gunaydin et al., 2014). While fibre photometry is not as specific as other Ca^{2+} imaging techniques, like two-photon microscopy, these require animals to be head-fixed instead of freely behaving (Svoboda & Yasuda, 2006), while the present technique does not. The size and weight of optical fibres provides more mobility whilst still providing vital information about network activity dynamics within a given circuit.

4.1.2 Brief summary of the proposed OFL circuit

Based upon prior research on fear learning and memory in combination with the anatomical mapping performed in Chapter 3, I investigate the role of the PL in receiving crucial information regarding threat and safety from cortico-limbic areas and output to the midbrain PAG to understand OFL expression.

OFL-related neuronal activation in the PL is of particular importance given prior evidence that this cortical region promotes cued-DFL, mainly through reciprocal connections with the BA (Burgos-Robles et al., 2009; Courtin et al., 2014; Franklin et al., 2017; Karalis et al., 2016; Klavir et al., 2017; Rozeske et al., 2015; Senn et al., 2014; Sotres-Bayon & Quirk, 2010; Tovote et al., 2015). In addition, recent research has shown vHPC projections to the mPFC are crucial for mediating anxiety-related behaviours and certain forms of directly-learned fear (Abbas et al., 2018; Jimenez et al., 2018; Marek et al., 2018; Padilla-Coreano et al., 2016, 2019; Park, Ganella, Perry, & Kim, 2020; Xu et al., 2016). Moreover, the rodent mPFC, which includes the PL, and its possible primate analogue, the pregenual ACC (Brodmann's Area 32) (Laubauch, Amarante, Swanson, & White, 2018), have been ascribed a role in social cognition and cued-OFL in human volunteers (Amodio & Frith, 2006; Apps, Rushworth, & Chang, 2016; Fenno et al., 2011; Franklin et al., 2017; Hill, Boorman, & Fried, 2016; Levy et al., 2018; Olsson et al., 2007; Scheggia et al., 2019). Lastly, inputs to the PAG have been shown not only to mediate the behavioural expression of fear, but also to participate in prediction error modelling contributing to learning (Grahl, Onat, & Büchel, 2018; McNally, Johansen, & Blair, 2011). Together with the PL

activity reported in the previous chapter in response to OFL, the PL appears to be especially well placed to assimilate threat and social cues to support cued-OFL. Functional inputs to and outputs from the PL will be demonstrated within this chapter.

4.2 Materials and Methods

4.2.1 Subjects

Unless otherwise specified, observers were adult male C57BL/6J mice obtained from The Jackson Laboratory (Bar Harbor, ME, USA) and housed in a temperature (72±5°F) and humidity (45±15%) controlled *vivarium* under a 12-hour light/dark cycle (lights on 0630 h). The C57BL/6J strain was chosen because it exhibits robust contextual OFL, as compared to other inbred mouse strains (Keum et al., 2016). For consistency, demonstrators were also adult male C57BL/6J-background mice unfamiliar (i.e., never cohoused) with the observer. Observer mice were singly-housed until at least 2 days before testing.

Additionally B6.Cg-*Pvalb*$^{tm1.1(cre)Aibs}$/J (PV-Cre) mice expressing Cre recombinase directed to PV-expressing cells (JAX strain 012358) were used in specific experiments. Mice were bred in-house via Cre+ male x C57BL/6J matings.

The number of mice used in each experiment is given in the figure legends. All experimental procedures were approved by the NIAAA Animal Care and Use Committee and followed the NIH guidelines outlined in 'Using Animals in Intramural Research' and the local Animal Care and Use Committees.

4.2.2 In vivo PL photosilencing during conditioning

4.2.2.1 Viral vector delivery and optrode implantation

Mice were placed in a stereotaxic alignment system (Kopf Instruments, Tujunga, CA, USA) to infuse virus and implant ferrules under isoflurane anesthesia. A viral construct containing the inhibitory opsin eArch3.0 (rAAV5/CaMKIIα-eArch3.0-

eYFP, titre: 4×10^{12} molecules/mL, obtained from the UNC Vector Core, Chapel Hill, NC, USA) or a control vector (rAAV5/CaMKIIα-eYFP, titre: 2×10^{12} molecules/mL) was bilaterally infused (0.30 µL/hemisphere) over 10 min using a Hamilton syringe and 33-gauge needle. The needle was left in place for a further 5 min to ensure diffusion. The infusion coordinates were AP +1.95 mm, ML ±1.00 mm, DV -1.90 mm, at a 20° angle relative to bregma. During the same surgeries, ceramic ferrules were implanted 0.15 mm above the viral injection site to direct optical fibres at the infected region. Ferrules were secured to the skull using cyanoacetate and acrylic cement. Optic fibres were produced according to (Bukalo et al., 2015; Sparta et al., 2012) (200 µm core, 0.39 NA, uncleaved fibre, Thorlabs, Newton, NJ, USA).

4.2.2.2 In vivo photosilencing

For this and other photosilencing experiments, all behavioural testing began no earlier than 3 weeks after surgery to allow for virus expression. Mice were handled for 2 min a day for 3 days and then habituated to being connected to the optical fibres for 30 min in the home-cage for 2 days prior to testing. Mice then underwent the standard OFL procedure as described in Chapter 2. Green light was shone into the PL of the observer during all 30 of the 30 s tone deliveries during conditioning. Light was delivered through a 65.5 µm bifurcated patch cable (Fiber Optics For Sale Co, Fremont, CA, USA) coupled to a 200-mW, 532-nm, laser system (Opto Engine, Midvale, UT, USA) interfaced with the Med Associates software to deliver TTL pulses to a laser driver in synchrony with the CS. Laser power at the tip of the fibre was measured before each test using a power meter (PM20, Thorlabs, Newton, NJ, USA) and adjusted to achieve 7 mW as was previously determined to effectively inhibit projections whilst reducing the potential of heating brain tissue around the tip of the optrode (Bergstrom et al., 2018). The estimated irradiance at 0.5 mm from the fibre tip was 4.87 mW/mm^2 based on 7 mW, N.A. 0.37 and 561 nm wavelength light (http://www.stanford.edu/group/dlab/cgi-bin/graph/chart.php). The observer was tested for OFL retrieval the next day with cabling attached to control for this potential procedural cue, but without light being shone.

4.2.2.3 *Verification of virus expression and efficacy*

To verify accurate ferrule placement, mice were perfused in the same manner as described in Chapter 3. After suspension in 4% PFA overnight followed by 0.1M PB at 4°C for 1-2 days, 50-μm coronal sections were cut with a vibratome (Leica VT1000 S, Leica Biosystems Inc, Buffalo Grove, IL, USA) and coverslipped with Vectashield HardSet mounting medium with DAPI (Vector Laboratories, Inc, Burlingame, CA, USA). Ferrule location was determined with the aid of an Olympus BX41 microscope. Mice without accurate viral expression or correct ferrule placement were removed from the analysis.

4.2.3 In vivo PL photosilencing during retrieval

The same photosilencing procedure described above was used, except the PL was silenced only during the 5 x 30 s tone presentations during the OFL retrieval day.

4.2.4 In vivo PL PV IN photosilencing

PV-Cre observer mice had 0.30 μl of a viral vector containing Cre-dependent eArch3.0 (rAAV5-DIO–e.Arch3.0-eYFP, titre: 3.4 x 10^{12}, UNC Vector Core) bilaterally injected into the PL (coordinates as above). During the same surgeries, ceramic ferrules were bilaterally implanted into the PL 0.15 mm above the viral injection. Three weeks later, observer mice underwent conditioning and retrieval, with green light shone in the PL during OFL conditioning CS-presentations (as above). At the completion of testing, virus expression and ferrule placement were verified (as above).

4.2.5 In vivo vHPC→PL photosilencing

Observer mice had 0.25-0.40 μL/hemisphere of a viral construct containing the inhibitory opsin eArch3.0 (rAAV5/CaMKIIα–e.Arch3.0-eYFP, titre: 1 x 10^{13} GC/mL, obtained from the University of Pennsylvania Vector Core, Philadelphia, PA, USA) or a control vector (rAAV5/CaMKIIα-eYFP, titre: 4 x 10^{12} GC/mL, obtained from the UNC

Vector Core) bilaterally injected into the vHPC (coordinates AP -3.10 mm, ML ±3.30 mm, DL -4.25 mm relative to bregma). During the same surgeries, ceramic ferrules were bilaterally implanted into the PL to direct fibre patch cords at infected vHPC inputs (coordinates as above). At least 5 weeks after surgery, observer mice underwent the standard OFL procedure, during which green light was shone in the PL throughout the 30 s duration of each conditioning CS, but not retrieval.

4.2.6 In vivo BA→PL photosilencing

Observer mice had 0.25-0.40 µL/hemisphere of a viral construct containing the inhibitory opsin eArch3.0 (rAAV5/CaMKIIα–e.ArchT3.0-eYFP, titre: 1 x 10^{13} GC/mL, from the UNC Vector Core) or a control vector (rAAV5/CaMKIIα-eYFP, titre: 4 x 10^{12} GC/mL, obtained from the UNC Vector Core) bilaterally infused into the BA (coordinates: AP -1.40 mm, ML ±3.23 mm, DL -4.85 mm relative to bregma). During the same surgeries, ceramic ferrules were bilaterally implanted into the PL to direct fibre patch cords at infected BA inputs (coordinates as above). At least 5 weeks after surgery, observer mice underwent the standard OFL procedure, during which green light was shone in the PL throughout the 30 s duration of each conditioning CS, but not retrieval.

4.2.7 In vivo vHPC→PL Ca^{2+} imaging via fibre photometry

Observer mice had 0.60 µL/hemisphere of a viral construct containing the calcium-indicator GCaMP6m (AAVdjEF1a-GCaMP6m, titre: 1 x 10^{13} GC/mL, obtained from Stanford University, Stanford, CA, USA), unilaterally infused into the vHPC (coordinates as above). During the same surgery, a fibre-optic cannula (ferrule: 1.25 mm, fibre: 3.5 mm long, 400 µm core, 0.66 NA, Doric Lenses) was chronically implanted, unilaterally in the same hemisphere, into the PL to direct fibre patch cords at infected vHPC inputs.

Three weeks later, observer mice underwent the standard OFL conditioning and retrieval protocol. Five min prior to retrieval testing, mice were attached to the fibre patch cord and allowed to rest in the homecage. To record fluorescence signals

during testing, a photometry system (Doric Lenses, Quebec, Canada) used 2 continuous sinusoidally-modulated LEDs (DC4100, Thorlabs) at 473 nm (211 Hz) and 405 nm (531 Hz) as a light source to excite GCaMP6s and as an isosbestic autofluorescence signal, respectively, based on previously described methods (Beas et al., 2018). The light intensity at the interface between the fibre tip and brain tissue ranged from 10-15 µW and was kept constant during testing. The LEDs were coupled to a single large core (400 µm), high NA (0.48) optical fibre patch cord and focused onto separate photoreceivers (model# 2151, Newport Corporation, CA, USA).

An acquisition system (RZ5P Processor, Tucker-Davis Technologies, Alachua, FL, USA), equipped with a real-time signal processor controlled the LEDs and independently demodulated the fluorescence brightness due to 473 nm and 405 nm excitation. Data were analysed by applying a least-squares linear fit to the 405 nm signal to align it to the 470 nm signal. The resulting fitted 405 nm signal was then used to normalize the 473 nm as follows: ΔF/F = (473 nm signal − fitted 405 nm signal) / fitted 405 nm signal. The resulting values were aligned to the 5 x CS-presentations and compared to a 30-s baseline period preceding the first CS. At the completion of testing, virus expression and ferrule placement were verified, as above.

4.2.8 In vivo PL→l/vlPAG Ca^{2+} imaging via fibre photometry and concurring vHPC→PL photosilencing with in vivo PL→l/vlPAG Ca^{2+} imaging

Observer mice were placed in a stereotaxic alignment system (Kopf Instruments) under isoflurane anesthesia. A viral vector a containing the Ca^{2+} indicator GCaMP7f (Dana et al., 2019) (pGP-AAVretro-syn-jGCaMP7f-WPRE, titre: 1.8 x 10^{13} vg/mL, Addgene plasmid #104488, kindly provided by Dr. D. Kim & the GENIE Project) was unilaterally infused (0.60 µL) into the l/vlPAG (coordinates AP -4.50 mm, ML +0.45 mm, DV -3.32 mm relative to bregma) and a fibre-optic cannula (ferrule: 1.25 mm, fibre: 3.5 mm long, 400 µm core, 0.66 NA, Doric Lenses) was bilaterally implanted to direct fibre patch cords at the PL (coordinates AP +1.95 mm, ML ±1.00 mm, DV -1.90 mm relative to bregma at a 20° angle). In an additional experiment, an inert GFP-fused vector (rAAV5/CaMKIIα-eYFP, titre: 4 x 10^{12} GC/mL, UNC Vector Core) or an eArch3.0 containing vector (rAAV5/CaMKIIα–e.Arch3.0-eYFP (titre: 1 x

10^{13} GC/mL, University of Pennsylvania Vector Core) was bilaterally infused (0.30 µL/hemisphere) into the vHPC (coordinates AP -3.10 mm, ML ±3.30 mm, DL -4.25 mm relative to bregma).

At least five weeks after surgery, mice were attached to the fibre patch cord in the home-cage for ~2-5 min prior to each testing session. Observers then underwent the standard OFL procedure, with the interval between CS-US presentations (conditioning) and CS presentations (retrieval) increased from 10 to 30 s to ensure CS-evoked Ca^{2+} recovered to baseline between presentations.

To record fluorescence signals, a photometry system (Doric Lenses, Quebec, Canada) used 2 continuous sinusoidally-modulated LEDs (Thorlabs) at 473 nm (511 Hz) and 405 nm (211 Hz) as a light source to excite GCaMP and an isosbestic autofluorescence signal, respectively. Light intensity at the tip of the patch cable (i.e., the interface of patch cable and fibre implant) was in the 50-100 µW range for each channel separately. Each wavelength was adjusted prior to recording to achieve a ~450 mV reading in the corresponding demodulated channels. The LEDs were connected to a mini cube (Doric Lenses) and each bandpass was filtered before being coupled to a single large core (400 µm), high NA (0.48) optical fibre patch cord. Emitted light was unilaterally (from the GCaMP infused hemisphere) projected through the same mini cube, passed through a GFP emission bandpass filter (500-525 nm), and then focused onto a Newport Visible Femtowatt Photoreceiver (Doric Lenses). A RZ5P Processor acquisition system (Tucker-Davis Technologies, Alachua, FL, USA), equipped with a real-time signal processor controlled the LEDs and independently demodulated the fluorescence brightness due to 473 nm and 405 nm excitation.

Additionally, a 200 mW, 594 nm laser (Opto Engine) was coupled via a 65.5 µm patch cable to the same mini cube (which passed 405 nm and 473 nm light, and back-projected fluorescence), and passed laser light through to the same patch cord and fibre implant (in the GCaMP-expressing hemisphere) through which emitted fluorescence was recorded. An optical fibre in the non-GCaMP-expressing hemisphere received 594 nm laser light directly from a second 200 mW laser (Opto Engine). The lasers were interfaced with the Med-PC software (Med Associates)

which delivered TTL pulses to a laser driver in synchrony with the CS. Laser power at the tip of the fibre was adjusted before each test using a PM20 power meter (Thorlabs) to achieve 2.5 mW so as not to interfere with the Ca2+ signalling nor cause photobleaching of the fluorophores (Beas et al., 2018) .

Fluorescence data were analysed by applying a least-squares linear fit to the 405 nm signal to align it to the 473 nm signal. The resulting fitted 405 nm signal was then used to normalize the 473 nm as follows: ΔF= (473 nm signal − fitted 405 nm signal). To assess CS- and US-related changes in activity, z-scores were calculated to compare ΔF values at each time point to the ΔF values during a 5 s period immediately preceding CS onset or US onset, respectively: z= [ΔF-mean(ΔF(t= -5 to 0))]/std, where std is the standard deviation of ΔF values during the pre-event period. To determine CS and US related changes in activity, the 2 s periods pre and post either event were compared. Values for area under the curve were calculated using MATLAB's built-in 'trapz' function, which uses trapezoidal numerical integration to calculate the area under a curve (z-score) between inputted x-values (time) on a graph of z-score versus time.

4.2.9 Behavioural analysis

All animal behaviour was hand-scored every 5 s. Only freezing behaviour is considered in this chapter. Despite recording all behaviours as introduced in section 2.2.3.4, additional behaviours beyond freezing were minimal and did not significantly differ between treatment groups and were not included in final analyses of the data.

4.2.10 Statistical analysis

Group differences were analysed using paired and unpaired Student's t-tests or analysis of variance (ANOVA) followed by Fisher's LSD *post hoc* tests. Mice with freezing scores greater or lesser than 3 standard deviations from the group mean were excluded from analyses. The threshold for statistical significance was set at $p < 0.05$. Information on Ca^{2+} imaging analyses is provided within the method descriptions for each experiment.

4.3 Results

Using a combination of optogenetics and photometry to investigate the functional circuitry underlying OFL, the findings presented here reveal an interacting network of brain structures, including the vHPC, PL, and PAG, that serves to calibrate socially learned fear.

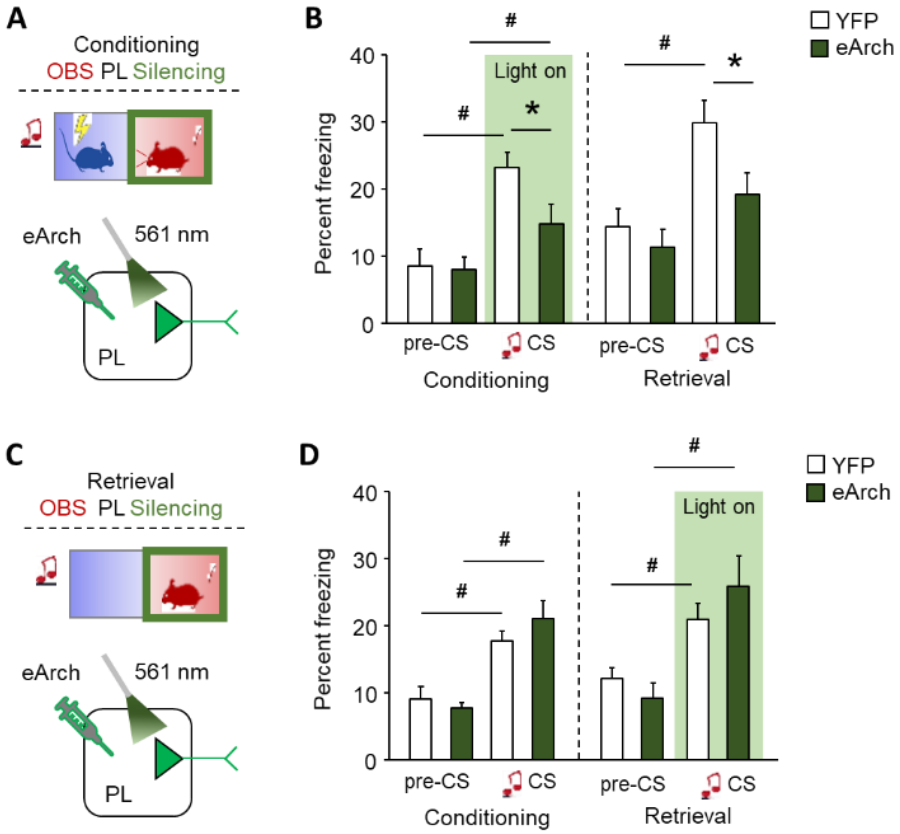
4.3.1 Cued-OFL recruits and requires PL

Based upon anatomical findings from the previous chapter combined with prior research on the role of PL in fear learning, I began by assessing the causal contribution of the PL to cued-OFL. Adopting an in vivo optogenetics approach, I expressed an AAV construct containing the inhibitory opsin, archaerhodopsin (eArch) or a control AAV construct, GFP, in the PL PNs of observer mice. These neurons were then silenced by shining green light, via chronically implanted optic fibres, during each CS presentation of OFL conditioning (Figure 4.1A). PL photosilencing in this manner led to a reduction in CS-evoked freezing, as compared to observers expressing GFP (unpaired t-test: $t(21) = 2.25$, $p = 0.035$; CS vs pre-CS in YFP group, paired t-test: $t(10) = 5.93$, $p < 0.001$; CS vs pre-CS in eArch group, paired t-test: $t(11) = 2.67$, $p = 0.022$). Additionally, this attenuation of freezing was sustained during the light-free retrieval test the following day (unpaired t-test: $t(21) = 2.27$, $p = 0.034$; CS vs pre-CS in YFP group, paired t-test: $t(10) = 6.09$, $p < 0.001$; CS vs pre-CS in eArch, paired t-test: $p > 0.05$) (Figure 4.1B).

The same procedure was then repeated in a separate experiment; however, PL neurons in observer mice were silenced only on retrieval CS presentations, rather than during conditioning (Figure 4.1C). There were no differences in CS-related freezing between eArch and YFP controls during the light-free conditioning test (unpaired t-test: $p > 0.05$; CS vs pre-CS in YFP group, paired t-test: $t(11) = 4.34$, $p = 0.001$; CS vs pre-CS in eArch group, paired t-test: $t(8) = 5.90$, $p < 0.001$). Interestingly, and in contrast to the reduction in freezing behaviour caused by silencing during conditioning, silencing during retrieval was without effect on behaviour (unpaired t-test; $p > 0.05$; CS vs pre-CS in YFP group, paired t-test: $t(11) = 3.17$, $p = 0.009$; CS vs

pre-CS in eArch group, paired t-test: $t(8) = 3.28$, $p = 0.011$) (Figure 4.1D). Together, these optogenetic experiments demonstrate that the PL is recruited and required for the formation, but not the retrieval, of a cued observational fear memory.

Figure 4.1. PL photosilencing during OFL

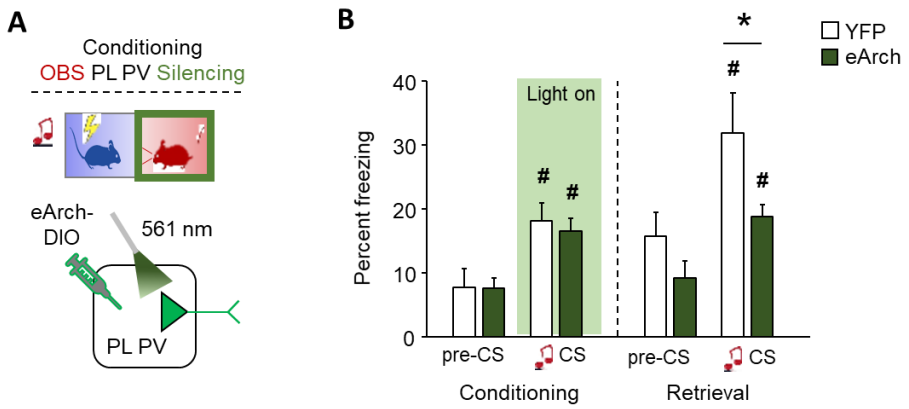


A. Schematic depiction of optogenetic silencing of PL PNs in observer mice (OBS) during CS presentations on OFL conditioning day. **B.** Percent freezing behaviour between YFP controls and eArch experimental OBS groups during PL silencing in conditioning followed by light-free retrieval (n = 11-12/group). **C.** Schematic depiction of optogenetic silencing of PL PNs in OBS during CS presentations on OFL retrieval test. **D.** Percent freezing behaviour between YFP and eArch OBS groups during light-free conditioning followed by PL silencing during OFL retrieval (n = 9-12/group). **p<0.001, *0.001<p<0.05 (2-tailed, between groups); #p<0.05 (2-tailed, within groups). Quantitative data are means ± SEM.

4.3.2 PV INs in PL are involved in the consolidation of OFL

Motivated by the finding that PL plays a causal role in the formation of an observed fear memory, coupled with anatomical data from the previous chapter showing that vHPC inputs to the PL target PV INs, I next investigated the role of PV INs in the PL during OFL. I used the same in vivo optogenetics approach as described in the previous section; however, PV-Cre mice were used. This required a different AAV with a DIO construct in order to target PV INs selectively. These INs were silenced by shining green light during each CS presentation of OFL conditioning (Figure 4.2A). In contrast to silencing PNs in the PL, there were no differences in CS-related freezing between eArch and YFP controls during the photosilenced conditioning day (unpaired t-test: $t(18) = 0.52$, $p > 0.05$; CS vs pre-CS in YFP group, paired t-test: $t(8) = 2.37$, $p = 0.045$; CS vs pre-CS in eArch group, paired t-test: $t(10) = 4.08$, $p = 0.002$) . However, there was a significant reduction in eArch freezing behaviour during the light-free retrieval test (unpaired t-test: $t(18) = 2.19$, $p = 0.042$; CS vs pre-CS in YFP group, paired t-test: $t(8) = 3.85$, $p = 0.005$; CS vs pre-CS in eArch, paired t-test: $t(10) = 2.98$ $p = 0.014$) (Figure 4.2B). This suggests that while PV INs in the PL play a role in the consolidation of OFL that does not affect behavioural freezing immediately, they do underpin memory formation.

Figure 4.2. PV IN in PL silencing during OFL conditioning



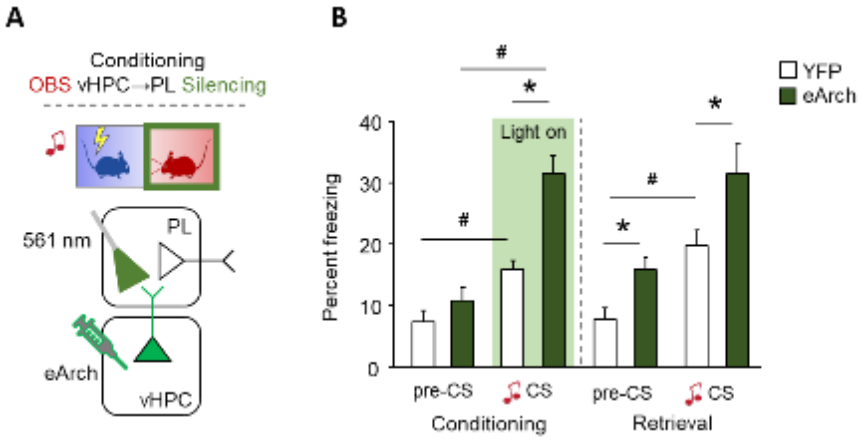
A. Schematic depiction of optogenetic silencing of PV INs in PL OBS during CS presentations on OFL conditioning day. **B.** Percent freezing behaviour between YFP controls and eArch experimental OBS groups during PV silencing in conditioning followed by light-free retrieval (n = 11-12/group).

