

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



## The neuropsychological mechanisms of fear learning and memory

Lam, Charlene

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

### END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

**THE NEUROPSYCHOLOGICAL MECHANISMS OF  
FEAR LEARNING AND MEMORY**

**LAM LOK MAN, CHARLENE**

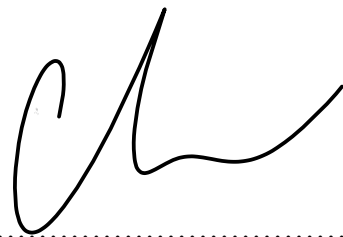
Thesis submitted for the Degree of Doctor of Philosophy  
at the University of Hong Kong and King's College London

December 2020



## Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to the University or to any other institution for a degree, diploma or other qualifications.



Signed .....

LAM Lok Man, Charlene

July, 2020

## Acknowledgements

I feel incredibly privileged to complete my PhD program at the University of Hong Kong and King's College London. The transition from a full-time clinical psychologist to a full-time research student was anxiety-provoking in the beginning, but in hindsight, I am pleased to have taken this step forward. The journey has undoubtedly been challenging and arduous at times, but I am fortunate enough to work with some of the kindest and most brilliant people I have ever met in this journey. First and foremost, I am extremely grateful to my three amazing supervisors Tatia Lee, Jenny Yiend, and Tom Barry for their immense knowledge, support, and guidance. I would not have been able to build the experiments from scratch and write the manuscripts without their dedicated supervision and wisdom. Their generosity in sharing their knowledge, along with their tremendous kindness and patience, are essential for shaping me to become more competent as a researcher and compassionate as a clinician.

I owe a very big thank you to my peers in the laboratory of Neuropsychology and the Institute of Clinical Neuropsychology. From graduated to current: Nichol Wong, Ruibin Zhang, Robin Shao, Gerard Yu, Emily Sin, JJ Wong, Idy Man, Qidi, Mengxia Gao and Zhongwan Liu: You have made the lab a friendly and supportive environment to work in. I am especially grateful to Helena Tam, Mandy Lo, Mandy Ng, Alice Chow for their unending support and laughter (and cakes!) at the ICN. Thank you to all people that have offered help and advice on this PhD, including Elias Mouchliantis, Christian Steinberg, Alicia Tse, Farah Mgaieth, Lawrence Kwan. My sincere thanks to Clive Wong and Robin Shao for helping me conduct my first brain-imaging study and handling my endless questions about e-prime, SPM and FSL. I also thank Chantel Leung for being such a wonderful buddy and sharing many fond memories with me at King's.

I owe a very special debt of gratitude to Kati Roesmann and Markus Junghöfer for their immense support and guidance throughout the Multi-CS project. From science to beers, I will always remember their kindness and great hospitality in Münster, my favourite city in Germany.

My deepest thanks to my friends in Hong Kong and afar for being my cheerleading squad; to my incredible significant other Berlin for being my data.table guru and showering me with his love and care; finally, to my dearest mom and dad, for their unconditional love and support.

## Abstract

Soon after experiencing an event, the memory is in an active state until it gets consolidated and encoded into long-term memory. Consolidated memories, once reactivated, become unstable and undergo a re-stabilisation process called reconsolidation, in which they are subject to modification. Interventions applied during the reconsolidation window may modify the original fear memory and prevent the spontaneous recovery and reinstatement of the fear response, leading to a more effective modulation of fear expressions than traditional approaches in which the fear memory is merely inhibited. The four studies reported in this thesis investigate the psychophysiological and neural mechanisms of extinction learning and fear recovery using a reconsolidation-based fear/threat conditioning paradigm. **Chapter 1** presents an overview of the thesis and its overarching goal to enhance extinction learning and reduce relapse in anxiety and fear-related disorder. **Chapter 2** reviews the literature on threat (or fear) conditioning, awareness and reconsolidation, and their potential to modulate maladaptive fear memories. **Chapter 3** presents an experimental study that investigates the extent of attenuation of fear responses, as indexed by pupillometry, during an implicit exposure to conditioned stimuli (N = 59). This study shows that explicit and implicit extinction modulated fear responses and the percentage of fear recovery was higher for implicit than explicit extinction. **Chapter 4** assesses the potential of using an implicit reminder cue to reactivate the original fear in a three-day reconsolidation conditioning experiment (N = 61). Although the findings do not support the use of a pre-extinction reminder cue in modulating the reinstatement of fear, the reminded conditioned stimulus (CS) was rated more unpleasant than the non-reminded CS and the safe stimulus following extinction, independent of its perceptual awareness during reactivation. **Chapter 5** examines the impact of the reminder-retrieval procedure using an implicit multi-CS conditioning paradigm. In Experiment 3 (N = 36), an unconditioned stimulus

reminder cue-induced a distinct pattern of pupil responses for the reminded relative to the non-reminded CS during early extinction. **Chapter 6** further investigates the impact of the retrieval-reminder procedure on the neural mechanisms of extinction and the return of fear. In Experiment 4 (N = 22), significant neural activation of the right dorsolateral prefrontal region was observed in the reminded CS relative to the non-reminded CS during early extinction. Importantly, the non-reminded CS evoked stronger neural responses in the right dorsolateral prefrontal cortex and the right hippocampus than the reminded CS following the reinstatement test. Overall, the four experiments provide encouraging physiological and neural evidence for the impact of the retrieval-reminder procedure on extinction learning. Finally, **Chapter 7** summarises and integrates the findings into the existing literature, and discusses the clinical implications, limitations and suggestions for future research on modulating maladaptive fear memories.



**The work presented in the current thesis gave rise to the following publications and submissions:**

Lam, C.L., Barry, T.J., Yiend, J., & Lee, T.M. (under review). Explicit extinction modulates defensive responses more effectively than implicit extinction

Lam, C.L., Höfig, A., Steinberg, C., Junghöfer, M., Yiend, J., Barry, T., Lee, T.M., Roesmann, K. (2019). Preventing return of fear through memory reconsolidation using a novel Multi-CS conditioning paradigm. Poster presented at the European Meeting on Human Fear Conditioning

Roesmann, K., Lam, C.L., Steinberg, C., Höfig, A., Barry, T., Lee, T.M., & Junghöfer, M. (2019) The Relationship of Memory Reconsolidation and Return of Fear: Clinical and Methodological Implications of a Novel MultiCS Conditioning Paradigm. Poster presented at the World Congress of Behavioural and Cognitive Therapies.

## Contents

<b>Declaration.....</b>	<b>1</b>
<b>Acknowledgements .....</b>	<b>2</b>
<b>Abstract.....</b>	<b>4</b>
<b>Contents .....</b>	<b>7</b>
<b>Table of Figures.....</b>	<b>10</b>
<b>Table of Tables .....</b>	<b>12</b>
<b>Abbreviations .....</b>	<b>14</b>
<b>Glossary .....</b>	<b>15</b>
<b>Chapter 1 Introduction.....</b>	<b>17</b>
1.1 Overview of thesis .....	18
<b>Chapter 2 Review of Key Concepts .....</b>	<b>20</b>
2.1 Pavlovian conditioning paradigm .....	21
2.1.1 Extinction from a learning perspective .....	22
2.1.2 Extinction from a neural perspective .....	27
2.2 Awareness and fear conditioning.....	30
2.2.1 Empirical evidence for the single-process model .....	32
2.2.2 Empirical evidence for the dual-process model.....	32
2.2.3 Single- or dual-processing model revisited.....	36
2.3 Memory reconsolidation .....	39
2.3.1 The origin of the reconsolidation hypothesis: the discovery and re-discovery.	41
2.3.2 Reconsolidation in humans: laboratory studies .....	42
2.3.3 Reconsolidation in humans: clinical studies .....	47
2.4 Overview of the studies.....	49

2.4.1 Methodological approach .....	49
-------------------------------------	----

**Chapter 3 Experiment 1: Explicit extinction modulates defensive responses more effectively than implicit extinction ..... 51**

3.1 Introduction .....	52
3.2 Methods .....	56
3.3 Results .....	63
3.4 Discussion.....	70
3.5 Conclusion .....	75

**Chapter 4 Experiment 2: No statistical evidence for implicit and explicit reminder cues on reducing the reinstatement of fear in a retrieval-extinction threat conditioning paradigm..... 77**

4.1 Introduction .....	78
4.2 Methods .....	81
4.3 Results .....	88
4.4 Discussion.....	98
4.5 Conclusion .....	103

**Chapter 5 Experiment 3: Evidence for differential extinction learning by disrupting reconsolidation in multi-CS conditioning ..... 105**

5.1 Introduction .....	106
5.2 Methods .....	109
5.3 Results .....	117
5.4 Discussion.....	122
5.5 Conclusion .....	126

**Chapter 6 Experiment 4: The impact of a US reminder cue on the neural mechanisms of extinction and the return of fear in multi-CS conditioning ..... 127**

6.1 Introduction .....	128
------------------------	-----

6.2 Methods.....	129
6.3 Results.....	138
6.4 Discussion .....	149
6.5 Conclusion .....	153
<b>Chapter 7 General summary and discussion.....</b>	<b>156</b>
7.1 Summary of the findings.....	157
7.2 Synthesis and discussion: Memory reconsolidation and extinction.....	159
7.3 Clinical implications .....	165
7.4 Limitation of fear conditioning/memory reconsolidation models.....	168
7.5 Definitions of fear .....	170
7.6 Future directions .....	175
7.7 Closing remark.....	177
<b>Chapter 8 References.....</b>	<b>178</b>
<b>Appendix A. ....</b>	<b>197</b>

## Table of Figures

Figure 2-1 (a) A single-process model; (b) a dual-process model.....	31
Figure 2-2 A Two-system model.....	38
Figure 3-1 (a) Timeline and (b) percept of the extinction learning .....	58
Figure 3-2 (a) Average change in pupil diameter in response to CS <sub>exp+</sub> , CS <sub>imp+</sub> and CS- across trials of threat acquisition phase with 95% confidence intervals. (b) Baseline-corrected pupillary responses in threat acquisition.....	65
Figure 3-3 Pupillary changes in the <i>aware</i> vs <i>unaware</i> trials during Extinction .....	67
Figure 3-4 CS unpleasantness rating after (a) threat acquisition, (b) extinction, and (c) re-extinction.....	69
Figure 3-5 (a) Baseline-corrected pupillary response after reinstatement. (b) Percentage of fear recovery after reinstatement .....	70
Figure 4-1 a) Overview of the experimental protocol. (1b) Percept of the reactivation trials with and without the interference of the continuous flash suppression. ....	86
Figure 4-2 Pupillary responses in the (a) acquisition, (b) extinction, and (c) re-extinction. ....	91
Figure 4-3 Valence rating in the (a) acquisition, (b) extinction, and (c) re-extinction.....	94
Figure 4-4 Likelihood rating in the (a) acquisition, (b) extinction, and (c) re-extinction. ....	95
Figure 4-5 Non-normalized pupillary responses in the (a) implicit reactivation group, and (b) explicit reactivation group across late-extinction and re-extinction.....	97
Figure 5-1 Experimental timeline and procedure during conditioning. In Multi-CS conditioning, reminded CS+ (rCS+), non-reminded CS+ (nrCS+) and CS- faces were presented in a pseudorandomized order whereby each CS was presented four times for 800ms (i.e., 72 trials per condition, 216 trials in total). The auditory USs started 600 ms after the CS onset.....	112
Figure 5-2 Pupil responses during the (a) acquisition, (b) early extinction, (c) late extinction, and (d) the first run of re-extinction .....	118
Figure 5-3 Results of Pair Comparison across days. A higher score on y-axis suggests a preference towards a particular type of CS (rCS, nrCS, CS-). ....	122
Figure 6-1 Experimental timeline during acquisition.....	133

Figure 6-2 Results of the Pair Comparison across each experimental day. ....	138
Figure 6-3 a) Right dorsolateral prefrontal region reactivity in the rCS > nrCS contrast during early extinction on Day 2 (30, 22, 46, $p_{\text{FWE-corr}} = .033$ ). b) Left dorsolateral prefrontal region reactivity in the nrCS > rCS contrast during late extinction on Day 2 (-44, 42, 22, $p_{\text{FWE-corr}} = .038$ ) .....	145
Figure 6-4 a) Enhanced right dorsolateral prefrontal region ( $z = 12$ , $p_{\text{FWE-corr}} = .036$ ) and b) hippocampus ( $y = -26$ , $p_{\text{FWE-corr}} = .007$ ) reactivity in the nrCS > rCS contrast following test of reinstatement on Day 3. ....	148
Figure 7-1 Gershman's model for predicting fear extinction with reference to the size of prediction error (x-axis) and the change in the latent cause inferred from the conditioning (y-axis). .....	162

## Table of Tables

Table 3-1 Result summary: Coefficient estimates, Standard Error, <i>t</i> statistics, and significance levels <i>p</i> for all predictors in the acquisition and re-extinction phase. Significant beta values suggest that pupil responses of the corresponding CS type were significantly different compared to those of the implicit CS+ (as the intercept). .....	64
Table 3-2 Estimated marginal means of the pupil responses and unpleasantness rating of CSs, their standard errors and confidence intervals in each experimental phase.....	64
Table 3-3 Coefficient estimates (beta), Standard Error, <i>t</i> statistics, and significance level <i>p</i> predicting pupillary responses in the extinction. ....	67
Table 4-1 Demographic Characteristics of the Sample (N = 59) .....	88
Table 4-2 Result summary: Coefficient estimates (beta), Standard Error, <i>t</i> statistics, and significance level <i>p</i> for each predictor in estimating pupillary responses in the LMM analyses.	89
Table 4-3 Estimated marginal means of the pupillary responses of CSs, their standard errors and confidence intervals in each experimental session .....	90
Table 4-4 Estimated marginal means of the valence and likelihood ratings of CSs, their standard errors and confident intervals after each experimental session .....	93
Table 5-1. Linear mixed-effects modelling (LMM) results for pupil responses on each experimental day (N = 36).....	119
Table 5-2 Estimated means differences and standard errors for pupil responses on each experimental day .....	119
Table 5-3 Linear mixed-effects modelling (LMM) results for pupil responses in computing the recovery of fear (N = 36).....	120
Table 6-1 Regions of Interest (ROIs) in the current study and their corresponding locations in the Brainetome Atlas.....	137
Table 6-2 Localization and statistics for whole-brain analysis for Day 1 (Acquisition). .....	140
Table 6-3 Localization and statistics for ROI-analyses for Day 1 (Acquisition). .....	142
Table 6-4 Localization and statistics for ROI-analyses for Day 2 (Extinction) .....	145
Table 6-5 Localization and statistics for whole-brain analysis for Day 3 (Re-extinction).....	147

Table 6-6 Localization and statistics for ROI-analyses for Day 3 (Re-extinction).....	148
Table 6-7 Localization and statistics for ROI-analysis for fear recovery from Day 2 to Day 3 (Extinction run 4 vs Re-Extinction run 1).....	149
Table 6-8 Localization and statistics for whole-brain analysis for Day 3 (Re-Extinction).....	154
Table 6-9 Localization and statistics for ROI-analyses for Day 2 (Extinction) and Day 3 (Re- Extinction) .....	155
Table 8-1 Supplemental analyses for Experiment 1 (a) Result summary: Coefficient estimates, Standard Error, <i>t</i> statistics, and significance levels <i>p</i> for all predictors in the acquisition and re- extinction phase. Significant beta values suggest that pupil responses of the corresponding CS type were significantly different compared to those of the implicit CS+ (as the intercept). (N = 59). (b) Estimated marginal means of the pupil responses and unpleasantness rating of CSs, their standard errors/deviations and confidence intervals in each experimental phase (N = 59) .....	197
Table 8-2. Supplemental Analyses for Experiment 2: Demographic Characteristics of the Sample (N = 59).....	198
Table 8-3. Experiment 2 result summary: Coefficient estimates (beta), Standard Error, <i>t</i> statistics, and significance level <i>p</i> for each predictor in estimating pupillary responses.....	198
Table 8-4. Pearson correlation coefficients between unpleasantness ratings (post-extinction) and pupil responses (post-reinstatement).....	199
Table 8-5. Supplemental analyses for Experiment 3: Estimated means differences and standard errors for pupil responses for participants with low detectability of CS-US associations (Day 1) ( <i>n</i> = 18) .....	200
Table 8-6. Supplemental analyses for Experiment 4: Localization and statistics for cerebellar ROI-analyses on Day 1 (Acquisition). The figure illustrates the bilateral cerebellar lobules VI and Crus 1 activation during early acquisition on Day 1 ( $p_{\text{FWE-corr}} < .05$ ).....	201
Table 8-7. Localization and statistics for cerebellar ROI-analyses on Day 3 (Re-Extinction). The figure illustrates the right cerebellar lobules VI deactivation in the $r\text{CS}+ > nr\text{CS}+$ contrast during re-extinction on Day 3. (16, -64, -18, $p_{\text{FEW-corr}} = .031$ ).....	202



## Abbreviations

BOLD	Blood oxygen level-dependent
CFS	Continuous Flash Suppression
CS	Conditioned stimulus
CS+	Conditioned stimulus plus
CS-	Conditioned stimulus minus
dACC	Dorsal anterior cingulate
dIPFC	Dorsolateral prefrontal cortex
fMRI	Functional magnetic resonance imaging
MEG	Magnetoencephalography
nrCS	Non-reminded conditioned stimulus
rCS	Reminded conditioned stimulus
vmPFC	Ventromedial prefrontal cortex
US	Unconditioned stimulus

## Glossary

Acquisition	A process of associating the conditioned stimulus and unconditioned stimulus in the Pavlovian conditioning theory.
Conditioning	A type of leaning that involves the learning of relations among events.
Conditioned stimulus	A stimulus, usually neutral in nature, comes to elicit a conditional response upon its previous pairings with another stimulus.
Conditioned response	The reaction elicited by a conditioned stimulus that has been paired with an unconditioned stimulus.
CS+	A conditioned stimulus that is paired with an aversive conditioned stimulus.
CS-	A conditioned stimulus that is not paired with an aversive unconditioned stimulus.
Consolidation	A hypothesis refers to a time-dependent process by which memory becomes stable and is reorganized and stored in the brain.
Contingency awareness	The explicit, declarative knowledge of the association between the conditioned stimulus and unconditioned stimulus.
Continuous Flash Suppression	A form of binocular rivalry wherein a visual stimulus presented to one eye is suppressed from awareness as a result of a rapidly changing sequence of high contrast viewed by the other eye.
Extinction	A decay of conditioned responses due to repeated pairings of the conditioned stimulus without the unconditioned stimulus.
Multi-CS conditioning	A learning phenomenon that occurs when a multitude of stimuli is paired with one or multiple unconditioned stimuli.
Reconsolidation	A hypothesis refers to the process by which retrieval and reactivation of a memory trace appears to destabilize the memory trace(s) and enable it/them to be updated and modified.
Reinstatement	The re-emergence of the association between the CS and US following the presentation of the unconditioned stimulus alone. Also a Return of Fear phenomenon.
Retrieval-extinction	A procedure in which the association of the CS and the US is retrieved before extinction takes place.
Return of fear	An increase of the conditioned responses or symptoms of fear following extinction.
Re-extinction	A decay of conditioned responses due to repeated pairings of the conditioned stimulus without the unconditioned stimulus following tests of return of fear.
rCS+	Reminded conditioned stimulus: A conditioned stimulus paired with the aversive conditioned stimulus and is later retrieved with a reminder cue before extinction.

nrCS+	Non-reminded conditioned stimulus: A conditioned stimulus paired with the aversive conditioned stimulus and is not retrieved with a reminder cue before extinction.
Single-CS conditioning	A learning phenomenon that occurs when one stimulus is paired with one unconditioned stimulus.
Unconditioned stimulus	A stimulus, usually aversive in nature, that elicits a response unconditionally (i.e. independent of pairings with other stimuli).

## **Chapter 1 Introduction**

## 1.1 Overview of thesis

Fear is an emotional experience deeply rooted in evolution that serves to protect us from dangers or potentially harmful situations. In healthy individuals, fear can facilitate action to maintain safety and well-being (Lang et al., 2000). Nevertheless, maladaptive form of fear resulted from traumatic experiences and chronic stress is implicated in the development and/or maintenance of anxiety- and fear-related disorders such as posttraumatic stress disorder. Anxiety and fear-related disorders are common mental health problems. With a 12-month prevalence of about 14.0%, they constitute the largest group of mental disorders (Wittchen et al., 2011). In 2017, anxiety and fear-related disorders affected over 284.3 million people worldwide (James et al., 2018) and accounted for 27.1 million disability-adjusted life-years, a composite measure of disease burden capturing prevalence of premature mortality and the number of years lost due to ill-health (Kyu et al., 2018). In short, these disorders are associated with high individual and societal burdens.

Exposure-based therapy involves repeated confrontations with feared stimuli and is the gold-standard clinical intervention for the treatment of anxiety or fear-related disorders; however, not every recipient responds to exposure therapy, and relapse is common. Depending on the type of anxiety disorder and the operationalization of relapse, the non-response rate to exposure therapy ranges from 10% to 30%, and the rate of relapse after initially successful therapy ranges from 19% to 62% (Craske & Mystkowski, 2006). A more recent longitudinal study involving 439 depression and anxiety patients showed that the relapse rate could reach up to 53% within one year of a standard course of psychological intervention with exposure components (Ali et al., 2017). From a clinical perspective, these non-response and relapse rates are far from optimal.

The overarching theme of this thesis is to study the neural and behavioural mechanisms underlying fear extinction and its recovery. To this end, the science behind the research in emotion, memory and consciousness was reviewed and considered to conceptualise the studies included in this thesis. Fear conditioning and memory reconsolidation are the two critical theories underlying the design of the experiments. The Pavlovian conditioning paradigm is one of the most widely researched models to examine the mechanisms underlying fear learning and extinction in the fields of psychology and neuroscience. Since its inception in the 1920s by early learning theorists such as Pavlov (1927) and Watson (1920), it has proved to be a robust framework for modelling the development and maintenance of pathological states of fear and anxiety (Bouton et al., 2001; Eelen & Vervliet, 2007). Memory reconsolidation, a putative process in which consolidated fear memories can be destabilised and are subject to modification, is positioned as another promising mechanism that may prevent the return of fear (Nader et al., 2000; Schiller et al., 2010a). Although fear memories can be modified behaviourally during reconsolidation, the conditions under which fear-related memories are reactivated have not been fully elucidated; therefore, the current thesis presents an investigation of the conditions that trigger memory reconsolidation in four experimental studies in humans.

This thesis has eight chapters. Chapter 1 describes the research questions and goals of this thesis. Chapter 2 reviews the theoretical and empirical developments in fear conditioning and memory reconsolidation. Chapter 3 outlines the four experiments conducted in this thesis. Chapters 4 to 6 describes the method and results of each study. Finally, Chapter 7 discusses the findings from these studies and their implications in research on human fear conditioning as well as their clinical applications.

## **Chapter 2 Review of Key Concepts**

## 2.1 Pavlovian conditioning paradigm

Pavlovian conditioning is a process of learning by which we learn the relations among events or stimuli; one stimulus is related to or predicts the occurrence of another (Pavlov, 1927; Rescorla, 1988). Developed by Russian physiologist Ivan Pavlov, winner of the Nobel Prize in Physiology or Medicine in 1904, conditioning research was first conducted to examine the digestive system in dogs. Pavlov and his colleagues later expanded the paradigm to study how experimental manipulation can evoke emotional responses in dogs. In one classic study, dogs were conditioned to salivate to a circle and an ellipse after repeatedly pairing the shapes with food. As Pavlov and colleagues made discrimination of the visual stimulus harder for the dogs to receive their food, the dogs began to wriggle and bark violently, displaying behaviours that Pavlov termed 'a condition of acute neurosis' (Pavlov, 1927, p. 291).

The findings on conditioned emotional reactions were further extended in human studies by Watson and Rayner (1920). In their seminal article entitled "Conditioned emotional reactions", Watson and Rayner demonstrated how initially neutral stimuli could be conditioned to elicit fear in a nine-month-old infant named Albert (known as 'Little Albert'). This discovery was revolutionary at the time as the prevailing view on pathological fear reactions in the early 1900s was mainly driven by psychoanalysis, where psychopathology was believed to develop as a result of a conflict between different internal states and structures (Wolpe, 1981). Contrary to the notions of the psychoanalytic tradition, conditioning experiments have demonstrated that emotional responses are highly influenced by learning processes and external contexts.

Inspired by Watson's work, Jones (1924) developed a treatment for phobic fears, which is considered a predecessor of modern-day exposure therapy. Wolpe and colleagues



(1981) further refined the treatment, and now exposure therapy is viewed as the treatment of choice for anxiety and fear-related disorders. Exposure therapy involves repeated approaches towards fear-provoking stimuli such that previously ‘dangerous’ stimuli are no longer threatening and are instead viewed as ‘safe’. Procedurally, this is equivalent to fear extinction in a Pavlovian conditioning paradigm in which the conditioned stimulus (CS) that was previously paired with an aversive stimulus (US) is now presented repeatedly without the US (Bouton, 2017).

### **2.1.1 Extinction from a learning perspective**

The contemporary conception of extinction refers to the gradual decay of anticipatory fear reactions (i.e., the conditioned response, CR) as an individual begins to learn that the CS no longer predicts the US (Bouton, 2017). Indeed, it has long been understood that extinction learning is relatively fragile and less durable than conditioning (Bouton, 2017; Pavlov, 1927). What factors contribute to this fragility has been a central question to scientists and clinicians alike. Several influential learning models have attempted to explain extinction learning in the Pavlovian conditioning framework. These models broadly fall into two categories: US processing models (e.g., Rescorla-Wagner's model) or CS processing models (e.g., Bouton's extinction model). In the next section, we will visit the basic tenets of these models, their empirical evidence, and the latest attempt to integrate these models.

#### **2.1.1.1 US processing model: the Rescorla-Wagner model**

According to Rescorla and Wagner (1972), both acquisition and extinction learning refer to the processes of associating the CS with an aversive US. They further proposed a

mathematical model to delineate the amount of learning on each trial of Pavlovian learning as follows:

$$\Delta V = \alpha\beta (\lambda - \Sigma V)$$

where

$\Delta V$  is the change in the associative strength on each trial ;

$\alpha$  is the salience of the CS;

$\beta$  is the salience of the US;

$\lambda$  is the maximum associative strength of the US;

$\Sigma V$  is the expected amount of associative strength for all stimuli present

The Rescorla-Wagner model considers conditioning a form of predictive learning and highlights two key aspects of learning. First, the amount of surprise (i.e.,  $(\lambda - \Sigma V)$  in the equation) generated from the difference between one's expectation and the reality is a determinant of the strength of the association; that is, learning occurs when there is a discrepancy between the predicted and actual outcome as a form of error-corrected learning. Second, simple contiguity of two events is insufficient to produce conditioning. Instead, information provided about the US is essential during the pairing (Rescorla, 1988).

Extinction, as viewed by Rescorla and Wagner, is considered a case of negative prediction error; the associative strength between the CS and the US decreases because of the absence of a predicted US and the gradual reduction in the associative value of the CS ( $\lambda \approx V$ ). Notably, extinction is viewed as a form of unlearning, and recovery of the learning is not predicted in this model.

The Rescorla-Wagner model has received support from behavioural (Culver et al., 2015; Leung et al., 2012; Rescorla, 2006) and neuroimaging studies (Montague et al., 1996;

Schultz et al., 1997). In primates, the neural substrate for prediction error signals has also been identified in the dopaminergic neurons located in the midbrain (Schultz et al., 1997). Although Rescorla-Wagner's model has enjoyed great success in explaining a variety of phenomena relevant to conditioning, it is insufficient for explaining the post-extinction recovery of fear.

#### **2.1.1.2 CS processing model: the Bouton model of extinction**

In contrast to the US processing model, CS processing models focus on how the CS is processed. Based on his systematic studies in rodents, Bouton argued that extinction does not erase the original CS-US learning (Bouton, 2002). Instead, it involves new inhibitory learning and the formation of a new CS-no US association. In other words, the CS both activates and deactivates the representation of the US, and the CS will no longer elicit a fear response when the newly formed inhibitory CS-US representation is stronger than the original CS-US representation. Bouton's model of extinction is the predominant model in the field as it is carefully validated in animal conditioning studies by rejecting alternative hypotheses such as incomplete extinction and generalization decrement, which are more parsimonious in nature (Vervliet et al., 2013).

Another important aspect of Bouton's model is its emphasis on context. Contexts refer to both external and internal (or interceptive states) information available at the time of learning, which serves as retrieval cues for specific CS-US relationships (Bouton, 2002). Bouton proposed new extinction learning as a form of context-dependent memory; the retrieval of it depends on where it was learned. Post-extinction recovery occurs because retrieval of such memory rarely survives a shift in context. If the context of the post-extinction test is dissimilar to the context in which extinction was learned, retrieval tends to

favour the original excitatory CS-US memory and thus, conditioned responses re-emerge in the new context. In addition to the physical environment, time and one's internal states are also factored into context.

There is a wealth of empirical support for Bouton's model of extinction (Bouton, 2017). One reliable finding of extinction studies is the resurgence of fear responses after extinction training, a phenomenon known as return of fear (ROF) (Rachman, 1979). Research into the ROF has reliably demonstrated that the original CS-US association can be renewed in a change of context (context renewal) (Bouton, 2002), reinstated through an unsignalled presentation of the US (reinstatement) (Haaker et al., 2014b) or recovered with the passage of time (spontaneous recovery) (Quirk, 2002). These ROF phenomena suggest that the original CS-US association is not completely abolished during extinction and support an inhibitory account of learning.

#### 2.1.1.3 The latent cause model

The US and CS processing models are posited to be rivals (Dunsmoor et al., 2015), but a recent perspective developing from statistical learning models may reconcile the two models (Gershman & Niv, 2012). According to the latent cause model, conditioned responses do not arise from direct CS-US associations, but from some explanatory constructs (i.e., latent causes) which are linked with the US. Individuals' predictions about the US are mediated by their belief about which latent cause(s) is/are active at the time. These latent causes are not directly observed but can be inferred by using the Bayesian rule, which is expressed in the following equation:

$$P(\text{cause} | \text{observations}) \propto P(\text{observations} | \text{cause}) P(\text{cause})$$

In fear conditioning experiments, the model assumes that each trial is caused by one latent cause, and each latent cause encompasses some characteristic probability that engenders the observed features in the CS, US or context. Once the active latent cause is determined in the current trial, the model can predict the probability of the occurrence of the US, and generate the conditioned response appropriately.

Importantly, the latent cause model does not assume only unlearning or inhibitory learning. During extinction, an individual continues to learn about the association of each latent cause, leading to either an update of an existing latent cause or the creation of a new one. As such, this model has unified the key features found in the Rescorla-Wagner and Bouton models mentioned above. Despite a statistical perspective, the latent cause model can be understood as a process in memory retrieval: one attempts to match their current observation with the prototypes stored in their memories (i.e., the latent cause). If a match is found, the memory is updated to reflect the current observation; if a match is not found, the current observation is encoded into a new memory. The model also puts forth a prediction that different extinction procedures lead to a different balance between updating the original fear memory and/or the formation of new extinction memory. The latent model is relatively new, and the conditions governing the balance are yet to be explored empirically.

In summary, decades of research in extinction has yielded several important mechanisms to account for the extinction process. Traditionally, the Rescorla-Wagner prediction error model and Bouton inhibitory learning model of extinction are considered mutually exclusive; the latter has galvanised much empirical evidence and is the predominant model in explaining the return of fear. The latent cause model has unified the two accounts

for extinction, showing that extinction can simultaneously erase and inhibit previously learned CS-US associations, but further empirical evidence is still needed.

### **2.1.2 Extinction from a neural perspective**

The brain circuitry underlying fear conditioning and extinction has been extensively studied in laboratory animals and humans. One significant insight drawn from these research efforts is that the neural network of fear-associated learning is relatively well conserved across species (Greco & Liberzon, 2016; Maren et al., 2013a; Maren & Holmes, 2016): animal studies of the brain circuits that govern fear conditioning and extinction are generally corroborated by human imaging studies. Over the past few decades, a detailed account of the fear conditioning neural network has been identified by means of lesions, pharmacological, and electrophysiological studies (Feinstein et al., 2011; J. LeDoux & Daw, 2018).

At the heart of the fear conditioning circuitry is the amygdala, which consists of a group of heterogeneous nuclei for detecting and responding to threat. Specifically, sensory information concerning the CS arrives through the lateral nucleus (LA) of the amygdala, where the association between the CS and US is encoded by means of synaptic plasticity. The integrated information concerning the CS and the US then flows to the central nucleus (CE) of the amygdala, either through a direct LA-CE projection or through a group of inhibitory GABAergic cells known as the intercalated nuclei (ITC). From the central nucleus, the information is further projected to the brainstem and hypothalamus, where a host of automatic behavioural responses (e.g., freezing) and autonomic/endocrine defensive reactions (e.g., increase in skin conductance responses) is triggered (LeDoux & Daw, 2018; Rajbhandari et al., 2017; Schafe & LeDoux, 2000). In addition to supporting fear learning,

synaptic plasticity in the amygdala is also important for encoding extinction memories during extinction learning (Trouche et al., 2013).