4.3.3 vHPC inputs to PL constrain OFL

Drawing upon the CTb+c-fos data from the previous chapter showing that vHPC inputs to the PL are activated in response to OFL led to my investigation of whether a vHPC→PL pathway is causally involved in OFL. Again, using an in vivo optogenetics approach, vHPC-projection neurons of observers were transfected with eArch (or a YFP control virus) and optic fibres were implanted to direct green light at the PL and photosilence vHPC→PL cells during each conditioning CS (Figure 4.3A). Strikingly, silencing the pathway caused a robust increase in CS-related freezing in eArch-expressing observers relative to YFP controls during CS-related freezing on OFL conditioning (unpaired t-test: $t(25) = 4.74$, $p < 0.001$; CS vs pre-CS in YFP group, paired t-test: $t(13) = 4.87$, $p < 0.003$; CS vs pre-CS in eArch group, paired t-test: $t(12) = 8.05$, $p < 0.001$). Moreover, a further increase in baseline (contextual) (unpaired t-test: $t(25) = 2.93$, $p = 0.007$) and CS-related freezing during the light-free retrieval test between groups was observed (unpaired t-test: $t(25) = 2.28$, $p = 0.032$; CS vs pre-CS in YFP group, paired t-test: $t(13) = 4.93$, $p < 0.001$; CS vs pre-CS in eArch group, paired t-test: $t(12) = 3.29$, $p = 0.006$) (Figure 4.3B). This reveals not only that vHPC inputs to

the PL are causally involved in OFL, but that this input normally serves to limit the degree of observational fear learning.

Figure 4.3. Silencing vHPC inputs to PL during OFL



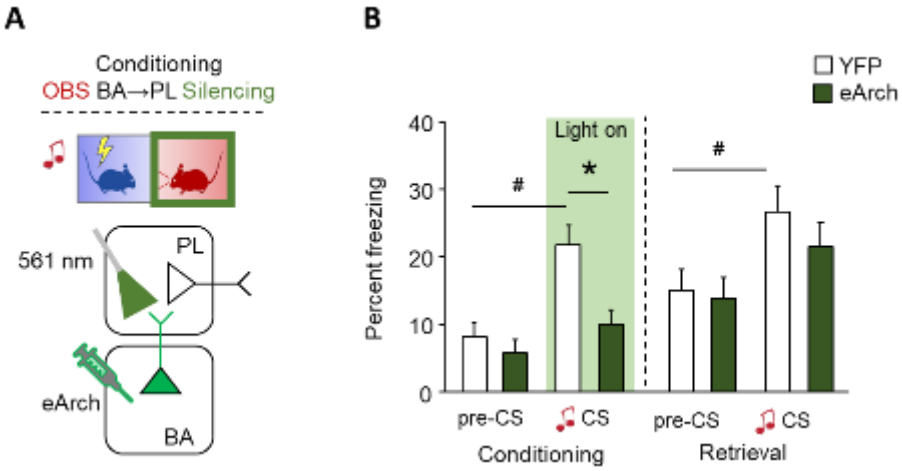
A. Schematic depiction of optogenetic photosilencing of PL-projecting vHPC neurons during conditioning. **B.** Percent freezing behaviour between YFP controls and eArch experimental OBS groups during vHPC→PL silencing in conditioning followed by light-free retrieval (n = 13-14/group).

4.3.4 Photosilencing BA inputs to PL does not have a lasting impact on OFL

I next asked whether the inhibitory influence of the vHPC→PL pathway was common to the BA input to the PL. I replicated the same optogenetic procedures as above, but with eArch bilaterally injected into the BA instead of the vHPC (Figure 4.4A). In contrast to the effects seen in the vHPC→PL circuit, silencing the BA→PL pathway produced a *decrease* in freezing behaviour during conditioning (unpaired t-test: $t(21) = 3.45$, $p = 0.002$; CS vs pre-CS in YFP group, paired t-test: $t(11) = 5.17$, $p < 0.001$; CS vs pre-CS in eArch group, paired t-test: $t(10) = 2.09$, $p = 0.064$), which did not persist into retrieval (unpaired t-test: $t(21) = 0.99$, $p = 0.329$; CS vs pre-CS in YFP group, paired t-test: $t(11) = 2.28$, $p = 0.044$; CS vs pre-CS in eArch group, paired t-test: $t(10) = 2.79$, $p = 0.019$) (Figure 4.4B). While the effect during conditioning is consistent with the pro-fear role ascribed to the BA→PL pathway in direct fear

(Burgos-Robles et al., 2009; Jin & Maren, 2015; Klavir et al., 2017; Senn et al., 2014; Sotres-Bayon et al., 2012), unlike in DFL, the inability of silencing to cause a lasting decrement in cued-OFL suggests that a loss of BA innervation can be compensated for by other inputs, for example, from the thalamus (Keum & Shin, 2019).

Figure 4.4. Silencing BA inputs to PL during OFL



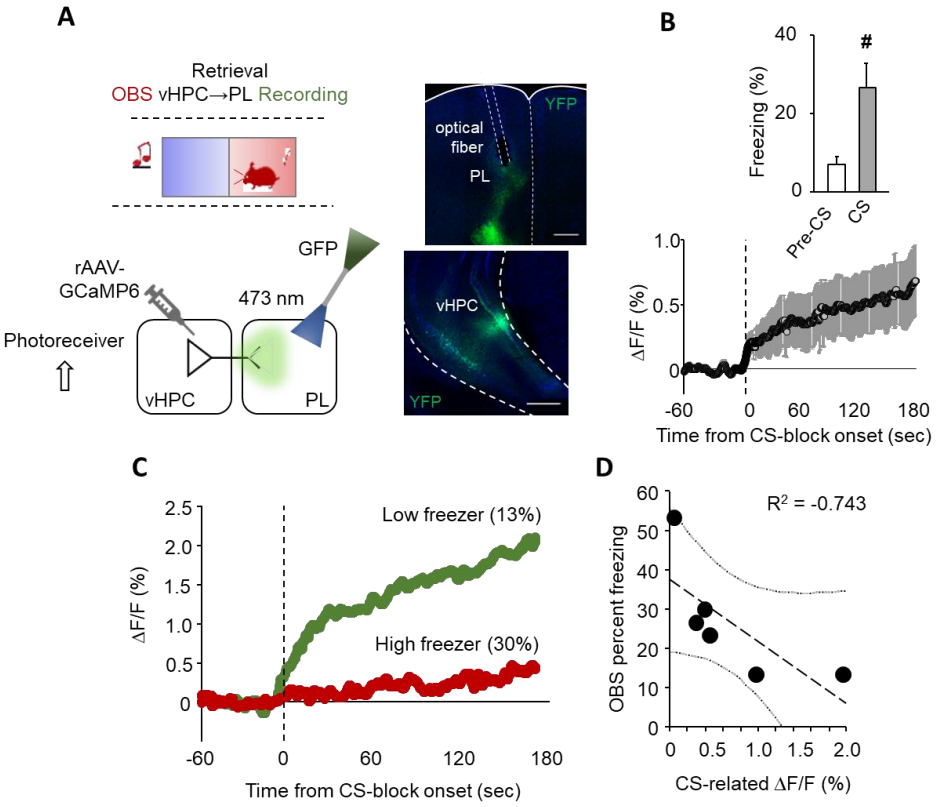
A. Schematic depiction of optogenetic photosilencing of PL-projecting BA neurons during conditioning. **B.** Percent freezing behaviour between YFP controls and eArch experimental OBS groups during BA→PL silencing in conditioning followed by light-free retrieval (n = 10-11/group).

4.3.5 vHPC→PL Ca^{2+} signals during retrieval negatively correlate with cued-OFL

To further support the optogenetic finding that photosilencing vHPC→PL amplifies OFL, I performed a subsequent experiment using Ca^{2+} imaging via fibre photometry recordings on vHPC input to PL neurons during OFL retrieval (Figure 4.5A). This analysis revealed a significant increase in Ca^{2+} responses at the onset of the first tone (first 5 s of CS vs pre-CS, repeated measures ANOVA: $t(21) = 3.45$, $p = 0.002$). This was followed by a sustained ramping of the response across the remainder of the session that did not dip during the 5 s interval between each of the 5 CS presentations (repeated measures ANOVA: $t(21) = 3.45$, $p = 0.002$) (Figure 4.5B).

Further examination of the Ca^{2+} response and level of CS-evoked freezing in each mouse revealed a significant negative correlation between the two measures, such that the highest Ca^{2+} signals were associated with the lowest freezing (R^2 = -0.74, p = 0.002) (Figure 4.5C-D). Interestingly, in vivo neuronal recording studies in rodents and non-human primates have detected population-level mPFC ramping responses, albeit on the order of seconds, not decaseconds, and linked these to timing and action preparedness (Narayanan, 2016).

Figure 4.5. Ca^{2+} imaging of vHPC inputs to PL during retrieval



A. Schematic depiction of Ca^{2+} imaging via fibre photometry of PL inputs from vHPC during retrieval (left) and example images of vHPC injection location and PL viral expression (scale bars = 500 µm) (right). **B.** Changes in freezing behaviour from baseline to averaged CS response (top) and change in Ca^{2+} signal from the onset of the first CS (bottom). **C.** Comparison of Ca^{2+} signals between OBS with high freezing behaviour (mean = 30%) vs low freezing behaviour (mean = 13%). **D.** Correlation between OBS freezing behaviour and Ca^{2+} signal changes in response to CS.

4.3.6 Ca^{2+} imaging shows OFL is signalled by PL projections to l/vlPAG

While the above data identify a critical contribution of PL projection neurons to OFL, the functional target of these neurons remains unknown. One possible target proposed in Chapter 3 is the midbrain PAG. Emerging evidence shows that l/vlPAG-projecting mPFC neurons regulate a number of processes that are likely important for OFL, including social stress, fear discrimination, and punished conflict (Franklin et al., 2017; Rozeske et al., 2015; Rozeske et al., 2018; Siciliano et al., 2019; Vander Weele et al., 2018). To test for a role of the PL→l/vlPAG pathway in OFL, endogenous in vivo correlates of behaviour were assessed using fibre photometry to image Ca^{2+} activity in l/vlPAG-projecting PL neurons during OFL. A retrogradely-traveling Ca^{2+} indicator, GCaMP7f (Dana et al., 2019), was injected into the l/vlPAG and optical fibres were chronically implanted in the PL (Figure 4.6A).

Remarkably, when the GCaMP signal (normalized to an isosbestic control) was aligned to the presentation of the CS and the US, robust event-related Ca^{2+} activity in the PL→l/vlPAG neurons of observers was detected (Figure 4.6B-E). Figure 4.6C demonstrates the population change in Ca^{2+} activity to the CS (paired t-test vs pre-CS: $t(4) = 3.74$, $p = 0.010$) while Figure 4.6E shows this activity change to the US (paired t-test vs pre-CS: $t(4) = 6.08$, $p = 0.002$) across all 30 OFL conditioning trials.

The Ca^{2+} activity to the CS could simply reflect a sensory response to the tone, or may represent the accumulation of associative strength to the CS as a predictor of threat. To parse these possibilities, CSs were subdivided into the first and last 5-CS blocks so Ca^{2+} activity could be examined during the early and late phases of learning. This revealed that while modest (and non-significant) CS-related Ca^{2+} activity was evident on early trials (paired t-test vs pre-CS: $t(4) = 1.81$, $p = 0.072$), activity was robust by late conditioning trials (paired t-test vs pre-CS: $t(4) = 2.52$, $p = 0.033$), consistent with a strengthening of the PL→l/vlPAG pathway engagement as observers assigned value to the CS (Figure 4.6F).

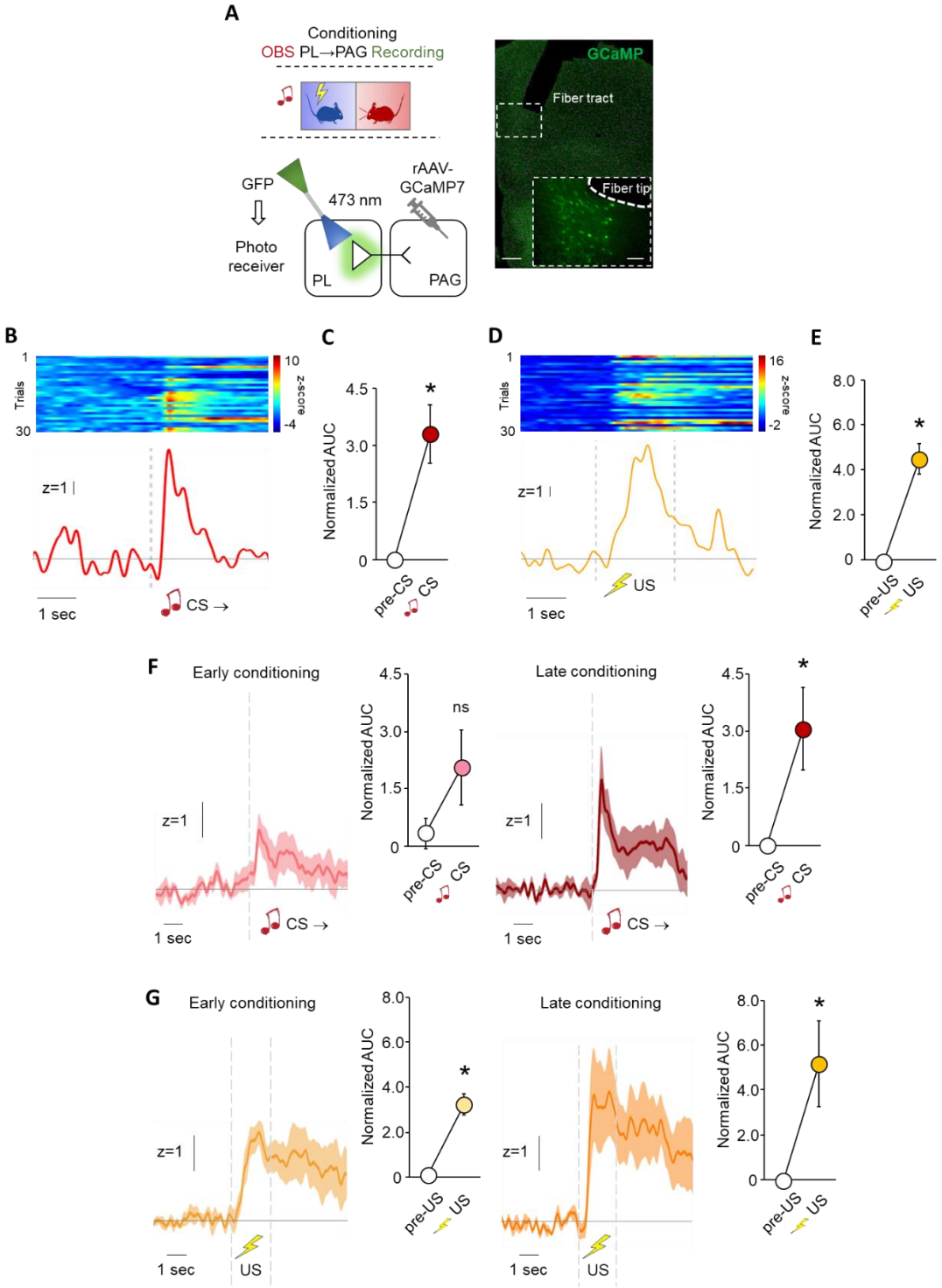
In contrast to the CS-related activity, the US-related activity of PL→l/vlPAG neurons was evident from early conditioning in observers. Conditioning USs were also subdivided into the last 5-US blocks and Ca^{2+} activity was examined in the early and late phases of learning (early, paired t-test vs pre-CS: $t(4) = 5.16$, $p = 0.003$; late,

paired t-test vs pre-CS: $t(4) = 2.83$, $p = 0.026$) (Figure 4.6G). This activity could reflect responses of these neurons to defensive reactions exhibited by demonstrators during shock delivery (e.g., flinching, jumping, vocalizing). A recent study detected enhanced Ca^{2+} responses to directly-experienced footshocks in mPFC neurons projecting to the PAG (Vander Weele et al., 2018). This is similar to what Amemori and Greybiel (2012) demonstrated with direct learning during a cost-benefit task using aversive air puffs with human participants. The present finding that these neurons are also responsive to input experienced vicariously speaks to the high sensitivity of this pathway to aversive stimuli. As with the US responses found in observers, CS-related activity could reflect the freezing behaviour of the demonstrators, or the observers to the CS.

4.3.7 OFL promotes vHPC inhibition of PL→l/vlPAG pathway

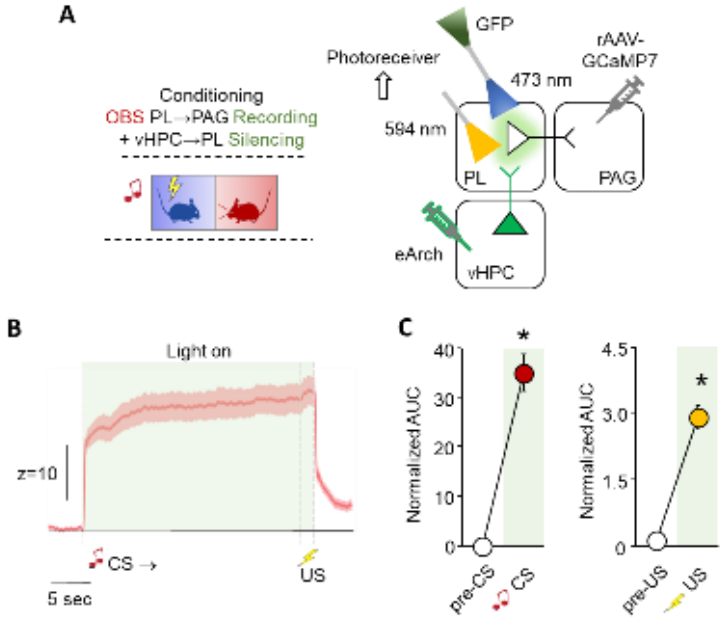
Finally, to test how inhibition of the vHPC→PL impacts output activity to the l/vlPAG, I used a combination of in vivo optogenetics and Ca^{2+} imaging (as described separately in sections 4.3.3 and 4.3.6, respectively). During OFL conditioning the outputs from the vHPC to the PL were photosilenced whilst Ca^{2+} imaging of the l/vlPAG inputs from the PL were simultaneously recorded (Figure 4.7A). A sustained population increase in Ca^{2+} activity was observed in outputs to the l/vlPAG in response to vHPC inhibition that is specific to the CS and US (Figure 4.7B). Moreover, Ca^{2+} activity showed a significant increase in response to the CS (paired t-test vs pre-CS: $t(11) = 9.34$, $p < 0.001$) and a further increase in response to the US (paired t-test vs pre-US: $t(11) = 9.98$, $p < 0.001$) (Figure 4.7C). This demonstrates a crucial role of OFL attenuation by the vHPC modulation of a cortico-midbrain circuit. Moreover, the combined use of both photosilencing and Ca^{2+} imaging exhibits that vHPC, PL, and l/vlPAG are recruited to calibrate OFL and expression by actively constraining observational conditioning.

Figure 4.6. Ca^{2+} imaging of PL output to l/vlPAG during OFL conditioning



A. Schematic depiction of retrograde GCaMP viral strategy for measuring activity in PAG-projecting PL neurons during OFL as well as representative images of GCaMP labelling in the l/vlPAG-projecting PL neurons proximal to the optic fibre (scale bar = 200 µm, inset = 50 µm). **B-E.** Example heat map and corresponding recording trace for Ca^{2+} activity aligned to the CS (**B**) and US (**D**). Population change in Ca^{2+} activity to the CS (**C**) and US (**E**) across all 30 conditioning trials. **F-G.** Population traces and quantified change in Ca^{2+} activity to the CS (**F**) and the US (**G**) on early and late conditioning trials (n = 5 mice).

Figure 4.7. Concurrent silencing vHPC→PL while recording from PL→l/vlPAG



A. Schematic depiction of eArch photosilencing vHPC inputs whilst using a retrograde GCaMP viral strategy for measuring activity in l/vlPAG-projecting PL neurons during OFL. **B.** Ca^{2+} activity response of PAG outputs from vHPC→PL silencing during 30 s CS+US delivery period. **C.** Averaged Ca^{2+} increase from PL outputs to PAG prior to CS and during CS presentation as well as prior to US and during US (n = 12).

4.4 Discussion

In the present chapter, I describe how a fear memory acquired solely through observation recruits multiple brain circuits, including pathways that serve to actively constrain such learning. Using both optogenetics and Ca^{2+} imaging, I found that the PL is activated and functionally necessary for the acquisition of cued-OFL. Moreover, PV INs in the PL may be involved in the consolidation of these memories as this population is necessary for the retrieval of OFL, which is further supported by the previous chapter's tracing study finding that vHPC inputs to the PL selectively target PV INs. In addition to the individual role of the PL in OFL, I demonstrated that PL inputs from the vHPC, rather than the BA, are critical for lasting constraint of OFL behaviour. This inhibitory control of vHPC over the PL is additionally responsible for calibrating PL→l/vlPAG driven expression of OFL.

4.4.1 PL recruitment to integrate threat and social information in OFL

Prior work has shown an important role for the PL in both cued- and contextual-DFL (Milad & Quirk, 2012; Rozeske et al., 2015; Tovote et al., 2015). There is also emerging evidence that the PL plays a particularly important part in learning and expressing appropriate fear associations under conditions of uncertainty and ambiguity. For example, disrupting the PL, or its outputs, impairs the ability to discriminate between safety and threat in behavioural settings such as fear extinction (Burgos-Robles et al., 2009; Fitzgerald, Seemann, & Maren, 2014; Livneh & Paz, 2012; Ye et al., 2017) and contextual- and cued-discrimination (Antoniadis & McDonald, 2006; Klavir, Genud-Gabai, & Paz, 2013; Likhtik, Stujenske, Topiwala, Harris, & Gordon, 2014; Rozeske et al., 2018; Xu & Südhof, 2013). These observations could reflect a major role for the PL in utilizing available environmental information to direct attention to appropriate predictors of threat, as well as to gate learning and behaviour accordingly (Furlong, Cole, Hamlin, & McNally, 2010; Marquis, Killcross, & Haddon, 2007; Sharpe & Killcross, 2014; Sharpe & Killcross, 2015a; Sharpe & Killcross, 2015b). This integrating-arbitrating function of the PL could be especially important in situations where there are conflicting indicators of danger and safety (Likhtik & Paz, 2015), such as with OFL where there are indicators of danger from the defensive responses of the demonstrator, yet no harm is physically experienced. The present data suggests OFL may be a special case of an ambiguous fear memory that is dependent upon the PL, and particularly liable to PL disruption.

Together with the c-fos data reported in Chapter 3, this research demonstrates that the PL is recruited and required for the formation, but not retrieval, of cued-observational fear memory. This dissociation between a requirement for memory acquisition, but not retrieval, contrasts with the necessity of the PL for the expression of DFL (Burgos-Robles et al., 2009), but agrees well with the reported contributions of the pACC and its inputs to the BA in both the cued- and contextual-versions OFL (Allsop et al., 2018; Jeon et al., 2010; Kim et al., 2012). These findings suggest the PL and pACC are recruited to integrate threat cues and social information to instantiate observational fear memory. However, once formed, these mPFC regions are dispensable for memory retrieval as other brain regions may fulfil this function.

4.4.2 PL inputs from vHPC restrict OFL

The set of optogenetics experiments examining BA and vHPC inputs to the PL showed that these regions have opposing influences over OFL. The finding that silencing vHPC inputs to the PL enhanced fear during OFL conditioning and retrieval demonstrates that this pathway normally exerts an inhibitory influence over PL-mediated OFL. This is further demonstrated by the Ca^{2+} imaging data during retrieval, which showed that activity of the vHPC→PL pathway is negatively correlated with freezing behaviour. These behavioural effects are in keeping with electrophysiological data indicating that HPC neurons exert a strong inhibitory influence on mPFC activity through the targeting of local cortical INs (Ishikawa & Nakamura, 2003; Tierney et al., 2004), although this is balanced by some degree of excitatory influence (Padilla-Coreano et al., 2016). It is also reminiscent of earlier evidence that vHPC projections to PL decreased DFL in the presence of safety cues and after extinction, but did not impact direct fear when such signals were absent (Burgos-Robles et al., 2009; Likhtik & Paz, 2015; Meyer-Mueller et al., 2020; Sotres-Bayon et al., 2012). Recent work has shown that vHPC projections to the IL, a cortical subregion that plays an opposite, fear-reducing role to the PL (Milad & Quirk, 2012), are recruited and necessary for renewal of extinguished fear by exposure to a novel context (Jin & Maren, 2015; Marek et al., 2018; Wang, Jin, & Maren, 2016). Therefore, vHPC→PL neurons may be recruited to reduce fear in situations where there is conflicting information about whether a cue predicts danger or safety. A similar conflict may likely occur in cued-OFL where the observer learns about a source of danger, but does so from a position of relative safety. It is thus suggested that this circuit may convey higher-order information that disambiguates threat signalled by the behaviour of the demonstrator from the absence of concomitant, directly experienced harm.

A reduction in OFL conditioning after silencing BA→PL neurons is generally consistent with the observation that activity in this pathway is associated with high fear states and emotional ambiguity (Burgos-Robles et al., 2017; Senn et al., 2014). There is also ample evidence that the reciprocal, PL→BA pathway, promotes fear (Courtin et al., 2013; Likhtik & Paz, 2015) and is strengthened after contextual-OFL

(Ito et al., 2015). Nonetheless, it is notable that BA→PL silencing during OFL conditioning only modestly attenuated cued-OFL retrieval the next day. This could reflect the inability to fully silence all BA→PL neurons or a potential contribution of other inputs, such as the thalamus, to support learning.

4.4.3 vHPC inhibition of PL→l/vlPAG pathway in OFL

Using Ca^{2+} imaging, this chapter identifies that PL projections to the l/vlPAG signal OFL. Emerging evidence shows that l/vlPAG-projecting mPFC neurons regulate a number of processes that are likely important for OFL, including social stress, fear discrimination, and punished conflict (Franklin et al., 2017; Rozeske et al., 2015; Rozeske et al., 2018; Siciliano et al., 2019; Vander Weele et al., 2018). The present finding illustrates the responsiveness of pathway-specific activity of PL→l/vlPAG in a vicarious setting, indicating the sensitivity of this connection to aversive information.

Moreover, when the vHPC-PL circuit was silenced during CS presentation and US delivery, the data indicated an increase in PL→l/vlPAG activity. This suggests that OFL biases vHPC inputs to l/vlPAG-projecting neurons in favour of inhibition. This is consistent with the finding that vHPC inputs to the PL are recruited to in order to limit OFL, together suggesting that this effect may occur by increasing inhibition of output to the l/vlPAG.

4.4.4 Conclusions

The work described in Chapters 2-4 defined a new behavioural assay to study cued-OFL. Moreover, combining immediate-early gene mapping, anatomical tracing, and in vivo optogenetics and Ca^{2+} imaging, I defined novel neuronal circuits contributing to OFL in mice. Together these findings revealed an interacting network of brain structures - including the vHPC, PL, and l/vlPAG - that served to calibrate socially-learned fear. Together, the data demonstrated that OFL was attenuated by vHPC modulation of the PL→l/vlPAG circuit. In effect, this disynaptic pathway would provide a safety signal during OFL which, when compromised, would amplify the formation of fear memories for witnessed traumatic events. The identification of this

critical pathway for gating cued-OFL opens the door for future research into the mechanistic aspects of the observed inhibitory control of vHPC inputs to the PL, as well as PL outputs to the l/vlPAG, on OFL expression. Moreover, there is great potential to inform translational research as human neuroimaging studies suggest that the neural circuits underlying OFL intersect closely with those in mice (Debiec & Olsson, 2017; Olsson et al., 2020). Whilst human neuroimaging falls beyond the scope of this thesis, the subsequent chapter will present the validation of an OFL paradigm for humans and examine the impact of dispositional traits on OFL. Ultimately, integrating OFL research across species can greatly advance our understanding of how OFL is acquired and processed leading us to best address stress- and trauma-related disorders born from observation of distressing events.

Chapter 5 Individual differences contributing to observational fear learning

5.1 Chapter Introduction

The previous experimental chapters have focused on behavioural and neural mechanisms underlying OFL in rodents. I began by behaviourally characterizing OFL and then identified novel regional and circuit contributions involving vHPC modulation of PL for OFL and behaviour. Additionally, I proposed a potential role of the PAG as an output region of the vHPC-PL circuit in modulating OFL discrimination.

In the present study the focus turns towards developing a behaviourally translational approach for human participants and validating that task. Fear learning is well studied across species, but the development of vicarious fear paradigms is in its relative infancy. Because of this, there is not yet a standardized paradigm that consistently permits the characterization of OFL. The first part of this chapter, addressed in Study 1, establishes and validates an OFL task for human participants with either a moderate or intense reaction to shocks by an actor on screen. Detailed computational modelling of trial-by-trial variation in behaviour is then applied to examine learning rate and choice variability for OFL as well as the degree to which these are influenced by the intensity of response in the shock reaction. The second aim of the chapter, addressed in Study 2, builds on this paradigm validation with data collection from a large participant sample in order to understand the association between characteristics of OFL (including learning rate variability generated from computational modelling) and dispositional traits related to emotionality (fear), namely anxiety and psychopathic traits.

5.1.1 OFL in humans

Learning from others' experiences is a highly evolutionarily conserved mechanism observed across animal species, including humans (Askew & Field, 2007; Debiec & Olsson, 2017; Gariépy et al., 2014; Helsen, Vlaeyen, & Goubert, 2015; Jeon

& Shin, 2011; Mineka & Öhman, 2002; Olsson & Phelps, 2007). Such social learning is a salient means of acquiring important information related to threat and safety (Adolphs, 2013) and may contribute to more complex behaviours, such as empathy (Bernhardt & Singer, 2012; de Waal & Preston, 2017; Keum et al., 2018; Lockwood, Apps, Valton, Viding, & Roiser, 2016; Olsson et al., 2016). It is also one potential mechanism underlying development of disordered behaviour in conditions such as anxiety and psychopathy, which are characterised by atypical patterns of learning and threat processing (Decety et al., 2013; Mkrtchian et al., 2017; Seara-Cardoso et al., 2015).

To date, the most commonly used OFL protocol in humans has deployed a setup where a participant (the observer) watches a video of an actor (the demonstrator) presented with two distinct stimuli (Golkar et al., 2015; Golkar & Olsson, 2016; Haaker et al., 2017; Kleberg et al., 2015; Lindström, Selbing, & Olsson, 2016; Olsson et al., 2016, 2007; Olsson & Phelps, 2004; Selbing & Olsson, 2019; Szczepanik et al., 2020). One stimulus, CS+, is always paired with an aversive shock delivery to the forearm of the demonstrator, while the other stimulus, CS-, is never paired with a shock. While the participant is watching the video, they are connected to electrodes for shock delivery like that of the demonstrator in the video. The learned threat from the video is then measured by directly presenting the two stimuli in the absence of the demonstrator and measuring the psychophysiological response. Transmission of OFL has mostly been measured by SCRs, a common proxy of fear response (Lonsdorf et al., 2017), reporting enhanced SCRs in response to the CS+ (Golkar et al., 2015; Golkar & Olsson, 2016; Lindström et al., 2016; Olsson et al., 2016, 2007; Olsson & Phelps, 2004; Selbing & Olsson, 2019; Szczepanik et al., 2020). Other measurements of physiological correlates of OFL include eye-tracking methods (Kleberg et al., 2015), fear-potentiated startle response (Lonsdorf et al., 2017), and fMRI to measure neural activation, which has implicated the AMG, mPFC, AI, and PAG in OFL (Golkar, Haaker, Selbing, & Olsson, 2016; Lindström et al., 2018; Olsson et al., 2007).

5.2 Study 1: Paradigm validation of a new OFL task

Whilst studies of OFL over the past 15 years have been largely based on the tasks developed by Olsson and colleagues (Golkar et al., 2015; Golkar & Olsson, 2016; Haaker et al., 2017; Kleberg et al., 2015; Lindström et al., 2016; Olsson et al., 2016, 2007; Olsson & Phelps, 2004; Selbing & Olsson, 2019; Szczepanik et al., 2020), my interest in understanding how OFL occurs led to two main changes to the task: i) using a reinforcement learning framework to model how learning occurs on a trial-by-trial basis; and ii) examining the role that the intensity of a demonstrator's reaction to being shocked has on OFL.

The present examples of OFL tasks serve as a strong foundation to study how fearful or threatening information can be relayed socially; however, I was specifically interested in understanding *how* OFL occurs. Because of this, the paradigm described in Study 1 relies on reinforcement learning where the two conditioned stimuli are probabilistically determined and, therefore, mostly predictive of an outcome (shock delivery or no shock), but not deterministically (in contrast with the protocols developed by Olsson and colleagues such as those reported by Haaker and colleagues (2017), where the conditioned stimuli are either completely predictive of an outcome or one stimulus is always predictive of an outcome whilst the other is predictive half of the time).

Through examination of the difference between expected and actual outcomes (prediction errors), which are thought to drive learning, *how* learning occurs can be characterized. A reinforcement learning model holds that when an actual outcome of a decision differs from what is expected, the associative value of a stimulus changes due to the influence of this new information (Sutton & Barto, 1998). In the present study, participants were instructed to predict the outcome of each CS without being instructed of their associations beforehand. The information gleaned from participants' predictions was analysed using a computational model, supported by Bayesian model comparison, to understand differences in fundamental reinforcement learning parameters: learning rate and temperature.

Learning rate refers to the extent to which the current expected value is updated by surprising information. A low learning rate suggests a minimization of

recent prediction errors, weighing information from the recent past less, while a high learning rate indicates a strong influence of recent prediction errors towards updating future predictions. The temperature parameter captures the participants' exploration and valuation in making choices. For example, a low temperature parameter indicates consistent behaviour by more frequently choosing the more probable stimulus outcome, while a high temperature parameter results in more random or exploratory choices, placing less emphasis on the expected values of each outcome (Lockwood et al., 2016).

In the context of the present study, two different stimuli are used; one stimulus is associated with a vicarious shock outcome 80% of the time, while a second stimulus is associated with a vicarious shock outcome only 20% of the time. Therefore, participants reliably choosing the high-probability CS as *always* relaying a shock outcome and always choosing the low-probability CS as *never* giving a shock would result in both a low learning rate as well as a low temperature parameter. However, if a participant frequently switches their outcome choice following trials where the expected outcome differs from the actual outcome, this updating of information would be evidenced with a high learning rate. If the participant guesses randomly from trial to trial with little to no consistency, this would result in a high temperature parameter. Together, this modification to the OFL paradigm allows for a more sensitive characterization of OFL and better understanding of what drives observational learning.

In addition to adding a reinforcement learning framework to the OFL task, I sought to understand if the intensity of the reaction to a shock by the actor on screen impacted OFL acquisition. Selbing and Olsson (2019) examined a variation of the impact of their actor's anxious behaviour on participant learning. In their study, participants viewed an actor who either displayed anxious anticipatory behaviour to both the CS+ and CS- or an actor who only demonstrated anxious anticipatory behaviour to the CS+. Contrary to expectations, participants viewing the more anxious actor had *better* OFL discrimination, as evidenced by slower extinction learning. This may be in part because the actor's anxious behaviour strengthened participant OFL as the indiscriminate anticipatory anxiety may have made the CS+

seem very dangerous, which ultimately made learning more impervious to extinction (Goldstein, 1979). This was further substantiated by participant's over-anticipating the US as compared with the less anxious learning model.

How humans use others' emotional cues to better understand danger and safety plays a significant role in the transmission of information (Olsson & Phelps, 2007). In Study 1 I focus on whether the intensity of pain response might enhance OFL by making the association between a shock predictive cue and a harmful outcome more obvious, or whether it might distract attention from associative learning. Therefore, in addition to validating an OFL paradigm based on a reinforcement learning framework, Study 1 aims to address how the intensity of a shock response might influence OFL.

5.3 Materials and Methods

5.3.1 Participants

Seventy-eight healthy male adults aged 18 and older were recruited though the University College London Psychology Subject Pool. Participants provided written informed consent and were compensated 7 GBP for their time, in line with institutional guidelines at the time of testing. Four participants were excluded from the analysis due to performance at or below chance level (~50%) on the OFL task, leaving a final sample size of 74. Thirty-four participants were exposed to a medium intensity OFL video, while 40 participants learned from a high intensity OFL video. UCL Division of Psychology and Language Sciences Ethics Committee provided ethics approval for the study (approval code BUCNI-BBK-16-002).

5.3.2 Materials

5.3.2.1 Generation of video stimuli

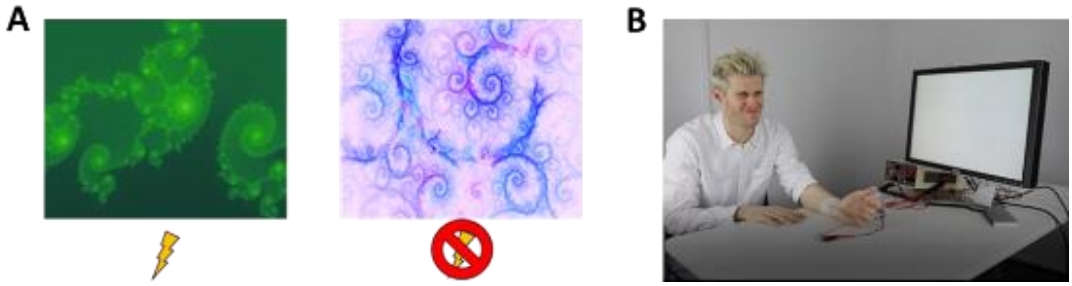
Two movies were created for the OFL task. Both movies showed a male actor from the Royal Shakespeare Company taking part in a probabilistically determined

fear learning task. Both movies displayed one of two different fractal images at a time on a computer screen in front of the actor; the green fractal was associated with an uncomfortable electrical shock, approximately 35 V, delivered to the actor's forearm 80% of the time, while the purple fractal predicted shock delivery 20% of the time (

Figure 5.1A). No audio was included for these videos. There were a total of 40 trials per video with equal presentations of the high and low predictive fractals. An electrical shock was delivered to the actor on 20 of the 40 trials, 250 ms after the disappearance of the fractal on the screen. Each fractal was presented on the actor's screen for 2-8 s in a randomized order with an inter-stimulus-interval between 2-8 s. The videos differed in the pain expression of the actor when receiving an electrical shock (

Figure 5.1B). The medium shock reaction video (6 min, 56 s) reflected a moderate reaction by the actor towards receiving an electrical shock, while the high shock reaction video (6 min, 46 s) displayed an exaggerated shock-response, despite the shock levels being the same. In both videos the actor actually received electrical shocks at a level adjusted to be uncomfortable, but not painful (based on subjective report), prior to recording the movies. Additional videos were filmed in which the actor was not actually shocked, but instead acted as if he had been; however, these videos were a bit more variable in their realism and were thus not used for testing.

Figure 5.1. Example stimuli



A. Two CSs presented to participants during the OFL task. Green fractal image associated with shock delivery outcome on 80% of trials; purple fractal image associated with shock delivery outcome on 20% of trials (shock and no-shock cartoons were not present during the task). **B.** Example of shock reaction by actor from the medium shock reaction video.

5.3.2.2 OFL Task

The OFL task assessed participants' abilities to acquire associative information from a social observation context. Participants were instructed to watch either the medium or high shock reaction video. During the 40 trials, participants were instructed to guess whether each fractal presented would result in a shock to the actor while the fractal remained on the screen (between 2-8 s). Failure to make a choice during the fractal presentation resulted in a no-response for the trial. The shock contingencies were not instructed, but needed to be learned throughout the task. All responses were recorded using Psytools software (Delosis Ltd).

5.3.2.3 Assessment of cognitive ability

Non-verbal fluid intelligence was measured using Scale 2, Form A, of the Cattell Cultural Fair IQ Test (CFIT; Cattell, 1987), presented and recorded in a web browser, rather than on paper. The different problems comprising this test cover nuances in abstract reasoning given in a multiple choice format. They contain figural information that participants need to complete in order to answer the questions correctly. Participants completed the task within a two-minute time frame and any answers provided within that time frame were recorded to be used for analyses - in line with

prior deployment of this task in large scale samples (Toledano et al., 2019). This format of task administration yields raw scores of cognitive ability, rather than a standardised estimate of intelligence, as the shortened task format precludes the use of the original standardisation process. However, the raw scores serve as a relative index of cognitive ability that can be used as a covariate in the task analyses.

5.3.3 Procedure

All tasks and questionnaires were presented on a Dell computer with a Windows operating system using Psytools software. Participants were either instructed to watch the medium or high intensity shock response OFL video followed by the completion of the cognitive ability assessment.

5.3.4 Analyses

5.3.4.1 Behavioural analyses

All behavioural analyses were performed using IBM SPSS Statistics 24 for Windows. Paired samples t-tests were used to assess the differences in predictions between high versus low probability conditions. Independent samples t-tests were used to test differences between the medium and high shock reaction conditions. Lastly, Pearson correlational analysis were used to relate learning rate and temperature generated from our computational model to cognitive ability.

5.3.4.2 Computational model

A reinforcement learning model was derived from the OFL task using MATLAB R2018a for Windows. A trial-by-trial analysis of shock/no-shock predictions using a reinforcement learning algorithm (Sutton & Barto, 1998) was used. This model presumes that the associative value '*Q*' of a stimulus changes when the predicted association differs from the actual outcome. For each trial '*t*', a prediction '*a*' has an expected value '$Q_t(a)$' that is updated by the discrepancy with the actual outcome of a trial 'r_t', or the prediction error '$r_t - Q_t(a)$', that updates future expectations of a

prediction '$Q_{t+1}(a)$'. The learning rate 'α' (between 0 and 1) is a scale of how much the prediction error updates future expectation, such that:

$$Q_{t+1}(a) = Q_t(a) + \alpha \times [r_t - Q_t(a)]$$

A low learning rate 'α' minimizes the influence of prediction errors and the degree to which associative values are updated. A participant's probability of choosing prediction '*a*' on trial '*t*', given the expected values of the available predictions '$Q_t(a)$', is given by the softmax link function:

$$P_t\,[a|Q_t(a)] = \frac{e^{[Q_t(a)/\beta]}}{\sum_{a'} e^{[Q_t(a')/\beta]}}$$

The temperature parameter 'β' controls the amount of exploration of a participant in making consistent (as opposed to noisy) predictions. A low β indicates consistent behaviour such that the participant is making their predictions based upon the more associated value on each trial. The softmax link function therefore estimates the trial-by-trial probability of each prediction by weighting expected values against the β parameter (Liu, Valton, Wang, Zhu, & Roiser, 2017; Lockwood et al., 2016).

To fit participant behaviour to an appropriate learning model, the maximum a posteriori (MAP), a hierarchical Bayesian approach, was used because of its increased accuracy in estimating the true underlying parameters and decreased sensitivity to outliers, both of which result in better estimation of individual differences across participants (Ahn, Krawitz, Kim, Busemeyer, & Brown, 2011; Daw, 2011). This process initially estimated α and β parameters for each individual participant using maximum likelihood estimation (MLE). The parameters were then re-estimated by applying priors defined using a Gaussian kernel with mean and standard deviation defined from the original distribution of α and β (Delgado, Phelps, & Robbins, 2011).

Four different learning models were compared to determine what best characterised OFL: i) a null model, which assumes participants show no learning and

seemingly make their choices at random; ii) a win-stay-lose-shift model, which is a simplistic strategy based solely on the success or failure of the previous decision; iii) a model that uses the same learning process for both high and low probability stimuli, with a single α and a single β; and a model that uses separate processes for the high- and low-probability stimuli, with separate α and β parameters.

Using Bayesian information criterion scores (BIC), the evidence for each of the above models was examined, based upon the number of parameters and likelihood of model fits (Kass & Raftery, 1995). This was performed separately for the medium- and high-intensity shock stimuli, as well as both stimuli together.

5.4 Results

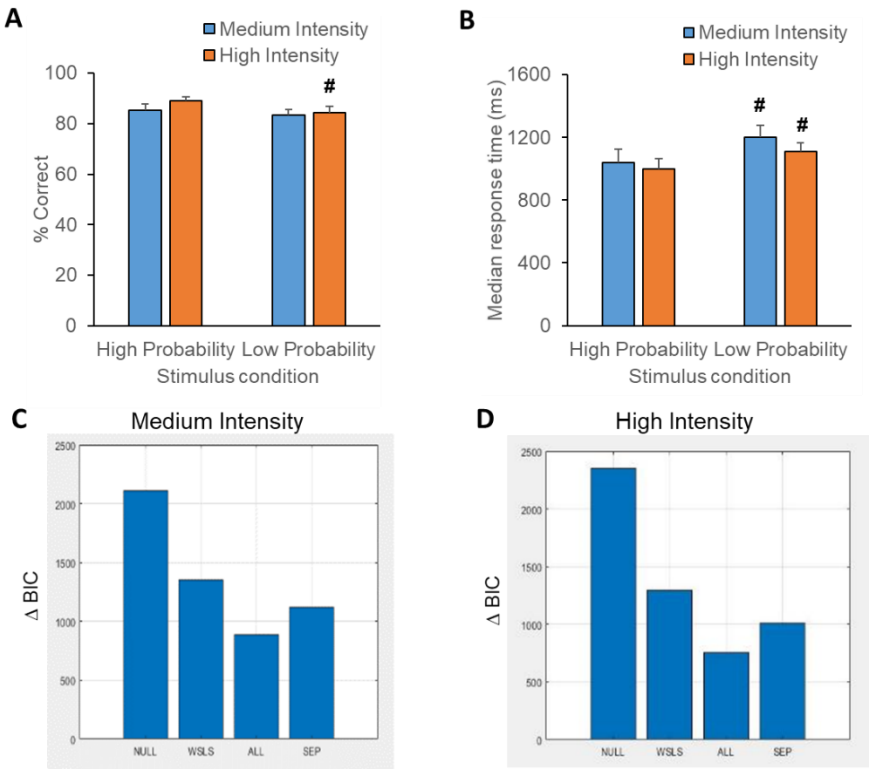
5.4.1 Characterization of OFL task to medium shock reaction

I first sought to establish an OFL paradigm utilizing reinforcement learning theory framework so that trial-by-trial variations in observed learning could be understood. Thirty-four participants were shown the OFL video with the actor on screen demonstrating a medium shock reaction. Participants were asked to make a prediction about shock expectancy to each CS presented. On average, participants were similarly accurate in predicting shock outcomes to the high-probability (85.37% ± 14.78) and low-probability (83.52% ± 12.03) CSs (paired t-test: $t(33) = 0.88$, $p = 0.387$); however, the median response time to make a prediction was significantly faster for the high-probability CS (1038.76 ± 466.39 ms) than the low-probability CS (1201.85 ± 529.80 ms; paired t-test: $t(33) = 4.170$, $p < 0.001$) (Figure 5.2A-B; blue bars). In other words, participants demonstrated that they were able to accurately predict shock-associated outcomes for the two CSs, with quicker responses for the high-probability CS, entirely through observation.

Additionally, a computational reinforcement learning model was applied in order to explain trial-by-trial variation in learning. The model fit was determined by comparing the three learning models against a null model, which assumed no learning occurred and participants behaved at random. Each participants BIC score

subtracted from the null model to determine which model best explained their behaviour. The winning model suggests participant choices were most parsimoniously characterized by a model with single learning rate ($\alpha = 0.145 \pm 0.068$) and temperature ($\beta = 0.057 \pm 0.066$) parameters ($\Delta BIC > 200$; Figure 5.2C) (note that the computational models did not incorporate response time and therefore are not affected by this difference). The low average learning rate and temperature parameters indicate a minimization of recent prediction errors and high consistency in predictions in line with their associative learning.

Figure 5.2. Comparison of medium versus high intensity shock response



A. Comparison of participant accuracy for high-probability and low-probability CSs between medium and high shock reaction groups. **B.** Comparison of median response times (ms) for high and low-probability CSs between medium and high shock reaction groups. **C.** Model outcome for medium shock reaction group (n=34) **D.** Model outcome for high shock reaction group (n=40). NULL = no model; WSLS = win-stay lose-shift; ALL = single learning rate and temperature parameters for high- and low-probability CSs; SEP = separate learning rate and temperature parameters for high- and low-probability CSs. #p<0.05 (2-tailed, within groups).

5.4.2 Characterization of OFL task with high shock expression

Next, I wanted to understand the impact of the intensity of the demonstrator's response on OFL. Forty participants performed the same task described in the previous section, however, the actor's reaction to receiving shocks was exaggerated. Participants were significantly more accurate in predicting shock outcomes for the high-probability CS (88.97% ± 11.13) than for the low-probability CS (84.21% ± 16.35; paired t-test: $t(39) = 2.61$, $p = 0.013$). Again, participants' median prediction response time was significantly quicker for the high-probability CS (1000.75 ± 365.28 ms) than the low-probability CS (1111.73 ± 435.68 ms; paired t-test: $t(39) = 3.16$, $p = 0.003$) (Figure 5.2A-B; orange bars).

A computational reinforcement learning model was then run for the high shock reaction group, finding that the model containing a single learning rate ($\alpha = 0.130 \pm 0.070$) and temperature ($\beta = 0.027 \pm 0.027$) parameter was again favoured ($\Delta BIC > 250$; Figure 5.2D). The mean parameters were comparable to the first dataset, again suggesting a relatively minor influence of recent prediction errors and a strong degree of determinism in predictions.

5.4.3 Behavioural comparison between medium and high intensity reaction stimuli

To determine whether the demonstrator's degree of pain expression had an effect on learning, I compared the accuracy and median response times between the participant group who watched the medium shock reaction video to the high shock reaction group. Accuracy was no better for the video with a high shock reaction than the medium shock reaction, $F(1, 72) = 1.09$, $p = 0.300$, $\eta 2 = 0.015$ (Figure 5.2A), nor

were there any differences in reaction time, $F(1, 72) = 0.99$, $p = 0.324$, $\eta 2 = 0.014$ (Figure 5.2B). Additionally, I compared the learning rate and temperature parameters of both groups and found that whilst the learning rate did not significantly differ based upon the demonstrator's shock reaction, $t(72) = 0.88$, $p = 0.38$, participants viewing the high shock reaction video were significantly more deterministic in their responding than the medium shock reaction group as evidenced by a lower temperature parameter, $t(72) = 2.60$, $p = 0.011$. Despite this difference in temperature parameters between groups, it does not appear that the intensity of a demonstrator's shock reaction impacted the learning process *per se*, as it is governed by the learning rate and, therefore, comparable between the groups.

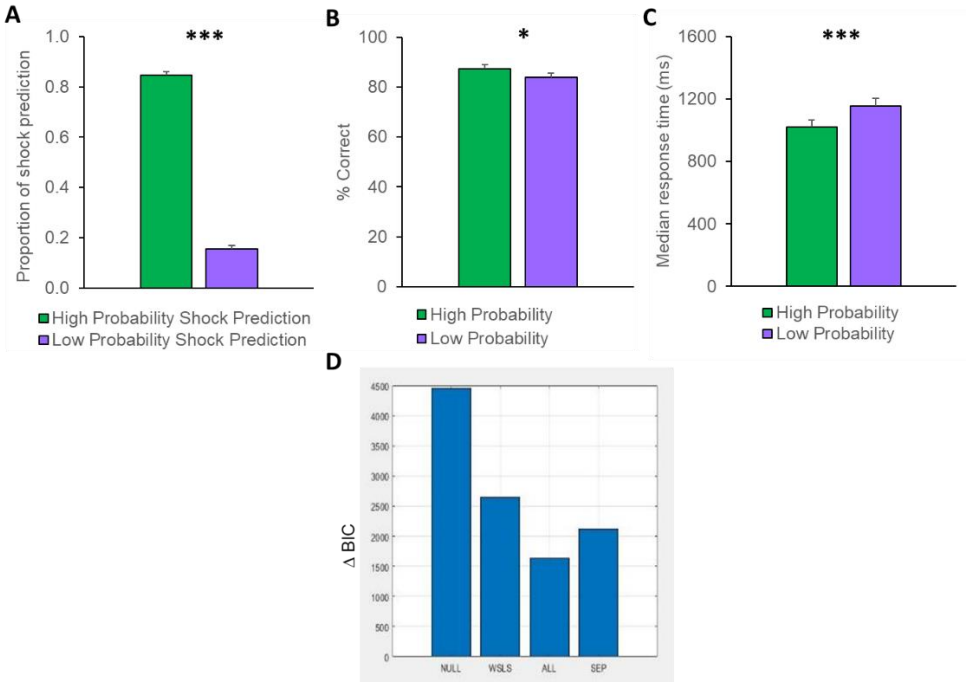
5.4.4 Behavioural data: differences between high and low probability stimuli

To increase power by analysing a larger number of participants together, the data from the medium- and high-shock reaction conditions were pooled for analysis. In this combined dataset, participants were able to learn the proportion of shock prediction between the high (0.85 ± 0.12) and low probability CSs (0.15 ± 0.12), $t(73) = 24.00$, $p < 0.001$ (Figure 5.3A). Moreover, participants were significantly better at predicting a shock outcome for the high-probability CS (87.32% ± 12.97) than a no-shock outcome for the low-probability CS (83.84% ± 14.44), $t(73) = 2.47$, $p = 0.020$ (Figure 5.3B) as well as significantly quicker at responding to the high-probability CS (median: 1018.22ms ± 412.29) than the low-probability CS (median: 1153.14ms ± 479.94), $t(73) = 5.16$, $p < 0.001$ (Figure 5.3C). This suggests that participants show slightly better learning of the high-probability CS than the low-probability CS. However, the computational reinforcement learning model indicated learning to best be described by a single learning rate ($\alpha = 0.14 \pm 0.07$) and temperature ($\beta = 0.04 \pm 0.05$; $\Delta BIC > 500$) (Figure 5.3D). Despite the significant differences between stimuli in terms of participant accuracy, the computational model suggests participants are not strongly weighing recent prediction errors in the value of their choices. Moreover, the low temperature parameter suggests participants are quite deterministic in their predictions.

5.4.5 Cognitive ability is not associated with OFL

To test the possibility that OFL performance might be related to cognitive ability, correlational analyses were performed between cognitive ability and OFL accuracy, prediction response time, learning rate, and temperature. There was a range of cognitive ability scores across participants from 12 – 44 (37.81 ± 5.24). No significant associations were found between cognitive ability and any of the OFL task performance measures, including learning rate, r = 0.10, p = 0.40, and temperature, r = -0.03 p = 0.82.

Figure 5.3. High- versus low-probability stimuli



A. Participants prediction of shock outcome association between the two stimuli. **B.** Averaged error rates for each stimuli across all 40 trials. **C.** Averaged time taken to make a choice response to two differing stimuli across all 40 trials; abbreviations as in Figure 5.2. ***p<0.001, *0.01<p<0.05 (2-tailed).