Although there is ample evidence that synaptic plasticity in the amygdala is critical for conditioning, the amygdala does not act alone. The bidirectional communication between the amygdala and hippocampus is integral in the encoding and processing of the contexts associated with fear (Fanselow, 2000; Sparta et al., 2014). Both amygdala and hippocampal activation have been reported during conditioning, suggesting its role in the acquisition of the conditioned response (Andreatta et al., 2015; Bach et al., 2011; Critchley, 2002). Moreover, it has been demonstrated that the activation of the amygdala and hippocampus typically decreases over the course of fear conditioning (Labar et al., 1998; Reinhardt et al., 2010).

In a large meta-analysis of fear conditioning experiments with 677 healthy participants, Fullana and colleagues (2016) reported a robust distributed activation during conditioning in several areas, including the anterior insular cortex (AIC), dorsal anterior cingulate cortex (dACC), dorsolateral prefrontal cortex, dorsal pons, dorsal precuneus, hypothalamus, thalamus, secondary somatosensory cortex, and ventral striatum. The ACC and insula are thought to be involved in the expression of the conditioned responses, and previous studies have demonstrated a more consistent activation in these areas over the course of conditioning. The role of the mPFC is still under debate, but its involvement is thought to play a role in anticipatory threat responses and the subjective appraisal of fear and anxiety (Kalisch & Gerlicher, 2014).

During extinction, many of the brain regions implicated in fear acquisition are activated (Fullana et al., 2018), along with additional regions such as areas of the prefrontal cortex (Dunsmoor et al., 2019; Milad & Quirk, 2012) and cerebellum (Kattoor et al., 2014).

In agreement with animal studies, amygdala activation shows a temporal reduction across extinction learning (Labar et al., 1998; Phelps, 2004).

The role of the ventromedial prefrontal cortex (vmPFC) in extinction has been examined extensively in the literature for its top-down modulatory control over subcortical structures such as the amygdala and hippocampus. Previous studies suggest that vmPFC activation is observed by the end of extinction learning, and its magnitude is correlated with the strength of the fear responses (Milad et al., 2007). Corroborative evidence has also been obtained using structural MRI, in which the cortical thickness of the vmPFC was positively correlated with extinction recall (Hartley & Phelps, 2010; Milad et al., 2005). The vmPFC-hippocampus network is also critical for encoding contextual information during extinction and recalling the safety signals acquired during extinction (Kalisch et al., 2006; Maren et al., 2013b; Milad et al., 2007; Orsini et al., 2011). In addition to the vmPFC, the extinction of conditioned responses may also involve the dorsolateral prefrontal cortex (dlPFC). The dlPFC, a region that is thought to underlie higher cognitive processes such as selective attention, working memory, beliefs and expectancies, showed robust activation during extinction in a meta-analysis (Fullana et al., 2018).

Collectively, functional neuroimaging studies of fear conditioning have been very useful in delineating the neural circuits of fear-associated learning and extinction. Extinction learning is associated with an intricate neural network that includes the amygdala, hippocampus, vmPFC, and dlPFC (Milad & Quirk, 2012). Importantly, aberrant activity in these brain regions has consistently been reported across various anxiety disorders (Greco & Liberzon, 2016; Milad et al., 2014). For instance, elevated fear learning and greater neural activation within the fear-associated network in response to spiders has been observed in individuals with arachnophobia (Schweckendiek et al., 2011); individuals with post-



traumatic stress disorder exhibit reduced activity in the vmPFC and hippocampus, but increased activity in the dACC during extinction recall and context renewal (Garfinkel et al., 2014). Moreover, extinction learning measured at the neural level has been shown to predict exposure therapy outcomes (Ball et al., 2017; Helpman et al., 2016).

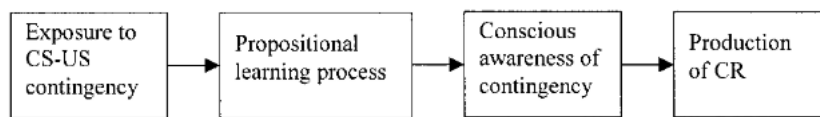
Taken together, both learning theories and neuroimaging studies have offered complementary information to understanding the intricate processes in fear extinction. Current research into the neural mechanisms may tend to support Bouton's extinction model because the neurocircuitry underlying fear extinction is distinct from fear conditioning. The precise neural circuits underlying the multiple pathways associated with return of fear require further investigation.

## **2.2 Awareness and fear conditioning**

Does information occurring outside of our awareness influence how we behave, think, and feel? For decades researchers have been trying to address this question in various fields within psychology. While it is widely accepted that non-conscious perception exists (Kouider & Dehaene, 2007), great controversy exists as to whether fear conditioning can occur outside of awareness. Currently, there are two main models accounting for the role of awareness in fear conditioning, the single-process model and the dual-process model. Proponents of the single-process model assert that there is only a single learning process in conditioning, and this learning process is propositional in nature (Lovibond & Shanks, 2002; Mitchell et al., 2009). Such a process requires a conscious effort in reasoning, produces conscious propositional knowledge, and elicits a conditioned response (Figure 2-1a). Conversely, proponents of the dual-process model postulated that conscious awareness does not have any causal role in the elicitation of conditioned responses (Clark et al., 2002; Wiens

& Öhman, 2002). Two independent learning processes could occur in parallel, a propositional learning process that gives rise to conscious awareness and a non-propositional process that leads to the production of conditioned responses (Figure 2-1b). In other words, production of the conditioned response is possible in the absence of awareness of the contingency between the CS and US.

(a)



(b)

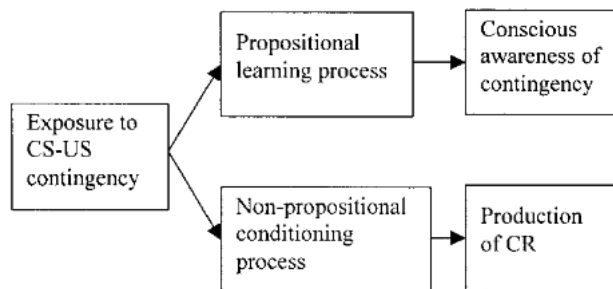


Figure 2-1 (a) A single-process model; (b) a dual-process model.  
 Source: Adapted from Lovibond and Shank (2002).

Critical to the understanding of the differences between these two models is the conditions under which learning occurs. First, the propositional approach assumes that learning involves hypothesis testing, and one would be aware of the CS-US contingency. In contrast, the dual-process model assumes that learning can take place in the absence of such awareness. Second, the single-process model posits that learning is effortful and dependent on the cognitive resources available at the time of learning, while the dual model regards

learning as an automatic process independent of cognitive resources. Third, the single-process model assumes that learning can be affected directly by verbal instruction, rules, and deductive reasoning, while the dual model assumes a minimal impact from these factors. In the sections below, I will review some relevant experimental paradigms and findings that have been employed to support each model.

### **2.2.1 Empirical evidence for the single-process model**

The main supporting evidence for the single process model comes from studies demonstrating a clear concordance between acquisition and contingency awareness. For instance, conditioned responses were only observed in participants who were aware of the CS-US contingency in a classical conditioning paradigm (Dawson et al., 1979). Using verbal instructions, Mertens and colleagues (2016) showed that by informing participants of the contingency between the CS and US in an instructed fear conditioning paradigm, participants' fear-potentiated startle responses would change accordingly regardless of their previous conditioning history, thereby demonstrating that propositional knowledge is required to form an association between a CS and US.

### **2.2.2 Empirical evidence for the dual-process model**

By contrast, there is ample empirical support for the dual-process model in fear conditioning experiments using techniques such as visual masking, binocular rivalry, and multiple CS presentations (Multi-CS conditioning). These methodologies and the associated findings are reviewed below as they serve as important background for the design of the experiments in the present thesis.

### 2.2.2.1 Visual masking

Visual masking refers to the condition in which the presence of one stimulus (e.g., a mask) affects the perception of the other stimulus (a target). Both forward and backward masks are commonly employed to prevent the perception of a target stimulus (Overgaard, 2015). By manipulating the timing of the presentation of the masks, the target stimuli can be suppressed from perceptual awareness. In backward masking, for instance, a target stimulus is first presented very briefly (e.g.,  $\leq 50$  ms; Kouider & Dehaene, 2007; Whalen et al., 1998), followed by a mask that shares similar features with the target stimulus at the same location. If the stimulus parameters are adjusted appropriately, observers would indicate being aware of the mask but not the preceding target stimulus.

Earlier visual masking experiments demonstrated that participants could acquire conditioned responses to masked stimuli and predict the occurrence of the US in a fear conditioning paradigm (Flykt et al., 2007; Katkin et al., 2001; Öhman & Soares, 1994). This differential conditioning, measured by skin conductance responses, is observed using fear-relevant stimuli such as pictures of spiders, snakes (Öhman & Soares, 1998) and angry faces (Olsson & Phelps, 2004). Subsequent studies have extended the investigation of visual masking on extinction learning. In a study conducted by Golkar and Ohman (2012), participants were first conditioned to faces with full perceptual awareness during acquisition, followed by extinction in which the CS+ was prevented from perception by masking. Importantly, participants did not show any significant differential fear responses at the end of extinction, suggesting that extinction learning might occur independently of perceptual awareness. A similar non-differential response to CS+ and CS- was also observed when masked pictures of weapons or animals were presented in the extinction phase (Flykt et al., 2007). Collectively, these visual masking studies provide evidence for unaware extinction

and pave the road for further investigation of extinction learning using other means of perceptual manipulations such as binocular rivalry.

#### 2.2.2.2 **Binocular Rivalry**

Under normal circumstances, our eyes perceive slightly different angles of the same image, and the visual system combines the two images to form a coherent percept via binocular fusion (Anderson & Nakayama, 1994); however, when two eyes receive different input at corresponding retinal locations, binocular rivalry occurs. The visual system cannot superimpose the two images into a coherent percept, and the resultant percept is one that alternates between the two images. Binocular rivalry is usually achieved by presenting different stimuli to each eye via a mirror stereoscope or a coloured anaglyph. Continuous flash suppression (CFS) is a variant of binocular rivalry that is a relatively newer experimental technique in the context of fear conditioning (Koch & Tsuchiya, 2007). CFS involves a process called dichoptic stimulation in which a stimulus is prevented from reaching awareness by presenting a strong dynamic noise in one eye relative to static images in another eye. Previous studies using this technique have shown that fearful faces are detected more quickly than neutral or happy faces (Yang et al., 2007).

Only three studies to date have employed the CFS as their masking procedure to study how fear is acquired without awareness in a fear conditioning paradigm (Mertens & Engelhard, 2020). Raio and colleagues (2012a) found that participants who viewed the CS with the CFS developed a greater skin conductance response to the CS+ compared with the CS- in the early acquisition phase. Similar learning was observed using fear-relevant (pictures of spiders) and fear non-relevant images (pictures of wallabies) (Lipp et al., 2014). These studies are in concert with studies using visual masking and provide evidence for fear

learning without perceptual awareness. A recent study has furthered the application of CFS in extinction learning. Oyarzun and colleagues (2019) examined the effect of awareness on extinction, where the conditioned stimulus (CS+) was suppressed from awareness using the CFS. Consistent with the hypothesis, they showed that fear responses could be modulated without conscious awareness.

The unconscious conditioning observation has been explored using functional imaging. Mounting evidence suggests that there is a distinct neural pattern in detecting threats in the absence of visual awareness. For instance, activation in the right amygdala was observed when participants viewed a masked, angry face (Morris et al., 1998; Whalen et al., 1998). In a meta-analytic study comparing the neuroimaging findings of subliminal and supraliminal stimuli presentation, Meneguzzo and colleagues (2014) reported that subliminal stimuli presentation is linked to increased activation in the right insula, the right fusiform gyrus and the anterior cingulate, where supraliminal stimuli presentation is linked to increased activation in the left rostral anterior cingulate. It is further proposed that recruitment of the right insula may support interceptive awareness, whereas engagement of the anterior cingulate may serve to integrate conscious and non-conscious processing.

#### **2.2.2.3 Multi-CS conditioning paradigm**

Multi-CS conditioning is a conditioning paradigm that manipulates awareness of the CS-US contingency. In this paradigm, many perceptually and physically similar stimuli (e.g., faces or tones) are either paired with one or multiple affective unconditioned stimuli (e.g., electric shock, aversive odour or sound), such that participants learn the CS-US associations under a very challenging condition. As a result, learning during acquisition occurs with strongly limited or even absent contingency awareness.

Conditioned responses as a result of conditioning were demonstrated in a series of electroencephalogram (EEG) and magnetoencephalography (MEG) studies (Brockelmann et al., 2011a; Rehbein et al., 2014; Steinberg et al., 2013). For instance, participants viewed a total of 312 faces in the acquisition phase, in which half of them were paired with an electric shock (CS+ faces) while the other half remained unpaired (CS- faces). Although participants could not differentiate whether the faces belonged to the CS+ or CS- category (as indicated by a low  $d'$  of 0.07), they showed increased prefrontal cortex activation towards the aversively paired CS+ as early as 50-80 ms following the presentation of the CS+ in the acquisition phase. Consistent with the neural findings, participants rated the non-paired CS- faces more pleasant than the shocked-paired CS+ faces in a post-learning behavioural task (Rehbein et al., 2014). Similar findings were observed in another Multi-CS conditioning study in which multiple faces were repeatedly paired with an aversive odour (Steinberg et al., 2012b) or when multiple natural tones were paired with aversive tones (Brockelmann et al., 2011a). Findings from Multi-CS conditioning studies have provided strong empirical support for the dual-process model, where propositional knowledge of the CS-US contingency may not be required in conditioning.

### **2.2.3 Single- or dual-processing model revisited**

The question over the role of awareness in conditioning remains unresolved and is still debated in the literature (Lovibond & Shanks, 2002; Mitchell et al., 2009). A recent meta-analytic study reviewing 30 empirical studies from 1970 to 2019 on conditioning independent of awareness suggested that there are significant publication biases and methodological limitations in the current literature. In the absence of quality evidence for dual or multiple systems of associative learning in conditioning, some authors argued that

‘there is very little to be lost, and much to be gained, by the rejection of the dual-system approach ...’ (Mitchell et al., 2009, p. 185).

It is worth noting that rejecting the role of awareness in conditioning might also be premature at this stage of science, given the reports of positive findings in physiological and neuroimaging studies. Outside the field of fear conditioning, contemporary theories of consciousness have advanced considerably in recent years and may shed light on this ongoing debate in fear conditioning.

Many of the contemporary theories of consciousness suggest a dissociation between conscious and unconscious processes in the brain that give rise to emotions such as fear. According to the Global Workspace Theory (Dehaene, 2011), domain-specific modules, which operate unconsciously in the cortical and subcortical regions of the brain, are attuned to the processing of a particular type of information, and a cortical global workspace connects multiple modules to give a subjective conscious experience. In the higher-order theory (e.g., Lau & Rosenthal, 2011), subjective conscious experience is based on both early sensory representations and a late-stage *re*-representation in the prefrontal cortex. How consciousness arises is still debatable; it is generally agreed among consciousness researchers that conscious experience is not completely independent from its unconscious physiological processes.

This has relevance for our conceptualisation of fear. Fear has been considered as a product of cortical circuits that underlie working memory and other cognitive functions, as well as subcortical circuits that control physiological responses and defensive behaviours (LeDoux & Pine, 2016). While the subcortical circuits operate without awareness, the cortical circuits receive inputs from the subcortical circuits, integrate them, and form the conscious feeling of fear (Figure 2-2).



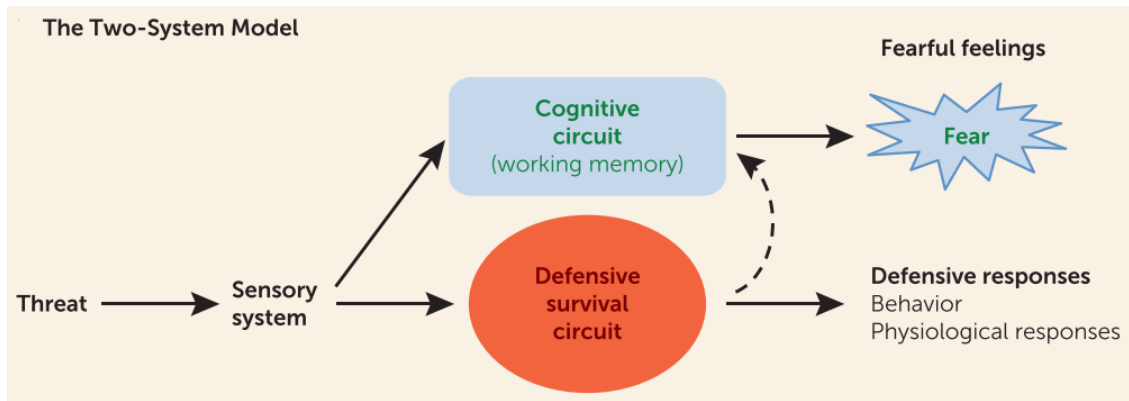


Figure 2-2 A Two-system model.

Source: Adopted from Ledoux and Pine (2016)

Consistent with the notion of dissociation between conscious and unconscious processing, a body of literature has shown that stimulus presentations outside of awareness increase physiological reactivity and affect conscious behavioural outcomes. There are now ample functional neuroimaging studies to suggest that subliminal visual stimuli increase activation in the physiological responses (Lipp et al., 2014; Raio et al., 2012a; but Hedger et al., 2016) and in various brain regions (Fang et al., 2016; Troiani & Schultz, 2013). For instance, combining CFS and fMRI, Troiani and colleagues (2014; 2013) demonstrated that unseen fearful faces resulted in greater amygdala and left parietal activity than unseen houses, as well as increased connectivity between the amygdala and multiple regions involved in the attention network including the bilateral pulvinar, bilateral insula, left inferior parietal sulcus, left frontal eye fields, and early visual cortex. Robust amygdala activation has also been observed in numerous fMRI studies comparing masked fearful stimuli to other neutral stimuli in both healthy and clinical populations (Diano et al., 2017; Ye et al., 2014).

Interestingly, several neuroimaging and behavioural studies have provided strong evidence for the impact of subliminal perception on behavioural outcomes (Gayet et al., 2016; Gomes et al., 2017). For instance, facial stimuli that are made invisible by the CFS

induce a congruency effect on a subsequent facial discrimination task or reaction time task (Lau & Passingham, 2007; Ye et al., 2014) and a reduced activity in the fusiform face area of the inferotemporal cortex (Kouider et al., 2009). Furthermore, subliminally perceived information may add to consciously visible information and facilitate subsequent decision making. In a perception decision task using intraocular suppression (Vlassova et al., 2014), participants were more accurate in judging the direction of the moving dots when the suppressed stimulus contained information that was congruent with the consciously perceived stimulus. Other empirical studies have shown that unconscious information can facilitate or influence higher cognitive functions such as mental arithmetic (Karpinski et al., 2019; Sklar et al., 2012) and response inhibition (Gaal et al., 2010; Parkinson & Haggard, 2014).

Taken together, a plethora of evidence supports the influence of unconscious processing on behaviours, perceptions, and cognitions. Furthering the two-system model of fear proposed by LeDoux (2016), targeting unconscious processes may affect higher-order circuits and the conscious feeling of fear. In other words, clinical interventions designed to target the unconscious processes in fear learning and extinction may ultimately reduce fear and anxiety at the conscious level. The present thesis adopts a view of fear that is line with the two-system framework and with this, investigates the impact of unconsciousness on extinction learning.

### **2.3 Memory reconsolidation**

The ability to maintain fear-related learning over long periods of time is an adaptive feature of the human memory system, and yet, memory is dynamic; it is constantly being updated during retrieval (Sara, 2008). The process in which memories are strengthened and

stabilised is known as memory consolidation. Decades of research from laboratory studies of human memory processes have confirmed that long-term memory is not instantly formed at the time of an experience but involves a gradual process at the cellular and system levels (Mcgaugh & Mcgaugh, 2012). At the cellular level, consolidation refers to the intracellular processes such as transcriptional activation and de novo protein synthesis, by which a cascade of molecular and cellular activities take place and give rise to lasting changes in the structure or function of a neuron in order to store the information (Alberini, 2008; Bisaz et al., 2014). At the system level, consolidation refers to the intercellular and inter-regional processes by which the activity in one brain area can influence that of another for storage of the information (Wang & Morris, 2010). Memories that are initially dependent upon the hippocampus undergo reorganisation and may become hippocampal-independent (Dudai, 2004).

Supports for the consolidation hypothesis of memory came from the research on retrograde amnesia in the late 1890s. Based on the observation from clinical patients with closed head injuries or other insults to the brains, French psychologist Théodule-Armand Ribot formulated '*loi de regression*' (the law of regression) in 1881, which held that new memories were more vulnerable to forgetting than old memories. A number of studies investigating retrograde amnesia have concluded that the formation of memories is a time-dependent process: short-term memories are formed within seconds to hours after the learning episode while long-term memories take hours to months to establish (Mcgaugh & Mcgaugh, 2012). Memories traces can also re-organize in the brain and this system consolidation is believed to last for weeks or even years in humans (Squire et al., 2015).

### **2.3.1 The origin of the reconsolidation hypothesis: the discovery and re-discovery**

For decades, memory researchers assumed that once the consolidation process is completed, memories are permanent and impervious to further interference; however, it was later discovered in the 1970s that amnesic agents not only impair the formation of new memories but also consolidated memories under certain circumstances. Misanin and colleagues (1968) were among the earliest researchers to report this observation, and their studies are considered forerunners of the reconsolidation hypothesis (Sara, 2008). Using a fear conditioning paradigm, they found that water-deprived rats showed more profound amnesia if they received a reminder of the conditioning before they received an electroconvulsive shock (ECS) relative to rats who did not receive a reminder. Based on this observation, Misanin and colleagues concluded that the state of the memory trace at the time of ECS administration was a primary determinant of the amnesic effect, which challenged the assumption of the consolidation hypothesis that consolidation occurs only once and that the consolidated memory trace is impervious to further disruption. Their finding was replicated by Terry and Holliday (1972), but the contemporary zeitgeist concerning memory loss and initial consolidation tempered the impact of the findings of Misanin et al. and others during this time. Research on memory reactivation and reconsolidation remained dormant for a few decades.

Interest in studying the labile nature of memory was rekindled by the work of Nader, Schafe, and LeDoux (2000), as well as Przybylski and Sara (1997) in the late 1990s. In their seminal study, Nader and colleagues (2000) showed that infusion of anisomycin, an inhibitor of protein synthesis, into the lateral amygdala shortly after reactivating the memory trace impaired the retention of a learned fear response in a group of fear-conditioned rats. In contrast, the infusion of anisomycin did not have any effect on the learned fear response six

hours after reactivating the memory. This finding provides compelling evidence for the lability of post-retrieval memory. Specifically, fear memories, once activated, undergo de-novo protein synthesis in order to persist and remain accessible at later times.

Research into memory reconsolidation has surged in the past two decades. The effect of amnesic agents on reactivated memories has been demonstrated in a range of animal models including crabs, snails, honeybees, chicks, mice and rats (Morris et al., 2006; Pedreira, 2004; Winters et al., 2009). Monfils and colleagues (2009) conducted a study in rodents to test whether extinction training after a brief reactivation might lead to direct integration of the extinction learning into the reactivated memory trace, a procedure known as retrieval-extinction. They found that rodents which underwent extinction training shortly after a brief reactivation showed a more persistent attenuation of fear responses relative to those who received extinction training alone when they were placed in a new context (renewal), exposed to the unconditioned stimulus (reinstatement), and after the passage of time (spontaneous recovery). The findings of this study are noteworthy; they suggest that the retrieval extinction procedure may potentially outperform the standard extinction training and provide substantial benefits for treatments of disorders associated with fear.

### **2.3.2 Reconsolidation in humans: laboratory studies**

Following Monfil's study, Schiller and colleagues (2010) provided the first evidence in humans that fear memories can be updated with a behavioural approach. In this seminal study, participants were conditioned to fear a coloured square and then underwent extinction in which all coloured squares were presented without a shock. Their results showed that participants who underwent the extinction training 10 minutes after a reactivation cue (i.e., within the reconsolidation window) showed no recovery of fear at a follow-up reinstatement

test relative to those participants who underwent extinction six hours after the reactivation cue (i.e., outside the reconsolidation window). Encouragingly, this reduction of fear responses lasted a year later (Schiller & Phelps, 2011).

Since the initial research by Schiller et al. (2010), there have been a few successful demonstrations of this reactivation-related modulation of fear in humans. For instance, Oyarzun and colleagues (2012) successfully replicated the result of Schiller's study using an auditory aversive stimulus in the fear conditioning paradigm. Steinfurth and colleagues (2014a) extended the experimental paradigm using seven-day-old memories and found that old fear memories can also be modified if extinction training is conducted during the reconsolidation window. Agren and colleagues (2012) showed that extinction during reconsolidation prevents the return of fear and that the process is mediated by the basolateral amygdala.

It is important to note that disrupting the reconsolidation of fear memories using behavioural interventions has not always yielded consistent findings (Fricchione et al., 2016; Golkar et al., 2012; Klucken et al., 2016; Drexler et al., 2014; van Schie et al., 2017). Some research groups have employed pharmacological chemicals to interrupt the reconsolidation process in humans. For instance, Kindt and colleagues have consistently demonstrated that post-reactivation administration of propranolol attenuates fear memory in healthy participants (Kindt et al., 2009; Soeter & Kindt, 2012; Soeter & Kindt, 2011, 2015a) as well as the clinical population (Soeter & Kindt, 2015b). Propranolol is a beta-adrenergic receptor antagonist traditionally used for the treatment of hypertension; it has been demonstrated that direct infusion of propranolol into the amygdala disrupts the reconsolidation of reactivated fear memories in rats (Dębiec et al., 2011). Interestingly, the propranolol manipulation

appears to attenuate the startle fear responses only, while declarative knowledge of the contingency remains intact (Soeter & Kindt, 2010).

Although pharmacological administration of propranolol may facilitate post-retrieval reconsolidation, such an amnesic effect is not always replicated (Bos et al., 2014; Schroyens et al., 2017; Spring et al., 2015; Thome et al., 2016). The contrasting results obtained across laboratory and clinical studies in reconsolidation suggest that some conditions or factors may render the triggering of reactivation and the reconsolidation process ineffective. These factors, termed boundary conditions, have been discussed extensively in several reviews (Else & Kindt, 2017b; Monfils & Holmes, 2018; Treanor et al., 2017; Zuccolo & Hunziker, 2019). Boundary conditions include the retrieval procedure, cue specificity, characteristics of participants, age and strength of memories, specificity of response systems, etc. I will discuss the first two factors in more detail as they are related to the experimental design of the studies in the present thesis.

#### 2.3.2.1 Prediction errors

One critical component needed for reconsolidation to occur is a novel perception or experience that mismatches the original activated memory, i.e., a prediction error. In an elegantly designed study, Sevenster and colleagues (2013a) systematically manipulated the percentage of prediction error participants received during the reactivation trial. Their study demonstrated that fear memory becomes labile and modifiable only when the outcome of the reminder trial is unpredictable. In other words, fear memory does not enter a labile state when no new learning occurs in the memory retrieval session. Similar observations have also been reported in animal models using *Chasmagnathus* crabs (López et al., 2016; Pedreira, 2004). These studies illustrate one key aspect of memory reactivation: retrieval of a fear

memory may not be sufficient for inducing its destabilization and reconsolidation. Instead, a discrepancy that challenges the original perception is required in order to destabilize the memory trace.

#### 2.3.2.2 Cue specificity

On top of prediction errors, the duration of a reminder trial is also implicated in the reconsolidation process. The animal model of memory suggests that reconsolidation is not initiated if the reminder is too short, whereas extinction learning is engaged if the reminder is too long (Pedreira & Maldonado, 2003), for instance, in rats. When the reminder trial is presented substantially longer than the initial acquisition, extinction could be instantiated (Alfei et al., 2015). Hence, it is not merely the duration of the CS reminder trial, but its relationship to the duration of CS exposure during acquisition is important for inducing reconsolidation. The optimal duration of a reminder has not been studied in humans, but it is speculated to depend on the history of learning (Elsey et al., 2018). Moreover, the boundary between extinction and reconsolidation might not be as absolute and rigid as previously proposed. There is a limbo state during the transition from reconsolidation to extinction, where memory lability remains low and is not sensitive to interruption (Cassini et al., 2017; Merlo et al., 2014).

In addition to the timing of a CS reminder trial, the type of reminder cue is also related to the process of reconsolidation. Most reconsolidation studies in humans employ a CS-reactivation paradigm (Klucken et al., 2016; Oyarzún et al., 2012). For instance, a picture of a spider is repeatedly paired with a shock during conditioning, and the same picture is used as a reminder cue during reactivation. Evidence to date has suggested that a CS reminder cue distinctly reactivate the relevant CS-US memory while leaving other US-related memory



intact (Doyère et al., 2007; Soeter & Kindt, 2011). While this selectivity of memory reconsolidation may protect the integrity of memory as a whole, it poses barriers to translating the research findings to the clinical setting. Traumatic memories are often composed of an extensive associative memory network involving multiple CSs. For instance, a road traffic accident survivor may be fearful of cars resembling the one that hit her, the smell of gasoline, or airbags. Often it is difficult to determine the precise nature of the CS in the real world or replicate each of the original CSs in a therapy room. In light of this, a few studies have employed a US-reactivation paradigm (i.e., presenting a US reminder cue) and demonstrated similar effectiveness in preventing the return of fear (Liu et al., 2014; Luo et al., 2015; Thompson & Lipp, 2017).

Despite the boundary conditions mentioned above, a meta-analytic study reviewing fear conditioning studies in humans reported an overall positive small-to-moderate effect size (Hedges'  $g = 0.40$ ) in favour of post-retrieval extinction over the traditional extinction approach (Kredlow et al., 2016). Notably, the effect size of post-retrieval behavioural extinction is similar in magnitude to that of propranolol for the reconsolidation of fear memories (Kredlow et al., 2016; Lonergan et al., 2013). Overall, the positive effect size points to a promising therapeutic intervention that provides a potentially long-term cure for patients suffering from anxiety and fear-related disorders. Clearly, more research is needed to elucidate the boundary conditions for triggering the reconsolidation process.

### 2.3.2.3 Neural correlates of memory reconsolidation

Several functional imagining studies have been conducted to capture the neural correlates of memory reconsolidation. Of those that employ fear conditioning as their experimental paradigm, increased activation in the amygdala was consistently reported in the

non-reminded CS versus the reminded CS contrast during memory retrieval (Agren, Engman, Frick, Björkstrand, et al., 2012; Björkstrand et al., 2016; Schiller et al., 2013a). Increased activation in the vmPFC from early to late extinction for the non-reminded CS was also observed (Schiller et al., 2013a), suggesting a pattern of diminished frontal involvement during reconsolidation of a reactivated threat memory. There is also a reported decrease in the amygdala-vmPFC functional connectivity as extinction progresses when a fear reminder cue is presented (Feng et al., 2015). In a recent attempt to replicate Schiller's findings, Klucken and colleagues (2016) failed to observe a statistically significant neuronal change in the vmPFC and the amygdala regions but found increased activation in the orbitofrontal and middle frontal gyrus during early extinction when comparing the non-reminded CS with the reminded CS.

### **2.3.3 Reconsolidation in humans: clinical studies**

A few studies to date have applied post-retrieval extinction to clinical populations, including anxiety disorders, post-traumatic stress disorder, and substance use disorders. Preliminary clinical studies have produced mixed results, with some studies reporting a benefit of the retrieval extinction procedure (Björkstrand et al., 2016; Soeter & Kindt, 2015b; Telch et al., 2017), while others failed to find any benefits in diminishing pathological anxiety or addictive behaviours. For instance, arachnophobia patients displayed more approach behaviours to a virtually presented spider when the propranolol was administered within the reconsolidation window (Soeter & Kindt, 2015b). Telch and colleagues (2017) observed a similar finding in a group of patients with spider or snake phobia: individuals who received a 10-second fear reactivation procedure prior to exposure therapy elicited lower phobic responses relative to those who received the reactivation procedure after the session.