5.5 Discussion

Study 1 described the development and testing of an OFL task using reinforcement learning theory in a community sample of males. I examined whether OFL performance or learning strategy was influenced by the intensity of shock reaction by the demonstrator on screen. Overall, the findings within this study indicated that participants were able to learn a threatening association through observation of another's response, were more accurate in their shock predictions to the high-probability CS, and were quicker at making a prediction response for the high-probability CS. Despite these differences in performance between the high- and low-probability CSs, the computational reinforcement learning model favoured a single learning rate and temperature parameter to characterize OFL. The low learning rate suggested that predictions were not heavily influenced by recent prediction errors and that participants updated their valuation of the CSs relatively slowly, whilst the low temperature parameter indicated consistency in prediction choice. Lastly, whilst the demonstrator's shock reaction did not impact learning rate, participants observing the high shock reaction condition were more deterministic in their predictions as evidenced by the difference in temperature parameters between groups.

Previous studies have looked at OFL in humans using a similar setup of participants viewing an actor on screen receiving uncomfortable electrical shocks (e.g., Olsson et al., 2007; Olsson & Phelps, 2004); however, these tasks typically include between 10-24 OFL trials (Lindström et al., 2018; Olsson et al., 2007), and the use of two different stimuli – a CS+ that would either provide a shock at least 50% of the time, and a CS- that would never be associated with a shock (Haaker et al., 2017). Moreover, the extent of learning is largely measured by SCRs, which demonstrate the physiological response during OFL, but tells little about the learning processes used for OFL. In contrast, the present study introduced a reinforcement learning framework to the task, such that the two CSs were probabilistically associated with either a shock to the demonstrator (green fractal – 80% shock-associated), or not (purple fractal – 20% shock-associated). OFL was then measured by participant outcome predictions over the course of the task using computational modelling to

derive learning rate and temperature parameters, which provided a better understanding of what drives observational learning.

The development of the OFL task presented within this chapter was designed to address this gap in the literature. The information gleaned through the present application of an instrumental reinforcement learning framework on the well-defined OFL paradigm allowed for the use of detailed computational modelling of behavioural variations on a trial-by-trial basis to assess individuals' abilities to learn from another's painful experience. Characterizing performance measures alone, such as accuracy and response time, is informative about whether OFL has occurred and to what degree, but it fails to take into account *how* OFL occurs in an uncertain external environment where feedback is constantly updating the weight of available information. The present findings indicated that prediction errors in the OFL task (i.e., when the high-probability CS *is not* followed by a shock to the demonstrator, or when the low-probability CS *is* followed by a shock) generated small revisions to the value estimate of either CS. This ultimately resulted in predictions that were more affected by a longer history of prediction outcomes reflected by the low learning rate value. The observation of a low temperature parameter further characterized prediction choice as quite deterministic.

Within the establishment of this paradigm I tested two groups of participants to assess whether the intensity of the shock reaction by the demonstrator made a difference to OFL. Participants either saw a medium shock reaction condition or a high shock reaction condition in order to determine if learning was more sensitive to a greater expression of pain. However, no significant impact of the actor's shock reaction on OFL task performance nor learning rate was found. Prior studies on OFL have generally used a single expression intensity to transmit information, however, Selbing & Olsson (2019) looked at the effect of anxious anticipatory behaviour of the demonstrator on participant OFL. The authors found more deterministic discrimination between aversive stimuli and neutral stimuli by participants observing a more fearful response to both stimuli, which is consistent with the present findings as evidenced by the low temperature parameter. The authors suggest that greater discrimination is required when an emotional cue is no longer predictive of potential

harm. While understanding the impact of anticipatory cues on OFL is critical to understanding social priming of a potentially dangerous situation, how an individual responds to harm or pain is dually important to learning.

Despite the nonsignificant findings of Study 1 on shock reaction and measures of OFL, the degree to which an individual expresses pain or threat in a more naturalistic situation may have a greater impact on OFL. Moreover, it may be that a greater difference in shock response is needed to better determine the critical contribution of another's aversive response on OFL. Future experiments could explore this by having a wider spectrum of shock reactivity beyond just moderate and exaggerated expressions. Additionally, a within-subject design presenting varying demonstrator reactions to a single participant may provide some insight into individual differences in OFL and the components of vicarious fear that are essential for transmission of information. These modifications may provide a better understanding of the threshold required to learn through observation and more nuanced understanding of which aspects of a demonstrator's reaction are most crucial for OFL.

Study 2 presented in this chapter was conducted to replicate and extend the findings from Study 1's paradigm validation by using a larger online sample. In addition, Study 2 also looked at how individual differences in dispositional traits, including anxiety and psychopathy, relate to variability in OFL.

5.6 Study 2: Study of dispositional traits and their association with individual differences in OFL

Both animal and human research have identified individual differences in reactivity and expression of OFL, which may contribute to individual differences in functioning and susceptibility to various psychiatric disorders (Debiec & Olsson, 2017; Jeon et al., 2010; Keum et al., 2016; Mikosz, Nowak, Werka, & Knapska, 2016; Olsson et al., 2016; Szczepanik et al., 2020). Research has shown individuals with mood and anxiety disorders may have a bias for interpreting situations as threatening, potentially making them hypervigilant towards socially derived fear (Debiec & Olsson, 2017; Helsen et al., 2011; Ueno et al., 2018). Conversely, attenuated fear responses in individuals with psychopathic traits may lead them to pay less attention to and become less aroused by another's distress. This in turn can result in fewer pairings between events that are threatening and the conspecific's fear response, potentially contributing to atypical development of empathy over time (Bird & Viding, 2014; Seara-Cardoso et al., 2016; van Dongen, 2020). Whilst clinical levels of anxiety and psychopathy exist in a small percentage of the population, the general populace displays these traits to a varying degree, thus landing somewhere along the spectrum for anxiety and psychopathy. Given their potential association with OFL, I investigate the relationship between these dispositional traits and OFL parameters.

5.6.1 The relationship between dispositional traits and learning

Dispositional traits are largely stable personality characteristics that can have a profound impact on a multitude of behaviours (Ajzen, 1987). Moreover, variation between individuals' dispositional traits can be indicative of personality and clinical disorders (Hong, 2013; Krueger, Caspi, Moffitt, Silva, & McGee, 1996; Watson, Kotov, & Gamez, 2006), making their characterization useful for better understanding contributing factors towards disparities in behaviour. While there are a multitude of considerations lending themselves to the diversity of individual responses to various stimuli (Adler & Clark, 2019; Hong & Paunonen, 2011; Samimy, Schettini, Fernhoff, Webster-Stratton, & Beauchaine, 2020; Steyer, Schmitt, & Michael, 1999), for the

purpose of the present study I focus on two traits and their association with OFL: anxiety and psychopathy.

5.6.1.1 Trait anxiety

Trait anxiety is a characteristic of individuals with clinically diagnosed anxiety disorders (Kennedy, Schwab, Morris, & Beldia, 2001), and reflects an anxious personality style reflective of an individual's propensity to perceive stimuli as threatening and dangerous by characteristically responding to aversive situations with transient physiological and social reactions (state anxiety) (Spielberger, 1983). Individuals with trait anxiety have been shown to exhibit enhanced avoidance behaviours, altered learning strategies, and hypervigilance to threat information (Aylward et al., 2019; Eysenck & Van Berkum, 1992; Mkrtchian et al., 2017; Pittig, Treanor, LeBeau, & Craske, 2018; Surcinelli, Codispoti, Montebarocci, Rossi, & Baldaro, 2006). With respect to fear learning, individuals with high-trait anxiety demonstrate greater fear response to safety stimuli compared to healthy controls, but not necessarily to aversive stimuli (Gazendam, Kamphuis, & Kindt, 2013; Haddad, Pritchett, Lissek, & Lau, 2012; Indovina, Robbins, Núñez-Elizalde, Dunn, & Bishop, 2011; Kindt & Soeter, 2014; Lonsdorf & Merz, 2017), suggesting possible impairments in inhibitory safety learning processes leading to enhanced fear generalization (Basten, Stelzel, & Fiebach, 2011; Berggren & Derakshan, 2014; Haaker et al., 2017; Jovanovic, Kazama, Bachevalier, & Davis, 2012.; Lonsdorf & Merz, 2017). Moreover, elevated levels of anxiety correlate with faster learning and inclination to update behaviour when faced with aversive information (Aylward et al., 2019).

Additionally, individuals with high-trait anxiety have been shown to have better processing of emotional faces and to do so more automatically compared to low anxiety participants (Holmes, Nielsen, Tipper, & Green, 2009; Surcinelli et al., 2006; Walentowska & Wronka, 2012). Specifically, Surcinelli and colleagues (2006) demonstrated that high-trait anxiety participants displayed a propensity towards recognizing fearful faces over any other emotional face presented. In contrast to angry faces, which directly signal threat and danger, fear faces are more ambiguous in the potential threat they represent and therefore may be interpreted by anxious

individuals as a strong stimulus of potential peril. However, despite the ample evidence of an association between trait anxiety and fear learning/threat processing, no study to date has investigated the association between individual differences in trait anxiety and OFL.

5.6.1.2 Psychopathic traits

In stark contrast to anxiety lies psychopathy, a serious personality disorder characterised by complex character and behavioural traits, which include a lack of empathy, unemotional/shallow affect, callousness, antisocial, and impulsive behaviours (Blair, Mitchell, & Blair, 2005; Frick & Viding, 2009; Hare, 2003; Hare & Neumann, 2009). Whilst diagnosed psychopaths represent a very small, but extreme subset of individuals, psychopathic traits are normally distributed throughout the general population and appear to be closely associated with the phenotypes present in clinical and incarcerated samples (Hare & Neumann, 2008; Seara-Cardoso & Viding, 2015). There are no studies to date looking specifically at the relationship between psychopathic traits and OFL, however, there are several characteristics of psychopathy that may lead us to expect an association with OFL. For example, individuals presenting high levels of psychopathic traits demonstrate atypical reactivity to others' fearful expressions and can present with atypical fear learning (Blair et al., 2005; Marsh et al., 2011; Patrick, Bradley, & Lang, 1993). López and colleagues (2013) identified that individuals who self-reported high levels of psychopathy-related fearlessness had a reduced physiological response to shock-associated CSs, possibly indicating that a lack of physiological arousal may lead to a blunted defensive response. Despite this finding, the authors report that no other associations were found between psychopathic traits and fear learning measures suggesting that individuals with greater levels of psychopathy might have atypical development of affective processing, but are not completely unable to respond to fear learning, consistent with other reports (Dolan & Fullam, 2004; Hare, 2003; Hare, 1965; Patrick, 1994; Richell et al., 2003).

Psychopathy is also linked with impairments in recognizing affective states, especially fearful and sad facial expressions (Blair et al., 2005; Dolan & Fullam, 2006;

Kosson, Suchy, Mayer, & Libby, 2002; Wilson, Juodis, & Porter, 2011) as well as atypical neural responses towards emotional stimuli (Dawel, O'Kearney, McKone, & Palermo, 2012; Decety et al., 2013; Lockwood, Bird, Bridge, & Viding, 2013; Seara-Cardoso et al., 2015). It appears that a possible lack of affective empathy grows from an atypical direct experience of threat, ultimately making it more difficult to recognize distress cues in another (Bird & Viding, 2014). It is possible that atypical responses to others' fear may lead to a poor development of empathy and, therefore, may be associated with OFL.

5.6.2 Dispositional traits in the context of OFL

As outlined in Study 1, using reinforcement learning theory and computational modelling, the present task enables the investigation into mechanisms that contribute to OFL in humans. In Study 2, I use the validated paradigm to assess whether individual differences in dispositional traits commonly associated with fear, namely trait anxiety and psychopathy, are associated with individual differences in OFL. To my knowledge, this area of enquiry has not received attention previously. Furthermore, I examine the possibility that an association between OFL and anxiety or psychopathic traits may be, in part, explained by differences in the propensity to resonate with others' emotional states, as assessed by a self-reported tendency to empathise with other people's affect ('affective empathy'). This ability is typically elevated in individuals with high-trait anxiety (Blair et al., 2016; Negd, Mallan, & Lipp, 2011; Serbic et al., 2020; Shu, Hassell, Weber, Ochsner, & Mobbs, 2017) and reduced in those with high levels of psychopathic traits (Decety et al., 2013; van Dongen, 2020; Lockwood et al., 2013; Marsh & Blair, 2008). Lastly, cognitive ability and state anxiety were measured to investigate their role as potential confounders to the association between the dispositional traits and OFL, should such associations emerge.

Based upon the supporting literature, I predicted that individuals with high-trait anxiety would be more likely to predict an aversive shock outcome and would be quicker to make their predictions due to increased attention to the fearful cues demonstrated on screen by the actor. Moreover, I predicted that high-trait anxiety individuals would be quicker to update their associations in response to prediction

errors. Conversely, I predicted that individuals with high-psychopathic traits would have worse accuracy than participants with lower psychopathic traits, due to decreased attention to the affective state of the demonstrator. Similarly, I hypothesized that high-psychopathic traits would be associated with lower learning rate and temperature parameters as poorer attention to observed fear may fail to engage learning following prediction errors, and also induce less consistent predictions. Lastly, I predicted that anxiety would positively correlate with affective empathy, given that anxious individuals are particularly aroused by threat and danger cues; and that psychopathic traits would be negatively associated with affective empathy, given that a lack of empathy is an important facet of psychopathy.

5.7 Materials and Methods

5.7.1 MTurk participants

Mechanical Turk (MTurk) is an Amazon overseen product established as a marketplace for work performed by humans. It is an online platform that allows participants, or 'workers,' the opportunity to complete a variety of tasks from their personal computers. To be eligible for the study, the "Human Intelligence Task" (HIT) approval rate, the rate workers have been approved for successfully completing other tasks, was set at minimum 97%, with the number of approved HITs from previous tasks required set to be greater than 50. An additional setting allowed for only males over 18 years of age to access the study in order to be as consistent as possible with the mouse research. Workers were compensated 7.50 USD for successfully completing the task and were informed at the start of the study that a failure to provide responses to 40% of trials during the OFL task would result in being rejected from the study without compensation. A total of 356 workers successfully completed the study and were compensated. Seventy-six of the workers were not compensated either because they represented a repeated attempt or for failing to provide a sufficient percentage of responses on the OFL task. A further 66 participants were excluded due to greater evidence for a null model than the wining

model for the OFL task, suggesting they failed to engage in learning from feedback, giving a final sample size of 290 workers. The study was approved by the UCL Division of Psychology and Language Sciences Ethics Committee (approval code BUCNI-BBK-16-002).

5.7.2 Materials

5.7.2.1 Video stimuli

Given that there were no differences in the OFL learning rates for medium- and high-shock reactions from Study 1, and that data from both videos provided reliable estimates of OFL, only the more naturalistic medium-shock reaction video was deployed in this MTurk study. This also kept the task length manageable for the workers. During the 40 trials of CS presentations, workers were instructed to guess whether each fractal would result in a shock or not while the fractal was visible. The shock contingencies were not instructed, but needed to be learned throughout the task, as in Study 1. All responses were recorded using Psytools software (Delosis Ltd).

5.7.2.2 Assessment of trait anxiety

Trait anxiety was assessed using the State-Trait Anxiety Inventory trait subscale (STAI-T; Spielberger, Gorsuch, & Lushene, 1970), a 40-item scale of statements used to self-report symptoms of anxiety. Statements were rated on a four-point scale with answers ranging from "Almost Never" to "Almost Always." These statements included questions such as "I calm, cool and collected" and "I worry too much over something that doesn't really matter." STAI state anxiety (STAI-S) subscale scores were also collected and checked against OFL task performance. There was a high level of internal consistency for both the 20 questions of the STAI-T ($\alpha = 0.61$) and the 20 questions of the STAI-S ($\alpha = 0.95$).

5.7.2.3 Assessment of psychopathic traits

Psychopathic traits were assessed with the Self-Report Psychopathy Scale-Short Form (SRP-4-SF; Paulhus, Neumann, & Hare, 2016), a 29-item scale designed to

measure psychopathic attributes in non-institutionalized samples. The SRP has been shown to have good construct validity and internal consistency and is strongly correlated with the PCL-R, a clinical measure of psychopathy (Paulhus et al., 2016; Sylvers, Lilienfeld, & LaPrairie, 2011). Questions were rated on a five-point scale from "Disagree Strongly" to "Agree Strongly" with some items presented as positive or negative statements. Statements such as "I haven't broken into a building or vehicle in order to steal something or vandalize" were reverse scored in such a way that higher scores indicated higher levels of psychopathic traits. Additionally, the SRP includes statements such as "Some people say I'm cold-hearted" and "I love violent sports and movies." There was a very high level of internal consistency for the present SRP questionnaire ($\alpha = 0.93$).

5.7.2.4 Assessment of affective empathy

Affective empathy was measured with the affective empathy subscale Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011), a 31-item questionnaire aimed at comprehending another's experience as well as share the emotional experience of another. The questionnaire is rated on a four-point scale from "Strongly disagree" to "Strongly agree" making statements such as "I am inclined to get nervous when others around me seem to be nervous" and "I usually stay emotionally detached when watching a film." There was a high level of internal consistency for the present affective empathy questionnaire ($\alpha = 0.81$).

5.7.2.5 Assessment of cognitive ability

Cognitive ability was assessed using the CFIT task previously described in Study 1.

5.7.3 Procedure

The task battery was posted on the MTurk platform and uploaded onto workers' personal computers. Workers watched and responded to the medium

intensity shock response OFL video before completing assessments of cognitive ability, anxiety, psychopathic traits, and affective empathy.

5.7.4 Computational model

The computational models described in Study 1 were implemented to characterize learning rate and temperature in this larger online sample. The learning model that best fit the data was determined by comparing the evidence for the same four possible learning models: a null model, a win-stay-lose-shift model, a model that uses the same learning rate for both high and low probability stimuli, and a model that uses separate learning rates for the different stimuli. This information was then used to calculate individual learning rates and temperature to characterize OFL. The model was run using MATLAB R2018a for Windows.

5.7.5 Analyses

All behavioural analyses were performed using IBM SPSS Statistics 24 for Windows, MATLAB R2018a for Windows, and RStudio 1.1.453. For the OFL reaction times and error rates, paired-samples t-tests were used to assess the differences between performance on high versus low probability stimuli predictions. Learning rate and temperature for OFL were obtained as described in Study 1. Pearson correlational analyses were used to examine associations between dispositional trait measures, OFL variables, and affective empathy. Benjamini-Hochberg False Discovery Rate (Benjamini & Hochberg, 1995) was used to control for the probability of making a Type I error due to multiple comparisons. Pearson correlational analyses were conducted to relate OFL measures with dispositional traits and affective empathy with state anxiety and cognitive ability to ascertain whether these needed to be accounted for in any examination of an association between dispositional traits, affective empathy and OFL.

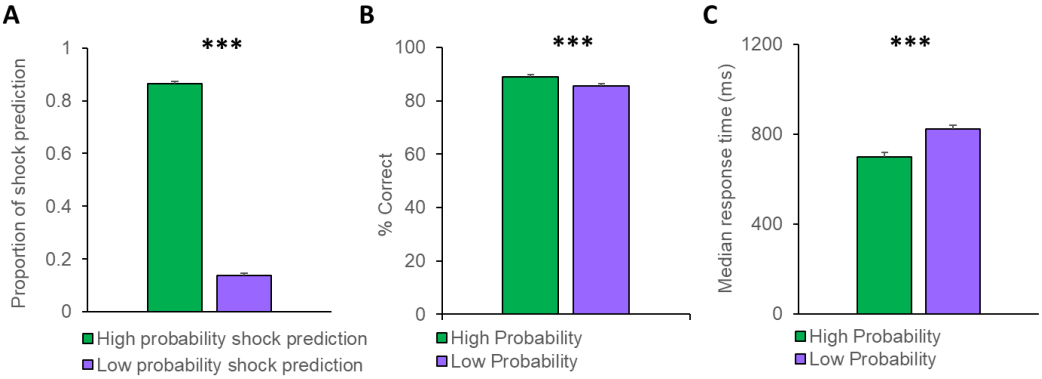
5.8 Results

5.8.1 Characterization of OFL

5.8.1.1 OFL behaviour

MTurk participants performed similarly to the Study 1 participants and were able to predict a correct shock outcome between the high- (0.87 ± 0.12) and low-probability stimuli (0.14 ± 0.12), $t(289) = 52.15$, $p < 0.001$ (Figure 5.4A). Moreover, participants were significantly more accurate in their predictions for the high-probability CS (89.03% ± 12.64) than for the low-probability CS (85.61% ± 13.46), $t(289) = 5.38$, $p < 0.001$ (Figure 5.4B), as well as significantly quicker at making predictions in response to the high-probability CS (median: 698.95ms ± 314.73) than the low-probability CS (820.83ms ± 342.55), $t(289) = 12.00$, $p < 0.001$ (Figure 5.4C). In other words, participants demonstrated greater accuracy and quicker prediction response times to the high probability shock CS, consistent with the overall findings in Study 1 when both groups were combined.

Figure 5.4. OFL behavioural characterization

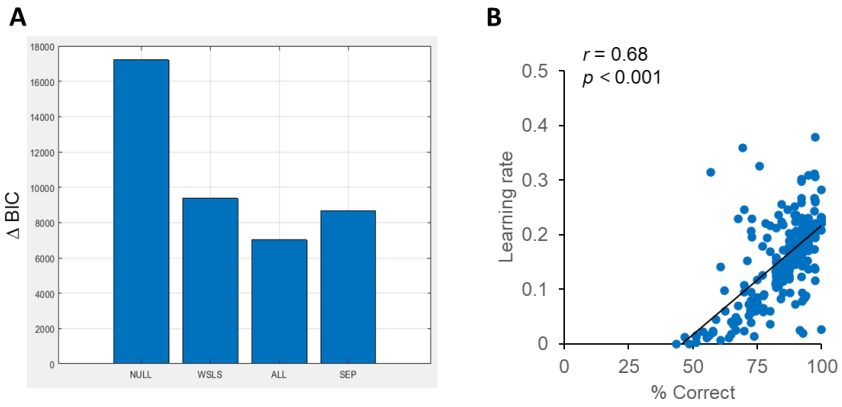


A. Participants prediction of shock outcome association between the two stimuli. **B.** Averaged error rates for each stimuli across all 40 trials. **C.** Averaged time taken to make a choice response to two differing stimuli across all 40 trials. ***p<0.001 (2-tailed).

5.8.1.2 Computational model for OFL

Using the same modelling approach as described in Study 1, participant choices were best characterized by a model with a single learning rate (α = 0.17 ± 0.07) and temperature (β = 0.07 ± 0.05) parameter (winning model evidence (ΔBIC) > 150) (Figure 5.5A). The low learning rate and temperature parameters demonstrate a minimization of prediction errors and that participants were very deterministic in their predictions, in line with what was observed in Study 1. Additionally, there was a strong significant positive correlation between total percent correct and learning rate (r = 0.68, p < 0.001, uncorrected), suggesting that the model does a good job of fitting the data (Figure 5.5B).

Figure 5.5. Computational model



A. Model outcome for favouring a learning model with a single learning rate (n=290); abbreviations as in Figure 5.2. **B.** Positive correlation between participant learning rate and OFL accuracy.

5.8.2 Associations between dispositional traits and OFL

The descriptive statistics of dispositional traits of participants included in the final analyses (n = 290) are presented in Table 5.1. These ranges are comparable to what would be expected from a community sample (Daniel-Watanabe, McLaughlin,

Gormley, & Robinson, 2020; Foulkes et al., 2014; Gordts, Uzieblo, Neumann, Van den Bussche, & Rossi, 2017; McCredie & Morey, 2019; Reniers et al., 2011; Toledano et al., 2019). The full table of correlations between all study measures can be found in Table 5.2. After FDR correction among the key variables of interest, only a modest significant correlation between trait anxiety and median response time (r = 0.22, p = 0.028) was observed, such that the higher the participant's trait anxiety, the slower they were in making predictions during the OFL task (Figure 5.6A). Additionally, there was a trend towards trait anxiety being modestly negatively correlated with accuracy (r = -0.16, p = 0.08) (Figure 5.6 Correlations between trait anxiety and OFLB), contrary to what was hypothesized, however this association did not survive correction for multiple comparisons. Beyond the single statistically significant association between trait anxiety and median response time, there were no significant correlations between participant dispositional traits and OFL.

I next examined whether affective empathy, which showed a modest positive correlation with trait anxiety (r = 0.12, p = 0.039), accounted for part of the association between trait anxiety and OFL median response time. The magnitude of the association remained the same, $r_{partial}(287) = 0.22$, $p < 0.001$, suggesting that it was not accounted for by affective empathy.

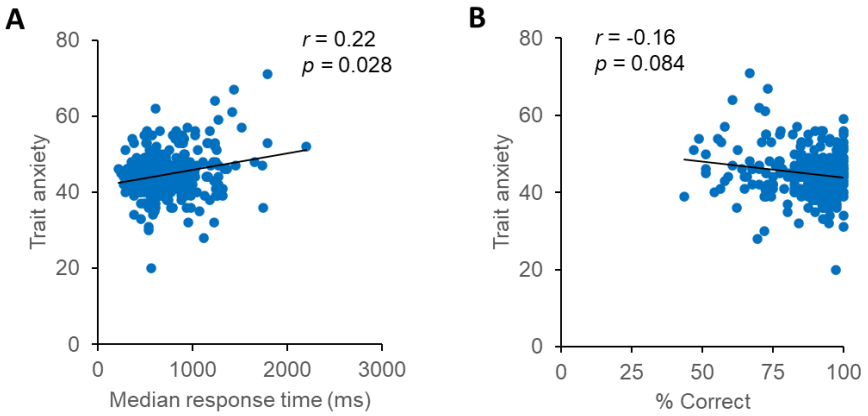
Because participant cognitive ability had a modest, but significant correlation with overall accuracy (r = 0.18, p = 0.002, uncorrected), median response time (r = -0.12, p = 0.041, uncorrected), and learning rate (r = 0.17, p = 0.005, uncorrected) in the OFL task, the correlational analysis between trait anxiety and median response time was repeated, controlling for cognitive ability. Pearson's partial correlation showed that the relationship between trait anxiety and response time was slightly weaker, but remained nominally significant, even after cognitive ability was controlled for, $r_{partial}(283) = 0.20$, p = 0.001 (uncorrected). Lastly, state anxiety did not relate to either trait anxiety nor OFL and was therefore not controlled for.

Table 5.1. Participants' descriptive statistics

	Mean	SD	Minimum	Maximum
CFIT/Cognitive ability*	35.13	6.15	6	44
STAI				
Trait anxiety	44.84	6.06	20	71
State anxiety	35.38	11.81	20	80
SRP				
SRP total	54	17.81	29	118
QCAE				
Affective empathy	30.9	5.26	15	43

*__Note:__ these are not standardised estimates of intelligence

Figure 5.6 Correlations between trait anxiety and OFL



A. Positive correlation between trait anxiety and median response time. **B.** Non-significant negative trend demonstrating the association between trait anxiety and OFL accuracy.

5.8.2.1 Sensitivity analysis: OFL task performance differences between highest and lowest participants according to trait anxiety

After performing FDR correction on the correlations between dispositional traits and OFL performance measures, only the association between trait anxiety and median response time remained significant. However, there were nominally

significant correlations between trait anxiety and total accuracy as well as learning rate that did not survive correction. Consequently, I performed post-hoc sensitivity analyses comparing the top and bottom 15% of participants for trait anxiety (n = 44 per group) in the OFL performance measures that showed a significant or trend association with trait anxiety. In line with the reported findings across all participants, low-trait anxiety participants (786.16ms ± 296.84) were significantly quicker at making outcome predictions than those with high-trait anxiety (937.89ms ± 410.15), $t(86) = 1.99$, $p = 0.050$. Additionally, there was a significant difference in overall accuracy between the low-trait anxiety (88.42% ± 11.38), and high-trait anxiety participants, (82.39% ± 15.27) $t(79.5) = 2.10$, $p = 0.038$; however, no significant differences were found for learning rate, (low anxiety: 0.17 ± 0.07; high anxiety: 0.15 ± 0.09; $t(83) = 1.61$, $p = 0.111$), nor temperature, (low anxiety: 0.07 ± 0.05; high anxiety: 0.08 ± 0.06; $t(86) = 0.92$, $p = 0.362$). Overall, it appears that participants with lower reported trait anxiety are both quicker and more accurate to make shock outcome predictions than those with the highest levels of trait anxiety, but I was not able to show that this pattern could be accounted for by the computational parameters.

Table 5.2 Correlations between all measures

	1	2	3	4	5	6	7	8	9	10
OFL										
1. Total percent correct										
2. Total median response time	-.25***									
3. Difference between high and low CS percent correct	-.08	-.02								
4. Difference between high and low CS median response time	-.01	.15*	.14*							
5. Learning rate	.68***	-.23***	-.10	-.05						
6. Temperature	-.04	.21***	.07	.01	.39***					
Intelligence										
7. Estimated IQ	.18**	-.12*	-.03	-.09	.17**	.03				
STAI										
8. Trait anxiety	-.16	.22*	-.03	.11	-.13	.08	-.02			
9. State anxiety	-.04	.04	-.03	.14	.01	.05	-.10	.09		
SRP										
10. SRP total	-.06	.12	.04	.13	-.02	-.02	-.09	.29***	.30***	
QCAE										
11. Affective empathy	.01	.08	.02	-.03	-.06	-.02	.02	.12*	.13*	-.13*

Notes: Pearson correlation coefficients are reported (2-tailed; FDR corrected within black box, other values are not corrected); ***p<0.001, **0.001<p<0.01, *0.01<p<0.05 (2-tailed).

5.9 Discussion

Study 2 examined the relationship between anxiety or psychopathic traits and individual differences in OFL. The study sample was drawn from a large sample of males posted on the MTurk platform, building upon the paradigm validation from the previous study. MTurk workers' OFL performance was consistent with what was reported previously indicating that the population of workers was similar to the community sample in Study 1. Contrary to what I had predicted, trait anxiety was positively correlated with prediction response time. This association was not accounted for by affective empathy, which showed a modest relationship with trait anxiety, and also remained significant after controlling for cognitive ability. Additionally, there were nominal significant correlations between trait anxiety and total accuracy as well as learning rate that did not survive correction, which were subsequently followed up with post-hoc sensitivity analyses comparing the top and bottom 15% of participants for trait anxiety. High-anxiety participants had significantly longer prediction response times and worse OFL accuracy than those with low-trait anxiety; however, there were no significant differences in learning rate between groups. Also contrary to my predictions, trait anxiety was not associated with any other OFL measures, including learning rate and temperature parameters nor did psychopathic traits correlate with any OFL measures.

A positive correlation between trait anxiety and prediction response time, while not expected, could be explained by a number of factors. Prior studies have shown that anxious individuals can be hypervigilant or respond more strongly in situations of threat and uncertainty (Grafton & MacLeod, 2014; Lonsdorf & Merz, 2017; Mogg, Millar, & Bradley, 2000; Rudaizky, Basanovic, & MacLeod, 2014). It may be that high-anxiety individuals are over-attending to CSs in such a way that interferes with prediction response time (and accuracy, as evidenced by the difference between the highest- and lowest-trait anxiety participants). Moreover, it is possible that the individuals with high-trait anxiety are generally worse at probabilistic learning (Berenbaum, Thompson, & Bredemeier, 2007). LaFreniere &

Newman (2019) reported that individuals with generalized anxiety disorder learned at a slower rate when performing positive and negative reinforcement probabilistic learning tasks than healthy controls. Additionally, participants with generalized anxiety disorder were more sensitive to aversive cues and more motivated to avoid negative outcomes than those without generalized anxiety. Whilst Study 2 did not record participant's clinical diagnoses, similar impairments may exist pre-clinically (Coles, Turk, & Heimberg, 2007) and might be detected despite this lack of information.

Similarly, no associations were found between psychopathic traits and OFL. Based upon the range in scores presented in the descriptive statistics, the lack of relationship is not due to these traits being either absent or invariant in this sample. Instead, it may be that the OFL task itself is not optimal for exploring the association between dispositional traits and individual differences in OFL. I will next consider why this might be.

Whilst I had anticipated more robust associations between dispositional traits and individual differences in OFL, there are a few limitations to the task that may have contributed to the failure to detect such associations. The crucial, naturalistic components that likely contribute towards the individual complexity of OFL are limited in the current context (Szczepanik et al., 2020). In stark contrast to the animal OFL paradigm, which provides an ecologically valid and emotionally salient environment, viewing a stranger on screen to predict whether they might receive an uncomfortable electrical shock in response to one image or another most likely will not elicit an experience of fear or threat in an observer, who is very much removed from the experience. Additionally, in the mouse OFL paradigm, effective transmission of fear is dependent upon the observer's ability to attend to the demonstrator's reaction to the US and the fear in response to the CS (Jeon et al., 2010; Jeon & Shin, 2011). However, in the present paradigm there is nothing but observation the entire task, which may be diluting crucial individual differences in how participants might react in a natural setting and attend to someone else's distress. Alterations to the task to enhance the realism of OFL and possibly better

engage the relationship between dispositional traits and OFL might be addressed by employing technologies like using a virtual reality setup (Parsons, Gaggioli, & Riva, 2017; Tarr & Warren, 2002), or having a demonstrator or environment that was familiar to participants (Golkar & Olsson, 2017). However, such features are nearly impossible to implement on an international platform like MTurk. However, small alterations to make learning more difficult, such as having the probability levels more similar for both CSs or including a reversal of associations, might reveal relationships between individual differences in OFL and anxiety and psychopathic traits.

Additionally, no demographic information was collected for these workers beyond gender and being at least 18 years of age. Therefore, there are numerous factors unaccounted for from this sampling that may confound these results. These include clinical diagnoses and treatment, ethnicity, nationality and residence, and English comprehension (Arditte, Çek, Shaw, & Timpano, 2016; Buhrmester, Kwang, & Gosling, 2015; Chmielewski & Kucker, 2020; Fleischer, Mead, & Huang, 2015; Golkar et al., 2015; Golkar & Olsson, 2017; Huff & Tingley, 2015; Kennedy et al., 2020; Kirmayer, 2001; Selbing & Olsson, 2017; Wittchen, Essau, & Krieg, 1991). Without having more information about the participants in the current study, it is difficult to make comprehensive conclusions as there are likely many subgroups represented by these MTurk workers that may wash out any strong correlations between dispositional traits and individual differences in OFL.

Furthermore, the online environment of MTurk in general may not be the best platform to conduct studies looking for relationships between traits and learning. The following points are not likely to be the reason for a lack of findings between dispositional traits and OFL, but may want to be considered if carrying out similar studies using MTurk. One consideration is that recent studies have reported issues of inattentive responding and declining data quality from MTurk workers (Chmielewski & Kucker, 2020; Fleischer et al., 2015; Kennedy et al., 2020), which could have consequences for how accurate responses to questionnaires of dispositional traits were and how the OFL task was completed. Another consideration is that recognizing fearful faces has been demonstrated to be

significantly worse in online studies (Bartneck, Duenser, Moltchanova, & Zawieska, 2015), which may mean that performing an OFL task online is not necessarily the most optimal way to study this form of learning. Lastly, there are conflicting reports about how accurately an MTurk sample represents the general population (Bartneck et al., 2015; Buhrmester et al., 2015; Huff & Tingley, 2015), particularly for studies where participants' psychopathologies are a crucial component of the research (Arditte et al., 2016); however, this seems unlikely for the present study as scores are in line with prior studies, but should be considered if running future OFL studies on MTurk. Ultimately, using MTurk as a platform for the current study allowed for easy data collection from a large sample of participants, but may have come at the cost of not being the most optimal means for conducting a study of this nature.

Study 2 replicated the results of what was found in Study 1 in a large online sample on the MTurk platform. I found a modest correlation between trait anxiety and median prediction response time; however, no other measures of OFL were found to be associated with trait anxiety nor psychopathy. It is, therefore, likely that this study was suboptimal for charting trait driven differences in OFL in humans. Future studies should consider modifying the task to make it more ecologically relevant and possibly make the learning more challenging. Addressing the discussed limitations in future studies could provide crucial information about how we learn from others in our environments and what may predispose some individuals to develop trauma-related disorders whilst others remain unaffected.

Chapter 6 General Discussion

Learning from observing how others engage with the environment provides significant adaptive value, especially when the environment may pose potential harm to the observer. Social transmission of information is a highly conserved mechanism that allows for learning about threats without having to directly experience them, therefore ensuring continued safety and survival (Debiec & Olsson, 2017; Kim et al., 2019; Olsson et al., 2020). This thesis was conducted with the overall aim of advancing our current understanding of the neurocircuitry and individual differences in trait factors subserving OFL in a behaviourally translational manner. The work within this thesis established and validated experimental OFL paradigms in both mice and humans. Moreover, a combination of behavioural analyses, immediate-early gene mapping, neuronal tracing, in vivo optogenetics, and Ca^{2+} imaging in mice were used to interrogate the neural circuits contributing to the learning and memory of observed fear. Additionally, computational modelling of trial-by-trial variation in behaviour as well as the relationship between OFL and dispositional traits were assessed in humans. This concluding chapter provides a summary of findings from Chapters 2-5, followed by a discussion of some of the main findings and how they contribute to current research in the field. Finally, future directions are explored bearing in mind some of the limitations of this work. This chapter concludes with discussion of the translational potential of this line of research.

6.1 Summary of findings

The first part of my PhD thesis focused on OFL in mice, covered in the in experimental chapters 2-4. Chapter 2 detailed the development, validation, and characterization of a cued-OFL behavioural assay in mice. Findings reported in Chapter 2 suggest that mice form a lasting, stimulus-specific memory solely through observing direct conditioning of an unfamiliar conspecific. I showed that the observation of footshock deliveries to the demonstrator was critical for the

acquisition and retrieval of OFL. Moreover, the cued-OFL paradigm proved to be specific to the observational learning of the CS-US association and not the result of generalization to non-US associated cues. Additionally, cued-OFL did not on average produce an enduring phenotype of anxiety-like behaviour or social disturbance.

Chapter 3 investigated the neurocircuitry underlying OFL using immediate-early gene mapping and neuronal tracing techniques. This revealed that OFL significantly engages several cortico-limbic areas involved in emotion processing, including the PL, pACC, CLA, BA, and vHPC. The latter four regions as well as the AI showed significant engagement of projections to the PL following OFL as compared to naïve controls. As compared with DFL mice, OFL recruitment of these five PL-projecting regions did not significantly differ from mice that had undergone DFL suggesting these pathways are similarly engaged regardless of direct or observed conditioning. However, while overall cellular activity within the cortical regions of interest were significantly lower for observers than demonstrators, this was not the case for the vHPC and BA. This in turn led to an investigation of how outputs from the vHPC and BA to PL were uniquely engaged in OFL. I then reported that trans-neuronal labelling demonstrated a vHPC→PL circuit that targeted glutamatergic PNs as well as PV INs in the PL. Additionally, this technology allowed for the visualization of output regions of the vHPC→PL pathway, which included the BA, AI, CLA, PAG, NAc, and NAs. This anatomical characterization provided critical direction for the functional experiments addressed in the following chapter.

Chapter 4 examined the functional neural networks subserving OFL using in vivo optogenetics and Ca^{2+} imaging via fibre photometry technologies. First, the PL was demonstrated to be required for the acquisition of cued-OFL. More specifically, the engagement of local PV INs were shown to be necessary for the consolidation of an observed fear memory. However, once an observational fear memory was formed, the PL was no longer required for the retrieval of OFL. Second, I demonstrated that upstream inputs to the PL from the vHPC and BA had opposing influences over OFL. Outputs from the vHPC to PL provided lasting constraint of OFL, whilst outputs from the BA to PL acutely promoted fear learning, but did not

contribute to the retrieval of the fear memory. Meanwhile, neuronal activity from the PL to l/vlPAG demonstrated responsivity to the CS experienced through observation. Specifically, Ca^{2+} activity to both the CS and US increased over the course of OFL, population activity to the CS ramped up steadily over conditioning while the circuit was already significantly engaged to the US from the start of OFL. Finally, using a combination of in vivo optogenetics and fibre photometry during OFL, vHPC→PL was photosilenced while Ca^{2+} activity was recorded from PL→l/vlPAG. This demonstrated that vHPC is recruited to actively constrain PL→l/vlPAG.

In Chapter 5 I characterized OFL in a human population. Study 1 detailed the development and validation of a cued-OFL task in humans using an instrumental reinforcement learning framework to model trial-by-trail variation in learning. I first demonstrated that human participants could successfully learn and predict a threat association through observation and were significantly more accurate and quicker in predicting a shock outcome to the high-probability CS. Moreover, this accuracy and speed was not influenced by the intensity of the demonstrator's shock reaction. Using a computational reinforcement learning model to characterize participant learning over the course of the task revealed a learning model favouring a single learning rate and temperature parameter to describe OFL. The low learning rate derived from the model prediction suggested learning was not heavily influenced by prediction errors and participants updated their valuation of the CSs relatively slowly. The low temperature parameter indicated consistency in prediction choice. Finally, while no differences in learning were apparent when comparing the medium-shock reaction group to the high-shock reaction group, the temperature parameter for the high group was significantly less than the medium group indicating a greater degree of determinism in predictions when observing exaggerated fear reactions.

Chapter 5, Study 2 applied the validated OFL task from Study 1 to a large sample (n=290) on the MTurk platform in order to assess the potential influence of dispositional traits on individual differences of OFL. No significant correlations were found between psychopathic traits and OFL. However, trait anxiety was positively correlated with median prediction response time in this sample, such that high levels

of trait anxiety were associated with slower prediction response time. This association was not accounted for by affective empathy, which showed a modest relationship with trait anxiety, and also remained nominally significant after controlling for cognitive ability. Additionally, there were nominal significant correlations between trait anxiety and total accuracy as well as learning rate that did not survive correction. These associations were subsequently followed up with post-hoc sensitivity analyses, which compared the top and bottom 15% of participants for trait anxiety. High-anxiety participants had significantly longer prediction response times and worse OFL accuracy than those with low-trait anxiety; however, there were no significant differences in learning rate nor temperature parameters between groups.

6.2 Synthesis, limitations, and future directions

6.2.1 How the current findings contribute to our understanding of OFL in mice

Despite the decades of research characterizing classical fear learning in rodent models and delving into the intricacies of the neurocircuitries contributing to fear learning and memory (Bukalo et al., 2015; Giustino & Maren, 2018; Holmes & Singewald, 2013; Marek, Sun, & Sah, 2019; Milton, 2019; Ressler, 2020), it has only been just over a decade since a defined model of OFL has entered into the field (Jeon et al., 2010; Jeon & Shin, 2011). Until that point, studies had mostly focused on fear-by-proxy designs, which, whilst providing ample evidence of inter-mouse communication of environmental information, the studies varied considerably in design (Church, 1959; Jones et al., 2018; Kavaliers et al., 2003; Knapska et al., 2010; Yusufishaq & Rosenkranz, 2013). The seminal protocol published by Jeon and Shin (2011) on OFL combined the decades of research on DFL, earlier findings of OFL in humans by Olsson and colleagues (2004, 2007) in humans, and the diverse research validating social communication of fear and threat in rodents to develop a contextual-OFL paradigm. With this, research on the behavioural intricacies and

underlying mechanisms involved could be examined in a standardized and comparable way across different studies. However, while certain research groups have found acceptably strong fear responses using contextual-OFL, others have reported a much greater degree of variability in behaviour (Allsop et al., 2018; Morozov & Ito, 2018). Additionally, while our understanding of OFL continues to grow, most studies have focused on the PFC, AMG, or the pathways between them (Allsop et al., 2018; Burgos-Robles et al., 2019; Chen & Hong, 2018; Ito et al., 2015; Ito & Morozov, 2019; Jeon et al., 2010), leaving out many other possible regions and circuits that may contribute to OFL and which should be systematically examined.

6.2.1.1 Advancing our understanding of the characteristics of cued-OFL in mice

In order to meaningfully contribute to the development of OFL research, I first established and characterized a cued version of the task in mice. The addition of CS-US pairings provides further associative information that can enhance learning more than context-only learning paradigms (Rustay et al., 2008). Natural examples of OFL, such as witnessing a bombing, can create strong, aversive memories to the sounds and smells of the event as well as to the location, ultimately leading to intense anxiety and persistent trauma. By including an additional sensory component to the task, in this case a white noise tone, I sought to better understand how a neutral trigger could evoke pathological levels of fear observationally.

In validating cued-OFL, I demonstrated that observation of repeated footshocks delivered to an unfamiliar demonstrator resulted in a lasting cued-fear memory. Freezing behaviour was impaired without access to visualization of the shock delivery itself, possibly suggesting that observer freezing may be more than a mimicry effect of fearful behaviour displayed by the demonstrator. A supplementary experiment could be done where observation to everything except for the US is blocked. This could provide additional information on how OFL is transferred. For example, it may be that OFL is dependent on the observed US, but more likely, OFL is a combination of watching the demonstrator's behaviour through the length of the task, in addition

to observing the US, which may provide contextual information or assigning meaning to the demonstrator's fear response.

Similarly, retrieval of OFL is specific to the CS. This was evidenced by observers' ability to significantly discriminate the CS from a novel stimulus. While other OFL studies have demonstrated contextual-OFL (Ito et al., 2015; Jeon et al., 2010; Jeon & Shin, 2011), I reported in Chapter 2 that it was the CS that evoked a freezing response during retrieval and not a return to the same conditioning environment nor a non-associated stimuli. There are a number of future experiments that could be performed to more closely assess how OFL is retrieved. A simple one would be to reverse the order of stimuli during retrieval such that the novel stimulus is presented first followed by the CS. Additionally, the strength of the CS could be considered by performing OFL retrieval in a novel context. This would provide information about whether the CS alone or the CS in addition to the conditioning context is necessary for OFL retrieval. For DFL studies, extinction training in a novel context has shown that a cued-fear memory is modulated by new learning that forms and strengthens neural circuits working to that reduce fear response in the presence of CS (Bukalo et al., 2015; Senn et al., 2014), but it remains uncertain if exposure to the CS in a novel context would have a similar effect on OFL.

Not only did I show the importance and specificity of the CS and US for acquiring and retrieving OFL, but I also provided evidence that observers' fear response was specific to OFL and did not cause a phenotype of heightened anxiety nor social aversion. Previous studies using different paradigms of socially-transmitted fear learning have demonstrated subsequent increases in anxiety and changes in sociability (Knapska et al., 2010; Krishnan et al., 2007), whilst anxiety and sociability following OFL were unchanged. This may be due to other tasks employing more aggressive versions of social learning that can be interpreted more as an imminent threat to a test subject, as is the case with social defeat. Additionally, OFL, as I have characterized it, is an acute occurrence. There is no reinforcement component of the learning beyond what is experienced during the 30 trials of conditioning. Perhaps repeating the OFL paradigm over the course of multiple days or providing a direct

experience of the US to the observer may produce phenotypic changes to anxiety and socialization; however, this lays beyond the bounds of this thesis's focus, but could serve as an interesting future direction to titrate how much observational fear exposure is required before impacting anxiety and socialization. It is therefore unsurprising that the OFL paradigm utilized in these experiments did not produce phenotypic changes.

Ultimately, while demonstrator freezing behaviour is significantly higher than that of observers, OFL behaviour demonstrates differing strategies to that of DFL. Direct CS-US exposure was almost entirely represented by freezing, in line with the literature (Fanselow, 1980), whereas OFL was most consistently measured by freezing response, but also displayed additional behaviours, such as grooming, rearing, and digging that may provide further information about the vicarious fear experience and learning strategies. Other social learning paradigms have reported pro-social behaviours, like allogrooming, increased pain sensitivity, or social buffering (Bartal et al., 2011; Bartal et al., 2014; Karakilic et al., 2018; Kikusui et al., 2006; Langford et al., 2006; Lu et al., 2018; Luo et al., 2020; Morozov & Ito, 2018). The variability in behaviours exhibited by observers reported within Chapter 2 might be indicative of individual differences among mice. Further comparative analyses addressing differences in proportions of behavioural activities across a large sample size of mice might reveal information processing differences not yet characterized in OFL. While a greater degree of variability exists in OFL behaviour as compared to DFL, the paradigm described in this thesis produced a robust and reliable fear memory that could be bidirectionally manipulated (Chapter 4).

6.2.1.2 Demonstrating a central role of the PL in acquiring OFL

Due to the relative recency of OFL research, the majority of what is understood about the underlying neural substrates of fear learning comes from the direct experience of an aversive event. From the well-established DFL literature, several key regions have been identified for their contribution to the processing and retrieval

of an associated fear memory. Particularly, the PL has received much attention for its critical involvement in acquiring and maintaining both cued- and contextual-fear (Burgos-Robles et al., 2009; Corcoran & Quirk, 2007; Milad & Quirk, 2012; Rozeske et al., 2015; Shibano et al., 2020; Sierra-Mercado et al., 2010; Tovote et al., 2015). Moreover, recent OFL studies have revealed potential involvement of the dmPFC, and more specifically, the adjacent ACC, in processing and regulating vicarious learning (Allsop et al., 2018; Burgos-Robles et al., 2019; Ito et al., 2015; Ito & Morozov, 2019; Jeon et al., 2010; Keum et al., 2016; Keum et al., 2018; Kim et al., 2014; Liu et al., 2017). Modulating OFL in this way, the dmPFC appears to contribute to the processing and perceiving pain in oneself and in others (Sivaselvachandran et al., 2018). Beyond these findings, however, and in marked contrast to the vast literature on DFL, the neural underpinnings of OFL remain unknown.

My next objective was to identify novel neural correlates of cued-OFL by using c-fos mapping to survey OFL-related neuronal activity in cortico-limbic regions involved in emotion regulation (Chapter 3). Several brain areas were identified as being significantly activated in response to OFL as compared to controls, including the pACC and BLA, which was in line with electrophysiological and optogenetic data previously reported (Allsop et al., 2018; Jeon et al., 2010). Additionally, I found OFL-related c-fos activity in brain regions involved in fear or affective pain processing, but not previously implicated in OFL, including the AI and vHPC. Given the known role of the PL in DFL and PFC in OFL as well as my finding that the PL is engaged by cued-OFL, suggests the dmPFC region could be especially well-placed to assimilate social and threat-related information to support cued-OFL.

This led me to assess the causal contribution of PL on OFL using an in vivo optogenetics approach to silence PL neuronal activity (Chapter 4). I first photosilenced PL activity in CaMKII-expressing neurons of observer mice during each conditioning CS (which included the US footshock delivery to the demonstrator) during OFL. This resulted in a reduction in CS-evoked freezing during conditioning as compared to a control group of GFP-expressing observers. Importantly, this attenuated CS-related freezing remained lower during a light-free retrieval test the

following day, which was consistent with a decrement in OFL memory formation. In a separate experiment, I found that silencing the PL only during retrieval CSs did not affect fear behaviour, indicating that the PL was necessary for the formation, but not the retrieval, of cued-OFL. This dissociation contrasts the reported involvement of PL for the expression of cued-DFL in rats (Burgos-Robles et al., 2009), but echoes prior OFL studies showing loss-of-function manipulations of the ACC that disrupt the acquisition, but not expression, of contextual-OFL and shock-primed OFL (Allsop et al., 2018; Jeon et al., 2010; Kim et al., 2012).

One potential mechanism for how PL encodes OFL is through sparsely populated inhibitory GABAergic INs. In contrast to the optogenetic photosilencing of CaMKII-expressing PNs in the previous experiment, which showed a critical contribution of the PL in acquiring OFL, INs are believed to inhibit PNs as a means of moderating the specificity of fear learning (Courtin et al., 2014). Specifically, PV INs in the dmPFC have been demonstrated to have a crucial role in driving fear expression as well as neuromodulation in social fear learning (Courtin et al., 2014; Zhou et al., 2018). To determine the role of PL PV INs on cued-OFL, I replicated the same in vivo optogenetics silencing experiment during CS-presentation of OFL, but instead used PV-Cre specific mice, which allowed for selective targeting of PV INs (Chapter 4). Unlike PL CaMKII-expressing PNs, I found that there were no differences in CS-related freezing between experimental and control groups during OFL; however, there was a significant attenuation of CS-evoked freezing during the retrieval session the following day between groups. This finding suggests PL PV INs play a more subtle contribution towards the consolidation of OFL memory formation. Together, this provides a more nuanced understanding of a potential local mechanism within the PL for how OFL is acquired and processed.

In a similar experiment, Zhou and colleagues (2018) demonstrated that chemogenetic inhibition of ACC PNs supressed contextual-OFL freezing behaviour, in line with our PL CaMKII-expressing inhibitory experiment; however, they showed that activating PV INs produced an acute attenuation of OFL. It should be noted that their OFL paradigm involved context-only conditioning and did not include an OFL retrieval

component, nor did they report bidirectionally testing PV INs, which may contribute to our contrary findings. Additionally, contextual-OFL studies have also demonstrated a modulatory role of Sst INs in the dmPFC (Keum et al., 2018; Xu et al., 2019). Further experiments addressing the role of Sst INs in cued-OFL as well as the effect of activating INs should be conducted to address the current inconsistencies in the literature and to develop a more fine grained understanding of the neural mechanisms underpinning OFL.

There is increasingly emergent evidence that the PL has a particularly important part in learning and expressing appropriate fear associations under conditions of uncertainty and ambiguity. For example, PL disruptions impair discrimination between threat and safety (Burgos-Robles et al., 2009; Fitzgerald et al., 2014; Livneh & Paz, 2012; Ye et al., 2017) as well as contextual- and cued-discrimination (Klavir et al., 2013; Likhtik et al., 2014; Rozeske et al., 2018; Xu & Südhof, 2013). The studies presented within this thesis support involvement of the PL in utilizing available environmental information to direct attention to the appropriate predictors of threat as well as in gating learning and behaviour accordingly (Furlong et al., 2010; Marquis et al., 2007; Sharpe & Killcross, 2015a; Sharpe & Killcross, 2015b). This mediation of the PL could be particularly necessary in situations of conflicting safety and threat information (Likhtik & Paz, 2015), as is the case with OFL where there are indicators of danger from the defensive responses of the demonstrator, yet no harm is physically experienced by the observer. OFL may therefore be a special case of ambiguous fear memory that is dependent upon the PL and particularly affected by PL dysfunction. Moreover, the literature on memory engrams presents a case for specific groups of cells in the mPFC that become activated during aversion learning and become strengthened over time as an episodic memory is consolidated into a long-term memory (Kitamura et al., 2017; Zelikowsky, Hersman, Chawla, Barnes, & Fanselow, 2014). Without proper functioning of the PL during OFL, as found in the optogenetic experiments presented in this thesis, it is possible that the fear memory is never formed, and therefore, cannot be recalled in the future. Together, these data reveal a novel and critical role for the PL in the instantiation of cued-OFL and

may reflect the ability to integrate environmental cues to disambiguate incomplete or ambiguous information supporting the selection between competing defensive response options.