There is also a growing body of literature examining the reconsolidation of appetitive memories associated with substance use disorders. For example, Xue and colleagues (2014) demonstrated that post-retrieval extinction reduced cue-induced heroin cravings in a group of heroin addicts after the intervention, and the effect was maintained six months later. In a randomised clinical trial, retrieval extinction using a five-minute video consisting of smoking content prior to extinction training was found to reduce cravings and smoking behaviours in humans with nicotine addiction (Germeroth et al., 2017). Post-retrieval intervention is also employed for individuals with drinking problems. In a post-retrieval counterbalancing procedure, Das and colleagues (2015) showed that post-intervention drinking and liking of alcohol stimuli were reduced in a group of hazardous, beer-preferring drinkers compared to their control counterparts.

Despite some success in translating the science of memory reconsolidation into a clinical intervention, there is also evidence against the clinical application of reconsolidation. Specifically, patients with arachnophobia did not benefit from virtual exposure to a spider as a reactivation cue before they underwent a standard course of exposure therapy (Shiban et al., 2015b). In three studies using pharmacological blockades of memory reconsolidation in a group of patients with PTSD, there were no significant group differences in physiological responses or changes in clinical symptoms between the reactivation groups and placebo counterparts (Wood et al., 2015). Finally, Maples-Keller and colleagues (2017) tested the efficacy of retrieval-exposure in people with a fear of flying and found no significant differences in the clinical measures of fear of flying between the reactivation cue group and the control group.

Taken together, the clinical evidence of memory reconsolidation is equivocal at this stage. The theory of memory reconsolidation has changed the way we understand the long-

term storage of memory, underscoring the lability and plasticity of a retrieved memory; however, the aforementioned boundary conditions may constrain the updating of a memory. The extent to which this theory is applicable to clinical intervention remains unknown at this stage. A thorough investigation of the boundary conditions for reconsolidation is warranted.

## **2.4 Overview of the studies**

Given the evidence for memory reconsolidation in human studies is encouraging, the current thesis presents an investigation of post-retrieval extinction strategy to attenuate fear-related defence response using fear conditioning paradigms in humans. Specifically, the thesis aims to answer the following question:

“What are the neural and behavioural mechanisms underlying  
extinction of fear memories and its recovery?”

### **2.4.1 Methodological approach**

The present thesis is comprised of four experiments that follow a fear conditioning paradigm, which consists of three phases: acquisition, extinction, and a test of return of fear. In experiments 2 and 3, the impact of a reminder-extinction procedure on the return of fear was further examined by inserting a reactivation phase between acquisition and extinction.

Across the studies, two loud tones (female scream and male scream) were used as the unconditioned stimulus. The conditioned stimuli were geometric figures in Experiments 1 and 2 and neutral faces in Experiments 3 and 4. Pupillary responses (Experiments 1-3) and neural activity (Experiment 4) were measured as an index of fear response. The research questions of the studies are detailed below:

---

<i>Experiment</i>	<i>Research Questions</i>
1	Does implicit exposure to a conditioned stimulus attenuate fear-related defensive responses?
2	Can implicit exposure to a reminder cue before extinction attenuate the recovery of fear?
3	How does an explicit reminder cue modulate the return of fear in an implicit learning paradigm?
4	What is the impact of an explicit reminder cue on the neural mechanisms of extinction and return of fear?

---

## **Chapter 3**

### **Experiment 1: Explicit extinction modulates defensive responses more effectively than implicit extinction**

### 3.1 Introduction

Studies of Pavlovian conditioning have advanced our understanding of the processes underlying threat learning and extinction. In Pavlovian conditioning, an initially neutral conditional stimulus (CS, e.g., a coloured figure) acquires an association with an aversive unconditional stimulus (US, e.g., electric shock) such that after multiple pairings, the CS begins to elicit a conditional defensive response (CR) on its own. The CS-US association can later be extinguished by repeatedly presenting the CS in the absence of the US such that a new CS–no US association forms that inhibits the original CS-US association and the CR. Extinction is an important process in exposure therapies for anxiety and fear-related disorders (Craske et al., 2014; Milad & Quirk, 2012; Vervliet, Craske, et al., 2013). However, patients receiving exposure therapies need to subject themselves to their feared objects during the process of therapy, which causes significant distress and may even result in the refusal to take part in therapy. One plausible way of mitigating this problem is to expose patients to their feared objects outside of conscious awareness.

Growing evidence exists that conditioned associations can be acquired outside of conscious awareness (Golkar & Öhman, 2012; Ho & Lipp, 2014; Olsson & Phelps, 2004; Raes & Raedt, 2011; Raio et al., 2012a; Vieira et al., 2017); however, not all researchers accept this interpretation (Mertens & Engelhard, 2020). In the field of conditioning, studies often employed various forms of masking techniques (Golkar & Öhman, 2012; Lipp et al., 2014; Olsson & Phelps, 2004) or binocular suppression (Raio et al., 2012b) to limit stimulus awareness amongst participants. The results of these studies have shown that threat conditioning occurs even when the conscious awareness of the stimulus is prevented. If the learning of threat associations can occur implicitly, it is conceivable that extinction learning can also be acquired implicitly outside of awareness. In this manuscript, the term conscious

awareness is used to denote perceptual awareness, i.e. conditions where ones do not perceive the visual stimuli with full perceptual awareness. Here, the term is not used to denote contingency awareness which describes the relationship between CS and US.

Few studies to date have examined the feasibility of implicit extinction. Among the available and clinically-relevant studies, Siegel and colleagues studied extinction learning via backward masking. In a procedure they called very brief exposure (VBE), they presented images of spiders very briefly (33 ms) to a group of spider-phobic participants such that these images were seen without their awareness. Participants receiving VBE showed reduced avoidance of a live *tarantula* and self-reported fear of the spider at the end of the experiment (Siegel & Warren, 2013b, 2013a; Weinberger et al., 2011). This modulating effect on their fear was evident a year later. In a subsequent functional magnetic resonance imaging (fMRI) study of VBE, a reduction of the BOLD response within the right amygdala was observed, whereas the BOLD response of the ventromedial prefrontal cortex remained unchanged across extinction trials (Siegel et al., 2017). In another recent fMRI experiment, participants' threat responses were reduced through the reinforcement of neural activities associated with the threat outside of their consciousness awareness (Koizumi et al., 2017a). Taken together, these studies provided support for extinction that takes place independent of perceptual awareness.

Recently, Oyarzun and colleagues (2019) expanded the evidence for implicit extinction using a human threat conditioning model. In their experiment, participants received implicit extinction using a binocular rivalry technique called continuous flash suppression (CFS), where fearful faces were prevented from entering conscious awareness through the flashing of fast-moving colourful Mondrians (arrays consisting of multi-coloured, high-contrast rectangles) to their dominant eyes. In the test of spontaneous



recovery where fearful faces were presented again, the group that underwent implicit extinction showed a reduced threat-potentiated startle response to the CS+ compared with the group that received explicit extinction. While their findings suggest a subtle modulation in the affective system during the implicit extinction, the robustness of implicit extinction has yet to be tested in terms of other return of fear (ROF) phenomena, such as reinstatement.

In the fear-conditioning literature, extinction is considered not as an unlearning of the original CS-US association but rather a competition between the existing association and the new CS+-no-US association that emerges during extinction (Bouton, 2004). Following extinction, conditioned responses can return spontaneously with time (spontaneous recovery), with exposure to an unsignaled US (reinstatement), or a CS-US pairing (rapid reacquisition) or following a contextual or stimulus change (context/stimulus renewal) (Bouton, 2017; Craske et al., 2014; Haaker et al., 2014b). These processes are thought to explain relapse after successful exposure-based therapy for anxiety disorders (Vervliet et al., 2013). They also each serve as a crucial test for the strength of extinction learning and the recovery of extinguished defensive responses in the laboratory (Hermans et al., 2005a, 2006).

The present study aimed to investigate implicit extinction by testing its effects on reinstatement of fear and by operationalizing fear using with a novel dependent variable: pupillometry. Although pupillometry, skin conductance responses (SCR) and electromyography (EMG) can each be employed as read-outs of defence response in fear conditioning studies (Lonsdorf et al., 2017), SCR and EMG have limitations that restrict their use in a test of implicit extinction. Conditioned responses measured by SCR are dependent on awareness of the contingency between the CS and US (Hamm & Weike, 2005) and arousal ratings (Lonsdorf et al., 2017). The assessment of EMG requires a loud white-noise to elicit a startle reflex which can serve as a secondary US (Lissek et al., 2005) and the inclusion of

it can interfere with the expression of fear for the CS (Sjouwerman et al., 2016). We examined pupillary responses as a physiological index of learning because of its sensitivity and reliability in measuring defensive responses to emotionally arousing (Bradley et al., 2008) and aversive stimuli (Sirois & Brisson, 2014; Wiemer et al., 2014). It has been shown as a reliable and valid index of CR in fear conditioning experiments (Leuchs et al., 2017a, 2019). Crucially, pupillary responses are not contingent on conscious awareness (Sperandio et al., 2018; Spering & Carrasco, 2015). Despite the advantages of using pupillary responses as a physiological read-out of fear, few studies to-date have indexed threat response in fear conditioning studies and no study has yet used pupillometry within a test of implicit extinction.

In the present study, we paired two geometric figures ( $CS_{exp+}$  and  $CS_{imp+}$ ) with an aversive female scream (US) during acquisition. A third geometric figure was never paired with the US and served as a control stimulus ( $CS-$ ). During extinction, all CSs were presented without the US. To manipulate the awareness of the CSs, we employed continuous flash suppression to create the  $CS_{imp+}$  condition, whereas the  $CS_{exp+}$  and the  $CS-$  were presented with full perceptual awareness. In addition, we compared the post-extinction defensive responses and their effectiveness of extinction learning in a reinstatement test. Pupillary responses and participants' self-reports of CS unpleasantness were recorded. Our main hypothesis was that both explicit and implicit exposure to a threatened  $CS+$  during extinction can attenuate the defensive response. In addition, we hypothesized that, irrespective of the mean of exposure during extinction, defensive responses to the  $CS_{exp+}$  and the  $CS_{imp+}$  would recover following extinction after four unsignaled US presentations.

## 3.2 Methods

### Participants

To determine required sample size, we conducted a power analysis using G\*power 3.1 (Faul et al., 2009) based on Raio et al (2012) study with a similar setup, in which the effect size for a CS+/CS- difference was (Cohen's)  $d = 0.87$ . Setting alpha at .05, a sample size of  $N = 20$  was required to achieve 95% power. As less than 50% of the participants was likely to acquire a differential conditioned response following conditioning using skin conductance measure (Schiller et al., 2010a), we recruited a total of 59 participants to allow for data analysis and attrition.

Fifty-nine participants with normal or corrected-to-normal vision, as well as normal hearing, were recruited for this study. Participants were excluded if they reported any current or historical psychiatric or neurological illnesses. Thirty-five participants were eliminated from the statistical analysis because they did not acquire threat conditioning as assessed by their differential pupillary responses to the CS+ and the CS- during acquisition. That is, they were excluded if their pupillary responses to the CS- were greater than either of the two CS+s. The final sample consisted of 24 participants ( $18.67 \pm 1.49$  years, Male:Female = 5:19). All participants signed the written informed consent forms, and course credits were given for their participation. All procedures performed were approved by the Human Research Ethics Committee of the University of Hong Kong, in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments.

### Stimuli

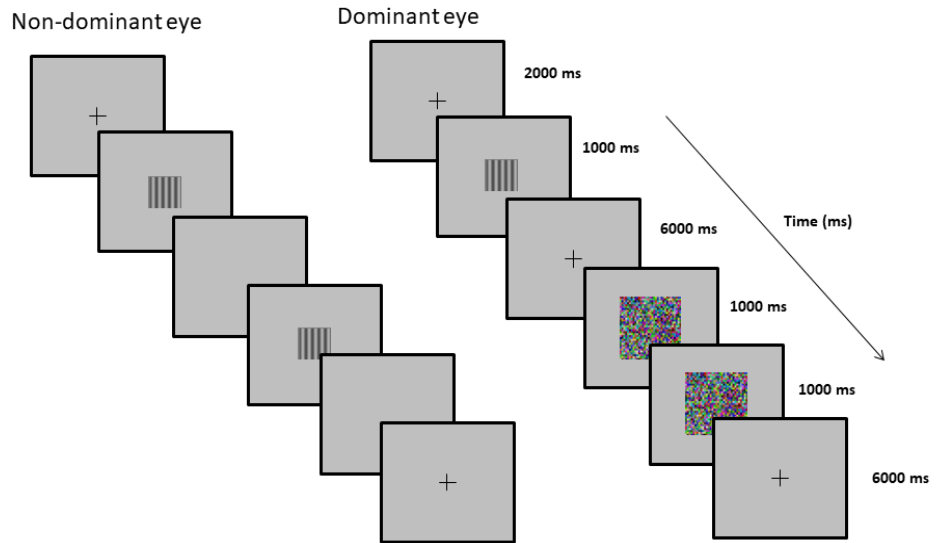
Three gray geometric figures (a square, a circle, and a diamond; 600 x 600 pixels) served as CS. We adjusted the brightness and the contrast of the figures, as well as the

background of the screen so that the luminance of the screen and the figures were of equal luminance. The CSs were presented over a gray background on two 17-inch computer monitors.

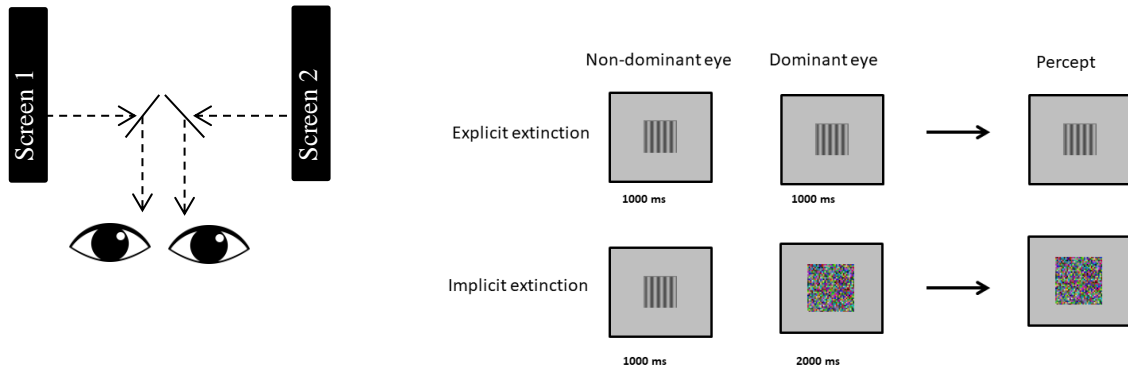
The US was a female scream presented for 1200 ms. The loudness of the US was normalized and resampled to 44100 Hz. It was presented at 90 db and was delivered through a stereo headset.

**Continuous flash suppression (CFS).** The CFS procedure was based on that of Brascamp and Naber (2016) using a dichoptic stimulation as follows. Two cold mirrors were placed at 45-degree angles relative to the participants' midlines to form a mirror stereoscope that allowed the views of two monitors to be projected onto participants' eyes. To render a stimulus implicit, we presented continuously flashing colourful Mondrians at a frequency of 10 Hz to participants' dominant eyes, as well as a stable low-contrast stimulus into their non-dominant eyes. Participants' ocular dominance was determined by the Hole-in-the-card Test (Dolman, 1919). The Mondrian was created using Psychtoolbox in MATLAB with reference to the CFS MATLAB toolbox (Nuutinen et al., 2017).

Figure 3-1 (a) Timeline and (b) percept of the extinction learning



(b)



**Measures**

**Questionnaires.**

*State-Trait Anxiety Inventory (STAI).* STAI “Form Y” consists of two 20-item subscales, namely the STAI-trait and the STAI-state subscales. The STAI-trait subscale measures relatively stable personal tendencies to experience anxiety symptoms, whereas the STAI-

state subscale evaluates the temporary anxiety symptoms associated with a specific situation or object. The responses were recorded on a four-point Likert scale. Total scores for both the STAI-state and the STAI-trait subscale were used. Higher scores indicated a greater level of anxiety.

**CS rating.** To index how the participants perceived the valence of the CSs after each phase, the participants rated the pleasantness of each CS on five-point Likert scales (1 *being pleasant* to 5 *very unpleasant*) after the acquisition, extinction, and re-extinction.

### **Physiological outcome**

**Pupillary response.** Pupillary responses were recorded at 250 Hz using a tower-mounted Eyelink 1000 plus (SR Research Ltd., Mississauga, Ontario, Canada). A three-point calibration procedure was conducted to locate the gaze position on the screen. The pupil data that the device reported were in arbitrary units. The data were transferred and preprocessed using MATLAB with in-house functions. First, blinks, defined as missing data, were linearly interpolated from 60 ms before the starting point until 60 ms after the endpoint of the blink. After interpolation, pupil responses were smoothed by a band-pass, third-order Butterworth filter between 0.02 Hz and 4 Hz. A low-pass filter removed high-frequency noise, whereas a high-pass filter decreased the basal slow drifts from the signals. All of the pupil data of each participant were z-transformed for further analysis and for comparison across participants. Furthermore, trials containing more than 50% of interpolated data points and pupillary responses greater than two standard deviations were removed (22 trials [13%] in acquisition; 23 trials [14%] in extinction; 13 trials [8%] in re-extinction). The baseline pupil diameter for each trial was determined by averaging the pupil diameter 500ms before the CS onset. To estimate the pupillary response related to the anticipation of the US, we set our window of

interest to 1.5 s – 2 s (i.e., 500 ms right before US onset) for locating the peak pupil size. The baseline-corrected pupil dilation was calculated by subtracting the average baseline from the maximum value identified in this window of interest.

## **Procedure**

Participants completed the questionnaires programmed in Inquisit 5 (Millisecond Software, Seattle, WA) on a computer, followed by the experimental conditioning task. The within-subject threat conditioning paradigm consisted of three phases: acquisition, extinction, and reinstatement and re-extinction.

**Acquisition.** Participants were familiarized with the experimental setup in a brief habituation phase, where six unreinforced CSs were presented in a random sequence. Immediately after habituation, CS<sub>exp+</sub> and CS<sub>imp+</sub> were presented six times co-terminating with the US, and two times without the US (i.e., a 75% partial reinforcement schedule). The CS- was presented eight times without the US. The CSs were presented for 4s with a 6-8s variable inter-trial-interval (ITI). The assignment of the CS and the first trial of each phase were counterbalanced across participants. The trial order within each phase was pseudo-randomized such that no CS was reinforced consecutively, at least one CS-US was presented before its corresponding CS–no US trials, and the first CS+ trial was always reinforced. Participants were instructed to pay attention to the mirrors in front of them and were informed that they would see some geometric figures and hear a scream occasionally. No explicit information was given on the contingencies between the CSs and the US. After Acquisition, participants completed the unpleasantness ratings of the CSs on a computer.

**Extinction.** Prior to the start of extinction, participants were instructed to recall the CS-US contingency in the acquisition session and no explicit instruction was given with respect to the change of CS-US contingency during extinction. After two practice trials, the participants underwent extinction. Only the Mondrian (i.e. no conditional or other stimuli) was shown in the practice trials in order to familiarize participants with the structure of this session. For implicit extinction trials, the Mondrian was presented for 1000 ms, followed by the onset of the CS<sub>imp+</sub> for another 1000 ms. The CS<sub>exp+</sub> and the CS<sub>-</sub> were presented to the participants without any interference from the CFS (Fig. 1). Manipulation checks were placed to determine the perceptual awareness of the conditioned stimuli and the suppression effect of the CFS. After each trial, participants were asked whether they had seen anything apart from the Mondrian (“*Did you see anything other than the Mondrians?*”), and if they answered “yes” to the first question, they were further asked to indicate which geometric figure (*square, diamond, circle, or other*) they saw. Participants completed two practice trials prior to the start of the extinction trials. Each CS was presented on the screen eight times for 1s with the same ITI as in Acquisition. All trials were presented without the US. After extinction, participants completed the unpleasantness rating of the CSs on the computer.

**Reinstatement and re-extinction.** Participants received four unsigned US presentations, followed by eight presentations of each CS without the US. The stimuli were presented for 4 s with a 6- to 8-s ITI. After re-extinction, they completed the unpleasantness ratings for the CSs.

## **Data analysis**



To compare the defensive responses of each CS, we employed linear mixed models (LMMs) with fixed and random effects in our analyses. LMMs were preferred because they were more robust to violations of the independence assumption among the data points, as well as more accurate estimates of the effects in the presence of random errors (Singmann & Kellen, in press). LMMs can also accommodate missing data in the sample and improve statistical power (Baayen et al., 2008).

To test whether *CS type* was an important parameter predicting pupillary responses, we compared the model with *CS type* as a fixed effect against a simpler model without this parameter. If *CS type* significantly improved the model fit, it would be included to estimate the effect of *CS type* in predicting the pupillary responses. Estimated marginal means (EMMs) of each *CS type* were computed to infer the statistical significance of the differences between *CS types*. Each comparison was adjusted for multiple comparisons using the Tukey method. Pupil data were averaged across the last two trials of the acquisition phase to evaluate threat learning and across the first two trials of re-extinction to infer the effect of extinction after the reinstatement procedure in the LMMs. We averaged the two trials in each phase for the analysis so as to retain more observable data values to infer the learning in each phase of the experiment. We did not directly compare the baseline-to-peak pupil responses of the CSs in extinction because CSimp+ was presented under the CFS and the length of the stimulus presentation was shorter during extinction than the acquisition and re-extinction sessions.

We computed the percentage of fear recovery using the following formula:  $100 \times \text{mean}(\text{first } 2\text{CS}^+ \text{ retention}) / \text{max}(\text{CS}^+ \text{ acquisition})$  to infer the effectiveness of extinction learning. A LMM consisting of *CS type* as the fixed effect, and participants as the random

effect was built to compute the differences of this index. A follow-up analysis was carried out to obtain the EMMs for each CS type.

To further explore the impact of perceptual awareness on their pupillary responses, we further separated these CS<sub>imp+</sub> trials into two conditions: *unaware* (trials in which participants could not detect the CS<sub>imp+</sub>) and *aware* (trials in which participants were conscious of the CS<sub>imp+</sub>). We examined the effect of the two conditions on the pupillary responses in the LMM. In this model, the fixed effect was awareness (*unaware* vs *aware*), and random effects were the variability of each participant at the intercept for the fixed effect.

To compare the difference between CS ratings, we examined the effect of *CS type* (CS<sub>imp+</sub>, CS<sub>exp+</sub>, and CS-) as the fixed effects, and participants as the random effect in the LMMs, followed by EMMs to delineate the contrast between CSs.

We performed the analyses in R 3.5.2 using the *lmerTest* (Kuznetsova et al., 2017), *emmeans* (Lenth et al., 2018), and *psych* (Revelle, 2019) packages.

### 3.3 Results

The results of the full sample (N = 59) were presented in Appendix A for references. The rest of the results were focused on the subset of the sample who demonstrated acquisition of threat in their pupil responses (n = 24). Table 3-1 depicts the main results of the CS type in the LMMs, including coefficient estimates (beta), standard error, t statistics, and significance level. Table 3-2 shows the estimated marginal means of the pupillary responses and unpleasantness rating of CSs, their standard errors, and CIs in each experimental phase. The mean state and trait anxiety of participants were 39.79 (SD = 6.29) and 49.08 (SD = 7.43), respectively.

Table 3-1 Result summary: Coefficient estimates, Standard Error,  $t$  statistics, and significance levels  $p$  for all predictors in the acquisition and re-extinction phase. Significant beta values suggest that pupil responses of the corresponding CS type were significantly different compared to those of the implicit CS+ (as the intercept).

Outcome	Phase	Parameters	$\beta$	SE	95% CI		$t$	$p$
Pupil response	Acquisition	intercept	0.84	0.12	0.60	1.08	7.03	<b>0.00</b>
		CS <sub>exp+</sub>	-0.07	0.14	-0.22	0.35	0.47	0.64
		CS -	-0.38	0.15	-0.67	-0.09	-2.57	<b>0.01</b>
	Re-extinction	intercept	0.71	0.11	0.48	0.93	6.25	<b>0.00</b>
		CS <sub>exp+</sub>	-0.29	0.15	-0.58	0.00	-1.97	<b>0.05</b>
		CS -	-0.06	0.15	-0.34	0.23	-0.39	0.70

Table 3-2 Estimated marginal means of the pupil responses and unpleasantness rating of CSs, their standard errors and confidence intervals in each experimental phase

Outcome	Phase	Type	EMMS	SE	95% CI	
<i>Pupil responses</i>						
	Acquisition	CS <sub>imp+</sub>	0.84	0.12	0.60	1.08
		CS <sub>exp+</sub>	0.91	0.12	0.67	1.14
		CS -	0.47	0.12	0.22	0.71
	Re-extinction	CS <sub>imp+</sub>	0.71	0.12	0.48	0.94
		CS <sub>exp+</sub>	0.42	0.12	0.19	0.65
		CS -	0.65	0.12	0.43	0.88
<i>CS unpleasantness rating</i>						
	Acquisition	CS <sub>imp+</sub>	3.04	0.24	2.57	3.51
		CS <sub>exp+</sub>	3.21	0.24	2.74	3.68
		CS -	1.71	0.24	1.24	2.18
	Extinction	CS <sub>imp+</sub>	2.23	0.26	1.70	2.76
		CS <sub>exp+</sub>	2.00	0.26	1.47	2.53
		CS -	1.73	0.26	1.20	2.26
	Re-extinction	CS <sub>imp+</sub>	1.96	0.17	1.61	2.30
		CS <sub>exp+</sub>	1.83	0.17	1.45	2.17
		CS -	1.17	0.17	0.83	1.52

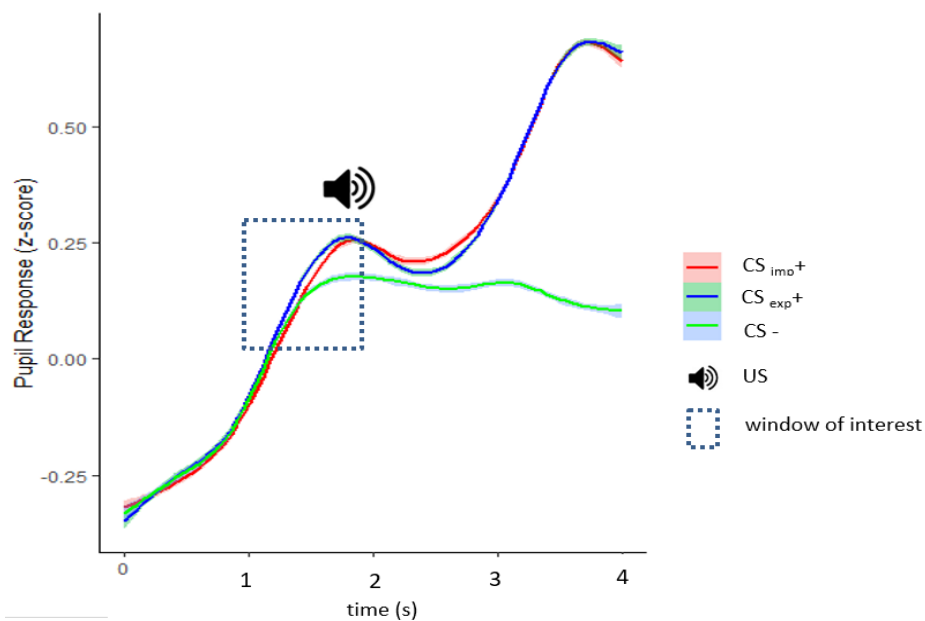
## Acquisition

**Pupillary responses.** Compared with the null model without *CS type* as a predictor, the mixed linear model with *CS type* affected pupillary responses significantly at the end of acquisition,  $\chi^2(2) = 10.41, p = .005$ . A follow-up analysis comparing the EMMs of each CS revealed that the  $CS_{exp+}$  elicited greater pupillary responses compared with the  $CS_-$ ,  $t(102) = 3.06, p = .008$ , and the  $CS_{imp+}$  also induced greater pupillary responses than did the  $CS_-$ ,  $t(103) = 2.54, p = .033$ . There was no significant difference in pupillary responses between the  $CS_{imp+}$  and the  $CS_{exp+}$ ,  $t(103) = -0.46, p = .888$ . Fig. 3-2 illustrates the mean pupillary changes of the last block for each type of CS.

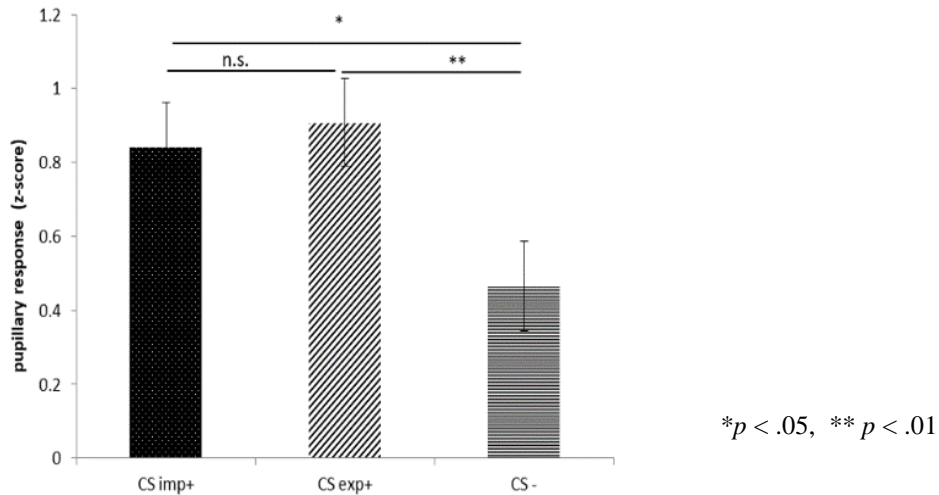
Figure 3-2 (a) Average change in pupil diameter in response to  $CS_{exp+}$ ,  $CS_{imp+}$  and  $CS_-$  across trials of threat acquisition phase with 95% confidence intervals. (b) Baseline-corrected pupillary responses in threat acquisition.

*Note:* The stimulus was present from 0 to 4 seconds; US administration occurred at 2s in  $CS+$  trials. For assessing pupil response per trial, pre-stimulus baseline average was subtracted from the maximum pupil diameter in the 1 to 2 seconds (marked as “window of interest”) before the US onset.

(a)



(b)



**Self-report CS rating.** After acquisition, significant differences were found in the participants' unpleasantness ratings between the CSs, ( $\chi^2(2) = 29.12, p < .001$ ). Consistent with our hypothesis, the participants rated the CS<sub>imp+</sub> and the CS<sub>exp+</sub> as being more unpleasant than the CS<sub>-</sub> (mean difference between the CS<sub>imp+</sub> and the CS<sub>-</sub> = -1.33, SE = 0.27,  $t(46) = 5.02, p < .001$ , Cohen's  $d = 0.95$ , 95% CI [0.31, 1.57]; mean difference between the CS<sub>exp+</sub> and the CS<sub>-</sub> = -1.50, SE = 0.27,  $t(46) = 5.65, p < .001$ , Cohen's  $d = 1.02$ , 95% CI [0.37, 1.65]). The CS<sub>imp+</sub> and the CS<sub>exp+</sub> were of equal unpleasantness on the scale,  $p = .806$ . Figure 3-4a shows the unpleasantness rating of the CSs after acquisition.

## Extinction

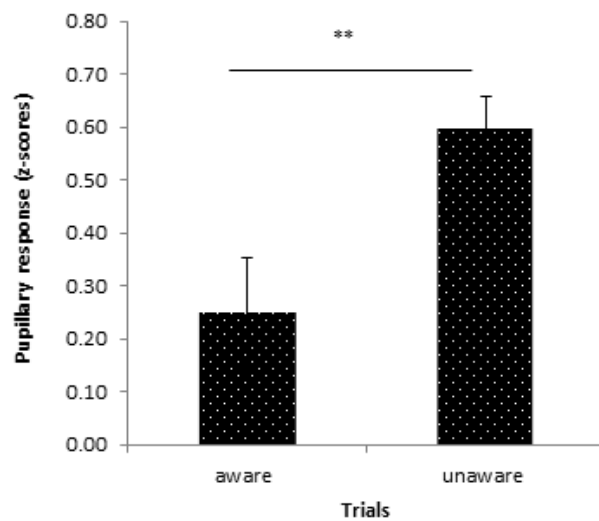
**CS awareness and pupillary responses.** Analyses of participants' awareness after each trial during extinction showed that some participants reported seeing something in addition to the Mondrians. Among the trials where the CS<sub>imp+s</sub> were presented under the CFS, 79.6% of the trials ( $n = 152$ ) were subsequently classified as *unaware*, whereas 20.4% of the trials ( $n = 39$ ) were deemed *aware*. The linear mixed model analysis revealed a significant effect

of awareness as a predictor on the pupillary responses compared with the null model,  $\chi^2(1) = 9.23$ ,  $p = .002$ . Further analysis comparing the pupil sizes in these trials indicated greater pupillary responses in the *unaware* trials compared with the *aware* trials (Table 3-3),  $t(113) = -3.08$ ,  $p = .003$ . Figure 3-3 shows the pupillary responses in the *aware* vs *unaware* trials, respectively.

Table 3-3 Coefficient estimates (beta), Standard Error,  $t$  statistics, and significance level  $p$  predicting pupillary responses in the extinction.

Phase	Parameters	$\beta$	SE	95% CI		$t$	$p$
Extinction	intercept	0.25	0.1	0.16	0.54	2.42	<b>0.02</b>
	unaware trials	0.35	0.11	-0.004	0.39	3.13	<b>&lt; .01</b>

Figure 3-3 Pupillary changes in the *aware* vs *unaware* trials during Extinction



\*\*  $p < .01$

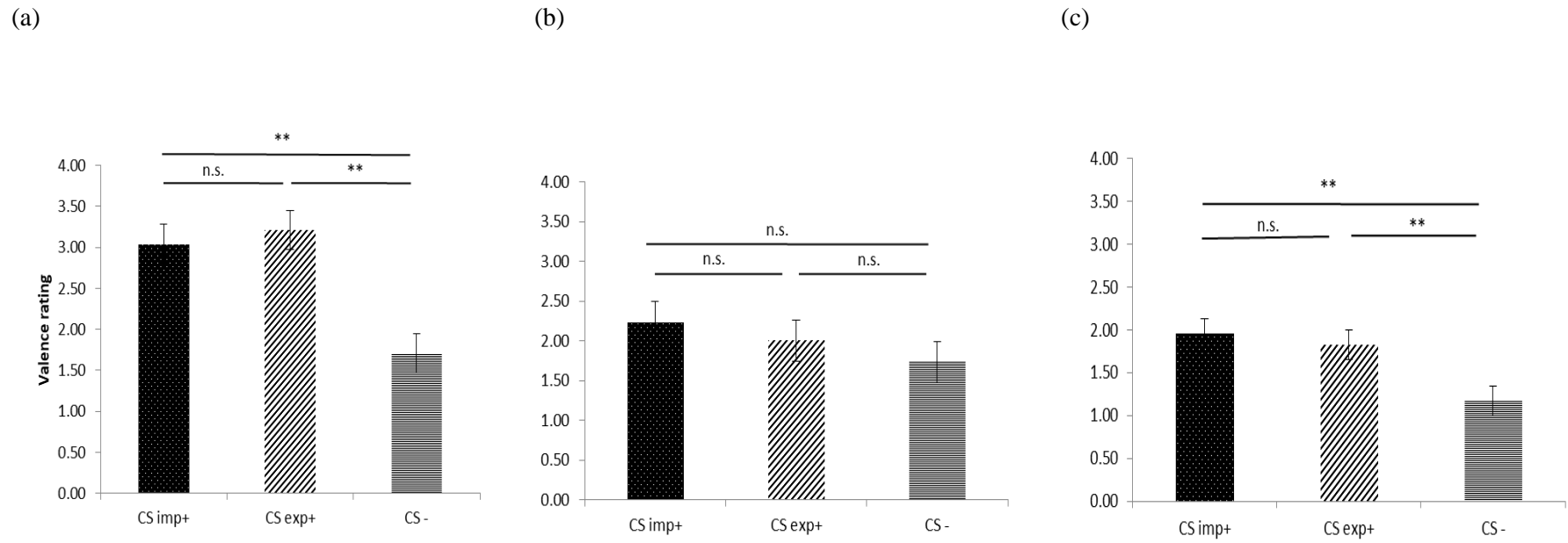
**Self-report CS rating.** After extinction, the mixed model with *CS type* as a fixed effect did not significantly differ from its null model ( $\chi^2(2) = 3.65, p = 0.162$ ), suggesting that no significant differences were found among the CSs with regard to their unpleasantness (Figure 3-4b).

### Reinstatement and re-extinction

**Pupillary responses.** Figure 3-5a illustrates the pupillary responses of each CS after the reinstatement procedure. Overall, the inclusion of *CS type* did not improve the model fit,  $\chi^2(2) = 4.32, p = .115$ , suggesting that there was no significant difference in the pupillary responses among the three CSs after reinstatement.

Interestingly, the inclusion of *CS type* in the model significantly predicted the recovery of fear ( $\chi^2(2) = 7.12, p = .028$ ). Specifically, the presentation of  $CS_{imp+}$  induced positive percentage of recovery of fear ( $\beta = 3.96, t[65.77] = 0.61, p = .530$ ) while the presentation of  $CS_{exp+}$  and  $CS-$  evoked a negative percentage of recovery of fear ( $CS_{exp+}: \beta = -22.86, t[43.89] = -2.72, p = .009$ ;  $CS-: \beta = -16.38, t[43.25] = -1.98, p = .055$ ). Furthermore, the percentages of fear recovery between the  $CS_{imp+}$  and  $CS_{exp+}$  were significantly different from each other (mean difference = 22.86, SE = 8.61,  $t(48.1) = 2.66, p = 0.028$ , Cohen's  $d = 0.67$ , 95% CI [0.03, 1.31]). In other words,  $CS_{imp+}$  evoked a higher percentage of fear recovery relative to the  $CS_{exp+}$  (Figure 3-5b).

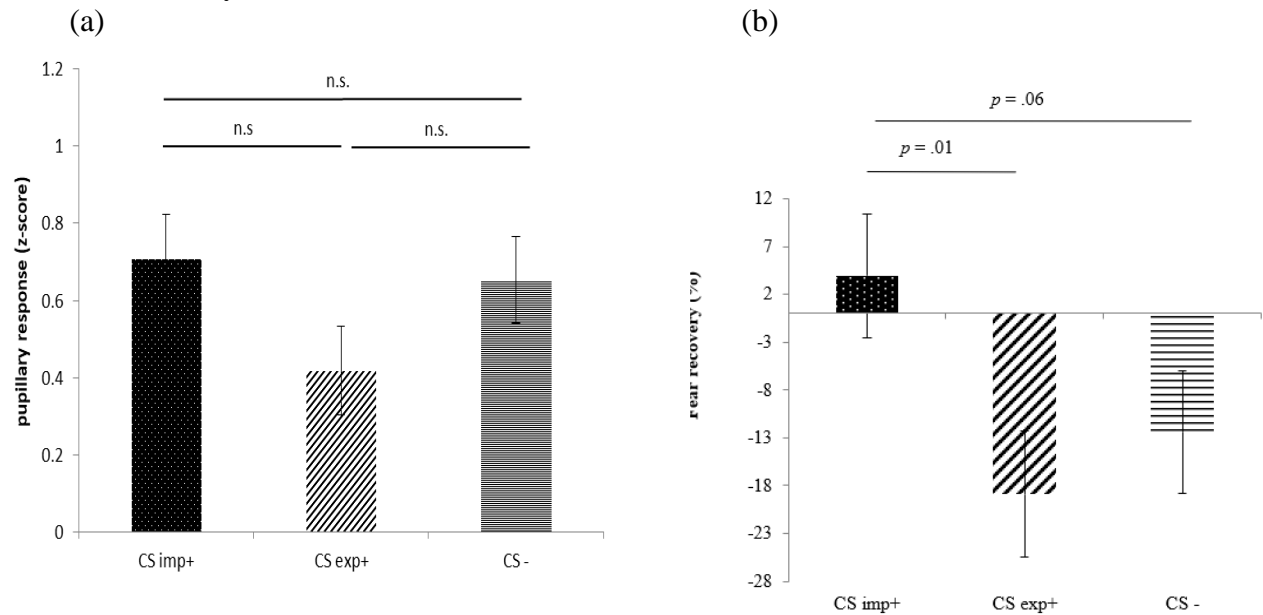
Figure 3-4 CS unpleasantness rating after (a) threat acquisition, (b) extinction, and (c) re-extinction



\* $p < .05$ , \*\*  $p < .01$



Figure 3-5 (a) Baseline-corrected pupillary response after reinstatement. (b) Percentage of fear recovery after reinstatement



**Self-report CS rating.** The unpleasantness rating of CS<sub>imp+</sub>, CS<sub>exp+</sub> and CS- differed significantly,  $\chi^2(2) = 16.14, p < .001$ . Participants rated CS<sub>imp+</sub> and CS<sub>exp+</sub> more unpleasant compared to CS- (mean difference between the CS<sub>imp+</sub> and the CS- = 0.78, SE = 0.20,  $t(48.1) = 4.01, p < .001$ , Cohen's  $d = 0.67$ , 95% CI [0.05, 1.28]; mean difference between the CS<sub>exp+</sub> and the CS- = 0.65, SE = 0.20,  $t(46) = 3.34, p < .005$ , Cohen's  $d = 0.78$ , 95% CI [0.15, 1.40]). There was no statistically significant difference between the unpleasantness rating of the CS<sub>imp+</sub> and the CS<sub>exp+</sub>,  $p = 0.783$  (Fig. 3-4c).

### 3.4 Discussion

The current study's aim was to investigate whether threat learning, operationalized in terms of pupillary responses, could be extinguished even in the absence of the awareness that the extinction of the conditioned threat is taking place. Our results show that both the implicitly-extinguished CS+ (CS<sub>imp+</sub>) and the explicitly-extinguished CS+ (CS<sub>exp+</sub>) evoked

similar pupillary responses to the CS- after reinstatement, but the percentage of fear recovery was greater for the CS<sub>imp+</sub> comparing with the CS<sub>exp+</sub>. These findings suggest that both explicit and implicit extinction may modulate defensive responses, though such modulation was weaker in implicit than explicit extinction.

The use of pupillary responses as an autonomic read out of threat learning and extinction is relatively new in the field of fear conditioning compared to other indices of fear such as SCR and EMG. A recent study comparing the conditioned response measured in skin conductance, pupillometry and startle electromyography suggested that skin conductance responses and startle responses habituated across learning, but pupillary responses did not (Leuchs et al., 2019). Hence the attenuated pupillary responses to the CS<sub>imp+</sub> and CS<sub>exp+</sub> after reinstatement in our study may not be explained by habituation of the autonomic responses. Rather, this result supports a case of implicit extinction where defensive responses are modulated unconsciously.

Our findings provide partial support for implicit extinction by Oyarzun and colleagues (2019). In their study, participants who underwent implicit extinction using CFS showed reduced threat-potentiated startle responses, but not skin conductance responses, to the threat-conditioned stimulus in a spontaneous recovery test. While we showed that both the CS<sub>exp+</sub> and the CS<sub>imp+</sub> evoked similar return of fear-related pupillary responses to the CS- after reinstatement, the percentage of fear recovery was higher for the CS<sub>imp+</sub> relative to the CS<sub>exp+</sub>. The percentage of fear recovery may represent a more reliable measure of extinction learning in our study because it takes into consideration the conditioned response during acquisition. Our findings neither invalidate previous findings nor do they suggest that implicit extinction does not exist. Rather, they suggest that implicit extinction learning is weaker than explicit extinction learning. The weak effect in implicit extinction is consistent

with the emerging view that unconscious processing is generally limited in scope (Raio et al., 2012a; Stein et al., 2020).

We followed a typical extinction retention calculation whereby mean CS+ responding during a retention test was divided by CS+ (max) responding during acquisition, and this score was used to index the strength of extinction learning (Li & Graham, 2016; Milligan-Saville & Graham, 2016). This approach is preferred in our study because we could not directly compare the pupillary responses among the stimuli due to the effect of the CFS during extinction. However, it is of note that many existing extinction retention calculations did not adjust for differences in the defensive responses during extinction (Lonsdorf et al., 2019). Lonsdorf and colleagues (2019) suggested that controlling for the differences in the conditioned responses during extinction is deemed necessary because extinction recall, according to Bouton's model of extinction (Bouton, 2002), is resulted from a constant competition between the extinction memory formed during extinction and the original memory formed during acquisition. While the current computation control of the differences in responding during acquisition, future research involving implicit extinction could consider experimental designs or incorporate other physiological measures that allow them to correct for the responding during extinction in computing the extinction retention.

Our work adds to the existing literature that pupillometry is also a sensitive psychophysiological marker for measuring defensive responses under implicit conditions. The physiology and the neural mechanisms of pupillometry are well studied. Expectation-induced pupillary responses are closely linked to the activation in the locus coeruleus, a subcortical structure that coordinates the noradrenergic system in the brain (Sirois & Brisson, 2014). In the conditioning literature, pupil responses to the CS+ and the CS- during fear conditioning have been shown to correlate with the activity of the dorsal anterior cingulate

(dACC) in the salience network (Leuchs et al., 2017b). The current evidence of pupil responses to the CSs is mainly drawn from the threat conditioning paradigm where the stimuli are presented with full awareness. Our work suggested that future research should examine the neural correlates of pupil responses during implicit extinction, with the goal of understanding how consciousness implicates in this process of learning.

Interestingly, when the  $CS_{imp+}$  was prevented from perceptual awareness via continuous flash suppression, participants' pupils dilated more in the trials in which they were unaware of the  $CS_{imp+}$  compared with those in which they were aware. We could not fully account for this observation, but our result suggests an early, subtle differentiation in stimulus processing that is, at least partly, contingent on the awareness of having seen the CS. Perhaps the trials that participants were unaware of were novel to them; they may be more arousing and require more mental effort to process what was seen and what was subsequently expected. Previous studies have shown that pupils dilate in response to increases in mental effort and arousal triggered by an external stimulus (Mathôt, 2018). Moreover, part of the pupil responses is voluntary and can be modulated by higher-level cognitions. In a study examining consciousness and pupil size, Sperandio and colleagues (2018) demonstrated that pupils constricted more when participants were aware of the content of pictures that were presented to them (e.g. pictures of the sun), suggesting that pupil responses could be modulated by the content of consciousness. Taken together, our results highlight the role of consciousness in modulating pupil responses. Future research could further investigate the mechanism underlying this unconscious process.

We observed a dissociation between self-report ratings and the pupillary change after re-extinction: whereas participants rated the  $CS_{imp+}$  and the  $CS_{exp+}$  as more unpleasant than the  $CS-$  was, their pupils responded similarly to the  $CSs+$  and the  $CS-$ . This discordance

between subjective ratings and psychophysiological responses may provide evidence for the two-system model of fear (LeDoux, 2014; LeDoux & Pine, 2016). In the two-system model, threats are expected to elicit both a non-conscious defensive response and a conscious pathway that gives rise to the feeling of fear. These processes interact but do not share the same pathways in the brain's fear system. In the current study, we used fear-irrelevant conditioned stimuli and non-clinical participants to examine the reinstatement effect after implicit extinction. More research is needed to elucidate the non-conscious process of threat learning and extinction using fear-relevant stimuli such that more effective interventions can be developed to modulate maladaptive threat responses in humans.

### **Limitations**

Three features of this work may limit the conclusions drawn regarding implicit extinction. First, the length of the trial windows in extinction was different from that in acquisition and re-extinction. We did so to enhance the suppression effect of the CFS in the extinction phase. Consequently, we could not directly compare the baseline-to-peak pupil responses of the same stimuli across the three experimental phases and our method of estimating the strength of extinction learning was different from previous studies (Schiller et al., 2012; Soeter & Kindt, 2011). Because of the CFS, we could not directly compare the pupil responses of the implicitly and explicitly viewed CS+ during Extinction as the pupil responses would be affected by the moving Mondrians of the CFS manipulation. Yet, we found evidence of the acquisition of conditional pupillary responses to the CSs+ compared with the CS-, and evidence of extinction in participants' subjective report on the unpleasantness rating of the CSs. Moreover, following previous research (e.g. Li & Graham, 2016; Milligan-Saville & Graham, 2016), we computed the percentage of fear recovery to

estimate the strength of extinction learning using the information in acquisition and re-extinction. Second, some participants were aware of the stimuli presented during extinction despite a relatively short window of CFS (2000ms) used in the present study. We chose this window length with reference to Oyazurin's study (5500 ms; 2019) and Yang's study (3000ms; Yang et al., 2007). Further research may overcome the CFS breakthrough by developing an individualized CFS threshold to achieve a better suppression effect. Third, due to practical constraints, acquisition, extinction, and the subsequent reinstatement test were completed in one day, a duration that might preclude the acquired threat memories from fully consolidating. However, there were reports on recovery effects measured by reinstatement even though extinction was conducted the same day as threat acquisition (e.g. Schiller et al., 2008). Researchers in future studies should consider conducting different phases on separate days to test the robustness of our findings. Lastly, the current study did not include a control condition to for the implicitly extinguished CS+, which may limit the overall findings of the study. Future study could consider including an additional CS+ without extinction or a CS- with CFS to directly compare the pupil responses during extinction.

### **3.5 Conclusion**

The present findings indicate that defensive threat responses can be modulated both explicitly and implicitly, but such modulation effect is weaker in implicit extinction. While the observed implicit modulating effect was weaker than the explicit pathway in the current study, these findings lead to important clinical implications for the treatment of fear-related and anxiety disorders. Patients receiving exposure therapies are required to repeatedly approach the fear-provoking stimuli explicitly, which causes significant distress to them. Implicit extinction can mitigate the initial fear that patients encounter when they undergo

exposure therapies. In addition to facing one's fear explicitly, clinicians can consider using implicit techniques such as a CFS-assisted extinction to lengthen one's exposure to fear-provoking stimuli in a therapy session. However, the long-term benefits of using implicit techniques need further examination as our data did not show robust support for the reduction of fear recovery. Clearly, additional empirical research is needed to maximize the implicit modulating effect by further understanding this implicit pathway of threat learning in humans and harnessing this pathway for the treatment of anxiety and fear-related disorders.

## **Chapter 4**

**Experiment 2: No statistical evidence for implicit and explicit reminder cues on reducing the reinstatement of fear in a retrieval-extinction threat conditioning paradigm**



## 4.1 Introduction

When memories are retrieved, they become labile and are subject to modification by a process known as memory reconsolidation (Nader et al., 2000; Przybylski & Sara, 1997). During this reconsolidation window, memories can be enhanced, maintained, or attenuated (Monfils & Holmes, 2018; Treanor et al., 2017; Ricco et al., 2006; Elsey et al., 2017). Research into the lability of reactivated memories has grown substantially in the past two decades, as it may present a promising avenue for reducing clinical relapse in fear-related and anxiety disorders.

Schiller and colleagues (2010a) provided the first evidence in humans that fear-related memories might be modified using behavioural extinction when the memory is retrieved and destabilized. Importantly, participants who received a reactivation cue before extinction training did not show recovery of fear after receiving the retrieval-extinction procedure. Since then, there are several successful demonstrations of reconsolidation-extinction in preventing the return of fear in humans. For instance, Oyarzun and colleagues (2012) directly replicated the results of Schiller's study using an auditory aversive stimulus in the fear conditioning paradigm. Steinfurth and colleagues (2014b) found that older fear memories could also be modified if the extinction training was conducted during reconsolidation. Using fMRI, Agren and colleagues (2012) showed that extinction during reconsolidation prevented the return of fear and that the process was mediated by the basolateral amygdala. A meta-analysis (Kredlow et al., 2016) of reconsolidation-extinction studies reported an overall positive small-to-moderate effect size (Hedges'  $g = 0.40$ ) in favour of retrieval-extinction over traditional extinction, supporting the modification of fear-related memory by means of reconsolidation.

Nonetheless, disrupting the reconsolidation of fear memories by behavioural intervention has not yielded consistent findings (Fricchione et al., 2016; Golkar & Öhman, 2012; Kindt & Soeter, 2013a; Drexler et al., 2016; Soeter & Kindt, 2011). The contrasting results obtained across laboratories suggest that there might be conditions or factors that render the reactivation and/or reconsolidation processes ineffective. These boundary conditions include the age and strength of memories, the type of reminders (CS or US), and the retrieval procedure (Treanor et al., 2017; Zuccolo & Hunziker, 2019). To our knowledge, no studies to date have evaluated the necessary perceptual characteristics of a reminder cue during memory reactivation.

Increasing evidence from neuroimaging and behavioural studies suggest that subliminal perception affects our thoughts, feelings, and behaviours (Gayet et al., 2016; Gomes et al., 2017). For instance, participants were more accurate in judging the direction of moving dots when the suppressed stimulus contained information congruent to the consciously perceived stimulus, suggesting that subliminally perceived information may add to consciously visible information and facilitate subsequent decision (Vlassova et al., 2014). Continuous flash suppression (CFS) is a form of binocular rivalry in which a highly salient and dynamic stimulus is presented to the dominant eye leading to a strong suppression of the stimulus presented to the non-dominant eye (Tsuchiya & Koch, 2005). Using this technique, Ye et al. (2014) demonstrated that facial stimuli that were made invisible by the CFS induced a congruency effect on a subsequent facial discrimination task. This congruency effect was also observed in the behavioural and neuroimaging study by Sid and colleagues (2009). Taken together, these studies provide empirical evidence that unconsciously perceived information may facilitate higher cognitive functions. Following this line of thought, it is

conceivable that a subliminally presented reminder cue before extinction might facilitate the retrieval of the original CS-US memory trace. Such implicit procedures might have an additional benefit for reconsolidation-based psychological interventions as it might reduce the initial discomfort associated with explicit recall of fear-related memories in a therapy session.

In the current study, we tested whether an implicitly presented reminder cue would activate and destabilize a memory for subsequent behavioural intervention. Using a Pavlovian conditioning paradigm, we conducted a retrieval extinction experiment using pupillary responses as the index of fear learning. We employed a mixed design, in which the memory of one of the two threat-conditioned stimuli (within-subject) was reactivated either implicitly or explicitly (between-groups) through a reminder cue before extinction learning. The experiment spanned three consecutive days: acquisition of the CS-US associations on Day 1, implicit/explicit reactivation of a CS, followed by standard extinction learning on Day 2, and a reinstatement test and re-extinction on Day 3. Subjective reports of the cognitive and affective aspects of fear learning were collected. Participants' self-reported trait anxiety levels were also measured as trait anxiety can modulate the extent of fear reduction in reconsolidation/extinction procedures (Soeter & Kindt, 2013). We hypothesized that pupillary responses and subjective reports of fear learning of the implicitly- and explicitly-reminded CS would be diminished following extinction, and the reinstatement of pupillary responses would be higher for the non-reminded CS relative to the reminded CS.

## 4.2 Methods

### Participants

Healthy participants with normal or corrected-to-normal vision and hearing were recruited for this study. We excluded participants with any self-reported, current or history of psychiatric or neurological disorders.

A power analysis to estimate the sample size for the current investigation was based on the effects evident in Schiller et al. (2010). Given that we expected conditioning to result in greater response to the CS that was paired with the aversive stimulus than the one without, a one-tailed t-test was used to estimate the sample size in G\*power 3.1 (Faul et al., 2009). Setting alpha at .05, a total sample of 36 would achieve an effect size of 0.5 with 90% power for a within-group comparison of mean differences. Based on the retentionrate (59%) observed in the Experiment 1 (Chapter 3), we recruited a total of 61 participants. Data from two participants were excluded because they discontinued participation after Day 1. The final sample consisted of 59 participants ( $22.20 \pm 4.72$  years, Male:Female = 15:44).

These participants were further randomised into two groups, namely the implicit reactivation group ( $n = 30$ ) and the explicit reactivation group ( $n = 29$ ). All participants signed the written informed consent forms and were compensated with course credits or money. All procedures were approved by the Human Research Ethics Committee of the University of Hong Kong (EA1709015), in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments.

## **Stimuli and measures**

### *Stimuli*

The conditioned stimuli (CS) consisted of a square, a circle, and a diamond (600 x 600 pixels), all in gray. We adjusted the brightness and the contrast of the shapes, as well as the background of the screen so that the luminance of the background and the shapes were equal. One of the conditioned stimuli served as a CS- and two served as CS+. All CS presentations were randomized among participants.

The unconditioned stimuli (US) were a female scream and a male scream (USa and USb, respectively). The loudness of the USs was normalized and resampled to 44100 Hz. It was delivered binaurally through headphones at 90db and lasted for 1.2 seconds. Both screams co-terminated with the conditioned stimuli.

### *Continuous flash suppression (CFS)*

Dynamic colourful Mondrians at a frequency of 10 Hz were projected to participants' dominant eyes, and the CS was presented to their non-dominant eye. Participants' eye dominance was identified by the Hole-in-the-card Test (Dolman, 1919) prior to the commencement of the main experimental procedure. The Mondrians were created using Psychtoolbox in MATLAB with reference to the CFS MATLAB toolbox (Nuutinen et al., 2017).

## **Measures**

### *Pupillary responses*

Pupillary responses were recorded at 250 Hz using a tower-mounted Eyelink 1000 plus (SR Research Ltd., Mississauga, Ontario, Canada). The experiment took place in a dark chamber. The outputs were transferred, preprocessed, and analysed using MATLAB (Version 2019b, Math-Works) and PsPM (Psychophysiological modelling, Version 4.2.1).

### *Subjective ratings*

**Likelihood rating.** Participants were asked to indicate their expectancy of experiencing the US associated with each CS after the acquisition, extinction, and re-extinction sessions. Specifically, they were asked to indicate how likely each CS would be followed by a scream on a 4-point Likert scale (1 = *not at all*, 4 = *very likely*). Their responses were recorded on a computer with *Inquisit 5* (Millisecond Software, 2005)

**Unpleasantness rating.** Participants were asked to rate the negative valence of each CS (“how unpleasant is the shape for you?”) after the likelihood rating. Offline rating was preferred as continuous online rating might boost the learning process itself and interfere with the process of learning (Lonsdorf et al., 2017). Answers were recorded on a 4-point Likert scale (1= *not at all*, 4 = *very unpleasant*) using *Inquisit 5*.

### *Questionnaire*

**State-Trait Anxiety Inventory** (STAI; Spielberger, Gorsuch, & Luschene, 1970). Participants completed the STAI-Trait subscale. It consists of 20 items that measure relatively stable personal tendencies to experience anxiety symptoms. Each item is rated on a 4-point Likert scale (from 1 = *almost never* to 4 = *almost always*), with scores ranging

from 20 to 80. Higher scores indicate greater levels of anxiety. The internal consistency of the STAI-T was high in the current study (Cronbach's  $\alpha = 0.92$ ).

### **Procedure**

The three experimental sessions took place across three consecutive days with a 24-hour interval between each. Before the experiment began on day 1, participants signed the informed consent, completed the questionnaires and the eye-dominance test. For all sessions, we did not instruct participants about the CS-US contingency. The experimental procedure (Figure 4-1) is detailed as follow:

**Day 1: Acquisition.** CSs were presented over a gray background on two 17-inch computer monitors using a dichoptic mirror stereoscope. Each CS was presented two times in a random sequence for participants to familiarize themselves with the experimental setup. Immediately following this habituation process, two of the CS+s were presented six times each, co-terminating with one of the USs and two times without the US (a 75% reinforcement schedule). A CS- was never paired with a scream. Each CS presentation lasted for four seconds and was separated by a white fixation cross during the inter-trial interval, which ranged from six to eight seconds. The assignment of the CS to each condition was counterbalanced across participants.

**Day 2: Reactivation and Extinction.** Twenty-four hours after the conditioning, participants were assigned into two groups. One group viewed a reminder cue for one CS+ with full perceptual awareness (i.e. explicit memory reactivation) and the other group viewed a

reminder cue for another CS+ with the CFS (i.e. implicit memory reactivation). In both conditions, the CS reminder cue was presented twice to reactivate the memory trace of associative learning. A manipulation check in relation to the perceptual awareness of the CS was placed following each reactivation trial. Specifically, participants were asked whether they had seen anything apart from the Mondrians (“*Did you see anything other than the Mondrians?*”), and if they answered “yes” to the first question, they were further asked to indicate which geometric shape (a *square, a diamond, a circle, or other*) they saw. Participants completed two practice trials before the start of the reactivation trials.

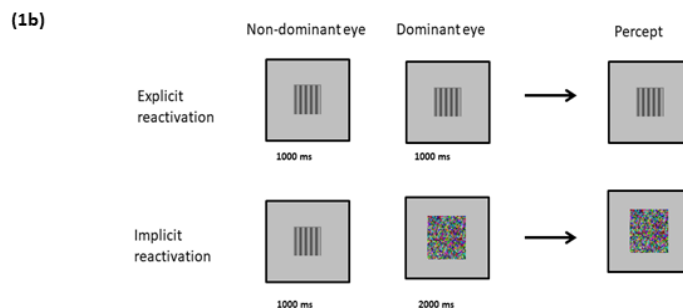
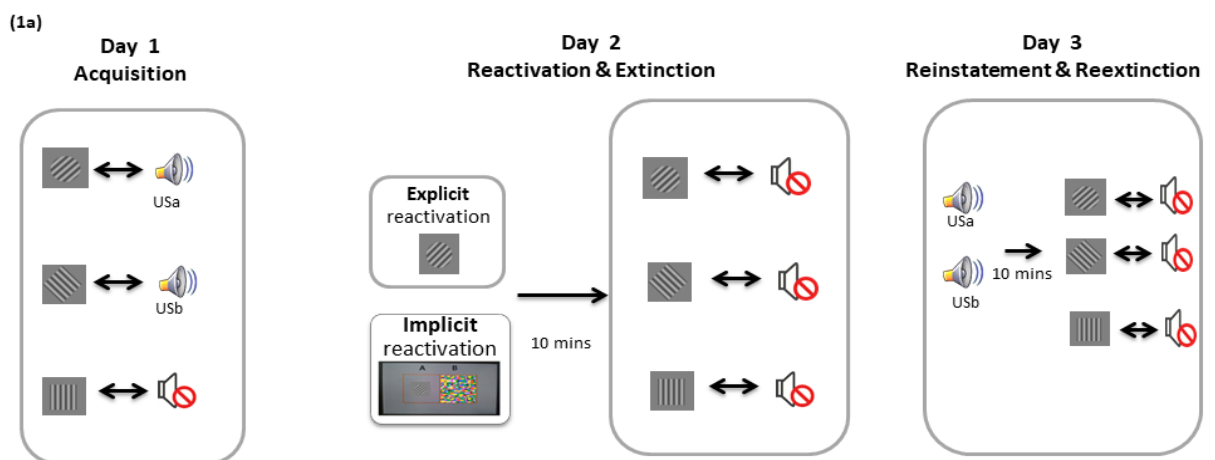
Immediately after the reactivation, all participants watched a neutral video (a BBC documentary on wildlife) for 10 minutes. Following the break, the non-reminded CS+ and CS- were presented eight times each without the USs. The reminded CS+ was presented seven times unreinforced. The stimulus presentation length and the ITI were identical to the acquisition phase.

**Day 3: Reinstatement and re-extinction.** The reinstatement test was conducted 24 hrs following extinction. The session began with four un-signalled US presentations, followed by a 10-minute video break. During re-extinction, participants viewed eight non-reinforced presentations of each CS. The duration of the stimulus presentation and of the ITI were identical to the acquisition session. The presentation of the CSs was randomised in the first trial to control for the potential confounding effects from the trial sequence on the reinstatement test.



Figure 4-1 a) Overview of the experimental protocol. (1b) Percept of the reactivation trials with and without the interference of the continuous flash suppression.

*Note:* During fear acquisition on Day 1, the to-be-explicitly reactivated CS+ and the to-be-implicitly reactivated CS+ were followed by either of the unconditional stimulus (USa: female scream; USb: male scream) on 6 out of 8 presentations, while the CS- was never paired with the US. During reactivation on Day 2, participants received an explicit reminder cue or an implicit reminder cue prior to a 10-minute break and the extinction session, where all CSs were presented without the US. For the implicit reactivation group, the reminder cue was rendered unconscious by the continuous flash suppression. On Day 3, four unsignalled USs were presented, followed by a reinstatement test 10 minutes later and re-extinction.



## Data Preprocessing and Statistical Analyses

Pupil responses were preprocessed in the PsPM toolbox according to Kret & Shie's (2018) recommendation. Bilateral pupil raw data were imported and the mean pupil size was generated. Pupil size samples outside of a predefined feasible range were rejected ( $< 500$  or  $> 10000$ ). Invalid data, as defined as contiguous missing data points larger than 75 ms, were removed. To increase the temporal resolution and smoothness of the data, the mean pupil size samples were resampled with interpolation to 1000Hz and smoothed with a zero-phase low-pass filter with a cutoff frequency of 4 Hz. The interpolated and filtered samples were then z-transformed within each experimental session. Non-normalized pupil samples were used for cross-session comparisons. To estimate the anticipatory pupil responses, we applied the general linear convolution model implemented in PsPM developed by Korn et al. (2017).

We then applied linear mixed models (LMMs) with fixed and random effects to compare the conditioned responses in each session, using the following R formula:  $\text{pupil responses} \sim CS\ type + 1|subject$ . In each linear mixed model, pupillary responses were predicted by *CS type* (reminded CS+, non-reminded CS+, CS-) as the fixed effect and *subject* as the random effect. If *CS type* significantly predicted the pupillary responses in any session, estimated marginal means (EMMs) were applied to infer the mean values of pupillary responses of each CS.