6.2.1.3 Demonstrating the importance of cortico-limbic inputs to PL

While I demonstrated PL is causally required for OFL acquisition, the PL is not singularly responsible for OFL. The PL is heavily innervated by a number of cortical and subcortical regions providing support for its role in cognitive processes (Hoover & Vertes, 2007) in addition to providing top-down regulation of emotional responses during DFL by integrating diverse neural inputs from various regions (Giustino & Maren, 2015; Marek et al., 2013; Padilla-Coreano et al., 2016; Parfitt et al., 2017; Senn et al., 2014; Sotres-Bayon et al., 2012; Ye et al., 2017; Yizhar & Klavir, 2018). In order to assess the PL projecting cortico-limbic circuits engaged in OFL, I used a combination of the fluorescent retrograde tracer, CTb, with c-fos to visualize and quantify PL projecting activity (Chapter 3). The three cortical-PL circuit regions (pACC, AI, CLA) and two limbic-PL regions (BA and vHPC) all had significant pathway-specific activation in response to OFL. While this PL-input activation following OFL suggests that OFL may tap into many of the same neural circuits cited above, these data only indicate a degree of equivalency in input recruitment between OFL and DFL. It does not necessarily show that the same projection neurons are engaged or that the same neuronal populations are targeted in the PL. Additionally, the recruitment of inputs from the pACC is noteworthy given this region has been shown to be a critical locus for OFL in prior studies (Burgos-Robles et al., 2019; Carrillo et al., 2019; Jeon et al., 2010; Keum et al., 2018; Pisansky et al., 2017; Sakaguchi et al., 2018). Interestingly however, the engagement of BA and vHPC projections to PL does not significantly differ between DFL and OFL, while there was significantly more pathway specific engagement in cortical projections in DFL mice than OFL mice. This suggests the limbic pathways are similarly engaged regardless of direct or indirect exposure.

To interrogate this finding further, I used an in vivo optogenetics approach to investigate whether the vHPC-PL and/or the BA-PL pathways are causally involved in

OFL (Chapter 4). Upon photosilencing vHPC inputs to the PL during CS presentations during the conditioning session of OFL, I found that inhibiting this pathway augmented CS-related freezing behaviour in experimental mice as compared to YFP controls during both OFL and retrieval. This indicates that the vHPC→PL circuit serves to limit the degree of OFL. On the other hand, when BA inputs to PL were silenced, there was an attenuation of freezing behaviour to the CS during conditioning, that did not persist into the retrieval test. Whilst the effect during conditioning was consistent with the pro-fear role ascribed to the BA→PL pathway in DFL (Burgos-Robles et al., 2009; Jin & Maren, 2015; Klavir et al., 2017; Senn et al., 2014; Sotres-Bayon et al., 2012), unlike in DFL, the inability of silencing to cause a lasting decrease in cued-OFL suggests that a loss of BA innervation can be compensated for by other inputs, such as from the mediodorsal and parafascicular nuclei of the thalamus, which serve to relay pain signals through the lateral and medial pain systems (Jeon et al., 2010; Keum & Shin, 2019). Freezing behaviour remains the most consistent measure of OFL across mice, however, these differences in circuitry between observed versus direct fear may highlight that freezing behaviour may be a weaker indicator of learning in OFL as it is in DFL. Additionally, these discrepancies may represent the more ambiguous nature of OFL and the differences in interpretation of vicarious threat, which ultimately could rely upon more nuanced contributions of many regions as opposed to largely being regulated by a single pathway or two. Despite these postulations, it appears that vHPC and BA have opposing influences over OFL, but additional work falling beyond the scope of this thesis is needed to calibrate the contributions of these pathways between OFL and DFL.

Due to the lasting response of the vHPC→PL pathway in constraining OFL behaviour using optogenetics, I next provided further evidence that the vHPC inhibits PL using Ca^{2+} imaging to measure the pathway activity of PL projecting vHPC neurons during retrieval (Chapter 4). vHPC→PL activity significantly increased in response to the first CS and steadily ramped up over the course of the 5 CS presentations, including the light-free interval between CSs. Moreover, when observers were divided between high- and low-CS-evoked freezing behaviour, low-freezing observers

were found to be driving the significant increase of Ca^{2+} activity in response to the CS. This was in opposition to observers with the highest freezing rates, which demonstrated non-significant changes in Ca^{2+} activity over the course of the experiment. This may indicate the degree of protective constraint the vHPC has over the PL by ramping up activity in order to suppress a fear memory in the absence of threat.

One possible mechanism of how vHPC might modulate PL-driven observationally acquired fear is through INs. To date, there are no social fear learning studies that have addressed the vHPC→PL pathway; however, excitation of vHPC PV INs has been shown to be crucial for discerning familiar versus novel conspecifics (Deng et al., 2019). In combination with my prior finding that PV INs in PL are involved with consolidation of OFL, I performed a tracing study to selectively visualize cell type specific labelling of vHPC inputs to PL (Chapter 3). Using a combination of Cre mice to target three prominent IN types in the PL (PV, Sst, and VIP) as well as PNs (via vGluT), I found that vHPC neurons preferentially targeted glutamatergic PNs and PV INs in the PL. This was further substantiated using a viral combination using Cre-dependent constructs with immunohistochemical staining to identify neuronal subtypes in PL receiving monosynaptic vHPC inputs. CaMKII expressing PNs and PV INs in PL receive direct inputs from vHPC neurons (Ährlund-Richter et al., 2019). Due to the monosynaptic connections onto inhibitory PV INs, the inhibitory and excitatory nature of this circuit suggests PV INs may exhibit an internal modulatory inhibition of the PL PNs to affect an inhibitory response. Future experiments should be designed to specifically investigate the causal role of inhibiting vHPC→PL PV INs to differentiate between glutamatergic and GABAergic constraint on PL.

Overall, these findings are in keeping with electrophysiological data indicating HPC neurons exert a strong inhibitory influence on mPFC activity through local cortical interneurons (Ishikawa & Nakamura, 2003; Tierney et al., 2004) concurrently balanced by some degree of excitatory influence (Padilla-Coreano et al., 2016). It is also in line with earlier evidence that vHPC projections to PL decreased DFL in the presence of safety cues as well as after extinction (Burgos-Robles et al., 2009; Likhtik

& Paz, 2015; Sotres-Bayon et al., 2012). Furthermore, the vHPC→IL pathway is recruited and necessary for renewal of extinguished fear by exposure to a novel context (Jin & Maren, 2015; Marek, Jin, et al., 2018; Wang et al., 2016). Additionally, it is likely that disrupting vHPC outputs might impact the encoding of HPC place cells during OFL, which have been shown to be activated specifically in response to social observation (Mou & Ji, 2016). Therefore, vHPC→PL neurons may be recruited to reduce fear in situations where there is conflicting information about danger/safety associated with a cue. This may be the case in cued-OFL where an observer learns about an associated threat, but does so from a position of relative safety. It is thus suggested that this circuit may convey higher-order information that disambiguates threat signalled by the behaviour of the demonstrator from the absence of concomitant, directly experienced harm. While the published research on OFL neurocircuitry has been almost entirely focused on BA-mPFC pathways, I have provided novel evidence for vHPC constraint of PL in OFL.

6.2.1.4 Demonstrating vHPC modulation of l/vlPAG-projecting PL neurons during OFL

As discussed above, limbic inputs to the PL are crucial for observational fear regulation, yet the outputs from this region are largely involved in the mediation of fear expression (Amorapanth et al., 1999; De Oca et al., 1998; Franklin, 2019; Rozeske et al., 2018). In order to address the downstream targets engaged by the vHPC→PL pathway, I performed a tracing experiment to specifically visualize regions receiving direct input from the vHPC through PL projections (Chapter 3). Several areas associated with emotion processing were identified, including those that have reciprocal inputs to the PL, like the BA, CLA, and AI as well as regions known to be involved in fear expression like the PAG, NAc, and NAs. Of these identified neuronal targets, the midbrain PAG was of particular interest based upon emergent evidence that PAG-projecting mPFC neurons regulate a number of processes that are likely important for OFL, including social stress, fear discrimination, and punished conflict (Franklin et al., 2017; Rozeske et al., 2015; Rozeske et al., 2018; Siciliano et al., 2019;

Vander Weele et al., 2018; Watson et al., 2016). For this reason, I sought to investigate the potential role of the l/vlPAG as engaged through vHPC-PL mediation.

To first address the contribution of l/vlPAG-projecting PL neurons in OFL, endogenous in vivo correlates of behaviour were assessed using fibre photometry to image Ca^{2+} activity (Chapter 4). I found an event-related alignment of significantly increased Ca^{2+} signal to both the CS and US presentation onset as compared to pre-stimuli baseline, which was evident for the entirety of conditioning. To better understand what this total change in activity might mean, I compared the first five events with the last five events so Ca^{2+} activity could be examined in the early and late stages of learning. Specifically, for the CS response, the Ca^{2+} activity observed might reflect a sensory response to the tone, or may represent the accumulation of associative strength to the CS as a predictor of threat. While Ca^{2+} activity was modest during early-learning, by late-learning the signal was robust, indicative of a strengthening of l/vlPAG-projecting PL neurons engagement as observers learn and assign value to the CS over time. In contrast, US-related activity was significantly elevated from the pre-US period during early- and late-learning. This may reflect the responsiveness of PL projecting neurons in the PAG to defensive reactions exhibited by demonstrators during shock delivery (e.g., flinching, jumping, vocalizing). This is supported by a recent study that detected enhanced Ca^{2+} response to directly experienced footshocks in mPFC neurons projecting to the PAG (Vander Weele et al., 2018). The present finding that the l/vlPAG-projecting PL neurons are also responsive to input experienced through observation speaks to the high sensitivity of these specific PL neurons to aversive stimuli. From these findings it seems that the PL neurons projecting to l/vlPAG are not only engaged by event-related stimuli during OFL, but are also involved in associative learning over time.

While the previous experiment suggests a role for l/vlPAG-projecting PL neurons in calibrating OFL, I wanted to understand how the constraint of vHPC on PL impacts downstream activity of PL cells projecting to the l/vlPAG. To address this, I used both in vivo optogenetics to photosilence vHPC terminals in the PL with concurrent Ca^{2+} imaging of l/vlPAG-projecting PL neurons during OFL (Chapter 4).

Silencing vHPC inputs to PL resulted in a robust disinhibition of PL→l/vlPAG cells. What remains unclear, however, is the effect of inhibiting vHPC on PL-l/vlPAG activity due to the CS. While it is demonstrated that photosilencing vHPC-PL neurons disinhibited l/vlPAG-projecting PL cells, because photosilencing and the CS occur concurrently, any Ca^{2+} activity change due to the CS *per se* is obscured. Despite this occlusion, this experiment illustrated a crucial role of vHPC modulation of a cortico-midbrain circuit. Moreover, the PL promoted OFL, whilst vHPC constrained this behaviour, including downstream activity of l/vlPAG-projecting neurons from the PL. In effect, this vHPC→PL→l/vlPAG disynaptic pathway may provide a safety signal during OFL which, when compromised, would amplify the formation of fear memories for witnessed traumatic events. The identification of this critical pathway for gating cued-OFL opens the door for additional research into the mechanistic aspects of the observed inhibitory control of vHPC inputs to the PL and PL outputs to the l/vlPAG in OFL expression.

In summary, the mouse work in this thesis has established and characterized a cued-OFL paradigm whereby a stimulus-specific memory for a discrete environmental cue paired with a footshock is acquired through observation. The PL, whilst on its own is critical for OFL acquisition, receives opposing signals from the BA (pro-freezing) and vHPC (pro-constraint), which negatively gate the observational fear response. Furthermore, l/vlPAG projecting PL cells may be involved in assigning associative value to the CS and are particularly sensitive to the more salient US; however, without vHPC constraint over PL, PL-l/vlPAG neurons are disinhibited. This not only provides novel evidence for vHPC modulation of a cortico-midbrain circuit gating OFL, but also that vHPC, PL, and l/vlPAG may be recruited to calibrate OFL.

As I have pointed out, there are some limitations to this work both in the imperfect nature of working with mice, technological challenges, and various control experiments that could be performed. One major limitation is measuring OFL entirely by freezing behaviour, despite there being several behaviours mice exhibit in response to OFL that the combination of may be more telling about the vicarious nature of fear learning. It may be that there are different behavioural strategies that

mice employ in response to a more ambiguous visual threat than that which is directly experienced. For example, the optogenetic studies, while impacting freezing behaviour, may also have a more nuanced effect on learning, such that a lack of freezing does not necessarily mean an absence of fear, but may indicate a change in threat perception and subsequent safety strategy weighing whether to flee or freeze. Moreover, while I demonstrated that mice require observation of the US delivery to induce freezing behaviour, it is not entirely clear whether mice are reacting in self-defence, out of empathy for the demonstrator, as mimicry, or a combination of all of the above. Whilst it will take some time for researchers to find adequate ways to understand the plight of mice in these cleverly devised behavioural paradigms, computational behaviour tracking has come a very long way from the start of this PhD project (Geuther et al., 2019; Mathis et al., 2018). Utilizing cutting-edge imaging analyses may provide profound insights into how OFL is perceived and interpreted, which in turn may deepen our understanding of the function of neuronal networks.

Another limitation was that due to the time constraints of a doctoral research programme, I relied heavily on the existing literature around DFL to inform what brain regions and circuits might be most related to OFL. For the c-fos studies, I limited the regions of interest to those that were involved in emotion processing and fear learning as opposed to doing a general study of the entire brain to find areas that were found to be specifically activated by OFL. Whilst I did identify a causal disynaptic circuit involved in OFL, it is always important to note that these isolated regions and circuits do not exist in a vacuum and there are unequivocally other networks involved in learning and processing observationally acquired fear beyond what this thesis presents. As research into OFL is still in its infancy, there are a number of different ways in which it is possible to decode this highly evolutionarily conserved learning mechanism. As technologies become available, the possibilities of interrogating the mechanisms underlying OFL also grow. Given the current state of validated techniques, future experiments should focus on whole brain activity quantification after OFL and retrieval, with a particular emphasis on how it is similar or differs from DFL.

Combining our anatomical knowledge with a brain-wide analysis of the unique and similar components of OFL, new regions, pathways, and networks could be discovered at a much more rapid pace and in a more systematic way. Future studies might also include further control experiments such as the bidirectional effect of stimulation, optogenetically manipulating the PL neurons projecting to PAG, and parsing apart the enhanced disinhibition effect from what is CS-related activity and what is the effect of photosilencing in PAG-projecting PL cell activity from inhibiting vHPC-PL neurons, just to name a few. Despite these future directions, the mouse work presented within this thesis has identified a disynaptic circuit not previously investigated within OFL nor within general fear learning. There are an abundance of questions generated from this work that will ultimately have a profound impact on understanding the neural circuitry of vicariously acquired trauma and in turn could meaningfully contribute to therapeutic interventions such as targeted pharmacological and behavioural therapies to help those suffering from anxiety and trauma related disorders originating from observing a traumatic event. I will next consider my research on OFL in humans, in which I investigated how OFL is learned and possible trait characteristics contributing to the spectrum of OFL as a behaviourally translational, complementary investigation into the involved neural network in order to deepen understanding of OFL.

6.2.2 How the current findings contribute to our understanding of OFL in humans

Whilst the mouse research within this thesis provided profound insight into the underlying neural processes of OFL, there are substantial limitations to the type of questions that can be asked and information that can be gleaned when exclusively relying on rodent models. While human research lacks in invasive, causal, interrogation of neurological networks, it offers an opportunity to study OFL in ways that are potentially directly relevant for humans, as well as the ability to relate OFL parameters to individual differences in dispositional traits thought to relate to fear learning. Devising complementary studies across animals and humans provides a powerful way gain a more complete biopsychosocial understanding of highly

conserved behaviours such as OFL. The research to date on OFL in humans has largely focused on the physiological and neurological response to witnessing a physically painful experience inflicted on another (Golkar et al., 2015; Golkar & Olsson, 2016; Lindström et al., 2016; Olsson et al., 2016, 2007; Olsson & Phelps, 2004; Selbing & Olsson, 2019; Szczepanik et al., 2020), but not on how OFL occurs nor the impact of individual differences on this behaviour.

6.2.2.1 Characterizing OFL and the discerning the role of demonstrator shock reaction on individual differences in learning

In order to study variations in OFL in humans, I first established and validated an observational fear task that utilized an instrumental reinforcement learning framework, which had not previously been addressed in the literature. By doing so, participant predictions about expected shock outcomes were modelled on a trial-by-trial basis as a means of characterizing learning. This modification of the foundational OFL task (Haaker et al., 2017; Olsson et al., 2007; Olsson & Phelps, 2004) not only demonstrated that OFL can occur from a video setup, but also the parameters of how that learning happens.

In presenting participants with two probabilistically determined stimuli predictive of a shock outcome to the actor either 80% or 20% of the time, I found that participants were more accurate and responded more quickly in their predictions to the high-probability CS in both a modest-sized community sample and a larger online sample. Additionally, accuracy and median prediction response time did not differ regardless of whether the demonstrator's shock reaction was moderate or exaggerated. This not only demonstrated that participants were able to successfully learn to predict shock outcomes through observation of a demonstrator, but there are performative differences in encountering a mostly negative CS as compared to a mostly neutral CS.

While participants presented differences in OFL performance to the two CSs, their overall learning was best described by a single learning rate and temperature

parameter for both stimuli. Prediction errors in the OFL task (i.e., when the high-probability CS *was not* followed by a shock to the demonstrator, or when the low-probability CS *was* followed by a shock) generated small revisions to the value estimate of either CS. Participant learning seemed to be more strongly affected by a longer history of prediction outcomes as reflected in the low learning rate value, while prediction choice was quite deterministic as evidenced by the low temperature parameter. This characterization was found was for both the high- and medium-shock reaction groups, which may indicate that the differences between shock reactions may not have differed significantly enough from one another to demonstrate differential OFL. Whilst prior studies on OFL have generally used a single expression intensity of the demonstrator to transmit information, Selbing and Olsson (2019) looked at the effect of anxious anticipatory behaviour of a demonstrator on participant OFL. The authors found greater deterministic discrimination between CS+ and CS- when undergoing an OFL task in which the demonstrator displayed anxious anticipatory behaviour to both CSs, rather than only anxiously anticipating the CS+. Whilst this study differs slightly from the work presented in this thesis, the enhanced determinism of discrimination is comparable with my finding that participants who observed the high-shock reaction video had a smaller temperature parameter and were thus more deterministic in their shock outcome predictions. Ultimately, however, the null findings of OFL performance measures and learning between the different shock reaction intensity conditions may have been due to not using an extreme enough difference in shock reaction between groups. Moreover, in order to more effectively look at the role of a demonstrator's pain response on differential OFL parameters might benefit from a within-subject study design. Future studies may want to have a participant undergo OFL from multiple demonstrator shock reaction intensities in order to best interrogate the relationship between a demonstrators pain and how OFL may be disrupted or enhanced accordingly. The role of the demonstrator in conveying fearful information can influence the degree of threat perceived by an observer and their subsequent

learning and memory of an event, but more research is required to interrogate the relationship between demonstrator's behaviour and OFL.

6.2.2.2 Demonstrating dispositional traits association with individual differences of OFL

Building upon the characterization of OFL in Study 1 of Chapter 5, I next investigated the relationship between anxiety and psychopathic traits with individual differences in OFL, as trait differences have been shown to be characterised by atypical patterns of learning and threat processing (Decety et al., 2013; Mkrtchian et al., 2017; Seara-Cardoso et al., 2015). Moreover, individual differences in reactivity and expression of OFL may explain individual differences in functioning and susceptibility to various psychiatric disorders, which has not been previously investigated in the context of OFL prior to this research (Debiec & Olsson, 2017; Olsson et al., 2016; Szczepanik et al., 2020). Overall, only a modest positive association between trait anxiety and median prediction response time was found. An additional post-hoc sensitivity analysis comparing the top and bottom 15% of participants for trait anxiety also revealed a negative correlation with OFL accuracy. There were no other significant associations found between trait anxiety and OFL nor any correlations between psychopathic traits and OFL.

As previous studies have shown, elevated anxiety levels have been connected to hypervigilance of threatening situations (Debiec & Olsson, 2017; Helsen et al., 2011; Ueno et al., 2018), the very modest findings between trait anxiety and OFL indicate that there may be fundamental design problems with the OFL task that need to be addressed for future studies. First, the low learning rate and temperature parameters may reflect that prediction errors had little impact on OFL, or it may indicate that the task itself was too simplistic and there was essentially a ceiling effect of learning. Altering the ratio of predictability of stimuli to be more ambiguous or including a reversal component might better engage learning. Additionally, by using a pre-recorded video on a screen means participants are unnaturally always paying

attention to the task while in the mouse paradigm, as in real life, the degree of learning is influenced by the extent of orienting and processing a demonstrator's fear.

More critically, however, is the question of ecological validity of the current OFL task for humans. Szczepanik and colleagues (2020), in a study conducted after commencement of the current thesis, addressed the ecological disconnect between learning from a stranger onscreen and more naturalistic forms of OFL that involve more realistic representations of experiences. Instead of using a pre-recorded video of a demonstrator receiving electrical shocks, the authors recruited friends to serve as observer-demonstrator pairs to undergo OFL via a live video stream. The authors reported that all observers responded strongly to the US delivery to their demonstrator. Moreover, they showed that whilst the physiological conditioned responses, as measured by SCR and startle response, were different between the learning phase and the direct-expression test phase, this difference depended on participants' declarative knowledge of the CS+/US association. Observers who were aware of the CS+/US contingency showed elevated levels of startle response to the CSs as compared to the neutral inter-trial intervals. Interestingly, however, general learning efficiency of participants correctly ascribing the CS+ to the US was much lower than in previous studies (Golkar & Olsson, 2016). The authors suggested that OFL may have two distinct components: an automatic psychophysiological response to watching the demonstrators' US response as well as a learned response about CS-US contingencies. This is consistent with the PL→l/vlPAG Ca^{2+} imaging findings in Chapter 4, which show a significant and sustained increase of activity in response to the US throughout OFL, whereas activity in response to the CS only becomes significant at the end of conditioning as the CS-US relationship is learnt. Whilst the realism of OFL was enhanced by using friends as conditioning pairs, demonstrators' responses to the shock varied widely, lessening the experimental constraint on demonstrator US reaction. This ambiguity in response clearly affected how well observers were able to learn the association between the CS+ and US. Despite these limitations, this work is the first reported effort to adapt the OFL protocol to be more ecologically relevant and raises important considerations for future designs.

Further developments to enhance the realism of OFL and better engage the relationship between dispositional traits and OFL might be addressed by using a demonstrator or environmental features that have some significance to participants. Additionally, technologies such as virtual reality might be useful to mimic a salient and realistic experience of OFL and should be considered for future research. Whilst the findings from this thesis represent only an initial step of research into individual differences in OFL in humans, they provide important groundwork and suggest ways for altering behavioural paradigms to be more translational across species and to better capture individual differences in fear learning.

6.2.3 Limitations and considerations of OFL as a behaviourally translational research task

OFL provides critical insight into how we learn from social interactions about the environment around us. It is a highly conserved mechanism ensuring safety and survival across species, and studied in both mice and humans within this thesis. Tremendous efforts have been made in less than two decades in developing and investigating OFL in different animal models (Burgos-Robles et al., 2019; Debiec & Olsson, 2017; Olsson et al., 2020; Olsson & Phelps, 2007), however, crucial differences of how OFL is researched in different species should be considered by those interested in conducting meaningful translational research.

The ecological disconnect discussed in the previous section extends beyond considerations of research on OFL for humans or for mice to include how OFL is studied using any animal model. This doctoral research has afforded me a unique perspective of researching a single behavioural task in both mice and humans, encountering a range of advantages and limitations of each. Ultimately, how OFL is modelled in mice is very different from how it is modelled in humans. Szczepanik and colleagues (2020) attempt to amend the human OFL task to be more reflective of the naturalistic and salient components of the mouse assay is a critical effort in advancing our understanding of observed trauma. The significance of behaviourally translational research is that it expands our understanding of a subject by capitalizing

on the strengths and capacity that are unique from one animal model to another in order to reveal a deeper understanding of the biopsychosocial components of our experiences.

With respect to the translatability of OFL between mice and humans, ecological alterations to the human task so OFL can be studied in a more natural way are crucial for the advancement of this learning mechanism. The fear component and sense of threat that is very real in the mouse assay is lacking in the human task. A pre-recorded video on a screen carries little threat of danger to participants. Moreover, how mice attend to the experience of OFL differs for the human setup. Human participants are unnaturally over-attentive to the task on the screen while the mouse paradigm, as in real life, the degree of learning is influenced by the extent of orienting and processing a demonstrator's fear. In contrast, however, there are important ethical and psychological constraints of having participants undergo OFL to the degree which is possible with mice. Additionally, there are physical restrictions on how and where OFL can be measured in humans that would sacrifice the ability to acquire neurophysiological data that is required for studying the neuroscience of OFL. Future experiments would benefit from considering and understanding how OFL is studied across species, but more importantly, how to ask questions and devise experiments that are collaborative across researchers using all species.

6.3 Conclusion

This thesis was conducted with the broad aim of creating a behaviourally translational OFL paradigm for mice and humans in order to understand the mechanisms subserving this highly conserved form of learning. I demonstrated that both mice and humans learn fearful associations through observation in laboratory settings. Limbic projections from the vHPC and BLA to the PL are comparably engaged regardless of where fear learning is directly experienced or obtained through observation. Similarly, the PL, a region identified as being critical for sustaining directly learned fears is also necessary for acquiring OFL. Fear learning is subsequently modulated by vHPC inputs to PL, constraining OFL, possibly in part

through inhibitory PV INs. Enhanced Ca^{2+} activity was detected in l/vlPAG-projecting PL neurons in response to OFL, which supports the idea that the PL→l/vlPAG pathway is highly sensitive to aversive information. Together, inhibitory control of PL outputs from the vHPC calibrates l/vlPAG-projecting PL neurons response to OFL. A vHPC→PL→l/vlPAG disynaptic pathway had not been identified prior to this research as being necessary for acquiring fear learning.

Additionally, I demonstrated that OFL occurs in humans similarly as to how it occurs in mice. Prior to this research, characterization of how OFL is acquired had not been defined. I found that participants were better and quicker at predicting shock outcomes to the high-probability CS. Moreover, OFL did not seem heavily influenced by prediction errors and CS valuation was updated relatively slowly. In exploring the relationship between dispositional traits and OFL, I found a modest correlation between trait anxiety and OFL prediction response time; however, no other significant associations were found between anxiety nor psychopathic traits and measures of OFL. Ultimately, the OFL task devised for human participants differs substantially in its ecological validity from the mouse assay, which should be addressed in future studies of OFL.