During reactivation, three participants from the implicit reactivation group indicated perceptual awareness of the CS reminder cue. They were re-grouped to the explicit reactivation group for analysis in the present study. The inclusion of the three participants in the reactivation group did not affect the demographic distribution between the two groups

and the overall pattern of acquisition on Day 1 (See Supplementary Tables 4-5 and 4-6 for the demographic and LMM results in Appendix A).

To estimate the fear recovery index according to Schiller et al. (2010), we performed a separate LMM including the non-normalized pupillary responses from the last two trials of extinction and the first two trials of re-extinction following reinstatement, using the following R formula:  $\text{pupil responses} \sim \text{CS type} * \text{session} + 1 | \text{subject}$ .

Significance was taken at  $p < .05$  and Cohen's  $d$  and its 95% confidence interval were reported as a measure of effect size. All analyses were conducted in R 3.5.2 using the lmerTest (Kuznetsova et al., 2017), emmeans (Lenth, 2016), and psych (Revelle, 2019) packages. Graphs were created using the ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2020) packages.

### 4.3 Results

#### Demographics

The demographic characteristics of the participants are presented in Table 4-1.

Table 4-1 Demographic Characteristics of the Sample (N = 59)

	Explicit group ( $n = 29$ )	Implicit group ( $n = 30$ )	$t$ - test ( $p$ )
Age	23.45 (5.90)	21.00 (3.59)	1.92 (.062)
Gender: male (female)	7 (22)	8 (22)	--
Education Level	14.69 (3.32)	14.40 (2.71)	0.37 (.715)
STAI-T	46.03 (9.79)	46.87 (9.65)	-0.33 (.743)

*Note.* STAI-T = State-Trait Anxiety Inventory- Trait

## Day 1: Acquisition

### *Pupillary responses*

LMMs revealed that CS type had a significant fixed effect on the pupillary responses in both groups (explicit reactivation group:  $\chi^2(2) = 32.81, p < .001$ ; implicit reactivation group:  $\chi^2(2) = 42.40, p < .001$ ). Overall, successful threat learning was supported by significantly elevated pupillary responses to the reminded CS+ and the non-reminded CS+ (Table 4-2). For the implicit reactivation group, pupillary responses were higher for the reminded CS+ than the CS-,  $t(62.1) = 6.07, p < .001; d = 0.92 [0.5, 1.37]$ , and non-reminded CS+ elicited higher pupillary responses than CS-,  $t(62.1) = 7.17, p < .001; d = 1.00 [0.43, 1.56]$ . Similar patterns were observed in the explicit reactivation group. Reminded CS+ evoked higher pupillary responses than CS-,  $t(60.1) = 4.86, p < .001; d = 0.79 [0.37, 1.22]$ . Pupillary responses were also higher for non-reminded CS+ than CS-,  $t(60.1) = 6.21, p < .001; d = 0.87 [0.3, 1.43]$ . There were no differences between reminded CS+ and non-reminded CS+ in both groups ( $ps > .05$ ) (Tables 4-2 & 4-3, Figure 4-2).

Table 4-2 Result summary: Coefficient estimates (beta), Standard Error, t statistics, and significance level  $p$  for each predictor in estimating pupillary responses in the LMM analyses.

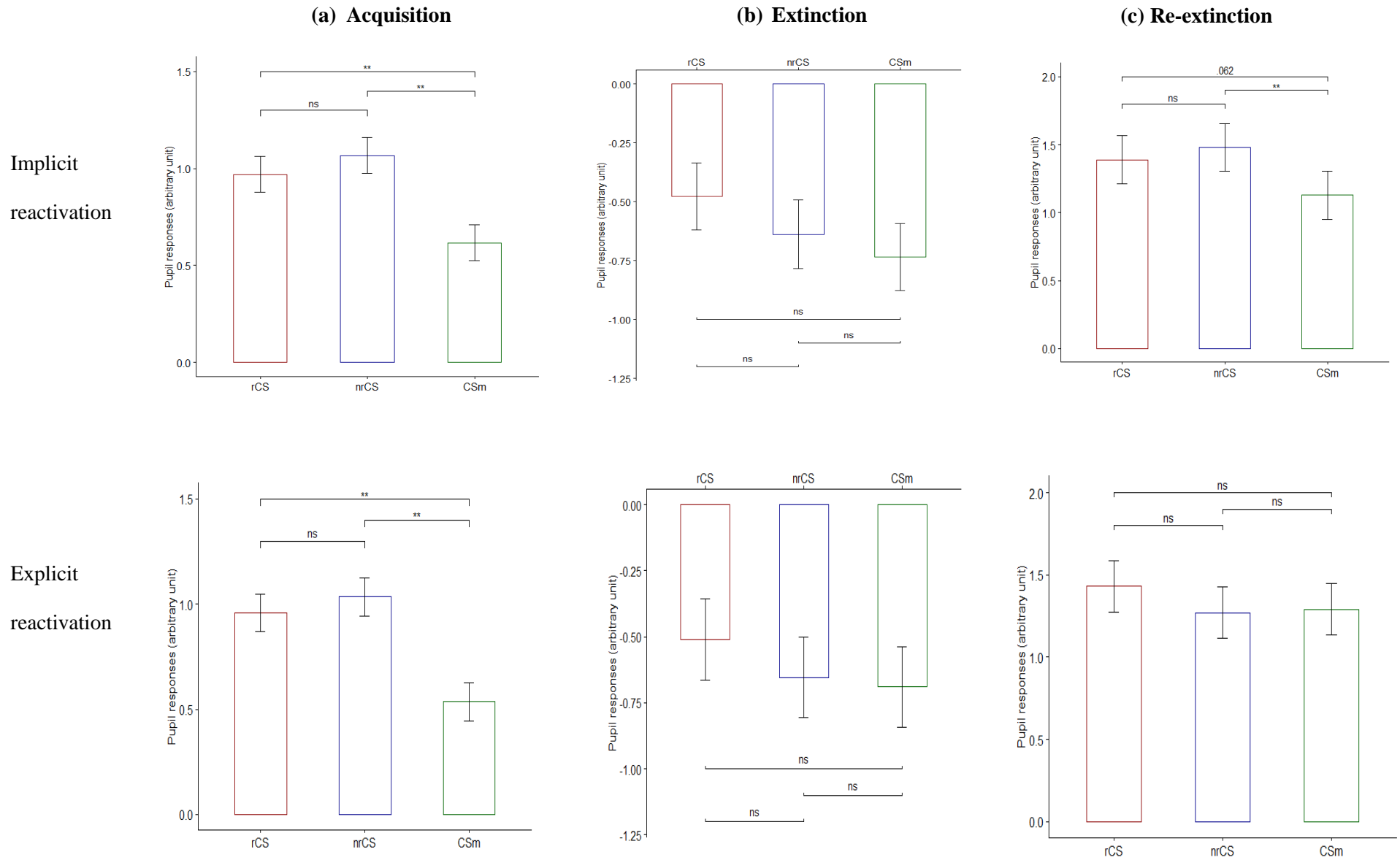
Group	Session	Parameters	$\beta$	SE	$t$	$p$
<i>Implicit reactivation</i>	Acquisition	CS -	0.54	0.09	6.07	< <b>0.001</b>
		Reminded CS+	0.42	0.07	6.17	< <b>0.001</b>
		Non-reminded CS+	0.50	0.07	7.29	< <b>0.001</b>
	Extinction	CS -	-0.73	0.74	1.02	0.310
		Reminded CS+	0.26	1.02	-1.37	0.178
		Non-reminded CS+	0.10	1.02	-1.46	0.150
	Re-extinction	CS -	1.13	0.17	6.48	< <b>0.001</b>
		Reminded CS+	0.26	0.14	1.94	<b>0.057</b>
		Non-reminded CS+	0.36	0.14	2.61	<b>0.012</b>

<i>Explicit reactivation</i>	Acquisition	CS -	0.62	0.09	6.78	<b>&lt;0.001</b>
		Reminded CS+	0.35	0.07	4.94	<b>&lt;0.001</b>
		Non-reminded CS+	0.45	0.07	6.32	<b>&lt;0.001</b>
	Extinction	CS -	-0.51	0.15	-3.40	<b>0.001</b>
		Reminded CS+	-0.14	0.16	-0.92	0.359
		Non-reminded CS+	-0.18	0.16	-1.16	0.252
	Re-extinction	CS -	1.29	0.15	8.39	<b>&lt;0.001</b>
		Reminded CS+	0.14	0.14	0.98	0.332
		Non-reminded CS+	-0.02	0.14	-0.11	0.910

Table 4-3 Estimated marginal means of the pupillary responses of CSs, their standard errors and confidence intervals in each experimental session

	Session	CS type	EMMS	SE	95% CI	
<i>Implicit reactivation</i>	Acquisition	Reminded CS+	0.97	0.09	0.78	1.16
		Non-reminded CS+	1.07	0.09	0.88	1.25
		CS -	0.62	0.09	0.43	0.80
	Extinction	Reminded CS+	-0.51	0.15	-0.82	-0.21
		Non-reminded CS+	-0.65	0.15	-0.96	-0.35
		CS -	-0.69	0.15	-1.00	-0.39
	Re-extinction	Reminded CS+	1.43	0.16	1.12	1.75
		Non-reminded CS+	1.27	0.16	0.96	1.59
		CS -	1.29	0.16	0.98	1.60
<i>Explicit reactivation</i>	Acquisition	Reminded CS+	0.96	0.09	0.78	1.14
		Non-reminded CS+	1.03	0.09	0.85	1.22
		CS -	0.54	0.09	0.36	0.72
	Extinction	Reminded CS+	-0.48	0.15	-0.77	-0.19
		Non-reminded CS+	-0.64	0.14	-0.92	-0.36
		CS -	-0.74	0.14	-1.02	-0.45
	Re-extinction	Reminded CS+	1.39	0.18	1.03	1.75
		Non-reminded CS+	1.48	0.18	1.13	1.84
		CS -	1.13	0.18	0.77	1.48

Figure 4-2 Pupillary responses in the (a) acquisition, (b) extinction, and (c) re-extinction.  
 Note: \*  $p < .05$ , \*\*  $p < .001$ . rCS: reminded CS+; nrCS: non-reminded CS+; CSm : CS-



### *Subjective ratings*

After the acquisition phase, participants in both implicit- and explicit-reactivation groups reported that the CSs+ were more likely to be followed by the US than the CS-: implicit reactivation group, *reminded CS+ vs CS-*  $t(62.1) = 6.01, p < .001, d = 1.10$  [0.67,1.56], *non-reminded CS+ vs CS-*  $t(62.1) = 4.51, p < .001, d = 0.65$  [0.27, 1.04]; explicit reactivation group, *reminded CS+ vs CS-*  $t(60.1) = 6.19, p < .001, d = 0.88$  [0.44, 1.35], *non-reminded CS+ vs CS-*  $t(60.1) = 6.54, p < .001, d = 1.16$  [0.68,1.68]. (Table 4-4 & Figure 4-3)

Similarly, participants reported higher levels of unpleasantness towards the reminded CS+ and non-reminded CS+ relative to CS-: implicit reactivation group, *reminded CS+ vs CS-*  $t(56) = 4.78, p < .001, d = 0.88$  [0.47,1.31], *non-reminded CS+ vs CS-*  $t(56) = 4.52, p < .001$ ; explicit reactivation group, *reminded CS+ vs CS-*  $t(58) = 5.09, p < .001, d = 0.88$  [0.47,1.31], *non-reminded CS+ vs CS-*  $t(58) = 3.81, p = .001, d = 0.83$  [0.39, 1.28] (Figure 4-4).

Taken together, the subjective ratings indicate that participants acquired threat conditioning after the acquisition.

Table 4-4 Estimated marginal means of the valence and likelihood ratings of CSs, their standard errors and confident intervals after each experimental session

	Session	CS type	<u>Valence rating</u>				<u>Likelihood rating</u>				
			EMMS	SE	95% CI		EMMS	SE	95% CI		
<i>Implicit reactivation</i>	Acquisition	Reminded CS+	3.20	0.24	2.72	3.68	3.63	0.20	3.23	4.04	
		Non-reminded CS+	2.83	0.24	2.35	3.31	3.20	0.20	2.80	3.60	
		CS -	1.73	0.24	1.25	2.21	1.90	0.20	1.50	2.30	
		Extinction	Reminded CS+	2.00	0.17	1.66	2.34	1.33	0.11	1.10	1.56
	Extinction	Non-reminded CS+	1.67	0.17	1.32	2.01	1.13	0.11	0.91	1.36	
		CS -	1.30	0.17	0.96	1.64	1.20	0.11	0.97	1.43	
		Re-extinction	Reminded CS+	2.00	0.19	1.62	2.38	1.43	0.15	1.14	1.72
			Non-reminded CS+	1.93	0.19	1.56	2.31	1.27	0.15	0.98	1.56
	CS -		1.53	0.19	1.16	1.91	1.17	0.15	0.88	1.46	
	<i>Explicit reactivation</i>	Acquisition	Reminded CS+	3.03	0.24	2.55	3.52	3.62	0.23	3.16	4.08
			Non-reminded CS+	2.97	0.24	2.48	3.45	3.72	0.23	3.27	4.18
			CS -	1.72	0.24	1.24	2.21	1.79	0.23	1.33	2.25
Extinction			Reminded CS+	1.83	0.16	1.51	2.15	1.41	0.16	1.09	1.74
		Non-reminded CS+	1.66	0.16	1.34	1.97	1.38	0.16	1.06	1.70	
		CS -	1.45	0.16	1.13	1.77	1.24	0.16	0.92	1.56	
		Re-extinction	Reminded CS+	1.62	0.11	1.31	1.94	1.48	0.17	1.14	1.82
Non-reminded CS+			1.52	0.11	1.20	1.83	1.34	0.17	1.00	1.69	
CS -			1.45	0.11	1.13	1.76	1.31	0.17	0.97	1.65	



Figure 4-3 Valence rating in the (a) acquisition, (b) extinction, and (c) re-extinction.

Note. \*  $p < .05$ , \*\*  $p < .001$ . imp\_rCS: implicitly-reminded CS+; ext\_rCS: explicitly-reminded CS; nrCS: non-reminded CS+; CSm : CS-

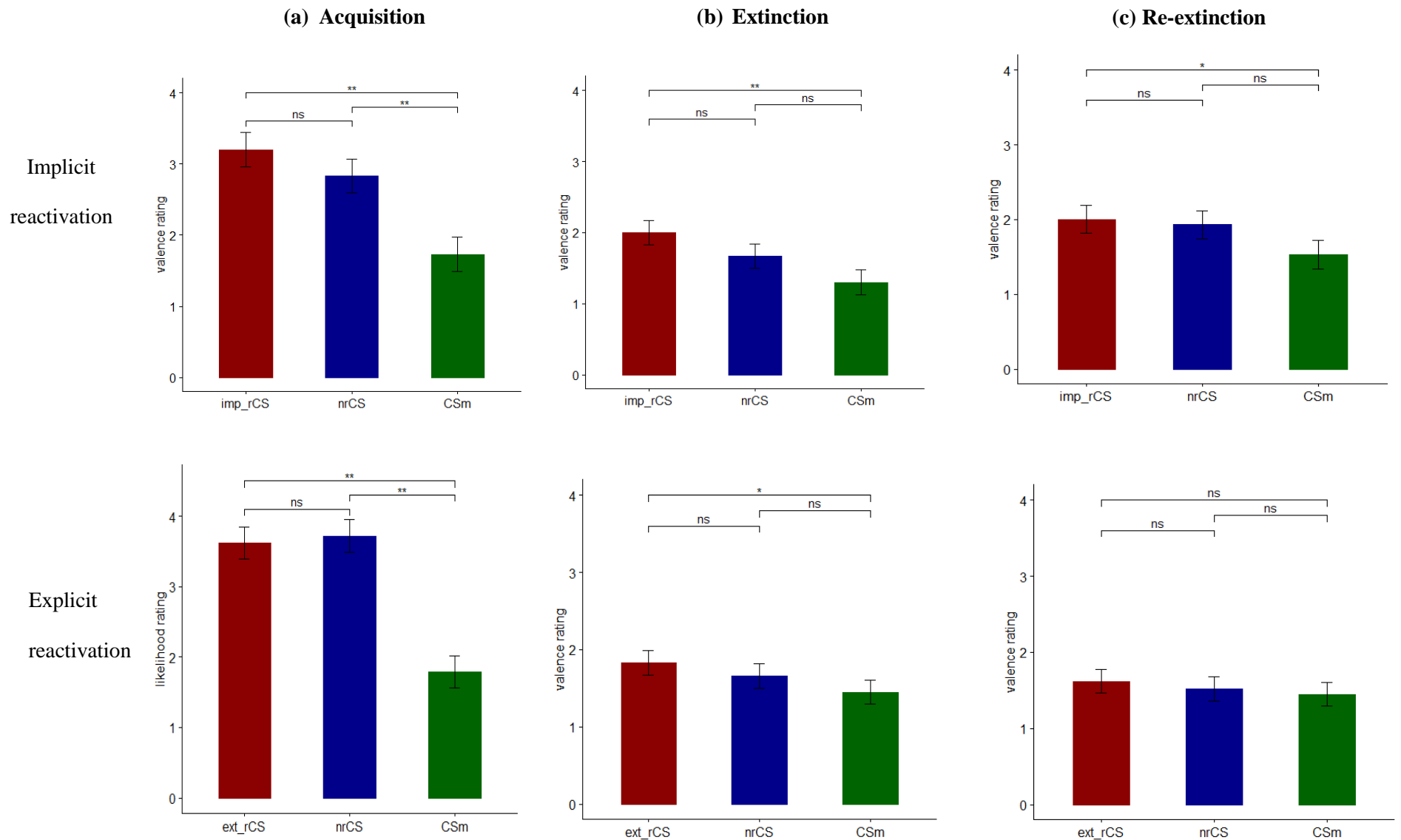
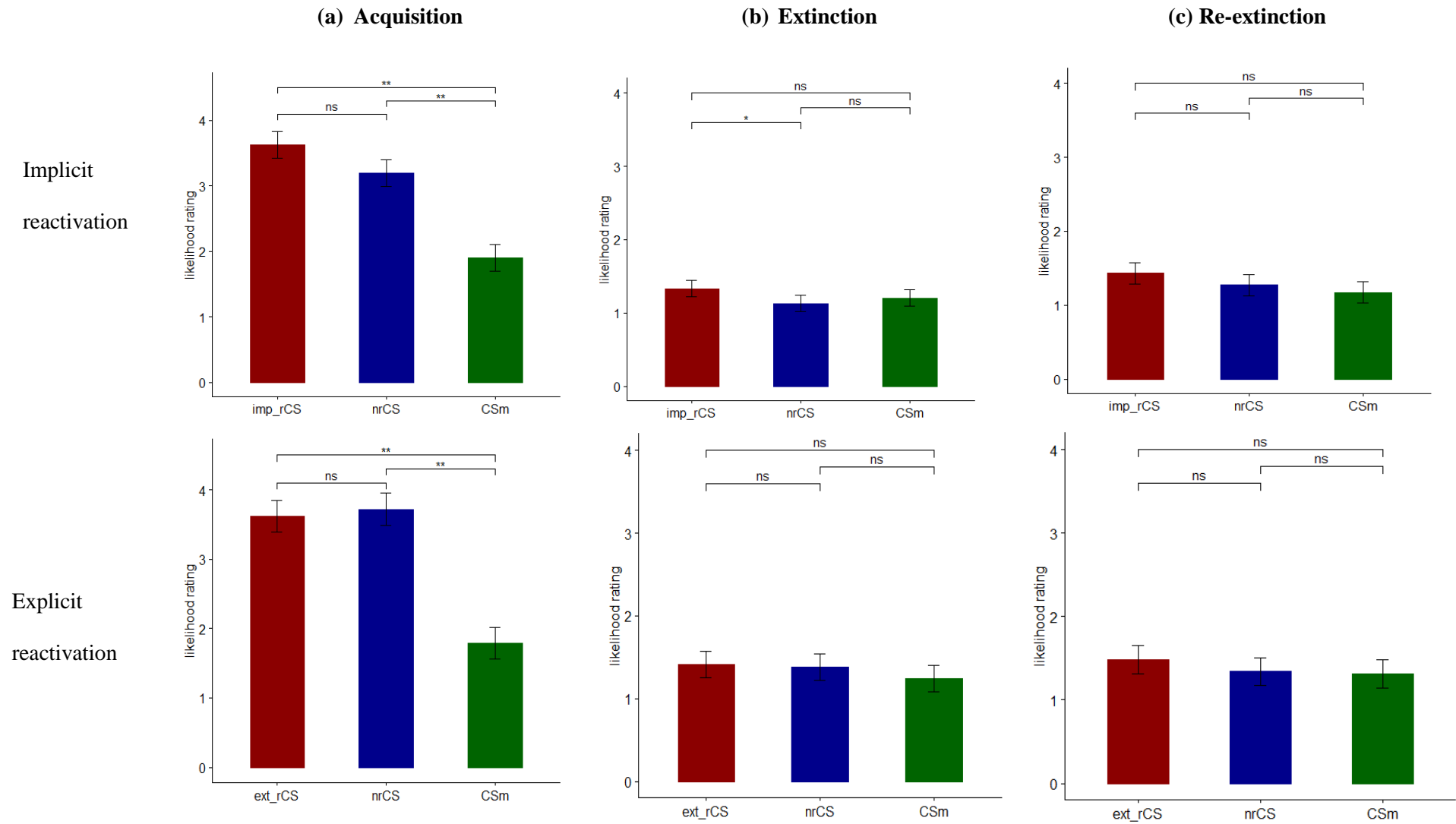


Figure 4-4 Likelihood rating in the (a) acquisition, (b) extinction, and (c) re-extinction.

Note: \*  $p < .05$ , \*\*  $p < .001$ . imp\_rCS: implicitly-reminded CS+; ext\_rCS: explicitly-reminded CS; nrCS: non-reminded CS+; CSm : CS-



## Day 2: Reactivation and Extinction

### *Pupillary responses*

We observed an overall negative pupil response averaged over all the trials across all CS type. The negative pupil responses were likely due to the constantly changing pupil sizes in response to unaroused stimuli. LMMs revealed evidence of extinction by the end of the session: *CS type* no longer had a significant fixed effect in the model predicting pupillary responses in the last two trials for both the implicit reactivation group,  $\chi^2(2) = 2.61, p = .271$ , and the explicit reactivation group,  $\chi^2(2) = 1.48, p = .478$ .

### *Subjective ratings*

After extinction, there was a reduction in US likelihood ratings. For both the explicit and implicit reactivation groups, the likelihood ratings were comparable between the CSs+ and the CS- ( $ps > .05$ ), independent of how the CSs+ were reactivated. Interestingly, comparison between the CSs+ revealed that the likelihood rating of the reminded CS+ was higher than that of the non-reminded CS+ in the implicit reactivation group,  $t(62.1) = 2.93, p = .012, d = 0.41 [0.04, 0.80]$ . This pattern was not observed in the explicit reactivation group.

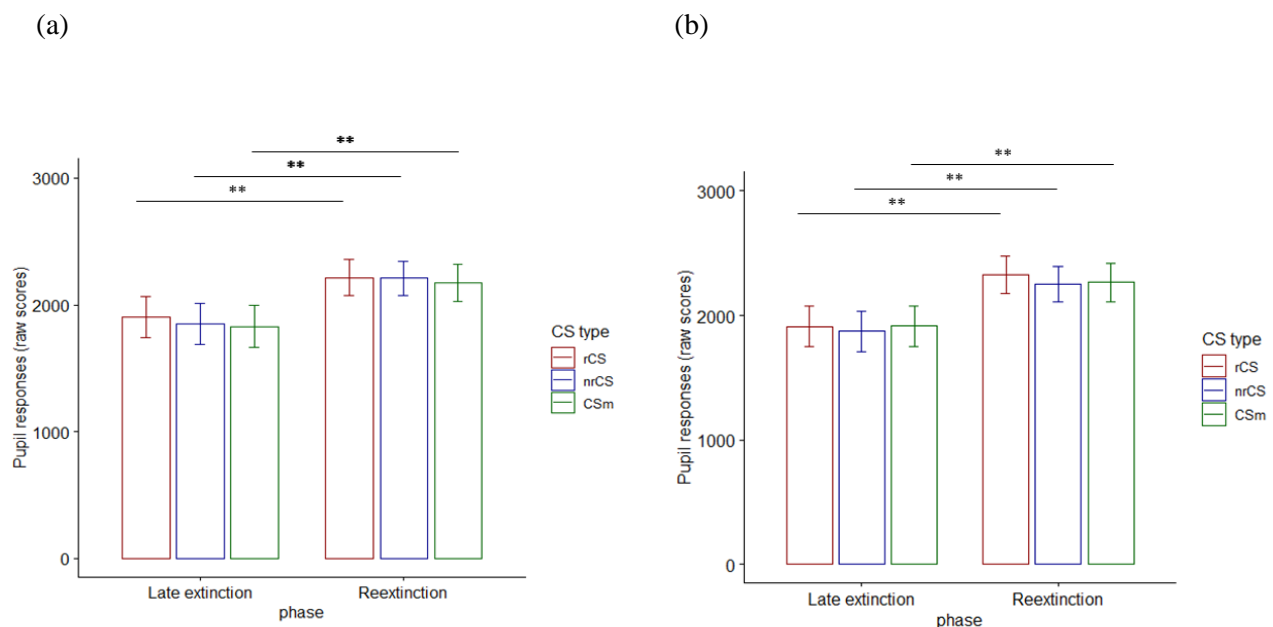
With respect to the unpleasantness rating, participants rated the reminded CS+ as more unpleasant than the CS- in both the implicit- and explicit-reactivation groups,  $t(58) = 3.73, p = .001, d = 0.80 [0.39, 1.23]$ , and  $t(56) = 2.34, p = .058, d = 0.39 [0.01, 0.78]$  respectively.

## Day 3: Reinstatement and Re-extinction

### *Pupillary responses*

The inclusion of experimental session (Extinction or re-extinction) in the LMM revealed that there was a significant increase in the pupillary responses from late extinction to early re-extinction for both the CS+ and CS- in the two groups,  $\chi^2(5) = 132.02, p < .001$  for the implicit reactivation group and  $\chi^2(5) = 134.53, p < .001$  for the explicit reactivation group. Fear responses recovered for all CSs after reinstatement in both groups of participants (Figure 4-5).

Figure 4-5 Non-normalized pupillary responses in the (a) implicit reactivation group, and (b) explicit reactivation group across late-extinction and re-extinction. Note. \*\*  $p < .001$



Crucially, the pattern of fear recovery was different between the two groups on Day 3 (Figure 4-2c). For the implicit reactivation group, the non-reminded CS+ evoked a significantly higher pupillary responses relative to the CS-,  $t(58.1) = 2.56, p = .013, d = 0.41$  [-0.13, 0.94], and the reminded CS+ elicited a higher pupillary response at a trend-level

relative to the CS-,  $t(58.1) = 1.91, p = .060, d = 0.35 [0.03, 0.74]$ . The reminded CS+ and non-reminded CS+ did not evoke a statistically different pupillary response,  $t(58.1) = -0.66, p = .513$ . For the explicit reactivation group, *CS type* did not have a significant fixed effect on the pupillary responses,  $\chi^2(2) = 1.43, p = .490$ . Given the significant effect of session, this suggests that fear responses recovered equally for all three CSs after reinstatement. Both the reminded and non-reminded CS+ evoked similar pupillary responses,  $t(60.1) = 1.07, p = .534$ .

#### *Subjective ratings*

Following reinstatement and re-extinction, US likelihood was rated as low and comparable for all CSs in both groups. With respect to the valence rating in the implicit reactivation group, the unpleasantness rating towards the reminded CS+ and the non-reminded CS+ remained higher relative to the CS-,  $t(58) = 3.11, p = .008, d = 0.50 [0.12, 0.89]$ , and  $t(58) = 2.67, p = .026, d = 0.45 [0.07, 0.83]$  respectively. The unpleasantness ratings were comparable for all three CSs after re-extinction in the explicit reactivation group,  $ps > .05$ .

#### **4.4 Discussion**

The current study investigated whether an implicitly presented reminder cue destabilizes the original CS-US associative memory trace for subsequent interference by behavioural extinction. Contrary to our predictions, presentations of implicit and explicit reminder cues before extinction did not provide evidence for reducing the reinstatement of fear measured by pupillary responses. In addition, participants rated the reminded CS+ as more unpleasant than the non-reminded CS+ and the CS- following extinction, and this

differential affective response is independent of the perceptual awareness of the CS during its reactivation. In sum, presentation of a pre-extinction reminder cue showed no statistical evidence for preventing the reinstatement of fear, but it may increase the subjective report of unpleasantness associated with the CS.

### **Return of pupillary responses following reinstatement**

Our findings did not provide support to the notion that reactivation-extinction prevents the return of fear in humans. Our findings, together with other replication studies using pupil responses (Zimmermann & Bach, 2020), skin conductance and fear-potentiated startle responses (Bos et al., 2014; Chalkia et al., 2020; Fricchione et al., 2016; Klucken et al., 2016; Schroyens et al., 2017; Soeter & Kindt, 2011; Spring et al., 2015; Thome et al., 2016) to index the return of fear in a retrieval-extinction paradigm fail to provide evidence for preventing the recovery of fear using a reminder cue before extinction. To our knowledge, no studies to date have investigated the impact of implicit reminder cues on subsequent extinction learning and the recovery of conditioned responses. Here we discuss two factors associated with implicit reactivation that may contribute to the return of fear observed in our study.

First, it is plausible that the presentation of implicit reminder cues might not have evoked sufficient prediction error in the present study. Previous research indicates that fear memory does not enter a labile state when no new learning takes place during memory reactivation (Sevenster et al., 2012, 2013b). While this explanation is tenable, it seems unlikely in the present study because the increase in affective ratings following extinction was specific to the reminded CSs only. Moreover, presenting an implicit CS reminder cue might have the advantage of circumventing the higher-level, cognitive pathway involved in

prediction errors by targeting the unconscious pathways involved in defensive responses (LeDoux & Pine, 2016).

Second, the strength of the unconscious processing of the cues might be limited in implicit reactivation. In the present study, we attempted to strengthen the impact of the implicit cue by having two reminder trials instead of a single reminder, which is more common in the conventional retrieval-extinction procedure (Oyarzún et al., 2012; Schiller et al., 2010a; Soeter & Kindt, 2012b); however, the transient effect of learning might not be sufficient to destabilize a CS-US association. This is echoed by one previous study demonstrating that unconscious fear learning tends to dissipate quickly after a few trials (Raio et al., 2012b). Following this line of thought, further research could vary the number of reactivation trials or lengthen its duration to rule out whether the strength of implicit reactivation is a boundary condition for triggering reconsolidation.

Third, there has been a call to re-evaluate the robustness of the findings in the reactivation-extinction literature (Chalkia et al., 2020). Since the publication of Schiller et al. (2010)'s study on the reactivation-extinction effect on ROF, there has been mixed findings in both conceptual (e.g. Soeter & Kindt, 2011; Meir Drexler et al., 2014) and methodological (Chalkia et al., 2020) replications of the original study. In a high-power, registered replication report of Schiller et al.'s study with a larger sample size ( $n = 124$ ), the authors observed spontaneous recovery of fear in both reactivation and no reactivation groups (Chalkia et al., 2020b). The authors further called for careful inspection of the inclusion and exclusion criteria for data analyses which were justified based on the researchers' degrees of freedom in the reconsolidation literature (e.g. Chalkia et al., 2020), and a general lack of clear standards for exclusions in human fear conditioning research (Lonsdorf et al., 2019).

Of interest for future investigations is also the reinstatement of fear in humans. We observed a differential reinstatement of pupillary responses only in the implicit reactivation group. While reinstatement of fear is a relatively consistent phenomenon in non-human animals, the return of differential responding in humans is not always evident. As reviewed by Haaker et al. (2014b), some studies reported enhanced responding to CS+ only (differential return of fear), some reported enhanced responding to both CS+ and CS- (non-differential return of fear), and some did not report any reinstatement effect. Furthermore, assessment of the return of fear is heterogeneous across laboratories and studies (Lonsdorf et al., 2017, 2019). Before there is a consensus for best capturing the recovery of fear in humans, future studies could consider additional experimental procedures such as spontaneous recovery and renewal of contexts to better model relapses and enhance the translational value of experimental research in humans.

### **Differential unpleasantness ratings following extinction**

In the present study, the reminded CSs+ were rated as more unpleasant than the CS- after extinction learning in both groups of participants. This differential CS unpleasantness rating may reflect a perseveration of the CS-US association after extinction. Several studies reported a similar observation (Vansteenwegen et al., 1998; Lipp et al., 2003; Dirikx et al., 2004; Zbozinek et al., 2014) and there are two possible accounts for this observation. First, the acquired negative valence of a CS+ may reflect a form of evaluative conditioning (De Houwer et al., 2001; Hofmann et al., 2010). One key characteristic of evaluative learning is its resistance to extinction (Hermans et al., 2002); therefore, the negative valence of a CS+ was likely to sustain despite the extinction procedure in the present study. Second, the



differential valence of CS reflects a renewal of conditioned responses, i.e. a post-extinction resurgence of fear if the CS is presented in a context that differs from the one in which extinction takes place (Bouton, 2002). In the present study, extinction training and the rating of the CSs were conducted in different stations in the same room, using different computers. It is conceivable that the transition from one station to the next station might introduce a change in context and a renewal of the CS-US association. Nonetheless, the renewal phenomenon was not observed in the likelihood ratings and pupil responses following extinction. Therefore, we believe the former account might be more applicable to explain the differential valence ratings observed in the present study.

Interestingly, the differential valence ratings were only found in the reminded CS+ in both groups of participants, suggesting that this effect is rather reminder specific. Several studies have shown that the valence of the CS+ at the end of extinction predicts the amount of conditioned responses following reinstatement (Dirikx et al., 2004, 2007; Hermans et al., 2005b). This valence-reinstatement model posits that negative valence plays an affective-motivational role in the re-emergence of fear (Hermans et al., 2002). Our data, however, did not support this notion as only weak associations between post-extinction unpleasantness ratings and post-reinstatement pupillary responses were found (Table 8-4 for supplementary analysis in Appendix A). Nonetheless, subjective evaluation of CS valence can be considered as a clinical proxy of subjective feelings reported by patients. The association between post-extinction valence ratings and reinstatement of fear warrants further investigation as it entails clinically relevant information to predict the long-term outcomes of exposure therapy (Zbozinek et al., 2015).

## **Limitations**

There are several limitations in the present study. First, we did not measure the presence of prediction error during reactivation. Sevensters and colleagues (Sevenster et al., 2013b, 2014) suggested that a rating of US expectancy following reactivation can be used as a behavioural index of prediction error and a proxy for the destabilisation of fear-related memories. We did not include such index because a rating of US expectancy may require participants to recall the learned fear explicitly, which may confound the findings in the implicit reactivation group. There are also studies demonstrating that a decrease in the US expectancies following reactivation may not be necessary for memory destabilization (Soeter & Kindt, 2012a, 2015a). Second, the number of extinction trials for the reminded CS+ was different from those of the non-reminded CS+ and the CS-. We acknowledge this limitation. Given that the US likelihood ratings and pupil responses following extinction were similar between the CS+ (reminded and non-reminded) and the CS-, we believe that extinction learning was sufficient in the present study. Third, we did not measure trial-based online ratings of CS valence or US expectancy. Although online-ratings may be considered as a valid and reliable measure of learning, online ratings may interfere with the process of learning itself, as it may enhance the conscious pathway of learning (Lonsdorf et al., 2017).

#### **4.5 Conclusion**

The prospect of using a reconsolidation-based approach in treatments of anxiety and fear-related disorder is exciting. Yet, questions regarding its mechanism of change and the optimal conditions for inducing reconsolidation remain largely unanswered. While our findings did not provide support for the use of reminder cues for preventing the reinstatement of fear, the present study shows that a reminder cue, implicitly or explicitly viewed, might increase the subjective report of unpleasantness associated with the CS following extinction.

Future reconsolidation research in humans should consider a comprehensive assessment of conditioned responses, including brain imaging, physiological measures, and subjective ratings, to further delineate the different aspects of reconsolidation in a retrieval-extinction conditioning paradigm.

## **Chapter 5**

### **Experiment 3: Evidence for differential extinction learning by disrupting reconsolidation in multi-CS conditioning**

## 5.1 Introduction

Memories of fear-evoking events and stimuli play an integral role in the emergence and maintenance of a variety of anxiety disorders, including panic disorder, specific phobias and posttraumatic stress disorder (Bisaz et al., 2014; Ledoux & Muller, 1997; Sylvers et al., 2011). Previously, researchers thought that memories, once consolidated, are stable and impervious to change for the lifetime of the memory trace (Mcgaugh & Mcgaugh, 2012). In the past two decades, researchers have discovered that memories undergo a new process of consolidation if they are reactivated (Nader et al., 2000; Karim Nader, 2015). This process, termed reconsolidation, provides a unique opportunity for interfering with the integrity of fear-related memory traces and offers translational value for the treatment of anxiety and fear-related disorders involving aversive and maladaptive fear memories.

One of the behavioural procedures to initiate the reconsolidation process in a Pavlovian conditioning paradigm involves a nonreinforced presentation of the conditioned stimulus (CS) to reactivate the related memory association, followed 10 minutes later by a standard extinction training, in which all conditioned stimuli are presented without the aversive stimulus (US) (Monfils et al., 2009; Schiller et al., 2010b). Without a reminder cue prior to extinction, extinction learning is considered a form of inhibitory learning in which a new CS-no US association is formed and competes with the existing CS-US association. As a consequence, the CS-US memory trace persists and conditioned responses associated with this memory trace re-emerge over time (spontaneous recovery), a change in the context (contextual renewal) or after exposure to the aversive stimulus (reinstatement) (Bouton, 2002). In contrast, a reminder cue 10 minutes before the standard extinction is thought to open a reconsolidation window for modifying the existing CS-US memory trace at its source,

therefore preventing the return of conditioned responses over time and after reinstatement in rats (Monfils et al., 2009) and humans (Schiller et al., 2010a).

The impact of a reminder-extinction procedure on the return of conditioned responses has been replicated in studies using healthy subjects (Oyarzún et al., 2012; Schiller et al., 2013b; Soeter & Kindt, 2015a; Thompson & Lipp, 2017b) and clinical populations (Björkstrand et al., 2016; Soeter & Kindt, 2015b; Telch et al., 2017; Xue et al., 2012). Nonetheless, several similarly powered studies have reported null findings (Golkar et al., 2012; Klucken et al., 2016; Schroyens et al., 2017; Thome et al., 2016). This discrepancy has been explained by several possible differences in the experimental protocol and the boundary conditions that govern the triggers of the reconsolidation processes (Elsey et al., 2018; Kredlow et al., 2016; Lee et al., 2017). Despite the inconsistent research findings, a meta-analysis examining 16 experiments involving the reconsolidation of human fear memories revealed a moderate effect size ( $g = 0.40$ ) of the reminder-extinction procedure over the standard extinction procedure (Kredlow et al., 2016). Further investigation of the neurocognitive processes underlying memory reconsolidation is therefore warranted to improve the translational value of reconsolidation research in humans.

To date, most of the existing studies in memory reconsolidation employ an explicit learning paradigm, whereby one stimulus is paired with one US in the conditioning process. However, a simple one CS to one US associative learning in real life is rare. In a scene of witnessing a traumatic motor vehicle accident, many associations related to the scene of the collision, such as the traffic lights, the inflated airbag, the broken glass, or a trapped passenger could be formed. These multiple affective associations could be modelled in a multi-CS conditioning paradigm. Multi-CS conditioning pairs a multitude of perceptually similar neutral stimuli (e.g. 52 distinct neutral faces) with one or multiple unconditioned

stimuli during acquisition, forming multiple conditioned stimuli (CS+). Because of the large number of perceptually similar and complex stimuli involved in learning, the multi-CS conditioning paradigm allows the investigation of the implicit processes in affective learning that are independent of explicit CS-US awareness (Steinberg et al., 2013). Previous multi-CS studies have yielded successful fear acquisition and extinction on both behavioural and neural levels (Brockelmann et al., 2011a; Junghofer et al., 2017; Rehbein et al., 2014; Roesmann et al., 2020; Steinberg et al., 2013). In these studies, behaviourally, CS+ were rated more unpleasant and arousing compared to the safe CSs that were never paired with the US (CS-). Neurally, CS+ evoked stronger activation in the right prefrontal cortex encompassing the lateral and orbital regions, as well as in the temporo-occipital regions as measured by magnetoencephalography (MEG) (Steinberg et al., 2012a).

Moreover, a US-reminder is effective in preventing the return of fear memory in a reminder-extinction procedure. In a series of studies, Liu and colleagues (2014; 2015) found that the presentation of a US before extinction disrupted the associations between the CS and US in both humans and rats. This US-triggered reconsolidation was selective to the reactivated US and persisted for six months in humans (Liu et al., 2014). The US-reminder-extinction procedure has also been applied and replicated in a recent study using both fear-related and fear-irrelevant stimuli (Thompson & Lipp, 2017). The presentation of a US reminder is particularly pragmatic in a multi-CS conditioning paradigm as the presentation of a US reminder could reactivate multiple CS-US associations.

In the present study, we investigated whether a US-reminder-retrieval procedure could trigger implicitly-learned fear memories for reconsolidation, and prevent the return of fear. Using a three-day multi-CS conditioning paradigm, we paired a multitude of neutral faces with two aversive tones during acquisition. Before extinction on Day 2, a US reminder

was given 10 minutes before all CS underwent extinction. We compared the conditioned responses, measured by the size of pupil dilation, following the reinstatement test on Day 3 to infer the return of fear. The main hypothesis of the present study was that the levels of conditioned responses of the reminded CS+ would be diminished relative to the non-reminded CS+ following the reinstatement.

## 5.2 Methods

### Participants

To determine required sample size, we conducted a power analysis using G\*power 3.1 (Faul et al., 2009) based on Leuchs et al (2017) study using pupillometry in a fear conditioning experiment, in which the effect size for a CS+/CS- difference was (Cohen's)  $d = 0.82$ . Setting alpha at .05, a sample size of  $N = 18$  was required to achieve 95% power. We recruited a total of 36 participants to allow counterbalancing of the stimuli in the current experiment.

Thirty-six healthy university students ( $19.83 \pm 2.48$  years; males: females = 12: 24) participated in the study. All participants reported normal hearing, normal or corrected-to-normal vision. The exclusion criteria of the study were self-report of current or history of psychiatric and/or neurological disorders. In the present study, participants showed low levels of trait anxiety ( $M = 44.37$ ,  $SD = 7.10$ ) as determined by the mean scores on the trait scale of the State-Trait Anxiety Inventory. Participants earned course credits or received a monetary reward for their participation. They gave written informed consent to the experimental protocol approved by the ethics committee of the university (EA170915) in accordance with the Declaration of Helsinki.



## Materials

**Unconditioned stimuli.** Two aversive tones (90 dB, duration: ~1200ms; USa: a female scream, USb: a male scream) were used. The screams were normalized and resampled to 44100 Hz. It was delivered binaurally through headphones. The assignment of USa and USb was balanced across participants.

**Conditioned stimuli.** Fifty-four images displaying faces (27 females) with neutral expressions were selected from existing face databases, including the Karolinska Directed Emotional Faces archive (Lundqvist et al., 1998) and the NimStim Face Stimulus Set (Tottenham et al., 2009). All faces were converted to grayscale images in Adobe® Photoshop® and were adjusted for its levels of brightness and contrast. They were pseudo-randomly split into three conditions: 18 CS+ (paired with USa), 18 CS+ (paired with USb) and 18 CS- faces (unpaired during conditioning). The attractiveness of the faces was rated by a pilot group of participants, who were recruited separately (N = 20); there were no significant differences in the rated attractiveness of the faces across the three conditions,  $F(2, 57) = 1.00, p = .376$ . The pictures were randomly assigned to three groups with equal gender ratio; and the condition of each group of pictures (i.e. whether they were reminded CS+, non-reminded CS+ or CS-) was counterbalanced across participants during acquisition.

**CS-US matching task.** Explicit knowledge of the stimulus category was assessed using a computerized CS-US matching task. In this task, all 54 CS were pseudo-randomly presented for 600 ms. Participants were asked to indicate for each face whether it was paired with a scream during conditioning (stimulus category: CS+ vs. CS-) on a Likert scale from -4 (*surely there was no scream*) to 4 (*surely there was a scream*), followed by a second question

in which they were asked to indicate whether the faces were paired with a male scream or a female scream on a Likert scale (  $-4 = surely\ female$  to  $4 = surely\ male$ ). For practice, they completed three trials prior to the start of the task.

**Pair comparison task.** Contingency awareness of CS-US pairings was also indirectly assessed by the Pair Comparison Task in which participants were presented with pairs of CS+ and CS- faces. They were asked to decide which face they preferred in a binary forced-choice format. It was expected that participants would show a preference for the CS- faces after conditioning as they were not followed by an aversive scream. Three distinct versions, each with 27 CS+ and CS- trials, were developed for this experiment such that each CS was only rated once after each experimental session.

**US rating task.** To identify the perceived valence and arousal of the USs, participants were asked to rate the valence and arousal of each tone on an 8-point Likert scale.

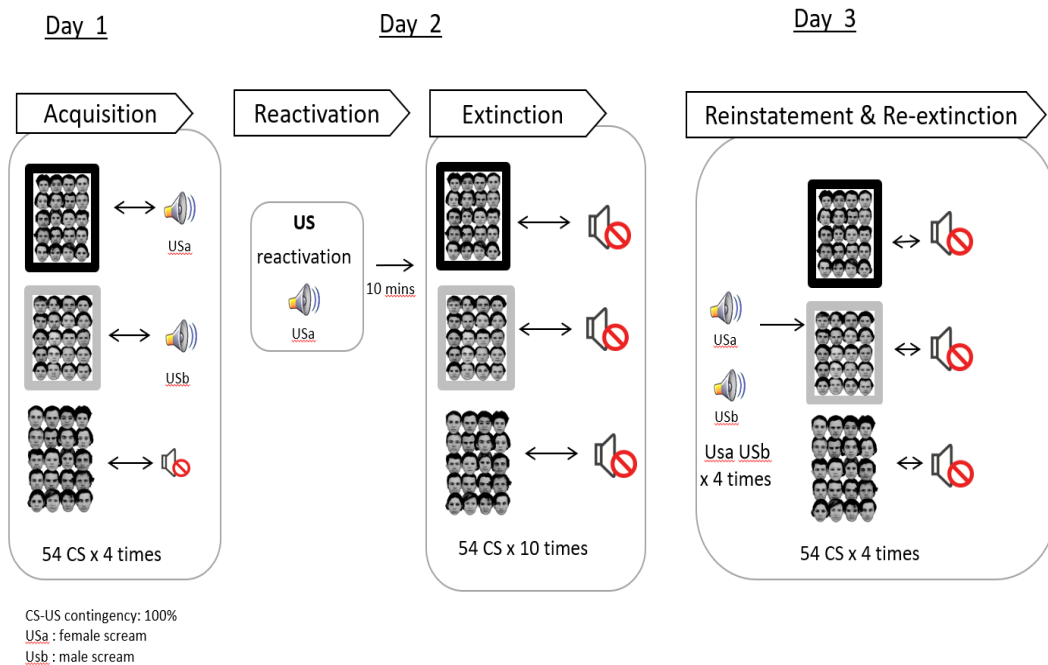
**Questionnaires.** Participants completed the Positive and Negative Affect Scale (PANAS; Watson et al., 1988) to assess their mood state before the start of each experimental session. This 20-item scale consists of a series of adjectives (e.g., Positive: content, Negative: Afraid), measuring positive and negative affective states. Participants were asked to indicate how they felt at the moment on a 5 -point Likert scale from 1 = *not at all* to 5 = *extremely*. Higher scores reflect higher positive or negative affect. The PANAS has good internal reliability in the Chinese population (overall Cronbach's  $\alpha = 0.82$ ) (Huang et al., 2003).

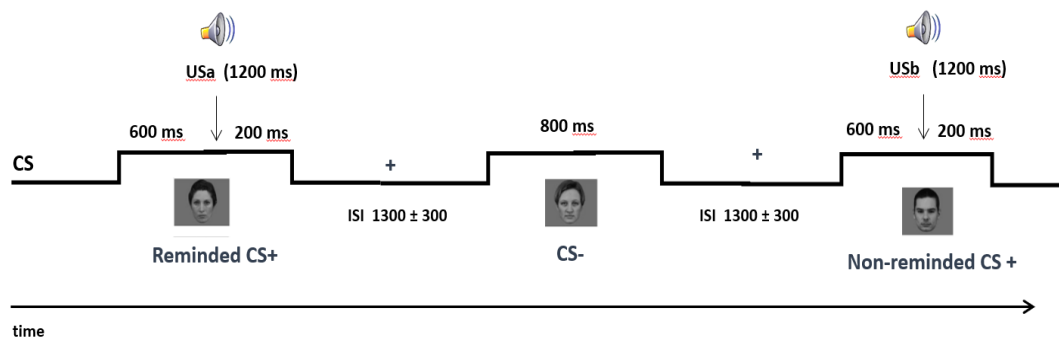
All behavioural tasks were conducted on a computer using Presentation (Neurobehavioral Systems, Albany, CA).

## Experimental Procedure

This experiment adopted a within-subject design and comprised three sessions which took place on three consecutive days. The experimental sessions included acquisition (Day1), reactivation and extinction (Day 2), and reinstatement and re-extinction (Day 3). Figure 5-1 shows an overview of the experimental procedure.

Figure 5-1 Experimental timeline and procedure during conditioning. In Multi-CS conditioning, reminded CS+ (rCS+), non-reminded CS+ (nrCS+) and CS- faces were presented in a pseudorandomized order whereby each CS was presented four times for 800ms (i.e., 72 trials per condition, 216 trials in total). The auditory USs started 600 ms after the CS onset.





### Day 1: acquisition

Prior to acquisition, participants completed a habituation phase in which all CSs were presented in a random order for 800 ms, each at the centre of the screen. Each CS was presented four times. A fixation cross was presented between trials, and the inter-trial interval (ITI) was  $1300 \pm 300$  ms. Participants were asked to view the faces passively on the screen.

**Fear acquisition.** rCS+, nrCS+, and CS- faces were presented in a pseudorandomized order whereby each CS was presented four times for 800ms (i.e., 72 trials per condition, 216 trials in total). rCS+ were followed by the USa, nrCS+ were followed by the USb, and the remaining CS were unpaired (CS-). The auditory USs started 600 ms after the CS onset. The ITI was  $1300 \pm 300$  ms. A 100% reinforcement schedule was adopted. Participants were instructed to pay attention to the centre of the screen. Their pupil responses were continuously tracked. After the acquisition phase, participants completed the CS-US matching task and the pair comparison task.

**Day 2: reactivation and extinction**

**Reactivation.** Participants returned to the station where acquisition took place. They put on the headset and completed a similar calibration procedure carried out by the eye tracker. A US reminder cue, randomized across participants, was given, followed by a 10-minute session in which participants watched a neutral documentary on wild wife animals.

**Extinction.** Immediately after the movie viewing participants received extinction in which all CSs were presented without any aversive stimuli. The duration of the stimulus presentation and ITIs were identical to Day 1. Each CS was presented ten times (i.e., 90 trials per condition, 270 trials in total) to ensure complete extinction. To avoid participants from unnecessary head movements due to prolonged sitting, a small break was put in place in the middle of extinction (i.e. after the 5<sup>th</sup> trial). Participants completed the pair comparison task after the extinction learning.

**Day 3: reinstatement and re-extinction**

**Reinstatement.** Each participant received a random presentation of un-signalled USa and USb (4 times each).

**Re-extinction.** Immediately after the US presentation, re-extinction took place in which all CS were presented without any US. The number of trials, duration of CS presentation, and the ITI were identical to Day 1. After re-extinction, participants completed the Pair Comparison task and the US rating task. They were then thanked and debriefed.

## Data recording and Statistical Analysis

### Pupil responses

Eye-tracking was performed with an Eyelink 1000plus (SR Research Ltd, Ottawa, Canada) with a tower-mount setting. A standard nine-point calibration was used to determine the gaze position on the screen, and pupil diameter (in arbitrary units) were recorded at a sampling rate of 1000 Hz.

Pupil responses were preprocessed in the PsPM toolbox according to Kret & Shie's (2019) recommendation. First, the right pupil raw data was imported; Pupil size samples outside of a predefined feasible range were rejected ( $< 500$  or  $> 10000$ ). Second, invalid data, defined as contiguous missing data points larger than 75 ms, were removed from the current analysis. Third, missing data were interpolated and smoothed with a zero-phase low-pass filter with a cut off frequency of 4 Hz. Finally, the interpolated and filtered samples were then z-transformed. Non-normalized pupil data were used for between-session comparisons.

We applied the general linear convolution model, implemented in PsPM Pupil Responses Module developed by Korn et al. (2017), to estimate the anticipatory pupil responses. To compare the conditioned responses in each session, linear mixed models (LMMs) with fixed and random effects were applied using the following R formula:  $\text{pupil responses} \sim \text{CS type} + 1|\text{subject}$ .

We compared this full model with a reduced model without *CS type* as a fixed factor using a Chi-squared test to infer the significance of the fixed factor in the models. Estimated marginal means (EMMs) were generated to compare and contrast the mean values of pupil responses of each CS. Extinction was divided into two parts with respect to the break during the task; early extinction consisted of 270 trials, whereas late extinction comprised the remaining 270 trials.

To examine recovery of fear after extinction, we performed a separate LMM including the non-normalized pupillary responses from late extinction and the first block of re-extinction trials, which consisted of the first 54 trials of CS presentations, following reinstatement. The following R formula was applied: pupil responses ~ CS type\*session + 1|subject.

This full model was compared with a reduced model without any fixed factors and a chi-squared test was conducted to infer the significance of the model. EMMs were applied to compare and contrast the estimated mean differences of pupil responses for each CS type across the two sessions.

**Contingency awareness.** The sensitivity index  $d'$  (Green & Swets, 1966) was employed to detect how well participants recognized the stimulus category of each CS in the CS-US matching task. A  $d'$  score of 0 indicates that the detectability was at a chance level. The index was tested against value 0 by one-sample t-test. In addition, repeated-measure ANOVAs and paired  $t$ -tests were applied to test the conditioning-induced changes in the CS on the pair comparison task.

**Ratings of US valence and arousal.** Paired  $t$ -tests were applied to evaluate the arousal and valence of the two USs.

Significance was taken at  $p < .05$  and Cohen's  $d$  and its 95% confidence interval were reported as a measure of effect size. All analyses were conducted in R 3.5.2 using the lmerTest (Kuznetsova et al., 2017), emmeans (Lenth, 2016), and psych (Revelle, 2019) packages. Graphs were created using the ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2020) packages.

## 5.3 Results

### Pupil responses

Pupil responses across the experimental sessions are illustrated in Figure 5-2. On Day 1, LMMs revealed a significant effect of CS type in the model,  $\chi^2(2) = 34.68, p < .001$ . The effect of conditioning was observed in which rCS+ and nrCS+ elicited significantly higher pupil responses than the CS- after repeated pairings (Table 5-1). rCS+ and nrCS+ did not differ significantly,  $t(74.1) = -0.47, p = .642$ . (Table 5-2).

On Day 2, the US associated with the rCS+ was reminded 10 minutes before extinction. After early extinction, we observed higher pupil responses for the nrCS+ relative to the CS-. The rCS+ invoked comparable pupil responses as the CS- (Table 5-1). Following late extinction, there were no differences across the CS,  $\chi^2(2) = 2.42, p = .298$ .



Figure 5-2 Pupil responses during the (a) acquisition, (b) early extinction, (c) late extinction, and (d) the first run of re-extinction

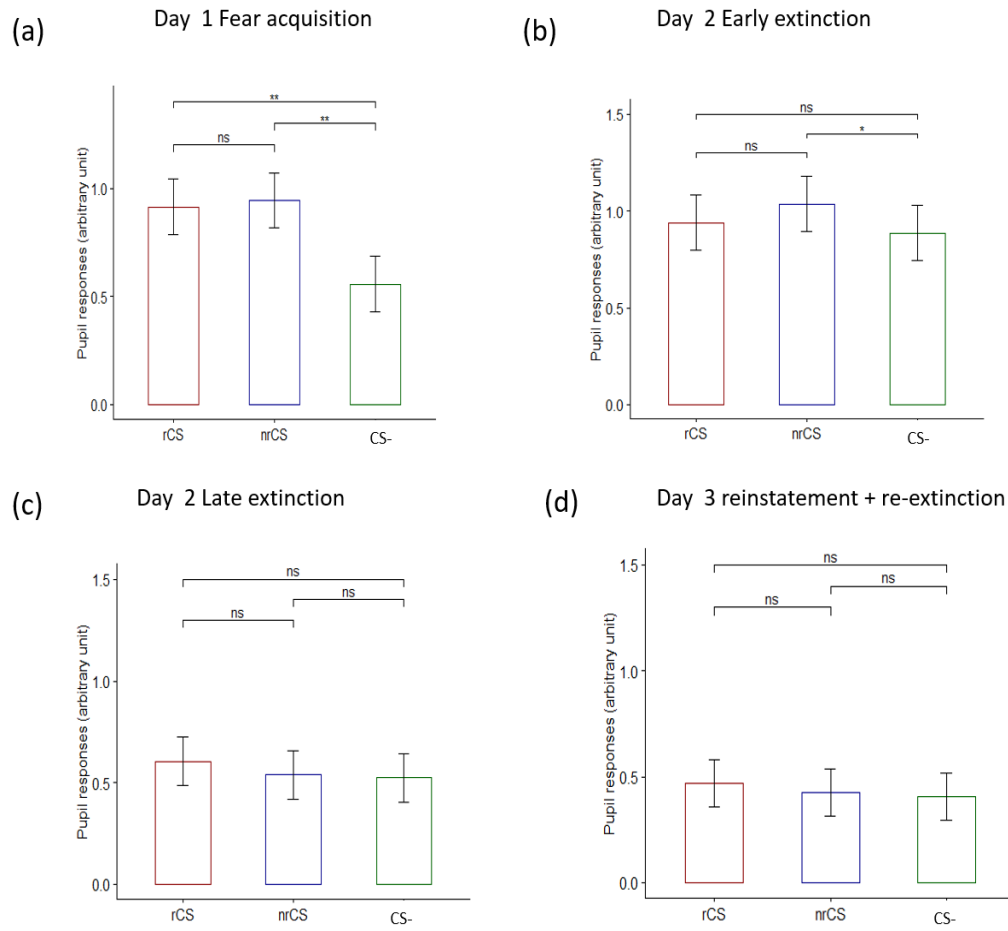


Table 5-1. Linear mixed-effects modelling (LMM) results for pupil responses on each experimental day (N = 36)

Day: session	Fixed factors	Parameter estimates	SE	df	t-value	p
Day 1: Acquisition	CS-	0.56	0.13	42.94	4.39	<.001
	rCS+	0.36	0.06	72.00	5.53	<.001
	nrCS+	0.39	0.06	72.00	6.03	<.001
Day 2: Early extinction	CS-	0.89	0.14	43.41	6.27	<.001
	rCS+	0.05	0.07	72.00	0.72	.477
	nrCS+	0.15	0.07	72.00	2.02	.046
Day 2: Late extinction	CS-	0.52	0.12	41.75	4.42	<.001
	rCS+	0.08	0.06	72.00	1.47	.147
	nrCS+	0.01	0.06	72.00	0.25	.802
Day 3: Re-extinction	CS-	0.41	0.11	61.52	3.64	<.001
	rCS+	0.06	0.10	72.00	0.64	.525
	nrCS+	0.02	0.10	72.00	0.20	.842

Table 5-2 Estimated means differences and standard errors for pupil responses on each experimental day

Day: session	contrast	Parameter estimate differences	SE	df	t-value	p
Day 1: Acquisition	CS- vs rCS	-0.36	0.07	74.1	-5.45	<.001
	CS- vs nrCS	-0.39	0.07	74.1	-5.92	<.001
	rCS+ vs nrCS+	-0.03	0.07	74.1	-0.47	.462
Day 2: Early extinction	CS- vs rCS	-0.05	0.08	74.1	-0.71	.483
	CS- vs nrCS	-0.15	0.08	74.1	-2.00	.045
	rCS+ vs nrCS+	-0.10	0.08	74.1	-1.30	.200
Day 2: Late extinction	CS- vs rCS	-0.08	0.06	74.1	-1.48	.322
	CS- vs nrCS	-0.01	0.06	74.1	-0.25	.967
	rCS+ vs nrCS+	0.06	0.06	74.1	1.20	.458
Day 3: Re-extinction	CS- vs rCS	-0.06	0.10	74.1	-0.63	.805
	CS- vs nrCS	-0.02	0.10	74.1	-0.02	.979
	rCS+ vs nrCS+	0.04	0.10	74.1	0.04	.902

On Day 3, pupil responses did not differ significantly across the type of CS,  $\chi^2(2) = 0.42, p = .809$ . rCS+ and nrCS+ did not evoke different pupil responses from CS- (Table 5-1) and they did not differ significantly from each other,  $t(74.1) = -0.43, p = .902$ . Fear recovery analysis did not reveal a significant difference including *session* or *session x CS type* in the full model,  $\chi^2(5) = 1.63, p = .897$ , suggesting that a reduced model without *CS type* or *session* as the fixed factors is preferred. The mean differences in pupil responses of each CS did not suggest a significant return of fear (Table 5-3).

Table 5-3 Linear mixed-effects modelling (LMM) results for pupil responses in computing the recovery of fear (N = 36)

<b>Fixed factors</b>	<b>Parameter estimates</b>	<b>SE</b>	<b>df</b>	<b>t-value</b>	<b>p</b>
CS-	157.59	37.27	163.30	4.23	<b>&lt;.001</b>
rCS+	17.07	45.62	178.19	0.37	.709
nrCS+	4.21	45.62	178.19	0.09	.927
Session	0.25	45.99	178.64	0.01	.996
rCS x session	-6.16	64.98	178.19	-0.10	.925
nrCS x session	-40.21	64.78	178.42	-0.62	.536

<i>Late extinction vs</i> <i>Re-extinction</i>	<b>Parameter estimate</b> <b>differences</b>				
CS-	-0.25	46.8	184	-0.01	.996
rCS+	5.91	46.7	184	0.13	.899
nrCS+	39.96	46.3	183	0.86	.389

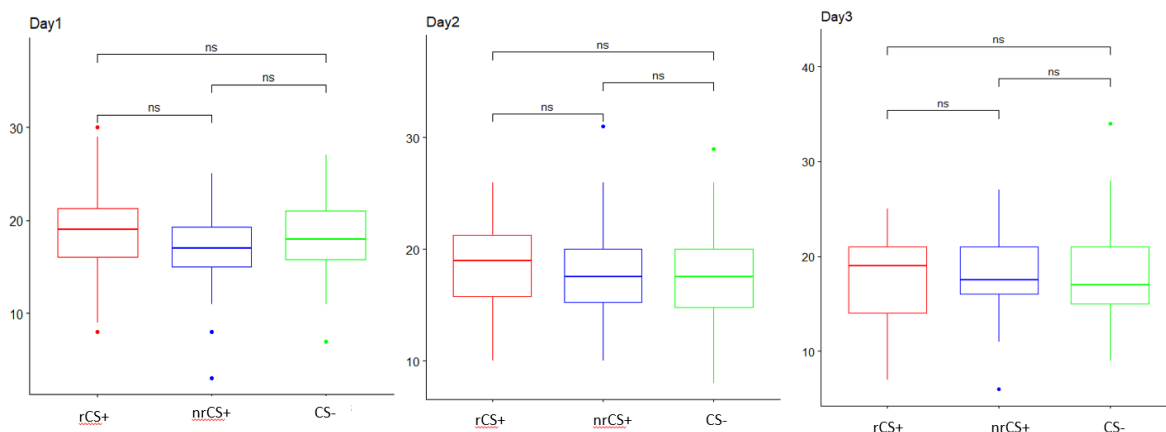
### **Contingency Awareness**

On the CS-US matching task, participants were able to report CS- US associations above chance,  $d' = 0.22$ , and it was differed significantly from zero  $t(33) = 3.05, p = .004$ .

The hit rates ( $M = 65.00\%$ ,  $S.D. = 15.48\%$ ) were on average higher than false alarm rates ( $M = 56.03\%$ ,  $S.D. = 20.23\%$ ). The  $d'$  of USa-paired-faces (0.13) and of USb-paired faces (0.15) were significantly different from zero,  $t(34) = 2.00$ ,  $p = .053$  &  $t(34) = 3.05$ ,  $p = .024$  respectively, suggesting that participants demonstrated some contingency awareness of the CS-US associations that were selective for the aversive stimuli. To evaluate whether CS-US awareness was a necessary condition for learning to take place, we included a subgroup of participants with very low detectability of the CS faces for comparing their pupil responses during Acquisition (Day1). This group of participants ( $n = 18$ ) reported the associations of the CS-US at a chance level,  $d' = -0.096$ ,  $SD = 0.27$ ,  $t(17) = -1.48$ ,  $p = 0.156$ ). The results of their pupil responses during Acquisition were presented in Appendix A Table 8-5..