In a world that is increasingly social, thanks in part to 24 hour news streams and social media networks, we are increasingly barraged by images and videos of social distress and trauma. Nothing is a more appropriate global example of OFL than the COVID-19 pandemic. As we turn a corner on more than a year of physical isolation, most of our memories are still rich with the pictures and stories depicting heaps of body bags and hospitals full beyond capacity, COVID-positive patients dying alone and medical staff with their faces completely obscured by masks, face-shields, goggles. We have observed our world be entirely upended with nearly 4 million people dead and close to 200 million infected. But as vaccines become more prevalent and we begin to gather again, the extent of our trauma is on full display. The research conducted within this thesis just barely begins to scratch the surface of the mechanisms of OFL, defining both a novel neural network and characterizing learning and the potential role of dispositional traits on individual differences of OFL

across animal models. Continuing this research with a multi-species, translational approach may provide critical information about helpful versus harmful social information. Individual differences in learning and memory might one day serve as risk factors for individuals at increased risk for developing certain psychopathologies both because of and in response to OFL. Future research should focus on progressing the OFL task in a more comparable and ecologically relevant way so as to have the greatest translational and clinical impact.

References

Abbas, A. I., Sundiang, M. J. M., Henoch, B., Morton, M. P., Bolkan, S. S., Park, A. J., Harris, A. Z., Kellendonk, C., & Gordon, J. A. (2018). Somatostatin interneurons facilitate hippocampal-prefrontal synchrony and prefrontal spatial encoding. *Neuron*, *100*(4), 926-939.

Adler, J. M., & Clark, L. A. (2019). Incorporating narrative identity into structural approaches to personality and psychopathology. *Journal of Research in Personality*, *82*, 103857.

Adolphs, R. (2013). The biology of fear. *Current Biology*, *23*(2), R79–R93.

Ahn, W. Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 95–110.

Ährlund-Richter, S., Xuan, Y., van Lunteren, J. A., Kim, H., Ortiz, C., Dorocic, I. P., Meletis, K., Carlén, M. (2019). A whole-brain atlas of monosynaptic input targeting four different cell types in the medial prefrontal cortex of the mouse. *Nature Neuroscience*, *22*(4), 657-668.

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. *Advances in Experimental Social Psychology*, *20*(C), 1–63.

Allsop, S. A., Wichmann, R., Mills, F., Burgos-Robles, A., Chang, C.-J., Felix-Ortiz, A. C., Vienne, A., Beyeler, A., Izadmehr, E. M., Glober, G., Cum, M. I., Stergiadou, J., Anandalingam, K. K., Farris, K., Leppla, C. A., Weddington, J. C., Nieh, E. H., Smith, A. C., … Tye, K. M. (2018). Corticoamygdala transfer of socially derived information gates observational learning. *Cell, 173*, 1-14.

Amemori, K. I., & Graybiel, A. M. (2012). Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nature Neuroscience*, *15*(5), 776–785.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders,* (5th ed.).

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–277.

Amorapanth, P., Nader, K., & Ledoux, J. E. (1999). Lesions of periaqueductal gray dissociate-conditioned freezing from conditioned suppression behavior in rats. *Learning and Memory*, *6*(5), 491–499.

Antoniadis, E. A., & McDonald, R. J. (2006). Fornix, medial prefrontal cortex, nucleus accumbens, and mediodorsal thalamic nucleus: Roles in a fear-based context discrimination task. *Neurobiology of Learning and Memory*, *85*(1), 71–85.

Aoued, H. S., Sannigrahi, S., Hunter, S. C., Doshi, N., Sathi, Z. S., Chan, A. W. S., Walum, H., & Dias, B. G. (2020). Proximate causes and consequences of intergenerational influences of salient sensory experience. *Genes, Brain and Behavior*, *19*(4), e12638.

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

Aravanis, A. M., Wang, L. P., Zhang, F., Meltzer, L. A., Mogri, M. Z., Schneider, M. B., & Deisseroth, K. (2007). An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *Journal of Neural Engineering*, *4*(3), S143.

Arditte, K. A., Çek, D., Shaw, A. M., & Timpano, K. R. (2016). The importance of assessing clinical phenomena in Mechanical Turk research. *Psychological Assessment*, *28*(6), 684–691.

Askew, C., & Field, A. P. (2007). Vicarious learning and the development of fears in childhood. *Behaviour Research and Therapy*, *45*(11), 2616–2627.

Atsak, P., Orre, M., Bakker, P., Cerliani, L., & Roozendaal, B. (2011). Experience modulates vicarious freezing in rats: A model for empathy. *PLoS ONE*, *6*(7), 21855.

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.

Bach, D. R., & Dayan, P. (2017). Algorithms for survival: A comparative perspective on emotions. *Nature Reviews Neuroscience, 18*, 311–319.

Bartal, I. B. A., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430.

Bartal, I. B. A., Rodgers, D. A., Bernardez Sarria, M. S., Decety, J., & Mason, P. (2014). Pro-social behavior in rats is modulated by social experience. *ELife*, *3*, e01385.

Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLoS ONE*, *10*(4), e0121595.

Basten, U., Stelzel, C., & Fiebach, C. J. (2011). Trait anxiety modulates the neural efficiency of inhibitory control. *Journal of Cognitive Neuroscience*, *23*(10), 3132–3145.

Beas, B. S., Wright, B. J., Skirzewski, M., Leng, Y., Hyun, J. H., Koita, O., Ringelberg, N., Kwon, H. -B., Buonanno, A., & Penzo, M. A. (2018). The locus coeruleus drives disinhibition in the midline thalamus via a dopaminergic mechanism. *Nature Neuroscience*, *21*(7), 963–973.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Berenbaum, H., Thompson, R. J., & Bredemeier, K. (2007). Perceived threat: Exploring its association with worry and its hypothesized antecedents. *Behaviour Research and Therapy*, *45*(10), 2473–2482.

Berggren, N., & Derakshan, N. (2014). Inhibitory deficits in trait anxiety: Increased stimulus-based or response-based interference? *Psychonomic Bulletin and Review*, *21*(5), 1339–1345.

Bergstrom, H.C, Lipkin, A. M., Lieberman, A. G., Pinard, C. R., Gunduz-Cinar, O., Brockway, E. T., … Holmes, A. (2018). Dorsolateral striatum engagement interferes with early discrimination learning. *Cell Reports, 23*(8), 2264-2272.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35*, 1–23.

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews, 47*, 520–532.

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

Blair, K. S., Otero, M., Teng, C., Geraci, M., Lewis, E., Hollon, N., Blair, R. J. R., Ernst, M., Grillon, C., & Pine, D. S. (2016). Learning from other people's fear: Amygdala-based social reference learning in social anxiety disorder. *Psychological Medicine*, *46*(14), 2943–2953.

Blanchard, D. C., & Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. *Journal of Comparative and Physiological Psychology*, *81*(2), 281–290.

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders, 113*, 1–20.

Bruchey, A. K., Jones, C. E., & Monfils, M. H. (2010). Fear conditioning by-proxy: Social transmission of fear during memory retrieval. *Behavioural Brain Research*, *214*(1), 80–84.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2015). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 133–139). American Psychological Association.

Bukalo, O., Pinard, C. R., Silverstein, S., Brehm, C., Hartley, N. D., Whittle, N., Colacicco, G., Busch, E., Patel., S., Singewald, N., & Holmes, A. (2015). Prefrontal inputs to the amygdala instruct fear extinction memory formation. *Science Advances*, *1*(6), e1500251.

Burgos-Robles, A., Gothard, K. M., Monfils, M. H., Morozov, A., & Vicentic, A. (2019). Conserved features of anterior cingulate networks support observational learning across species. *Neuroscience and Biobehavioral Reviews*, *107*, 215–228.

Burgos-Robles, A., Kimchi, E. Y., Izadmehr, E. M., Porzenheim, M. J., Ramos-Guasp, W. A., Nieh, E. H., … Tye, K. M. (2017). Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nature Neuroscience, 20*(6), 824-835.

Burgos-Robles, A., Vidal-Gonzalez, I., & Quirk, G. J. (2009). Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *Journal of Neuroscience*, *29*(26), 8474–8482.

Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., & Keysers, C. (2019). Emotional mirror neurons in the rat's anterior cingulate cortex. *Current Biology*, *29*(8), 1301-1312.

Chen, P., & Hong, W. (2018). Neural circuit mechanisms of social behavior. *Neuron*, *98*(1), 16–30.

Chen, Q., Panksepp, J. B., & Lahvis, G. P. (2009). Empathy is moderated by genetic background in mice. *PLoS ONE*, *4*(2), e4387.

Chen, T. W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., … Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473.

Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., … Boyden, E. S. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature*, *463*(7277), 98–102.

Church, R. M. (1959). Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*, *52*(2), 132–134.

Coles, M. E., Turk, C. L., & Heimberg, R. G. (2007). Memory bias for threat in generalized anxiety disorder: The potential importance of stimulus relevance. *Cognitive Behaviour Therapy*, *36*(2), 65–73.

Cook, M., & Mineka, S. (1990). Selective associations in the observational conditioning of fear in rhesus monkeys. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*(4), 372–389.

Corcoran, K. A., & Quirk, G. J. (2007). Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *Journal of Neuroscience, 27*(4), 840-844.

Courtin, J., Chaudun, F., Rozeske, R. R., Karalis, N., Gonzalez-Campo, C., Wurtz, H., … Herry, C. (2014). Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature*, *505*(7481), 92–96.

Cummings, K. A., & Clem, R. L. (2020). Prefrontal somatostatin interneurons encode fear memory. *Nature Neuroscience*, *23*(1), 61–74.

Dana, H., Sun, Y., Mohar, B., Hulse, B. K., Kerlin, A. M., Hasseman, J. P., … Kim, D. S. (2019). High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature Methods*, *16*(7), 649–657.

Daniel-Watanabe, L., McLaughlin, M., Gormley, S., & Robinson, O. J. (2020). Association between a directly translated cognitive measure of negative bias and self-reported psychiatric symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2020.02.010

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning: Attention and Performance XXIII*. Oxford University Press.

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, *36*, 2288–2304.

De Oca, B. M., DeCola, J. P., Maren, S., & Fanselow, M. S. (1998). Distinct regions of the periaqueductal gray are involved in the acquisition and expression of defensive responses. *Journal of Neuroscience*, *18*(9), 3426–3432.

de Voogd, L. D., Murray, Y. P. J., Barte, R. M., van der Heide, A., Fernández, G., Doeller, C. F., & Hermans, E. J. (2020). The role of hippocampal spatial representations in contextualization and generalization of fear: Fear memory contextualization. *NeuroImage*, *206*.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*(1), 279–300.

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509.

Debiec, J., & Olsson, A. (2017). Social Fear Learning: from Animal Models to Human Function. *Trends in Cognitive Sciences*, *21*(7), 546–555.

Decety, J., Skelly, L. R., & Kiehl, K. A. (2013). Brain response to empathy-eliciting scenarios involving pain in incarcerated individuals with psychopathy. *JAMA Psychiatry*, *70*(6), 638–645.

Deisseroth, K. (2011). Optogenetics. *Nature Methods, 8*, 26–29.

Delgado, M. R., Phelps, E. A., & Robbins, T. W. (2011). Decision Making, Affect, and Learning: Attention and Performance XXIII. In *Decision Making, Affect, and Learning: Attention and Performance XXIII*. Oxford University Press.

Deng, X., Gu, L., Sui, N., Guo, J., & Liang, J. (2019). Parvalbumin interneuron in the ventral hippocampus functions as a discriminator in social memory. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(33), 16583–16592.

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine*, *34*(6), 1093–1102.

Dolan, M., & Fullam, R. (2006). Face affect recognition deficits in personality-disordered offenders: Association with psychopathy. *Psychological Medicine*, *36*(11), 1563–1569.

Eysenck, M. W., & Van Berkum, J. (1992). Trait anxiety, defensiveness, and the structure of worry. *Personality and Individual Differences*, *13*(12), 1285–1290.

Fanselow, M. S. (1980). Conditional and unconditional components of post-shock freezing. *The Pavlovian Journal of Biological Science : Official Journal of the Pavlovian*, *15*(4), 177–182.

Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, *1*(4), 429–438.

Fanselow, M. S., & Dong, H. W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron,* 65, 7–19.

Faturi, C. B., Rangel, M. J., Baldo, M. V. C., & Canteras, N. S. (2014). Functional mapping of the circuits involved in the expression of contextual fear responses in socially defeated animals. *Brain Structure and Function*, *219*(3), 931–946.

Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The development and application of optogenetics. *Annual Review of Neuroscience*, *34*(1), 389–412.

Feyder, M., Karlsson, R.-M., Mathur, P., Lyman, M., Bock, R., Momenan, R., … Holmes, A. (2010). Association of mouse *Dlg4* (PSD-95) gene deletion and human *DLG4* gene variation with phenotypes relevant to autism spectrum disorders and Williams' Syndrome. *American Journal of Psychiatry*, *167*(12), 1508–1517.

Fitzgerald, P. J., Seemann, J. R., & Maren, S. (2014). Can fear extinction be enhanced? A review of pharmacological and behavioral findings. *Brain Research Bulletin, 105*, 46–60.

Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology, 8*, 196–202.

Foulkes, L., McCrory, E. J., Neumann, C. S., & Viding, E. (2014). Inverted social reward: Associations between psychopathic traits and self-report and experimental measures of social reward. *PLoS ONE*, *9*(8), e106000.

Franklin, K. B. J., & Paxinos, G. (2008). *The mouse brain in stereotaxic coordinates* (3rd ed.). Elsevier Academic Press.

Franklin, T. B. (2019). Recent advancements surrounding the role of the periaqueductal gray in predators and prey. *Frontiers in Behavioral Neuroscience*, *13*, 60.

Franklin, T. B., Silva, B. A., Perova, Z., Marrone, L., Masferrer, M. E., Zhan, Y., … Gross, C. T. (2017). Prefrontal cortical control of a brainstem social behavior circuit. *Nature Neuroscience*, *20*(2), 260–270.

Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology, 21*, 1111–1131.

Furlong, T. M., Cole, S., Hamlin, A. S., & McNally, G. P. (2010). The role of prefrontal cortex in predictive fear learning. *Behavioral Neuroscience*, *124*(5), 574–586.

Gariépy, J. F., Watson, K. K., Du, E., Xie, D. L., Erb, J., Amasino, D., & Platt, M. L. (2014). Social learning in humans and other animals. *Frontiers in Neuroscience, 8,* 58.

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biological Psychology*, *92*(2), 342–352.

Gerull, F. C., & Rapee, R. M. (2002). Mother knows best: Effects of maternal modelling on the acquisition of fear and avoidance behaviour in toddlers. *Behaviour Research and Therapy*, *40*(3), 279–287.

Geuther, B. Q., Deats, S. P., Fox, K. J., Murray, S. A., Braun, R. E., White, J. K., … Kumar, V. (2019). Robust mouse tracking in complex environments using neural networks. *Communications Biology*, *2*(1), 1–11.

Giustino, T. F., Fitzgerald, P. J., & Maren, S. (2016). Fear expression suppresses medial prefrontal cortical firing in rats. *PLoS ONE*, *11*(10).

Giustino, T. F., & Maren, S. (2015). The role of the medial prefrontal cortex in the conditioning and extinction of fear. *Frontiers in Behavioral Neuroscience, 9*.

Giustino, T. F., & Maren, S. (2018). Noradrenergic modulation of fear conditioning and extinction. *Frontiers in Behavioral Neuroscience, 12*.

Goldstein, M. L. (1979). The effect of UCS intensity on the long-term retention of a classically conditioned fear response. *Bulletin of the Psychonomic Society*, *13*(6), 357–358.

Golkar, A., Castro, V., & Olsson, A. (2015). Social learning of fear and safety is determined by the demonstrator's racial group. *Biology Letters*, *11*(1), 20140817.

Golkar, A., Haaker, J., Selbing, I., & Olsson, A. (2016). Neural signals of vicarious extinction learning. *Social Cognitive and Affective Neuroscience*, *11*(10), 1541–1549.

Golkar, A., & Olsson, A. (2016). Immunization against social fear learning. *Journal of Experimental Psychology: General*, *145*(6), 665–671.

Golkar, A., & Olsson, A. (2017). The interplay of social group biases in social threat learning. *Scientific Reports*, *7*(1).

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & 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.

Grafton, B., & MacLeod, C. (2014). Enhanced probing of attentional bias: The independence of anxiety-linked selectivity in attentional engagement with and disengagement from negative information. *Cognition and Emotion*, *28*(7), 1287–1302.

Grahl, A., Onat, S., & Büchel, C. (2018). The periaqueductal gray and Bayesian integration in placebo analgesia. *ELife*, *7*.

Gruene, T. M., Flick, K., Stefano, A., Shea, S. D., & Shansky, R. M. (2015). Sexually divergent expression of active and passive conditioned fear responses in rats. *ELife*, *4*.

Gunaydin, L. A., Grosenick, L., Finkelstein, J. C., Kauvar, I. V., Fenno, L. E., Adhikari, A., … Deisseroth, K. (2014). Natural neural projection dynamics underlying social behavior. *Cell*, *157*(7), 1535–1551.

Haaker, J., Golkar, A., Selbing, I., & Olsson, A. (2017). Assessment of social transmission of threats in humans using observational fear conditioning. *Nature Protocols*, *12*(7), 1378–1386.

Haddad, A. D. M., Pritchett, D., Lissek, S., & Lau, J. Y. F. (2012). Trait anxiety and fear responses to safety cues: Stimulus generalization or sensitization? *Journal of Psychopathology and Behavioral Assessment*, *34*(3), 323–331.

Hallock, H. L., Quillian, H. M., Maynard, K. R., Mai, Y., Chen, H.-Y., Hamersky, G. R., … Martinowich, K. (2019). Molecularly-Defined Hippocampal Inputs Regulate Population Dynamics in the Prelimbic Cortex to Suppress Context Fear Memory Recall. *BioRxiv*, 802967.

Han, X., Chow, B. Y., Zhou, H., Klapoetke, N. C., Chuong, A., Rajimehr, R., … Boyden, E. S. (2011). A high-light sensitivity optical neural silencer: Development and application to optogenetic control of non-human primate cortex. *Frontiers in Systems Neuroscience*, *5*.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto, ON: Multi-Health Syst.

Hare, R. D. (1965). Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal Psychology*, *70*(6), 442–445.

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology, 4*, 217–246.

Hare, R. D., & Neumann, C. S. (2009). Psychopathy and its measurement. In *The Cambridge Handbook of Personality Psychology* (pp. 660–686). Cambridge University Press.

Harnett, N. G., Ference, E. W., Wood, K. H., Wheelock, M. D., Knight, A. J., & Knight, D. C. (2018). Trauma exposure acutely alters neural function during Pavlovian fear conditioning. *Cortex*, *109*, 1–13.

Helsen, K., Goubert, L., Peters, M. L., & Vlaeyen, J. W. S. (2011). Observational learning and pain-related fear: An experimental study with colored cold pressor tasks. *The Journal of Pain*, *12*(12), 1230–1239.

Helsen, K., Vlaeyen, J. W. S., & Goubert, L. (2015). Indirect acquisition of pain-related fear: an experimental study of observational learning using coloured cold metal bars. *PloS One*, *10*(3), e0117236.

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., & Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature*, *454*(7204), 600–606.

Hill, M. R., Boorman, E. D., & Fried, I. (2016). Observational learning computations in neurons of the human anterior cingulate cortex. *Nature Communications*, *7*(1), 1–12.

Holmes, A., & Rodgers, R. J. (1998). Responses of Swiss-Webster mice to repeated plus-maze experience: Further evidence for a qualitative shift in emotional state? *Pharmacology Biochemistry and Behavior*, *60*(2), 473–488.

Holmes, A., Nielsen, M. K., Tipper, S., & Green, S. (2009). An electrophysiological investigation into the automaticity of emotional face processing in high versus low trait anxious individuals. *Cognitive, Affective and Behavioral Neuroscience*, *9*(3), 323–334.

Holmes, A., & Rodgers, R. J. (2003). Prior exposure to the elevated plus-maze sensitizes mice to the acute behavioral effects of fluoxetine and phenelzine. *European Journal of Pharmacology*, *459*(2–3), 221–230.

Holmes, A., & Singewald, N. (2013). Individual differences in recovery from traumatic fear. *Trends in Neurosciences, 36*, 23–31.

Hong, R. Y. (2013). From dispositional traits to psychopathological symptoms: Social-cognitive vulnerabilities as intervening mechanisms. *Journal of Psychopathology and Behavioral Assessment, 35*(4), 407–420.

Hong, R. Y., & Paunonen, S. V. (2011). Personality vulnerabilities to psychopathology: relations between trait structure and affective-cognitive processes. *Journal of Personality*, *79*(3), 527–562.

Hoover, W. B., & Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure and Function*, *212*(2), 149–179.

Hopwood, T. L., & Schutte, N. S. (2017). Psychological outcomes in reaction to media exposure to disasters and large-scale violence: A meta-analysis. *Psychology of Violence*, *7*(2), 316–327.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research and Politics*, *2*(3).

Indovina, I., Robbins, T. W., Núñez-Elizalde, A. O., Dunn, B. D., & Bishop, S. J. (2011). Fear-conditioning mechanisms associated with trait vulnerability to anxiety in humans. *Neuron*, *69*(3), 563–571.

Ishikawa, A., & Nakamura, S. (2003). Convergence and interaction of hippocampal and amygdalar projections within the prefrontal cortex in the rat. *Journal of Neuroscience*, *23*(31), 9987–9995.

Ito, W., Erisir, A., & Morozov, A. (2015). Observation of distressed conspecific as a model of emotional trauma generates silent synapses in the prefrontal-amygdala pathway and enhances fear learning, but Ketamine abolishes those effects. *Neuropsychopharmacology*, *40100*(10), 2536–2545.

Ito, W., & Morozov, A. (2019). Prefrontal-amygdala plasticity enabled by observational fear. *Neuropsychopharmacology*, *44*(10), 1778–1787.

Jackson, J., Karnani, M. M., Zemelman, B. V., Burdakov, D., & Lee, A. K. (2018). Inhibitory control of prefrontal cortex by the claustrum. *Neuron*, *99*(5), 1029-1039.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., … Shin, H. S. (2010). Observational fear learning involves affective pain system and Ca v 1.2 Ca 2+ channels in ACC. *Nature Neuroscience*, *13*(4), 482–488.

Jeon, D., & Shin, H. S. (2011). A mouse model for observational fear learning and the empathetic response. *Current Protocols in Neuroscience*, (SUPPL.57).

Jimenez, J. C., Su, K., Goldberg, A. R., Luna, V. M., Biane, J. S., Ordek, G., … Kheirbek, M. A. (2018). Anxiety cells in a hippocampal-hypothalamic circuit. *Neuron*, *97*(3), 670-683.

Jimenez, S. A., & Maren, S. (2009). Nuclear disconnection within the amygdala reveals a direct pathway to fear. *Learning & Memory*, *16*(12), 766–768.

Jin, J., & Maren, S. (2015). Fear renewal preferentially activates ventral hippocampal neurons projecting to both amygdala and prefrontal cortex in rats. *Scientific Reports, 5*.

Jones, C. E., Agee, L., & Monfils, M.-H. (2018). Fear conditioning by proxy: Social transmission of fear between interacting conspecifics. *Current Protocols in Neuroscience*, *83*(1), e43.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience, 16*, 521–534.

Jovanovic, T., Kazama, A., Bachevalier, J., & Davis, M. (2012). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology, 62*(2), 695-704.

Karakilic, A., Kizildag, S., Kandis, S., Guvendi, G., Koc, B., Camsari, G. B., … Uysal, N. (2018). The effects of acute foot shock stress on empathy levels in rats. *Behavioural Brain Research*, *349*, 31–36.

Karalis, N., Dejean, C., Chaudun, F., Khoder, S., R Rozeske, R., Wurtz, H., … Herry, C. (2016). 4-Hz oscillations synchronize prefrontal-amygdala circuits during fear behavior. *Nature Neuroscience*, *19*(4), 605–612.

Karlsson, R.-M., Tanaka, K., Heilig, M., & Holmes, A. (2008). Loss of glial glutamate and aspartate transporter (Excitatory Amino Acid Transporter 1) causes locomotor hyperactivity and exaggerated responses to psychotomimetics: Rescue by haloperidol and metabotropic Glutamate 2/3 agonist. *Biological Psychiatry*, *64*(9), 810–814.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kavaliers, M., Colwell, D., & Choleris, E. (2003). Learning to fear and cope with a natural stressor: Individually and socially acquired corticosterone and avoidance responses to biting flies. *Hormones and Behavior*, *43*(1), 99–107.

Kennedy, B. L., Schwab, J. J., Morris, R. L., & Beldia, G. (2001). Assessment of state and trait anxiety in subjects with anxiety and depressive disorders. *Psychiatric Quarterly, 72*(3), 263-276.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, *8*(4), 614–629.

Keum, S., Park, J., Kim, A., Park, J., Kim, K. K., Jeong, J., & Shin, H.-S. (2016). Variability in empathic fear response among 11 inbred strains of mice. *Genes, Brain and Behavior*, *15*(2), 231–242.

Keum, S., Kim, A., Shin, J. J., Kim, J.-H., Park, J., & Shin, H.-S. (2018). A missense variant at the Nrxn3 locus enhances empathy fear in the mouse. *Neuron*, *98*(3), 588-601.

Keum, S., & Shin, H.-S. (2016). Rodent models for studying empathy. *Neurobiology of Learning and Memory*, *135*, 22–26.

Keum, S., & Shin, H.-S. (2019). Genetic factors associated with empathy in humans and mice. *Neuropharmacology*, *159*, 107514.

Kikusui, T., Winslow, J. T., & Mori, Y. (2006). Social buffering: Relief from stress and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*, 2215–2228.

Kim, A., Keum, S., & Shin, H.-S. (2019). Observational fear behavior in rodents as a model for empathy. *Genes, Brain and Behavior*, *18*(1), e12521.

Kim, B. S., Lee, J., Bang, M., Seo, B. A., Khalid, A., Jung, M. W., & Jeon, D. (2014). Differential regulation of observational fear and neural oscillations by serotonin and dopamine in the mouse anterior cingulate cortex. *Psychopharmacology*, *231*(22), 4371–4381.

Kim, E. J., Kim, E. S., Covey, E., & Kim, J. J. (2010). Social Transmission of Fear in Rats: The Role of 22-kHz Ultrasonic Distress Vocalization. *PLoS ONE*, *5*(12), e15077.

Kim, S., Mátyás, F., Lee, S., Acsády, L., & Shin, H.-S. (2012). Lateralization of observational fear learning at the cortical but not thalamic level in mice. *Proceedings of the National Academy of Sciences*, *109*(38), 15497–15501.

Kindt, M., & Soeter, M. (2014). Fear inhibition in high trait anxiety. *PLoS ONE*, *9*(1), 86462.

Kirmayer, L. (2001). Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of Clinical Psychiatry*, *62*, 22–28.

Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., … Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science*, *356*(6333), 73.

Klavir, O., Genud-Gabai, R., & Paz, R. (2013). Functional connectivity between amygdala and cingulate cortex for adaptive aversive learning. *Neuron*, *80*(5), 1290–1300.

Klavir, O., Prigge, M., Sarel, A., Paz, R., & Yizhar, O. (2017). Manipulating fear associations via optogenetic modulation of amygdala inputs to prefrontal cortex. *Nature Neuroscience, 20*(6), 836-844.