On the pair comparison task, participants did not report preferences towards a specific type of CS after conditioning on Day 1,  $F(2,70) = 0.76$ ,  $p = .472$ . Exploratory analyses revealed that 27 participants (out of 36) acquired a preference for the CS- over the rCS+ and nrCS+,  $F(2,52) = 3.57$ ,  $p = .035$ . Participants did not show any preferences for either type of the faces after extinction on Day 2,  $F(2, 70) = 0.74$ ,  $p = 0.482$ . and re-extinction on Day 3,  $F(2,70) = 0.024$ ,  $p = .977$  (Figure 5-3).

Figure 5-3 Results of Pair Comparison across days. A higher score on y-axis suggests a preference towards a particular type of CS (rCS, nrCS, CS-).



Note. rCS+ : reminded CS+; nrCS+: non-reminded CS+ ; CS- : CS minus

## US rating

The male scream was rated more aversive than the female scream,  $t(35) = -3.86$ ,  $p < .001$ . However, the female scream was rated more aroused than the male scream,  $t(35) = 3.35$ ,  $p = .001$ .

## Questionnaires

With respect to the PANAS, participants reported a mean score of 23.62 (SD = 5.93) on the positive affect scale, and a mean score of 16.15 (SD = 6.37) on the negative affect scale. There were no significant changes in positive affect and negative affect across the three experimental days,  $F(2,98) = 1.82$ ,  $p = .168$  and  $F(2,98) = 0.25$ ,  $p = .779$  respectively.

## 5.4 Discussion

The current study investigated the effect of a retrieval-reminder procedure in preventing the return of fear in a multi-CS paradigm. Multiple different faces were paired

with aversive human screams during conditioning, and the scream was used to reactivate the fear-associated memories 10 minutes before extinction learning the following day. Return of fear was tested in a reinstatement test on Day 3. Our findings did not show evidence for the reinstatement of differential pupil responses of the conditioned stimulus following the reinstatement test.

Despite a multitude of CS-US and CS-no US pairings during conditioning, participants were able to demonstrate a remarkable affective associative learning measured by the CS-US matching task and pupil responses. Participants' awareness of the CS-US association was weak in the present sample ( $d' = 0.22$ ) but was comparable to multi-CS studies with  $d'$  ranging from 0.01 to 0.78 (Rehbein et al., 2014; Steinberg et al., 2012a; Steinberg et al., 2013). Our study gives support to the notion that associative learning of emotional materials could take place under limited, (Junghöfer et al., 2017) or even the absence of contingency awareness (Brockelmann et al., 2011a; Roesmann et al., 2020; Steinberg et al., 2012a). Although contingency awareness was not statistically demonstrated in the pair comparison task using the full sample on Day 1, the majority of the participants (27 out of 36) showed a categorical preference for the faces that were not paired with any scream. We performed additional analyses on these 27 participants to compute the conditioning, extinction, and return of fear index using pupil responses. The subset sample yielded the same patterns of results, and we presented the results with a full sample here.

The role of awareness in Pavlovian conditioning has been a heated debate since the early 2000s when Lovibond and Shank (2002) proposed contingency awareness as a necessary condition for effective conditioning. The debate remains unresolved until now. In a recent meta-analytic review, Merten et al. (2020) concluded that there was no convincing evidence for fear conditioning outside of awareness. Conversely, investigations of neural

activity, physiological measures, and behavioural ratings have consistently shown that affective learning could occur without conscious awareness (Brockelmann et al., 2011b; Lipp et al., 2014; Oyarzún et al., 2019; Tabbert et al., 2005). The current study provides evidence for the latter, supporting that full awareness of the CS-US association may not be necessary for affective learning, and that there is a dissociation of affective learning in different measures.

During early extinction, the non-reminded CS+ continued to elicit a higher pupil response than the CS- during early extinction, whereas the reminded CS+ evoked a comparable response relative to the CS-. However, we did not observe a significant statistical difference between the non-reminded CS+ and the reminded CS+ in their pupil responses. While Schiller et al. (2013b) reported diminished prefrontal involvement during late extinction in an fMRI single CS conditioning paradigm, suggesting a reminder cue might evoke a distinct pattern of extinction. Our findings did not provide support for this notion. Since fMRI is costly and labour-intensive, and is not pragmatic for use in a therapy room, there is a need to find a real-time indicator of reconsolidation following memory reactivation, such that researchers and clinicians could gauge the process of reconsolidation and the timing of this window for implementing psychological interventions more effectively.

Contrary to our hypothesis, there was no evidence for the return of fear measured by pupil responses on Day 3 in the present study. The study of the return of fear phenomenon in a multi-CS conditioning paradigm is relatively new. It is possible that there was a genuine absence of return of fear because the affective learning was weak and implicit in the multi-CS conditioning paradigm. Nevertheless, a similar null finding in the pupil response has been reported by Rosesmann et al. (2019) using the same Multi-CS experimental protocol,

but they observed differential neural activation pattern measured by the MEG between the CS+ and the CS- following a reinstatement test on Day 3, independent of how the CS+ was reactivated on Day 2. It is, therefore, speculated that pupil responses, despite being a sensitive measure of conditioned responses (Leuchs et al., 2017b, 2019), might not be sufficiently sensitive to capture the subtle reinstatement phenomenon, particularly within a multi-CS paradigm. The dissociation between neural and physiological measures has also been reported in other research groups. For instance, Lonsdorf et al. (2014) did not find a reinstatement effect measured by skin conductance responses but observed an increased activation of the fear network in the fMRI data. Given the dissociation between neural correlates and behavioural responses noted in previous studies (Lonsdorf et al., 2014), further investigation could consider using neuroimaging such as fMRI to unveil the pathways underlying reinstatement of fear in a multi-CS implicit learning paradigm.

### **Limitations**

First, the valence and arousal of the female scream were rated differently from the male scream, which might affect the perception of threat and the behavioural or physiological measures of fear responses. We acknowledge this limitation and have minimised this effect by fully randomising the allocation and presentation order of the two screams during conditioning, reactivation and reinstatement. Second, the use of Caucasian faces in an Asian sample might induce a cross-race effect, the relative ease when recognizing the same race as compared to cross-race faces (Meissner & Brigham, 2001). Selecting Caucasian faces as stimuli in the present study was intended to allow a direct comparison of this experiment to other experiments included within this thesis that were conducted amongst Caucasian participants. Nevertheless, the participants of the current study were able to demonstrate the



conditioning effect measured by pupil responses on Day 1, hence we believe the impact of the cross-race effect on learning is minimal in the present study. Third, we acknowledge that the behavioural measures (CS-US matching task, and Pair Comparison) conducted on Day 1 reflect participants' memories of the CS-US associations after the conditioning, but not the contingency awareness occurred *during* the learning. While we believe that the relationship between memory performance after conditioning and contingency awareness during conditioning are correlated, future studies could consider including an intermittent online US expectancy measure to capture the awareness during the encoding.

## **5.5 Conclusion**

Our findings suggest that presenting a US reminder cue before extinction may facilitate extinction learning for the reminded conditioned stimuli in the early phase. The impact of the reminder-extinction procedure on the return of fear remains inconclusive in the present study as there was no reinstatement of fear in the CS+. Given the differential pupil responses were no longer present for the reminded CS+ in the early extinction phase, we seek to further examine the neural mechanisms of reconsolidation-extinction and reinstatement using fMRI in the next experiment.

## **Chapter 6**

**Experiment 4: The impact of a US reminder cue on the neural mechanisms of extinction and the return of fear in multi-CS conditioning**

## 6.1 Introduction

Processing of fear-related memory is central to understanding debilitating fear- and anxiety-related disorders (Bisaz et al., 2014; LeDoux & Pine, 2016). In the past few decades, much progress has been made in uncovering the neural circuits underlying the formation and storage of fear memories. Previously we have reviewed the neural correlates of fear acquisition and extinction in Chapter 2. In brief, the amygdala and its subnuclei, as well as the hippocampus, are key nodes for encoding threat-relevant information, while the hippocampus, ventromedial prefrontal cortex (vmPFC), and dorsolateral prefrontal cortex (dlPFC) are implicated in extinction learning (Milad & Quirk, 2012). Multi-CS conditioning studies also report the involvement of the dlPFC during fear learning and its extinction (Steinberg et al., 2013a; Steinberg et al., 2013b).

In Experiment 3 (Chapter 5), we found a reminder-specific differential pupil response following early extinction and suggested that a reminder cue might evoke a distinct pattern of learning in early extinction specifically for the reminded CS. Based on this finding, the current experiment examined the neural mechanisms underlying the reminder-specific extinction learning in a multi-CS conditioning paradigm. In addition, we investigated further whether there was neural evidence for the return of fear in the multi-CS conditioning paradigm because we did not find the reinstatement of fear measured by pupillometry in the previous experiment. It was hypothesized that enhanced prefrontal activity, including the vmPFC and dlPFC in the contrast between non-reminded CS versus reminded CS during extinction learning would be observed in the present study. In addition, increased blood-oxygen-level-dependent (BOLD) signal in the amygdala and the hippocampus during re-extinction in the same contrast would be observed.

## 6.2 Methods

### *Participants*

As this is the first study to examine the neural mechanism of retrieval-extinction in a multi-CS conditioning paradigm, we did not conduct a power analysis in advance but we referenced a previous EEG/fMRI study in multi-CS conditioning in which a sample of 21 subjects was sufficient to detect the signals related to our task of interest (Steinberg et al., 2013).

Twenty-two healthy Chinese university students ( $25.68 \pm 3.93$  years; males: females = 11: 11) were recruited in this study. All participants reported normal hearing, normal or corrected-to-normal vision. They were excluded if they self-reported current or history of psychiatric/neurological illnesses. To ensure that participants did not show contraindications for undergoing magnetic resonance imaging, the following additional exclusion criteria were applied: 1) installation of metallic implants or medical apparatus (e.g. pacemakers or artificial joints) in the body, 2) claustrophobia, 3) pregnancy (for women); 4) having tattoos. The study protocol was approved by the ethics committee of the university (EA170915) and was implemented in accordance with the Declaration of Helsinki. Written informed consent forms were obtained from all participants.

### *Materials*

**Unconditioned stimuli.** Two aversive tones (duration: 1000ms; USa: a female scream, USb: a male scream) were used. Human screams were used in previous multi-CS conditioning paradigm (e.g. Roesmann et al., 2020) and were chosen in our experiment because they had

relevant association with the CS in our experiment, which were all female or male faces. They were delivered binaurally through MRI-compatible headphones inside the scanner. The assignment of US was balanced across participants.

**Conditioned stimuli.** Fifty-four grayscale images displaying faces (27 females) with neutral expressions were pseudo-randomly split into three conditions: 18 CS+ (paired with USa), 18 CS+ (paired with USb) and 18 CS- faces (unpaired during conditioning). The pictures were randomly assigned to three groups with equal gender ratio; and the condition of each group of pictures (i.e. whether they were reminded CS+, non-reminded CS+ or CS-) was counterbalanced across participants during acquisition.

**CS-US matching task.** Explicit knowledge of the stimulus category was assessed using a computerized CS-US matching task after acquisition on Day 1. All 54 CS were pseudo-randomly presented for 600 ms, followed by two questions. First, participants were asked to indicate for each face whether it was paired with a scream during conditioning (stimulus category: CS+ vs. CS-) on a Likert scale from -4 (*surely there was no scream*) to 4 (*surely there was a scream*). Second, they were asked to indicate whether the faces were paired with a male scream or a female scream on a Likert scale ( -4 = *surely female* to 4 = *surely male*). For practice, they completed three trials prior to the start of the task.

**Pair comparison task.** It was an indirect measure to evaluate participants' awareness of CS-US contingency pairings. On this task, pairs of CS+ and CS- faces were shown on the computer screen. For each pair of faces, participants were asked to decide which face they preferred in a binary forced-choice format. Three distinct versions, each with 27 CS+ and

CS- trials, were developed such that each CS was only rated once after each experimental session.

**US rating task.** To identify the perceived valence and arousal of the USs, participants were asked to rate the valence and arousal of each tone on a 4-point Likert scale.

**Questionnaires.** Participants completed a computerized Positive and Negative Affect Scale (PANAS; Watson et al., 1988) to evaluate their mood state before the start of each experimental session. The PANAS consisted of 20 adjectives of positive and negative affective states (e.g., Positive: content, Negative: Afraid). Participants were asked to indicate how they felt at the moment on a 5 -point Likert scale from 1 = *not at all* to 5 = *extremely*. Higher scores reflect higher positive or negative affect.

The CS and US were presented using E-Prime 1.0 software (Psychology Software Tools, Pittsburgh, PA). CS-US matching task and Pair comparison task were conducted on a computer using Presentation (Neurobehavioral Systems, Albany, CA). Responses of the US rating task and questionnaires were collected using Inquisit 5 (Millisecond Software, Seattle, WA).

### *Brain imaging*

Data were acquired using a 3T Philips Achieva MR scanner (Philips Medical Systems, The Netherlands) equipped with an 8 - channel SENSE head coil. Head movement was restricted using foam cushions. High resolution anatomical T1-weighted images were acquired using a magnetization-prepared rapid gradient echo (MPRAGE) sequence with the

following parameters: 160 sagittal slices, repetition time (TR) = 6.9 ms, echo time (TE) = 3.2 ms, matrix = 240 x 240, FOV = 240 × 240 x 160 mm<sup>3</sup>, flip angle = 8°, voxel size = 1 × 1 × 1 mm<sup>3</sup>). During visual presentations, task-based BOLD imaging was collected using a T2\*-weighted echo-planar imaging (EPI) sequence (39 slices, TE = 30 ms, TR = 2000 ms, matrix = 124 × 124, FOV = 230 × 230 mm<sup>2</sup>, flip angle = 90°, voxel size = 1.6 × 1.6 × 3.5 mm<sup>3</sup>). The duration of the scans were 33.9 minutes, 33.4 minutes, and 16.8 minutes on Day 1(Acquisition), Day 2(Reactivation) and Day 3(Re-extinction) respectively.

### **Design and Procedure**

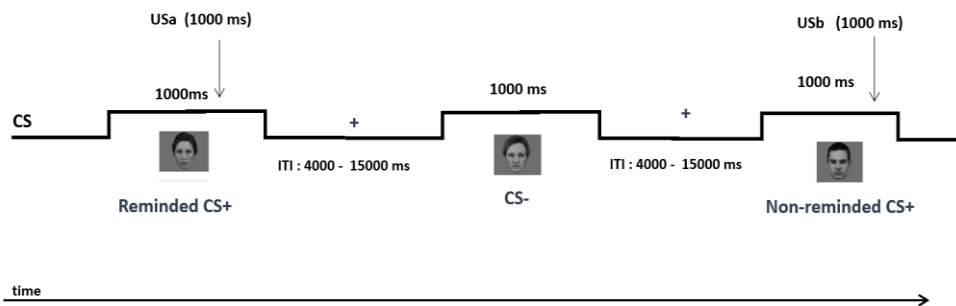
The experiment consisted of three sessions that took place on three successive days: Acquisition (Day 1), Reactivation and Extinction (Day2), and Reinstatement and Re-extinction (Day 3). The total experimental time was about 5 hours. The detailed procedures are described below:

#### Day 1: acquisition

Participants first completed the questionnaires and habituation of the CS outside the scanner, followed by acquisition inside the scanner. To meet the specific requirement of fMRI analysis, each CS was presented for 1000 ms, with an inter-trial interval of 4000 to 15000 ms (Figure 6-1). The CS+-US contingency was maintained at 100% and the CS- was presented without the US. We jittered the onset of the US from 200 to 700 ms following the onset of the CSs. The presentation order, the time of onset of the CS, and the time of onset of the US were arranged with reference to the Functional MRI of the Brain (FMRIB) Software Library (FSL)'s design efficiency (Jenkinson et al., 2012), in which 5000 sequences were produced, and the model with the highest efficiency for estimation of the BOLD

response was chosen. During acquisition, each CS was presented four times (216 trials in total), and the overall presentations were divided into four runs (8.1 to 8.7 min each, 33.9 mins in total). Following the acquisition, they completed the Pair Comparison Task and CS-US matching task outside the scanner.

Figure 6-1 Experimental timeline during acquisition in the scanner. A total of 54 grayscale images displaying faces (27 females) with neutral expressions were pseudo-randomly split into three conditions: 18 reminded CS+ (paired with USa), 18 non-reminded CS+ (paired with USb) and 18 CS- faces (unpaired during conditioning). Each CS was presented 4 times ( $54 \times 4 = 216$  trials). All CS were presented for 1000 ms. The onset of the US was jittered from 200 to 700 ms following the onset of the CS, and it lasted for 1000 ms. The inter-stimulus interval varied pseudo-randomly between 4000 to 15000 ms.



## Day 2: reactivation and extinction

Participants completed the PANAS outside the scanner upon arrival. They then completed the reactivation task inside the scanner, followed by resting-state scanning. During the resting state, participants were instructed to keep their eyes open, relax their mind, and keep their head still. The resting-state scan lasted for nine minutes.

Following the resting-state scan, extinction continued inside the scanner. During extinction, each CS was presented eight times without reinforcement. The ITI was between 2000 ms and 6000 ms. The presentation of the CSs (a total of 432 trials) was divided into four runs. Each run lasted from 8.2 to 8.6 minutes, and the extinction lasted for 33.4 minutes



in total. Following extinction, participants completed post-task resting-state scan and completed the Pair Comparison outside the scanner.

### Day 3: reinstatement and re-extinction

Participants first completed the PANAS outside the scanner and then completed the reinstatement and re-extinction inside the scanner. During the reinstatement, each US was presented four times alone. Immediately after this, the CS were presented four times each, unreinforced. The presentation was divided into two runs and lasted for 16.8 minutes. After re-extinction, participants' preferences for the CS were obtained in the Pair Comparison task outside the scanner. They were then debriefed, thanked, and paid HK\$400 as remuneration.

All picture presentations in the MRI chamber were implemented using E-prime 1.0 (Psychology Software Tools, Pittsburgh, PA). Behavioural tasks were conducted on a computer using Presentation (Neurobehavioral Systems, Albany, CA).

### **fMRI preprocessing and processing**

Image preprocessing was carried out using FM RIPREP, an fMRI preprocessing pipeline recommended by Esteban et al. 2019. Each T1-weighted volume (T1w) was corrected for INU (intensity non-uniformity) and skull-stripped. Spatial normalization to the ICBM 152 nonlinear asymmetrical template was performed through nonlinear registration using brain-extracted versions of both T1 weighted volume and template. Brain tissues of cerebrospinal fluid, white matter, and gray matter were performed on the extracted T1w. Functional data were slice time corrected, and motion-corrected. This was followed by co-registration to the corresponding T1w using boundary-based registration with six degrees of

freedom. Motion correcting transformations, BOLD-toT1w transformation, and T1w-to-template warp were concatenated and applied using Lanczos interpolation. Physiological noise regressors were extracted using the component-based noise correction method (Behzadi, 2007). ICA-based Automatic Removal Of Motion Artifacts (ICA-AROMA) was used to remove motion artefacts (Pruim et al., 2015).

### *First-level analysis*

Functional data was first modelled at the subject level by fitting a voxel-wise General Linear Model (GLM) to the BOLD data acquired for each run. Each run was modelled separately and included the following task regressors: the time of onset of the CS, the time of onset of the US and six motion regressors. The US regressors were orthogonalized with respect to the reminded CS+ and the non-reminded CS+ regressors. Task regressors were modelled as event-related designs and convolved with a canonical gamma hemodynamic response function. The main contrast was to assess the effect of the *CS type*; specifically, the potential differences between the reminded CS+ (rCS) and the non-reminded CS+ (nrCS) during extinction and re-extinction. To this end, the following contrast maps were constructed:  $rCS > nrCS$  for each run on each experimental day. For complementary and exploratory analysis,  $rCS > CS-$  and  $nrCS > CS-$  contrasts were computed. In addition, CS+ was created by joining rCS and nrCS together and was compared against the CS- (i.e.  $CS+ > CS-$ ). As the primary aim of the present experiment was to explore extinction learning and re-extinction, the BOLD responses for each *CS type* were averaged across two runs to ensure that rCS and nrCS did not differ during early and late acquisition. The BOLD responses of each *CS type* were averaged in each run during extinction and re-extinction for more detailed comparison.

*Second-level analysis*

The contrast of parameter estimate (COPE) images of the CS+ > CS-, rCS > nrCS, rCS > CS-, nrCS > CS- were entered into a group mean model using FSL's randomize with 5000 permutations. Correction for multiple comparisons was performed using FSL's threshold-free cluster enhancement (TFCE) tool. Resulting contrast maps were thresholded at  $p_{\text{FWE}} < .05$  with a minimum cluster size of 10 voxels. Only statistically significant contrast maps were reported in the result section.

On top of whole-brain analyses, *a priori* regions of interest (ROIs) analyses on the amygdala, hippocampus, ventromedial prefrontal cortex, and dorsolateral prefrontal cortex were conducted. These ROIs were chosen because these structures were crucially involved in fear acquisition, extinction, and re-extinction (Agren, Engman, Frick, Björkstrand, et al., 2012; Schiller et al., 2013a). ROI anatomical masks were constructed based on the Brainnetome Atlas (Fan et al., 2016), encompassing the following regions (Table 6-1). ROI analyses were performed with a threshold of  $k > 10$  and  $p_{\text{FWE}} < .05$ , corrected for family-wise errors within the specified ROIs.

Table 6-1 Regions of Interest (ROIs) in the current study and their corresponding locations in the Brainetome Atlas

<i>ROIs</i>	<i>Corresponding regions in the Brainnetome Atlas</i>
Left dlPFC	A8vl_L, A946d_L, A946v_L
Right dlPFC	A8vl_R, A946d_R, A946v_R
Left vmPFC	A11m_L, A14m_L, A32sg_L
Right vmPFC	A11m_R, A14m_R, A32sg_R
Left hippocampus	cHipp_L, rHipp_L
Right hippocampus	cHipp_R, rHipp_R
Left amygdala	lAmyg_L, mAmyg_L
Right amygdala	lAmyg_R, mAmyg_R

With reference to Ernst et al (2019)'s findings which suggest that bilateral cerebellar lobules Crust 1 and VI are implicated fear acquisition, we ran additional analyses including these two brain regions as ROIs and the results were presented in Appendix A: Table 8-6 and Table 8-7. To estimate the return of fear recovery with reference to Schiller et al. (2013), we conducted two separate analyses exploring the differences between the reminded CS and the non-reminded CS in two ways. First, we compared the contrast of the reminded CS+ and non-reminded CS+ within the first run during re-extinction. Second, we compared this contrast between the trials in the last run of extinction (i.e., Extinction run 4) and the first run of re-extinction.

### 6.3 Results

#### Day 1: acquisition

##### *Contingency awareness*

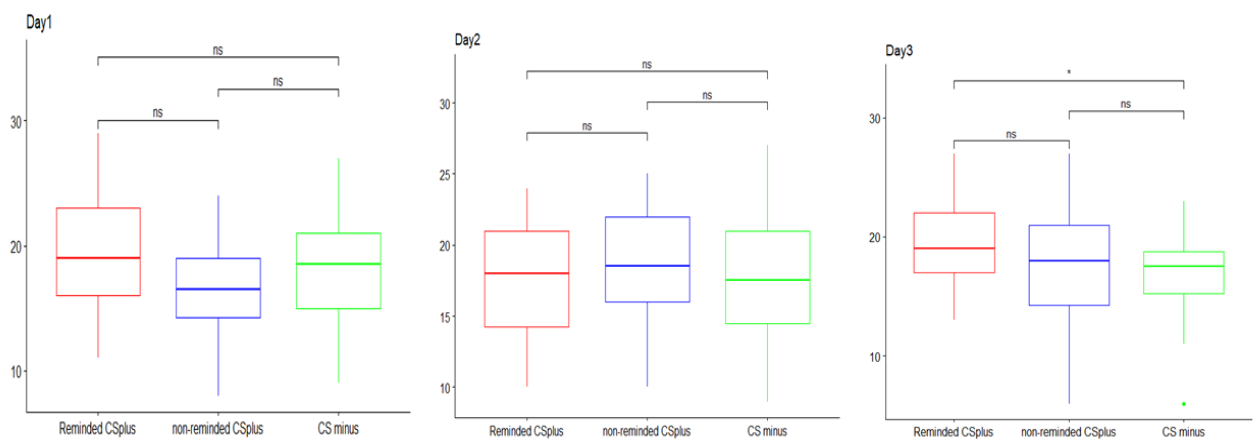
On the CS-US matching task, participants reported a low detectability of the CS-US association,  $d' = -0.11$ , which was at a chance level,  $t(21) = -1.38, p = .180$ . The detectability for female-scream-paired and male-scream-paired-faces were low,  $d' = -0.13$  and  $d' = 0.03$  respectively, and were not significantly different from zero,  $t(21) = 1.71, p = .101$  and  $t(21) = 0.60, p = .554$ .

On the Pair Comparison task, participants did not report preferences towards any groups of CS after conditioning,  $F(2, 42) = 1.63, p = .208$  (Figure 6-2). A subset of the participants (15 out of 22, i.e. 68% )acquired the hypothesized behavioural preferences to CS- faces,  $F( 2,28) = 3.34, p = .050$ .

No differences in the arousal and unpleasantness between the two USs were observed on the US rating task.

Figure 6-2 Results of the Pair Comparison across each experimental day.

*Note:* On each day, a pair of CS+ and CS- faces were shown side by side and participants were asked to indicate their preference to either face. A higher score on y-axis suggests a preference towards a particular type of CS (rCS, nrCS, CS-). \*  $p < .05$



*fMRI: Whole-brain analyses*

Regions that showed statistically significant differences in activation/deactivation across CS type, including MNI coordinates, family-wise corrected *p*-values for peak voxels for all statistically significant clusters, are summarized in Table 6-2. Because the primary goal of the present study was to explore extinction learning and the return of fear, the fear acquisition results were briefly described to ensure that the rCS+ and nrCS+ did not differ, and they were both different from the CS-. Overall, increased BOLD responses were found for the CS+ compared to the CS- in the areas of middle temporal gyrus encompassing the hippocampus, the medial orbital part of the superior frontal gyrus, and the inferior frontal gyrus in the early acquisition phase. Increased activity remained in the middle temporal gyrus and the inferior frontal gyrus for the CS+ > CS- contrast in the late acquisition phase. In addition, the superior temporal gyrus and the cuneus showed increased activation in the late acquisition phase. Throughout acquisition, the fusiform area was significantly deactivated during CS+ presentations relative to the CS-.

*fMRI: ROI- analyses*

The CS+ relative to the CS- evoked higher activation in the bilateral amygdala, hippocampus, and dorsolateral prefrontal cortex in the early acquisition phase (Table 6-3). The left hippocampus remained to have stronger activation in the CS+ compared to the CS- in the late acquisition phase. Consistent with our hypothesis, rCS+ and nrCS+ did not differ throughout the acquisition.

Table 6-2 Localization and statistics for whole-brain analysis for Day 1 (Acquisition).

	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{FWE-corr}$	
Acquisition (Run 1 & 2)	CS+ > CS-	Middle temporal gyrus	L	32841	-22	-68	-56	3.54	0.008	
		Middle temporal gyrus	R		70	-18	-6	3.54		
		Cerebellum	L		-10	-76	-14	3.54		
		Superior temporal gyrus	R		52	-12	-6	3.54		
		Hippocampus	R		16	-4	-16	3.54		
			R		32	-6	-16	3.54		
			L		-26	-12	-6	3.54		
		Middle frontal gyrus, orbital part	L	31	-26	32	-22	3.16	0.033	
		Middle frontal gyrus, orbital part	L		-26	32	-22	3.16		
		Superior frontal gyrus, orbital part	L		-22	30	-24	3.04		
		Inferior frontal gyrus, orbital part	L		-32	36	-20	2.75		
		Inferior frontal gyrus, orbital part	L	26	-42	42	-8	3.09	0.048	
		Inferior frontal gyrus, orbital part	L		-42	42	-8	3.09		
		Middle frontal gyrus, orbital part	L		-34	44	-6	3.09		
			CS- > CS+	Fusiform gyrus	R	942	38	-42	-24	3.54
		Fusiform gyrus	R		42	-56	-20	3.54		
		Inferior occipital gyrus	R		34	-88	-16	3.54		
		Lingual gyrus	R		20	-90	-10	3.54		
		Middle occipital gyrus	R		28	-92	2	3.54		
Acquisition (Run3 & 4)	CS+ > CS-	Middle temporal gyrus	L	4243	-42	-2	-16	3.54	0.008	
			L		-38	-44	16	3.54		
			L		-36	-38	8	3.54		
		Middle temporal gyrus	L		-66	-48	8	3.54		
		Middle temporal gyrus	L		-64	-40	6	3.54		

	Middle temporal gyrus	L		-68	-34	4	3.54	
	Middle temporal gyrus	L		-68	-22	2	3.54	
	Cerebellum	L	4127	-20	-54	-24	3.54	0.034
	Cuneus	L		-6	-94	26	3.54	
	Calcarine	L		-6	-94	12	3.54	
	Calcarine	R		12	-92	10	3.54	
	Calcarine	R		14	-74	8	3.54	
	Calcarine	R		26	-74	6	3.54	
	Calcarine	L		-8	-92	6	3.54	
	Superior temporal gyrus	R	4030	52	0	-8	3.54	0.038
		R		34	-36	16	3.54	
	Middle temporal gyrus	R		66	-50	16	3.54	
		R		62	12	0	3.54	
	Middle temporal gyrus	R		70	-44	-2	3.54	
	Superior temporal gyrus	R		46	-16	-4	3.54	
		R		42	-26	-4	3.54	
	Inferior frontal gyrus, triangular part	R	31	58	24	0	3.35	0.027
	Inferior frontal gyrus, triangular part			58	24	0	3.35	
CS- > CS+	Fusiform gyrus	R	183	42	-40	-24	3.54	0.046
	Inferior Occipital gyrus	R	121	30	-88	-12	3.54	0.043

*Note.* No differences were observed for the rCS+ > nrCS+ contrast during the acquisition phase.



Table 6-3 Localization and statistics for ROI-analyses for Day 1 (Acquisition).

	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{\text{FWE-corr}}$
Acquisition (Run 1 & 2)	CS+ vs CS-	Amygdala	L	362	-26	6	-26	3.54	0.004
					-26	6	-26	3.54	
					-26	2	-24	3.54	
					-8	-8	-12	3.54	
					-16	-8	-12	3.54	
					-12	-6	-12	3.54	
		Amygdala	R	516	-26	-10	-8	3.35	0.001
					38	-2	-20	3.54	
					38	-2	-20	3.54	
					34	-2	-20	3.54	
					32	-6	-16	3.54	
					16	-4	-16	3.54	
		Dorsolateral PFC	L	228	16	-8	-14	3.54	0.030
					20	-16	-12	3.54	
					-44	34	20	3.54	
					-44	34	20	3.54	
					-34	26	24	3.54	
					-52	32	26	3.35	
		Dorsolateral PFC	R	64	-48	24	26	3.24	0.030
					-50	30	16	3.04	
52	38				26	3.35			
52	38				26	3.35			
48	38				24	3.16			
Hippocampus	L				115	38	56	30	
		56	30	20		3.04			
		56	22	30		2.73			
					-8	-8	-12	3.54	0.011

---

					-8	-8	-12	3.54	
					-16	-8	-12	3.54	
					-12	-6	-12	3.54	
					-34	-6	-12	3.35	
				93	-16	-28	-10	3.54	0.022
					-16	-28	-10	3.54	
				78	-36	-24	-8	3.54	0.010
					-36	-24	-8	3.54	
				32	-12	-44	-4	3.54	0.018
					-12	-44	-4	3.54	
					-18	-46	-4	3.35	
		Hippocampus	R	626	36	-4	-18	3.54	0.001
					36	-4	-18	3.54	
					42	-10	-16	3.54	
					32	-6	-16	3.54	
					16	-4	-16	3.54	
					18	-18	-14	3.54	
					16	-8	-14	3.54	
Acquisition (Run 3 & 4)	CS+ vs CS-	Hippocampus	L	30	-40	-18	-10	3.54	0.041
					-40	-18	-10	3.54	
					-40	-26	-8	3.54	

---

## Day 2: reactivation and extinction

### *Contingency awareness*

After extinction, participants did not report any preferences to either group of the faces,  $F(2,60) = 1.44$ ,  $p = 0.245$  on the pair comparison task (Figure 6-1).

### *fMRI: Whole-brain analyses*

No significant differences were found in all contrasts (CS+ vs CS-, rCS+ vs rCS-, rCS+ vs CS-, and nrCS+ vs CS-) during early extinction and late extinction.