Kleberg, J. L., Selbing, I., Lundqvist, D., Hofvander, B., & Olsson, A. (2015). Spontaneous eye movements and trait empathy predict vicarious learning of fear. *International Journal of Psychophysiology*, *98*(3), 577–583.

Knapska, E., Mikosz, M., Werka, T., & Maren, S. (2010). Social modulation of learning in rats. *Learning & Memory*, *17*(1), 824–831.

Knapska, E., Nikolaev, E., Boguszewski, P., Walasek, G., Blaszczyk, J., Kaczmarek, L., & Werka, T. (2006). Between-subject transfer of emotional information evokes specific pattern of amygdala activation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3858–3862.

Kondrakiewicz, K., Rokosz-Andraka, K., Nikolaev, T., Górkiewicz, T., Danielewski, K., Gruszczyńska, A., … Knapska, E. (2019). Social transfer of fear in rodents. *Current Protocols in Neuroscience*, *90*(1).

Kosson, D. S., Suchy, Y., Mayer, A. R., & Libby, J. (2002). Facial affect recognition in criminal psychopaths. *Emotion*, *2*(4), 398–411.

Krishnan, V., Han, M.-H., Graham, D. L., Berton, O., Renthal, W., Russo, S. J., … Nestler, E. J. (2007). Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell*, *131*(2), 391–404.

Krueger, R. F., Caspi, A., Moffitt, T. E., Silva, P. A., & McGee, R. (1996). Personality traits are differentially linked to mental disorders: A multitrait-multidiagnosis study of an adolescent birth cohort. *Journal of Abnormal Psychology*, *105*(3), 299–312.

Laubach, M., Amarante, L. M., Swanson, K., & White, S. R. (2018). What, if anything, is rodent prefrontal cortex? *ENeuro*, *5*(5).

LaFreniere, L. S., & Newman, M. G. (2019). Probabilistic learning by positive and negative reinforcement in generalized anxiety disorder. *Clinical Psychological Science*, *7*(3), 502–515.

Langford, D. J., Crager, S. E., Shehzad, Z., Smith, S. B., Sotocinal, S. G., Levenstadt, J. S., … Mogil, J. S. (2006). Social modulation of pain as evidence for empathy in mice. *Science (New York, N.Y.)*, *312*(5782), 1967–1970.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*(1), 155–184.

LeDoux, J. E., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus: Sensory interface of the amygdala in fear conditioning. *Journal of Neuroscience*, *10*(4), 1062–1069.

Levy, D. R., Tamir, T., Kaufman, M., Weissbrod, A., Schneidman, E., & Yizhar, O. (2018). Dynamics of social representation in the mouse prefrontal cortex. *BioRxiv*, 321182.

Li, Z., Lu, Y.-F., Li, C.-L., Wang, Y., Sun, W., He, T., … Chen, J. (2014). Social interaction with a cagemate in pain facilitates subsequent spinal nociception via activation of the medial prefrontal cortex in rats. *PAIN*, *155*(7), 1253–1261.

Likhtik, E., & Paz, R. (2015). Amygdala-prefrontal interactions in (mal)adaptive learning. *Trends in Neurosciences, 38*, 158–166.

Likhtik, E., Stujenske, J. M., Topiwala, M. A., Harris, A. Z., & Gordon, J. A. (2014). Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. *Nature Neuroscience*, *17*(1), 106–113.

Lindström, B., Golkar, A., Jangard, S., Tobler, P. N., & Olsson, A. (2019). Social threat learning transfers to decision making in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4732–4737.

Lindström, B., Haaker, J., & Olsson, A. (2018). A common neural network differentially mediates direct and social fear learning. *NeuroImage*, *167*, 121–129.

Lindström, B., Selbing, I., & Olsson, A. (2016). Co-evolution of social learning and evolutionary preparedness in dangerous environments. *PLOS ONE*, *11*(8), e0160245.

Liu, L., Ito, W., & Morozov, A. (2017). GABAb receptor mediates opposing adaptations of GABA release from two types of prefrontal interneurons after observational Fear. *Neuropsychopharmacology*, *42*(6), 1272–1283.

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

Livneh, U., & Paz, R. (2012). Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. *Neuron*, *75*(1), 133–142.

Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E., & Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(35), 9763–9768.

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: High levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Frontiers in Human Neuroscience*, *1112*, 36–1.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247–285.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience and Biobehavioral Reviews, 80*, 703–728.

López, R., Poy, R., Patrick, C. J., & Moltó, J. (2013). Deficient fear conditioning and self-reported psychopathy: The role of fearless dominance. *Psychophysiology*, *50*(2), 210–218.

Lu, Y. F., Ren, B., Ling, B. F., Zhang, J., Xu, C., & Li, Z. (2018). Social interaction with a cagemate in pain increases allogrooming and induces pain hypersensitivity in the observer rats. *Neuroscience Letters*, *662*, 385–388.

Luo, W. J., Li, C. L., Geng, K. W., Wang, X. L., Du, R., Yu, Y., … Chen, J. (2020). The similar past pain experience evokes both observational contagious pain and consolation in stranger rat observers. *Neuroscience Letters*, *722*, 134840.

Mahn, M., Prigge, M., Ron, S., Levy, R., & Yizhar, O. (2016). Biophysical constraints of optogenetic inhibition at presynaptic terminals. *Nature Neuroscience*, *19*(4), 554–556.

Marek, R., Jin, J., Goode, T. D., Giustino, T. F., Wang, Q., Acca, G. M., … Sah, P. (2018). Hippocampus-driven feed-forward inhibition of the prefrontal cortex mediates relapse of extinguished fear. *Nature Neuroscience*, *21*(3), 384–392.

Marek, R., Strobel, C., Bredy, T. W., & Sah, P. (2013). The amygdala and medial prefrontal cortex: partners in the fear circuit. *The Journal of Physiology*, *591*(10), 2381–2391.

Marek, R., Sun, Y., & Sah, P. (2019). Neural circuits for a top-down control of fear and extinction. *Psychopharmacology, 236*, 313–320.

Marek, R., Xu, L., Sullivan, R. K. P., & Sah, P. (2018). Excitatory connections between the prelimbic and infralimbic medial prefrontal cortex show a role for the prelimbic cortex in fear extinction. *Nature Neuroscience*, *21*(5), 654–658.

Maren, S. (2001). Neurobiology of pavlovian fear conditioning. *Annual Review of Neuroscience*, *24*(1), 897–931.

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience, 14*, 417–428.

Marquis, J.-P., Killcross, S., & Haddon, J. E. (2007). Inactivation of the prelimbic, but not infralimbic, prefrontal cortex impairs the contextual control of response conflict in rats. *European Journal of Neuroscience*, *25*(2), 559–566.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 32*, 454–465.

Marsh, A. A., Finger, E. C., Schechter, J. C., Jurkowitz, I. T. N., Reid, M. E., & Blair, R. J. R. (2011). Adolescents with psychopathic traits report reductions in physiological responses to fear. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(8), 834–841.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289.

McCredie, M. N., & Morey, L. C. (2019). Who are the Turkers? A characterization of MTurk workers using the Personality Assessment Inventory. *Assessment*, *26*(5), 759–766.

McDonald, A. J. (1991). Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience*, *44*(1), 1–14.

McGinty, V. B., & Grace, A. A. (2008). Selective activation of medial prefrontal-to-accumbens projection neurons by amygdala stimulation and pavlovian conditioned stimuli. *Cerebral Cortex*, *18*(8), 1961–1972.

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences, 34*, 283–292.

Meyer-Mueller, C., Jacob, P. Y., Montenay, J. Y., Poitreau, J., Poucet, B., & Chaillan, F. A. (2020). Dorsal, but not ventral, hippocampal inactivation alters deliberation in rats. *Behavioural Brain Research*, *390*, 112622.

Mikosz, M., Nowak, A., Werka, T., & Knapska, E. (2016). Sex differences in social modulation of learning in rats. *Scientific Reports*, *5*(1), 18114.

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: Ten years of progress. *Annual Review of Psychology,63*(1), 129–151.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Milton, A. L. (2019). Fear not: Recent advances in understanding the neural basis of fear memories and implications for treatment development. *F1000Research*, *8*, 1948.

Mineka, S., & Öhman, A. (2002). Phobias and preparedness: the selective, automatic, and encapsulated nature of fear. *Biological Psychiatry*, *52*(10), 927–937.

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.

Mogg, K., Millar, N., & Bradley, B. P. (2000). Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder. *Journal of Abnormal Psychology*, *109*(4), 695–704.

Montagrin, A., Saiote, C., & Schiller, D. (2018). The social hippocampus. *Hippocampus*, *28*(9), 672–679.

Morozov, A., & Ito, W. (2018). Social modulation of fear: Facilitation vs buffering. *Genes, Brain and Behavior*. https://doi.org/10.1111/gbb.12491

Moscarello, J. M., & Maren, S. (2018). Flexibility in the face of fear: Hippocampal–prefrontal regulation of fear and avoidance. *Current Opinion in Behavioral Sciences, 19*, 44–49.

Mou, X., & Ji, D. (2016). Social observation enhances cross-environment activation of hippocampal place cell patterns. *ELife*, *5.*

Murugan, M., Jang, H. J., Park, M., Miller, E. M., Cox, J., Taliaferro, J. P., … Witten, I. B. (2017). Combined social and spatial coding in a descending projection from the prefrontal cortex. *Cell*, *171*(7), 1663-1677.

Narayanan, N. S. (2016). Ramping activity is a cortical mechanism of temporal control of action. *Current Opinion in Behavioral Sciences, 8*, 226–230.

Negd, M., Mallan, K. M., & Lipp, O. V. (2011). The role of anxiety and perspective-taking strategy on affective empathic responses. *Behaviour Research and Therapy*, *49*(12), 852–857.

Okuyama, T. (2018). Social memory engram in the hippocampus. *Neuroscience Research*, *129*, 17–23.

Okuyama, T., Kitamura, T., Roy, D. S., Itohara, S., & Tonegawa, S. (2016). Ventral CA1 neurons store social memory. *Science*, *353*(6307), 1536–1541.

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, *21*(4), 197–212.

Olsson, A., McMahon, K., Papenberg, G., Zaki, J., Bolger, N., & Ochsner, K. N. (2016). Vicarious lear learning depends on empathic appraisals and trait empathy. *Psychological Science, 27*(1), 25-33.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience*, *2*(1), 3–11.

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102.

Padilla-Coreano, N., Bolkan, S. S., Pierce, G. M., Blackman, D. R., Hardin, W. D., Garcia-Garcia, A. L., … Gordon, J. A. (2016). Direct ventral hippocampal-prefrontal input is required for anxiety-related neural activity and behavior. *Neuron*, *89*(4), 857–866.

Padilla-Coreano, N., Canetta, S., Mikofsky, R. M., Alway, E., Passecker, J., Myroshnychenko, M. V., … Gordon, J. A. (2019). Hippocampal-prefrontal theta transmission regulates avoidance behavior. *Neuron*, *104*(3), 601-610.

Parfitt, G. M., Nguyen, R., Bang, J. Y., Aqrabawi, A. J., Tran, M. M., Seo, D. K., … Kim, J. C. (2017). Bidirectional control of anxiety-related behaviors in mice: Role of inputs arising from the ventral hippocampus to the lateral septum and medial prefrontal cortex. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *42*(8), 1715–1728.

Park, C. H. J., Ganella, D. E., Perry, C. J., & Kim, J. H. (2020). Dissociated roles of dorsal and ventral hippocampus in recall and extinction of conditioned fear in male and female juvenile rats. *Experimental Neurology*, *329*, 113306.

Pärnamets, P., Espinosa, L., & Olsson, A. (2020). Physiological synchrony predicts observational threat learning in humans. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1927), 20192779.

Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, *7*(4), 42.

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology, 31*, 319–330.

Patrick, C. J., Bradley, M. M., & Lang, P. J. (1993). Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology*, *102*(1), 82–92.

Paulhus, D., Neumann, C., & Hare, R. (2016). *Self-report psychopathy scale 4th edition (SRP 4) manual* (4th ed.). Multi-Health System.

Pereira, A. G., Cruz, A., Lima, S. Q., & Moita, M. A. (2012). Silence resulting from the cessation of movement signals danger. *Current Biology, 22*, R627–R628.

Pisansky, M. T., Hanson, L. R., Gottesman, I. I., & Gewirtz, J. C. (2017). Oxytocin enhances observational fear in mice. *Nature Communications*, *8*(1), 2102.

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience and Biobehavioral Reviews, 88*, 117–140.

Rachman, S. (1977). The conditioning theory of fearacquisition: A critical examination. *Behaviour Research and Therapy*, *15*(5), 375–387.

Reissner, C., Runkel, F., & Missler, M. (2013). Neurexins. *Genome Biology*, *14*(9), 1–15.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95.

Ressler, K. J. (2020). Translating across circuits and genetics toward progress in fear- And anxiety-related disorders. *American Journal of Psychiatry*, *177*(3), 214–222.

Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. R. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the "language of the eyes"? *Neuropsychologia*, *41*(5), 523–526.

Rodgers, R. J., Cao, B. J., Dalvi, A., & Holmes, A. (1997). Animal models of anxiety: An ethological perspective. *Brazilian Journal of Medical and Biological Research, 30*, 289–304.

Rozeske, R. R., Valerio, S., Chaudun, F., & Herry, C. (2015). Prefrontal neuronal circuits of contextual fear conditioning. *Genes, Brain and Behavior, 14*, 22–36.

Rozeske, R. R., Jercog, D., Karalis, N., Chaudun, F., Khoder, S., Girard, D., … Herry, C. (2018). Prefrontal-periaqueductal gray-projecting neurons mediate context fear discrimination. *Neuron*, *97*(4), 898-910.

Rudaizky, D., Basanovic, J., & MacLeod, C. (2014). Biased attentional engagement with, and disengagement from, negative information: Independent cognitive pathways to anxiety vulnerability? *Cognition and Emotion*, *28*(2), 245–259.

Rudy, B., Fishell, G., Lee, S. H., & Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental Neurobiology*, *71*(1), 45–61.

Rustay, N., Browman, K., & Curzon, P. (2008). *Cued and Contextual Fear Conditioning for Rodents* (pp. 19-37).

Sah, P., Faber, E. S. L., De Armentia, M. L., & Power, J. (2003). The amygdaloid complex: Anatomy and physiology. *Physiological Reviews, 83*, 803–834.

Sakaguchi, T., Iwasaki, S., Okada, M., Okamoto, K., & Ikegaya, Y. (2018). Ethanol facilitates socially evoked memory recall in mice by recruiting pain-sensitive anterior cingulate cortical neurons. *Nature Communications*, *9*(1), 1–10.

Samimy, S. M., Schettini, E., Fernhoff, K., Webster-Stratton, C., & Beauchaine, T. P. (2020). Parent training for childhood conduct problems. In *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier.

Sankoorikal, G. M. V., Kaercher, K. A., Boon, C. J., Lee, J. K., & Brodkin, E. S. (2006). A mouse model system for genetic analysis of sociability: C57BL/6J versus BALB/cJ inbred mouse strains. *Biological Psychiatry*, *59*(5), 415–423.

Scheggia, D., Managò, F., Maltese, F., Bruni, S., Nigro, M., Dautan, D., … Papaleo, F. (2019). Somatostatin interneurons in the prefrontal cortex control affective state discrimination in mice. *Nature Neuroscience*, 1–14.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., … Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

Seara-Cardoso, A., Sebastian, C. L., Viding, E., & Roiser, J. P. (2016). Affective resonance in response to others' emotional faces varies with affective ratings and psychopathic traits in amygdala and anterior insula. *Social Neuroscience*, *11*(2), 140–152.

Seara-Cardoso, A., & Viding, E. (2015). Functional neuroscience of psychopathic personality in adults. *Journal of Personality*, *83*(6), 723–737.

Seara-Cardoso, A., Viding, E., Lickley, R. A., & Sebastian, C. L. (2015). Neural responses to others' pain vary with psychopathic traits in healthy adult males. *Cognitive, Affective and Behavioral Neuroscience*, *15*(3), 578–588.

Selbing, I., & Olsson, A. (2017). Beliefs about others' abilities alter learning from observation. *Scientific Reports*, *7*(1), 16173.

Selbing, I., & Olsson, A. (2019). Anxious behaviour in a demonstrator affects observational learning. *Scientific Reports*, *9*(1), 9181.

Senn, V., Wolff, S. B. E., Herry, C., Grenier, F., Ehrlich, I., Gründemann, J., … Lüthi, A. (2014). Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron*, *81*(2), 428–437.

Serbic, D., Ferguson, L., Nichols, G., Smith, M., Thomas, G., & Pincus, T. (2020). The role of observer's fear of pain and health anxiety in empathy for pain: An experimental study. *British Journal of Pain*, *14*(2), 74–81.

Sharpe, M. J., & Killcross, S. (2014). The prelimbic cortex Contributes to the down-regulation of attention toward redundant cues. *Cerebral Cortex*, *24*(4), 1066–1074.

Sharpe, M. J., & Killcross, S. (2015a). The prelimbic cortex uses higher-order cues to modulate both the acquisition and expression of conditioned fear. *Frontiers in Systems Neuroscience*, *8*, 235.

Sharpe, M. J., & Killcross, S. (2015b). The prelimbic cortex directs attention toward predictive cues during fear learning. *Learning & Memory*, *22*(6), 289–293.

Shibano, N., Yamazaki, M., Arima, T., Abe, K., Kuroda, M., Kobayashi, Y., … Sano, Y. (2020). Excitation of prefrontal cortical neurons during conditioning enhances fear memory formation. *Scientific Reports*, *10*(1).

Shu, J., Hassell, S., Weber, J., Ochsner, K. N., & Mobbs, D. (2017). The role of empathy in experiencing vicarious anxiety. *Journal of Experimental Psychology: General*, *146*(8), 1164–1188.

Siciliano, C. A., Noamany, H., Chang, C. J., Brown, A. R., Chen, X., Leible, D., … Tye, K. M. (2019). A cortical-brainstem circuit predicts and governs compulsive alcohol drinking. *Science*, *366*(6468), 1008–1012.

Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2010). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology*, *36*(2), 529–538.

Sivaselvachandran, S., Acland, E. L., Abdallah, S., & Martin, L. J. (2018). Behavioral and mechanistic insight into rodent empathy. *Neuroscience & Biobehavioral Reviews*, *91*, 130–137.

Sotres-Bayon, F., & Quirk, G. J. (2010). Prefrontal control of fear: More than just extinction. *Current Opinion in Neurobiology, 20*, 231–235.

Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E., & Quirk, G. J. (2012). Gating of fear in prelimbic cortex by hippocampal and amygdala inputs. *Neuron, 76*(4), 804-812.

Sparta, D. R., Stamatakis, A. M., Phillips, J. L., Hovelsø, N., van Zessen, R., & Stuber, G. D. (2012). Construction of implantable optical fibers for long-term optogenetic manipulation of neural circuits. *Nature Protocols*, *7*(1), 12–23.

Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory. In *Self-Evaluation Questionnaire*. Consulting Psychologists Press.

Spielberger, C., Gorsuch, R., & Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory,* Consulting Psychologists Press.

Sterley, T. L., Baimoukhametova, D., Füzesi, T., Zurek, A. A., Daviu, N., Rasiah, N. P., … Bains, J. S. (2018). Social transmission and buffering of synaptic changes after stress. *Nature Neuroscience*, *21*(3), 393–403.

Steyer, R., Schmitt, M., & Michael, E. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408.

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience, 15*, 655–669.

Südhof, T. C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature, 455*, 903–911.

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders*, *20*(1), 110–117.

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

Svoboda, K., & Yasuda, R. (2006). Principles of two-photon excitation Mmicroscopy and its applications to neuroscience. *Neuron, 50*, 823–839.

Sylvers, P., Lilienfeld, S. O., & LaPrairie, J. L. (2011). Differences between trait fear and trait anxiety: Implications for psychopathology. *Clinical Psychology Review, 31*, 122–137.

Szczepanik, M., Kaźmierowska, A. M., Michałowski, J. M., Wypych, M., Olsson, A., & Knapska, E. (2020). Observational learning of fear in real time procedure. *Scientific Reports*, *10*(1), 16960.

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience, 5*, 1089–1092.

Tierney, P. L., Degenetais, E., Thierry, A.-M., Glowinski, J., & Gioanni, Y. (2004). Influence of the hippocampus on interneurons of the rat prefrontal cortex. *European Journal of Neuroscience*, *20*(2), 514–524.

Toledano, M. B., Mutz, J., Röösli, M., Thomas, M. S. C., Dumontheil, I., & Elliott, P. (2019). Cohort profile: The study of cognition, adolescents and mobile phones (SCAMP). *International Journal of Epidemiology*, *48*(1), 25-26I.

Tovote, P., Esposito, M. S., Botta, P., Chaudun, F., Fadok, J. P., Markovic, M., … Lüthi, A. (2016). Midbrain circuits for defensive behaviour. *Nature*, *534*(7606), 206–212.

Tovote, P., Fadok, J. P., & Lüthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews Neuroscience*, *16*(6).

Ueno, H., Suemitsu, S., Murakami, S., Kitamura, N., Wani, K., Okamoto, M., … Ishihara, T. (2018). Empathic behavior according to the state of others in mice. *Brain and Behavior*, *8*(7), e00986.

Van De Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science*, *340*(6131), 483–485.

van Dongen, J. D. M. (2020). The Empathic Brain of Psychopaths: From Social Science to Neuroscience in Empathy. *Frontiers in Psychology*. Frontiers Media S.A.

Vander Weele, C. M., Siciliano, C. A., Matthews, G. A., Namburi, P., Izadmehr, E. M., Espinel, I. C., … Tye, K. M. (2018). Dopamine enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli. *Nature*, *563*(7731), 397–401.

Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L., & Quirk, G. J. (2006). Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learning & Memory*, *13*(6), 728–733.

Walentowska, W., & Wronka, E. (2012). Trait anxiety and involuntary processing of facial emotions. *International Journal of Psychophysiology*, *85*(1), 27–36.

Wang, Q., Jin, J., & Maren, S. (2016). Renewal of extinguished fear activates ventral hippocampal neurons projecting to the prelimbic and infralimbic cortices in rats. *Neurobiology of Learning and Memory*, *134*, 38–43.

Wang, Y., DeMarco, E. M., Witzel, L. S., & Keighron, J. D. (2021). A selected review of recent advances in the study of neuronal circuits using fiber photometry. *Pharmacology Biochemistry and Behavior, 201*, 173113.

Warren, B. L., Vialou, V. F., Iñiguez, S. D., Alcantara, L. F., Wright, K. N., Feng, J., … Bolaños-Guzmán, C. A. (2013). Neurobiological sequelae of witnessing stressful events in adult mice. *Biological Psychiatry*, *73*(1), 7–14.

Watson, D., Kotov, R., & Gamez, W. (2006). Basic Dimensions of Temperment in Relation to Personality and Psychopathy. In R. Krueger & J. Tackett (Eds.), *Personality and Psychopathology* (pp. 7–38).

Watson, T. C., Cerminara, N. L., Lumb, B. M., & Apps, R. (2016). Neural correlates of fear in the periaqueductal gray. *The Journal of Neuroscience*, *36*(50), 12707.

Wilson, K., Juodis, M., & Porter, S. (2011). Fear and loathing in psychopaths: A meta-analytic investigation of the Facial Affect Recognition Deficit. *Criminal Justice and Behavior*, *38*(7), 659–668.

Wiltgen, B. J., Sanders, M. J., Behne, N. S., & Fanselow, M. S. (2001). Sex differences, context preexposure, and the immediate shock deficit in Pavlovian context conditioning with mice. *Behavioral Neuroscience*, *115*(1), 26–32.

Wittchen, H.-U., Essau, C. A., & Krieg, J.-C. (1991). Anxiety disorders: Similarities and differences of comorbidity in treated and untreated groups. *British Journal of Psychiatry, 159*(SEPT.SUPPL.12), 23-33.

Wright, K. M., & McDannald, M. A. (2019). Ventrolateral periaqueductal gray neurons prioritize threat probability over fear output. *ELife*, *8*, e45013.

Xu, C., Krabbe, S., Gründemann, J., Botta, P., Fadok, J. P., Osakada, F., … Lüthi, A. (2016). Distinct hippocampal pathways mediate dissociable roles of context in memory retrieval. *Cell*, *167*(4), 961-972.

Xu, H., Liu, L., Tian, Y., Wang, J., Li, J., Zheng, J., … Xu, H. (2019). A disinhibitory microcircuit mediates conditioned social fear in the prefrontal cortex. *Neuron*, *102*(3), 668-682.

Xu, W., & Südhof, T. C. (2013). A neural circuit for memory specificity and generalization. *Science*, *339*(6125), 1290–1295.

Yamanashi, T., Maki, M., Kojima, K., Shibukawa, A., Tsukamoto, T., Chowdhury, S., … Sudo, Y. (2019). Quantitation of the neural silencing activity of anion channelrhodopsins in Caenorhabditis elegans and their applicability for long-term illumination. *Scientific Reports*, *9*(1), 1–11.

Ye, X., Kapeller-Libermann, D., Travaglia, A., Inda, M. C., & Alberini, C. M. (2017). Direct dorsal hippocampal–prelimbic cortex connections strengthen fear memories. *Nature Neuroscience*, *20*(1), 52–61.

Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M., & Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron, 71*, 9–34.

Yizhar, O., & Klavir, O. (2018). Reciprocal amygdala–prefrontal interactions in learning. *Current Opinion in Neurobiology*, *52*, 149–155.

Yusufishaq, S., & Rosenkranz, J. A. (2013). Post-weaning social isolation impairs observational fear conditioning. *Behavioural Brain Research*, *242*(1), 142–149.

Zelikowsky, M., Hersman, S., Chawla, M. K., Barnes, C. A., & Fanselow, M. S. (2014). Neuronal ensembles in amygdala, hippocampus, and prefrontal cortex track differential components of contextual fear. *Journal of Neuroscience*, *34*(25), 8462–8466.

Zhong, J., Liang, M., Akther, S., Higashida, C., Tsuji, T., & Higashida, H. (2014). C-Fos expression in the paternal mouse brain induced by communicative interaction with maternal mates. *Molecular Brain*, *7*(1), 1–11.

Zhou, C., Zhou, Z., Han, Y., Lei, Z., Li, L., Montardy, Q., … Wang, L. (2018). Activation of parvalbumin interneurons in anterior cingulate cortex impairs observational fear. *Science Bulletin*, *63*(12), 771–778.

Zinbarg, R. E., & Mohlman, J. (1998). Individual differences in the acquisition of affectively valenced associations. *Journal of Personality and Social Psychology*, *74*(4), 1024–1040.

Zingg, B., Chou, X., Zhang, Z., Mesik, L., Liang, F., Tao, H. W., & Zhang, L. I. (2017). AAV-mediated anterograde transsynaptic tagging: Mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron*, *93*(1), 33–47.

Zinn, C. G., Clairis, N., Cavalcante, L. E. S., Furini, C. R. G., De Carvalho Myskiw, J., & Izquierdo, I. (2016). Major neurotransmitter systems in dorsal hippocampus and basolateral amygdala control social recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(33), E4914–E4919.