### *fMRI: ROI-analyses*

During early extinction, increased BOLD responses were observed in the right dorsolateral prefrontal region in the rCS+ relative to the nrCS+ (Figure 6-2). No significant differences were found in the CS+ vs CS- contrast.

During late extinction, increased BOLD responses were found in the amygdala for the CS+ compared to the CS-. The increased activation was likely driven by responses to the nrCS+, as the enhanced BOLD responses were only observed in the contrast between nrCS+ and CS-. Importantly, reduced activation was found for the rCS+ relative to the nrCS+ in the left dorsolateral prefrontal region (Table 6-4).

Figure 6-3 a) Right dorsolateral prefrontal region reactivity in the rCS > nrCS contrast during early extinction on Day 2 (30, 22, 46,  $p_{\text{FWE-corr}} = .033$ ). b) Left dorsolateral prefrontal region reactivity in the nrCS > rCS contrast during late extinction on Day 2 (-44, 42, 22,  $p_{\text{FWE-corr}} = .038$ )

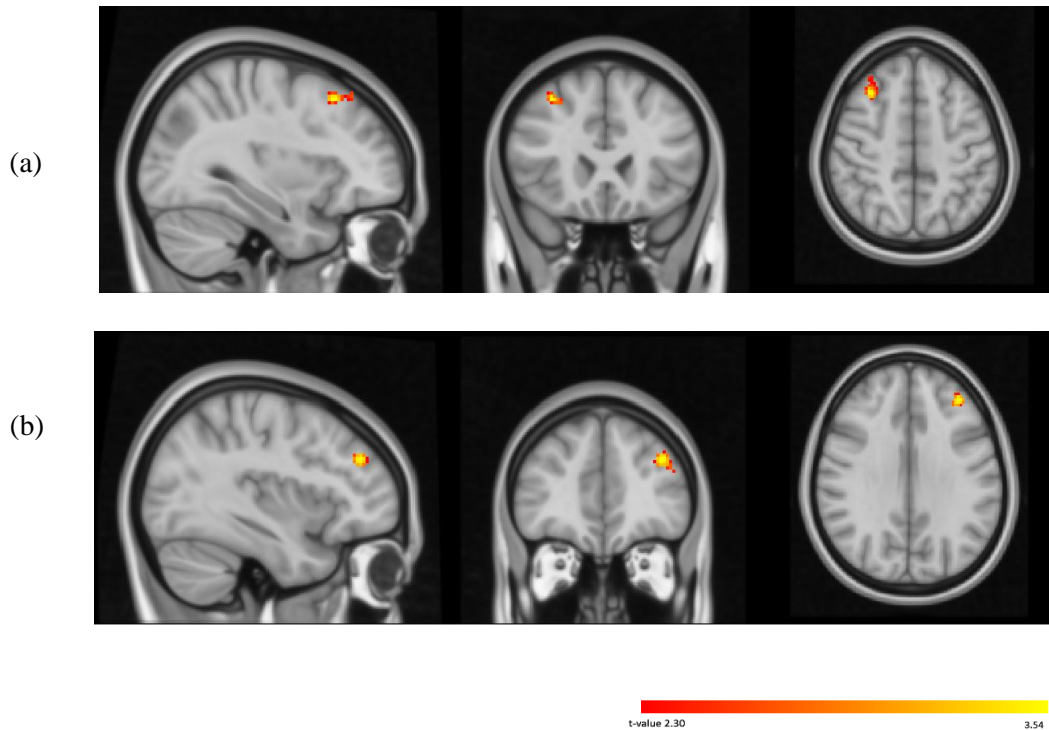


Table 6-4 Localization and statistics for ROI-analyses for Day 2 (Extinction)  
*Note:* PFC = prefrontal cortex; rCS = reminded CS; nrCS = non-reminded CS

	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{\text{FWE-corr}}$
Extinction (Run 2)	rCS > nrCS	Dorsolateral PFC	R	84	30	22	46	3.35	0.033
					30	22	46	3.35	
					34	20	48	3.35	
					34	34	48	3.09	
					24	28	54	3.04	
Extinction (Run 4)	nrCS > rCS	Dorsolateral PFC	L	80	-44	42	22	3.54	0.038
					-44	42	22	3.54	
					-34	38	26	3.54	

### Day 3: reinstatement and re-extinction

#### *Contingency awareness*

After re-extinction, participants reported a differential preference to the CS,  $F(2,60) = 3.89$ ,  $p = .026$ . Specifically, rCS+ were preferred over CS-,  $t(21) = 2.75$ ,  $p = .012$ . No significant differences were found between the rCS+ and the nrCS+,  $p = .719$ , and between the nrCS+ and the CS-,  $p = .266$ .

#### *fMRI: Whole-brain analyses*

No significant differences were observed in the CS+ vs CS- contrast. Importantly, decreased BOLD responses were found in the rCS+ relative to the nrCS+ in the areas of the cerebellum, thalamus, fusiform gyrus, paracentral lobule, precuneus, and the middle temporal gyrus (Table 6-5). In addition, higher activation was found for the nrCS+ compared to the CS- in the areas including the parahippocampal gyrus, cerebellum, thalamus, supplementary motor area, dorsolateral part of the superior frontal gyrus, as well as the inferior parietal gyrus (Table 6-8).

#### *fMRI: ROI- analyses*

Significant and reduced activations were found in the right dorsolateral prefrontal region and the right hippocampus in the rCS+ compared to the nrCS+. nrCS+ showed stronger BOLD responses in the right amygdala and the left hippocampus relative to the CS- (Figure 6-4). The BOLD responses of the CS+ and CS- did not differ significantly (Table 6-9).

Table 6-5 Localization and statistics for whole-brain analysis for Day 3 (Re-extinction)

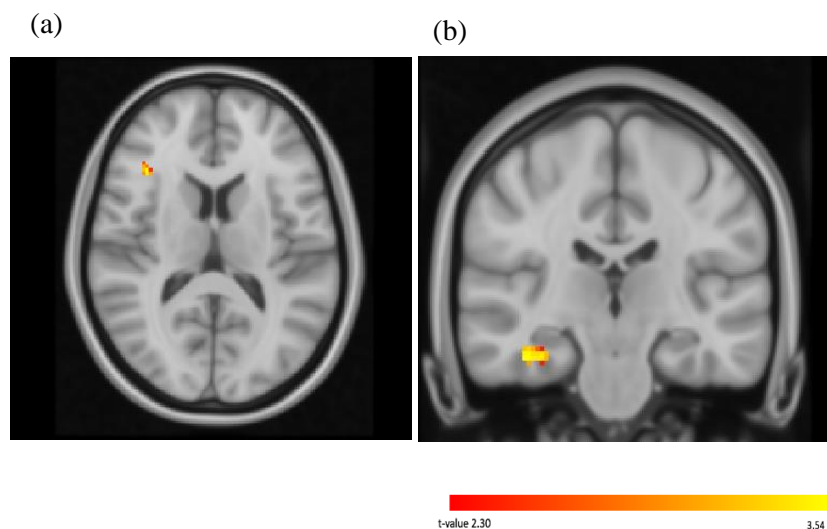
	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	<i>p</i> FWE-corr	
Re- extinction  (Run1)	nrCS > rCS	Cerebellum (lobule 6)	R	373	16	-64	-18	3.54	0.031	
		Cerebellum (lobule 6)	R		16	-64	-18	3.54		
		Lingual gyrus	R		20	-60	-10	3.54		
		Parahippocampus	R		28	-42	-8	3.54		
		Lingual gyrus	R		22	-48	-4	3.54		
		Lingual gyrus	R		16	-48	-10	2.99		
			R	154	26	-30	18	3.54	0.040	
			R		26	-30	18	3.54		
		Thalamus	R		16	-16	18	3.54		
			R		32	-18	22	3.54		
			R		30	-28	14	3.35		
			R		32	-26	24	3.35		
			R		32	-30	20	3.16		
		Fusiform gyrus	R	125	36	-32	-26	3.54	0.045	
		Fusiform gyrus	R		36	-32	-26	3.54		
		Fusiform gyrus	R		42	-26	-22	3.54		
		Parahippocampus	R		34	-24	-22	3.54		
		Fusiform gyrus	R		40	-18	-20	3.54		
		Parahippocampus	R		28	-14	-24	3.35		
		Hippocampus	R		36	-8	-20	3.24		
		Paracentral lobule	R	103	10	-30	66	3.54	0.045	
		Paracentral lobule	R		10	-30	66	3.54		
		Supplementary Motor Area	R		6	-22	70	3.54		
		Supplementary Motor Area	L			-4	-12	70	3.54	
		Paracentral lobule	R			6	-26	70	3.35	
		Supplementary Motor Area	R			10	-24	64	3.24	
			R	102	12	-40	-28	3.54	0.036	
			R		12	-40	-28	3.54		
		Cerebellum (lobule 3)	R		14	-34	-28	3.54		
		Cerebellum (lobule 4, 5)	R		16	-52	-24	3.35		
		Paracentral lobule	L	91	-2	-36	68	3.54	0.044	
		Paracentral lobule			-2	-36	68	3.54		
Precuneus			-4	-40	70	3.54				
Middle temporal gyrus	R	38	40	-60	12	3.54	0.046			
Middle temporal gyrus			40	-60	12	3.54				
Middle temporal gyrus			42	-56	8	3.35				

Table 6-6 Localization and statistics for ROI-analyses for Day 3 (Re-extinction).

Note: PFC = prefrontal cortex; rCS = reminded CS; nrCS = non-reminded CS

	Contrast	Structure	Side	Size (voxels)	x	y	z	Z <sub>max</sub>	$p_{\text{FWE-corr}}$
Re-extinction (Run 1)	nrCS > rCS	Dorsolateral PFC	R	16	38	26	12	3.54	0.036
			L	16	38	26	12	3.54	0.036
		Hippocampus	R	398	42	-26	-22	3.54	0.007
					42	-26	-22	3.54	
					34	-24	-22	3.54	
					40	-18	-20	3.54	
					28	-42	-8	3.54	
					24	-44	-4	3.54	
	28	-14	-24	3.35					

Figure 6-4 a) Enhanced right dorsolateral prefrontal region ( $z = 12$ ,  $p_{\text{FWE-corr}} = .036$ ) and b) hippocampus ( $y = -26$ ,  $p_{\text{FWE-corr}} = .007$ ) reactivity in the nrCS > rCS contrast following test of reinstatement on Day 3.



#### *Fear recovery from Day 2 to Day 3 (Extinction 4<sup>th</sup> run vs Re-Extinction 1<sup>st</sup> run)*

Whole-brain analyses did not reveal significant differences in the BOLD responses in all contrasts comparing the first run of Re-extinction and the last run of extinction.

Regarding the ROI-analyses, increased BOLD responses in the right dorsolateral prefrontal region was found in the nrCS+ relative to the CS-. The activations in the rCS+ and CS- did not differ significantly (Table 6-7).

Table 6-7 Localization and statistics for ROI-analysis for fear recovery from Day 2 to Day 3 (Extinction run 4 vs Re-Extinction run 1)

Contrast	Structure	Side	Size (voxels)	x	y	z	Z <sub>max</sub>	<i>p</i> <sub>FWE-corr</sub>
nrCS > CS-	dorsolateral PFC	R	82	34	28	52	3.54	0.017
				34	28	52	3.54	
			11	36	40	44	3.24	0.044
				36	40	44	3.24	

## 6.4 Discussion

The present study examined the impact of a single US presentation before extinction on the neural correlates of extinction and the return of fear in a three-day multi-CS conditioning paradigm. It was hypothesized that enhanced prefrontal activity, including the vmPFC and dlPFC in the contrast between non-reminded CS versus reminded CS would be observed during extinction learning and increased BOLD signals would be observed in the amygdala and the hippocampus during re-extinction in the same contrast. We found that (1) during acquisition, reminded CS+ and non-reminded CS+ did not differ and that the CS+ relative to the CS- evoked higher activation in the bilateral amygdala, hippocampus, and dorsolateral prefrontal cortex in the early acquisition phase; (2) during early extinction, significant activation of the right dorsolateral prefrontal regions (dlPFC) in the reminded CS+ relative to the non-reminded CS+ was observed, (3) following the reinstatement test,



stronger BOLD responses in the right dlPFC and the right hippocampus in the non-reminded CS+ relative to the reminded CS+ was observed. Moreover, only the non-reminded CS+ showed stronger BOLD responses in the right amygdala and the left hippocampus relative to the CS-; the activation of these brain regions in the reminded CS+ was similar to the CS-.

The neural activations during acquisition on Day 1 are consistent with previous research findings that the bilateral amygdala, hippocampus, and dorsolateral prefrontal cortex are implicated in the acquisition phase (Fullana et al., 2016; LaBar & Cabeza, 2006; Rajbhandari et al., 2017b). Since the detectability of the CS-US association was at a chance level ( $d' = -0.11$ ), our findings provide further neural evidence that fear-related learning can be acquired even when participants were not aware of the contingency between the faces and aversive screams in the present study.

Regarding the underlying neural correlates of post-reminder extinction, our findings was partially consistent with Schiller et al (2013b)'s study in which the vmPFC BOLD responses were higher in the reminded CS+ relative to the non-reminded CS+ during early extinction. Such increased activity in the vmPFC was absent in our sample, but we observed an enhanced BOLD response in the dorsolateral PFC to the reminded CS+ relative to the non-reminded CS- during early extinction. Dorsolateral PFC is thought to underlie higher cognitive processes such as working memory and attention which are important for conscious appraisal of threat (Gilmartin et al., 2014). While dlPFC does not directly project to the amygdala, it may control amygdala activity indirectly through its projections to the vmPFC and the lateral temporal cortex (Buhle et al., 2014; Ochsner & Gross, 2005). The involvement of the dlPFC may suggest that frontal region is recruited early during extinction process if the CS is reminded prior to extinction learning.

Following the reinstatement test, we found an increased amygdala BOLD response in the non-reminded CS+ relative to the CS-, a pattern that was absent in the reminded CS+ > CS- contrast. This finding provides partial support to previous studies reporting a significant reduction in the amygdala activity comparing the reminded CS+ with the non-reminded CS+ during memory retrieval (Agren et al., 2012; Björkstrand et al., 2015; Schiller et al., 2013b). The enhanced amygdala activity for the non-reminded CS+ is thought to reflect the recovery of the original CS-US memory, which agrees with the notion that extinction is a form of new CS-no US learning (Bouton, 2002). Conversely, the comparable amygdala activity between the reminded CS+ and the CS- following reinstatement agree with the reconsolidation hypothesis, which states that the original CS-US memories are modified and inhibitory regulation from the prefrontal region is no longer necessary (Nader, 2015; Schiller et al., 2013b). Behaviourally, the preference towards the reminded CS+ relative to the CS- on the pair comparison task might suggest that the acquired fear was abolished for the faces that were once paired with the aversive scream. Nevertheless, the null finding in the same task on Day 1 may limit the extent of this conclusion.

Interestingly, the whole-brain analyses revealed that multiple regions, including the hippocampus and the cerebellum, were also engaged during early re-extinction. Hippocampus and cerebellum are both implicated in the fear extinction in humans (Kattoor et al., 2014; Milad et al., 2014). It seems possible that the recruitment of these brain areas was required for the non-reminded CS+ to retrieve the inhibitory CS-US memory trace acquired in extinction; whereas for the reminded CS+, the recruitment of these regions was reduced as the extinction took place during the window of reconsolidation. In sum, our findings complemented and extended existing evidence that a reminder-extinction procedure may be effective in preventing the return of fear in humans via the putative mechanism of

reconsolidation. We herein showed that implicitly acquired fear associations were also subject to modification by reconsolidation if the memories were reactivated by the associated reminder cue before extinction.

Dysfunctional prefrontal regulation is a hallmark of anxiety disorders (Craske et al., 2017; Ball et al., 2013). Several studies to-date have explored the clinical utility of non-invasive brain stimulation techniques in the treatment of anxiety disorders by targeting the lateral PFC region (Kar & Sarkar, 2016; Shiozawa & Sato, 2016). For instance, an anodal stimulation to the right dlPFC by transcranial direct current stimulation (tDCS) enhanced the recall of episodic memory compared to sham or cathodal stimulation (Sandrini et al., 2013). In a fear conditioning study, Mungee and colleagues (2013) showed that a facilitatory anodal tDCS, applied shortly after memory reactivation, over the right lateral PFC enhanced fear expression at test. The potential role of an inhibitory cathodal tDCS over the right dlPFC on the modulation of fear responses has not been thoroughly explored in the fear conditioning literature. Future studies are warranted to clarify the clinical value of non-invasive brain stimulations in the treatment of anxiety disorders, and how the putative process of reconsolidation affects the effectiveness of these interventions.

### *Limitations*

One limitation of the study is that we employed a 100% reinforcement in the learning phase and the presence of the US might confound the BOLD responses relevant to conditioning. To mitigate this problem, we optimized the design efficiency of the experiment by testing 5000 different combinations of sequence (in terms of trial orders, the onset time of the CS and US presentation) and selecting the design with the highest efficacy. In addition, we included an explicit model of the responses related to the US in the first-level analyses.

Second, we did not observe explicit learning on the behavioural tasks (CS-US matching task or Pair comparison task). Although such absence of learning was also demonstrated in previous multi-CS conditioning studies, future studies could consider integrating MRI-compatible measures such as pupil responses or skin conductance responses to corroborate the learning phenomenon. Third, the current sample size is similar to other previous multi-CS conditioning studies that had between 20 and 24 participants (Steinberg et al., 2013) and reported detectable neural signals. Samples of this size are relatively common in the neuroscience literature where many studies contain fewer than 20 participants. Nevertheless, we cannot exclude the possibility of false positive and negative effects due to lack of power or the sample might not be representative of the population as a whole. To mitigate this, we followed a robust pipeline and statistical analysis in the preprocessing and analysis, including fMRI preprocessing pipeline, nonparametric permutation test for statistical inference, and corrected for multiple comparisons in the reported results (Eklund et al., 2016)

## **6.5 Conclusion**

We tested the impact of a reminder-extinction procedure on the neural mechanisms of extinction and the return of fear in multi-CS conditioning. Our findings suggested that presenting a US-reminder cue before behavioural extinction might reduce the return of fear by diminishing the involvement of the dlPFC. The prospect of preventing the return of fear through disrupting reconsolidation in humans is exciting. Future studies should clarify the neural mechanisms of this evolutionary-adaptive process in mitigating the dysfunctional prefrontal regulation underlying fear-related and anxiety disorders.

Table 6-8 Localization and statistics for whole-brain analysis for Day 3 (Re-Extinction)

	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{\text{FWE-corr}}$		
Re-extinction (Run 1)	nrCS > CS-	Lingual gyrus	L	157	-16	-46	-8	3.54	0.040		
		Lingual gyrus	L		-16	-46	-8	3.54			
		Lingual gyrus	L		-20	-50	-6	3.54			
		Cerebellum (lobule 4 and 5)	L		-6	-44	-12	3.35			
		Lingual gyrus	L		-12	-46	-10	3.35			
		Parahippocampus	L		-20	-38	-10	3.16			
		Parahippocampus	L	-16	-36	-8	3.16	0.033			
		Cerebellum (lobule 4 and 5)	R	130	16	-44	-12		3.54		
		Cerebellum (lobule 4 and 5)	R		16	-44	-12		3.54		
		Lingual gyrus	R		20	-50	-4		3.54		
		Lingual gyrus	R		10	-38	-2	3.09			
					L	119	-38	-34	4	3.54	0.039
					L		-38	-34	4	3.54	
					L		-26	-32	10	3.54	
					L		-16	-32	10	3.04	
				Thalamus	L		-12	-30	12	2.99	
					L		95	-14	-18	-4	
					L	-14		-18	-4	3.54	
				Thalamus	L	-16		-26	-2	3.54	
					L	-24		-22	2	3.54	
				Supplementary Motor area	L	71	-8	-2	58	3.54	0.042
				Supplementary Motor area	L		-8	-2	58	3.54	
				Supplementary Motor area	R		2	-4	64	3.35	
				Precentral gyrus	L	32	-44	-2	42	3.54	0.042
				Precentral gyrus	L		-44	-2	42	3.54	
				Superior frontal gyrus, dorsolateral	L	32	-20	-6	52	3.54	0.042
		Superior frontal gyrus, dorsolateral	L	-20	-6		52	3.54			
		Inferior parietal gyrus	L	10	-54	-32	46	3.54	0.042		

Table 6-9 Localization and statistics for ROI-analyses for Day 2 (Extinction) and Day 3 (Re-Extinction)

	Contrast	Structure	Side	Size(voxels)	x	y	z	Zmax	<i>p</i> <sub>FWE-corr</sub>
Extinction (Run 4)	CS+ > CS-	Amygdala	R	103	16	-4	-14	3.54	0.003
					16	-4	-14	3.54	
	nrCS > CS-	Amygdala	R	193	18	4	-24	3.54	0.017
					18	4	-24	3.54	
					18	-4	-14	3.54	
					18	-4	-14	3.54	
		Dorsolateral PFC	R	65	34	26	52	3.54	0.033
					34	26	52	3.54	
					40	28	52	3.54	
					40	28	52	3.54	
				19	42	54	22	3.35	0.030
				42	54	22	3.35		
Re-extinction (Run 1)	nrCS > CS-	Amygdala	R	11	20	-10	-10	3.54	0.026
					20	-10	-10	3.54	
		Hippocampus	L	130	-18	-44	-6	3.54	0.024
					-18	-44	-6	3.54	
					-22	-44	-8	3.24	
					-20	-38	-10	3.16	
					-16	-36	-8	3.16	
					-16	-30	-2	3.09	
			-24	-34	-12	2.91			

## **Chapter 7 General summary and discussion**

## 7.1 Summary of the findings

Since the (re)discovery of memory reconsolidation at the turn of the twenty-first century, scientific research into this mechanism of reconsolidation both in laboratory animals and humans has grown substantially. The recognition of the fundamental, plastic nature of consolidated memories is profound and it has opened a new era for engineering human memories in the past two decades. However, the practical translation of this scientific discovery into humans is still in infancy. The review in Chapter 1 highlighted the inconsistencies of findings on the reconsolidation of human fear memories and identified the challenges associated with human reconsolidation research. It is hoped that the four experiments described in the current thesis contribute to the existing literature on the extinction and reconsolidation of human fear memory. The empirical findings address the question posed at the outset of the thesis:

“What are the neural and behavioural mechanisms underlying  
extinction of fear memories and their recovery?”

Using physiological, behavioural and neural measures in a Pavlovian conditioning paradigm, the thesis examined the following research questions:

1) Does implicit exposure to a conditioned stimulus attenuate fear-related defensive responses?

2) Can implicit exposure to a reminder cue before extinction attenuate the recovery of fear?

3) How does an explicit reminder cue facilitate extinction and modulate the return of fear in a multi-CS conditioning paradigm?



4) What is the impact of an explicit reminder cue on the neural mechanisms of extinction and return of fear?

The existing literature suggests that fear learning can be acquired and modulated with or without conscious awareness. Using continuous flash suppression to perceptually suppress awareness of the presented stimuli (Experiment 1), both the implicitly and explicitly viewed conditioned stimuli evoked similar pupil responses to the safe stimuli after reinstatement. This observation suggests that both explicit and implicit extinction may modulate fear-related defensive response. Moreover, further analysis revealed that the percentage of fear recovery was greater for the implicit than the explicit extinction.

Building on the results of Experiment 1, which suggested that implicit exposure to a conditioned stimulus may attenuate fear responses, Study 2 examined whether an implicitly presented reminder cue would also destabilise the original CS-US association and render the memories susceptible to subsequent behavioural extinction. It was found that the presentation of implicit and explicit reminder cues before extinction did not reduce the return of fear, as measured by pupillary responses following a reinstatement test. Moreover, participants rated the reminded CS as more unpleasant than the non-reminded conditioned stimulus and the safe stimulus following extinction. This differential affective response was independent of the perceptual awareness of the conditioned stimulus during its reactivation. In sum, Experiment 2 did not provide evidence for the use of a pre-extinction reminder cue to modulate the reinstatement of fear.

The notion that return of fear can be effectively modulated via memory reconsolidation was further tested in a multi-CS conditioning paradigm, an experimental design that has a higher signal-to-noise ratio and greater ecological validity. Together,

experiments 3 and 4 support the notion that a US-reminder cue may trigger the reactivation of CS-US memories, making them susceptible to subsequent reconsolidation-interference by extinction. In Experiment 3, the reinstatement of fear measured by pupil responses was absent for all CS regardless of whether it was reminded before extinction or not. However, we observed a reminder-specific pupil response following early extinction, suggesting that a US-reminder cue might evoke a distinct pattern of extinction learning for the reminded CS+ relative to the non-reminded CS+. In Experiment 4, ROI-analyses revealed significant activation of the right dorsolateral prefrontal area in the reminded CS+ relative to the non-reminded CS+ during early extinction. This observation suggests stronger recruitment of right dorsolateral prefrontal region early in the extinction. Moreover, stronger BOLD responses in the right dorsolateral prefrontal cortex and the right hippocampus were evident in the non-reminded CS relative to the reminded CS after the reinstatement test.

## **7.2 Synthesis and discussion: Memory reconsolidation and extinction**

Disentangling the reasons why a manipulation fails to trigger the reconsolidation of memory is challenging for inconsistent findings, and null results present a conundrum for researchers in the field of memory reconsolidation (Kindt, 2018). However, careful evaluation of the results from the current thesis may extend our understandings of the mechanisms underlying memory reactivation and modification.

### *In search of an optimal means for reactivation*

The principles of consolidation and reconsolidation imply that memory is plastic and malleable, allowing memories to remain relevant and guide future behaviours (Lee, 2009; Lee et al., 2017). This fundamental nature of memory updates is evolutionarily adaptive and

important for our survival in an ever-changing environment. However, there are also systems in place, termed boundary conditions, to protect the integrity of a memory from being updated or reconsolidated incessantly and indiscriminately (Treanor et al., 2017; Zuccolo & Hunziker, 2019). Given that implicit exposure could modulate fear-related defensive behaviours in Experiment 1, the failure to modulate the return of fear using an implicit reminder cue in Experiment 2 might relate to unsuccessful memory reactivation. At present, the optimal retrieval protocol for triggering memory reconsolidation is largely unknown (Eley et al., 2018; Kindt, 2018; Visser et al., 2018). The duration, times of repetition, dosage as well as timings of the reactivation procedure are yet to be defined in the literature, although they are likely to depend on previous encoding history (Visser et al., 2018), the strength and age of memories (Eley & Kindt, 2017a), and individuals' genetic makeup or psychiatric vulnerability (Zuccolo & Hunziker, 2019). The duration of reactivation trials has been studied more systematically in laboratory animals, where longer and repeated retrievals of memory may risk initiating a new extinction learning (Pedreira & Maldonado, 2003), or entering a transient state between extinction and reconsolidation (Merlo et al., 2014) instead of memory destabilisation.

Another boundary condition related to the reactivation of memory is prediction error, a mismatch between what is expected and what happens. In Experiment 2, the reinforcement schedule in the acquisition phase was 75%, and presenting two unreinforced trials of the CS in the reactivation phase might not be strong enough to violate the expectancy of the previous learning and inform participants that there is something new to learn, regardless of whether they were in the explicit or implicit reactivation group. Previously, a study has shown that in a case in which a CS was certain to predict a shock during acquisition, a single unreinforced presentation of the CS during reactivation led to a violation of expectancy and triggered

memory destabilisation. (Sevenster et al., 2013b). However, in the case of uncertainty where a CS predicted the shock 50% of the time, more unreinforced trials were needed to create this prediction error (Elsley & Kindt, 2017a).

The putative prediction errors defining the boundary between reconsolidation and extinction can also be accounted for by a computational model of memory modification (Gershman et al., 2017). In the case of a small prediction error, the posterior probability of the acquisition latent cause is high, and modification of the original memory takes place. In the case of a large prediction error, the posterior probability of the acquisition latent cause is low, a formation of a new memory occurs (Figure 7-1). According to Gershman (2017), there is a ‘sweet spot’ in which the prediction errors are large enough to induce a weight change in the latent cause of the CS-US association but small enough to prevent the formation of a new latent cause. Ideally, the key to a persistent reduction of fear is to assign acquisition and extinction learning to the same latent cause. Building on Gershman’s computational model of memory modification, Moris and colleagues recently demonstrated that a partial extinction procedure aimed at modifying the original latent structure of the CS-US association reduced the rate of reacquisition in a fear conditioning paradigm (Morís et al., 2017)

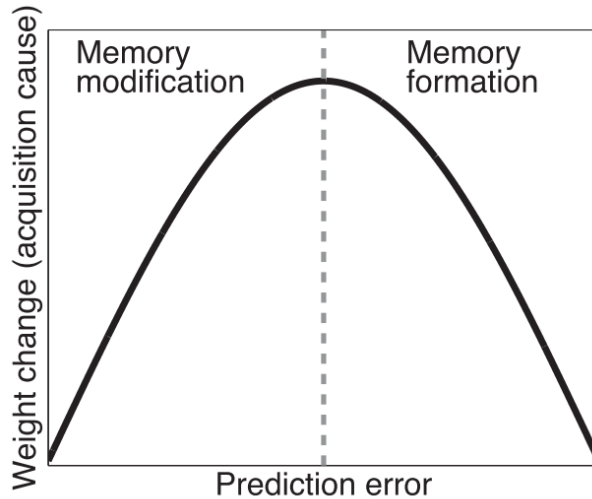


Figure 7-1 Gershman’s model for predicting fear extinction with reference to the size of prediction error (x-axis) and the change in the latent cause inferred from the conditioning (y-axis).

*Source:* Adopted from Gershman (2017).

In the absence of an established and reliable reactivation protocol, Kindt and colleagues (2018) illustrated their development of a reactivation protocol in a series of pilot cases. In brief, they modified their reactivation protocol multiple times by carrying out a few series of pilot testing and carefully observing participants’ responses. Using this bi-directional translational approach, they adjusted participants’ level of exposure to a spider (e.g., from passive observation to touching a live tarantula in a terrarium, and from varied exposure time to a standard brief exposure), and arrived at a final protocol used in their randomised control study in which participants were asked to touch a tarantula, after a brief two-minute exposure to it, without actually touching the spider (Soeter & Kindt, 2015b). Their experience highlights the importance of incorporating clinical observations into a testing protocol, because of the absence of a real-time assessment of reconsolidation or prediction errors in the state of the science.

*In search of a real-time indicator for reconsolidation*

Apart from the optimal means of reactivating the target memory, a real-time indicator for the mechanism of change in reconsolidation is under-investigated in the literature. To this end, the findings of distinct BOLD signals and pupil responses during extinction learning (Experiments 3 and 4) might shed some light on the issue. While reduced prefrontal recruitment has been reported before (Schiller et al., 2013b), to the author's knowledge, Experiment 3 is the first study reporting a diminished differential pupil response for the reminded CS following reactivation in the early extinction phase in a multi-CS conditioning paradigm. Zimmermann and colleagues (2020) attempted to test the reminder-extinction procedure with the use of pupil dilation and skin conductance responses, but their findings failed to support the reconsolidation effect on the reinstatement of fear and no differences in pupil responses were noted during extinction. One major difference between Experiment 3 and Zimmermann's study is the experimental design (multi-CS conditioning *versus* a single-CS conditioning). The multi-CS conditioning paradigm is known for its higher signal-to-noise ratio because of the higher number of trials for analysis. Further investigation is warranted to substantiate whether pupil responses are a sensitive real-time read-out for reconsolidation-related changes during extinction.

The conventional methods for measuring conditioned responses in humans include skin conductance responses (SCR), fear-potentiated startle responses (FPS), subjective ratings and reaction-time tasks (Haaker et al., 2014a). Pupillometry is a relatively novel method in the field of human fear conditioning and a few recent studies have supported its sensitivity in quantifying the conditioned responses (Leuchs et al., 2017b, 2019). Anatomically, there is no direct cortical connection to the pupillary dilator or constrictor muscles that control the dilation of pupils, but the activity of the anterior cingulate cortex (ACC), supramarginal gyrus, thalamus, and insula are found to correlate with pupil responses

during fear conditioning (Leuchs et al., 2017b). It is also of note that the read-out measures of conditioned responses do not always covary in different read-out measures (Kindt & Soeter, 2013b; Roesmann et al., 2020; Steinberg et al., 2012a). The experiments presented in the current thesis demonstrated a similar dissociation between subjective verbal reports and pupil responses (Experiments 2 and 3), and between verbal reports and neural signatures (Experiment 4). It is reasonable to assume that different measures correspond to distinct response systems reflecting a threat-related state or feelings of fear. For instance, conditioning of SCR is observed only in participants who are aware of the CS-US contingency, whereas conditioning of FPS is found irrespective of contingency awareness (Weike et al., 2007). In an implicit learning paradigm, there is also evidence of a dissociation between US expectancy ratings and neuronal responses (Steinberg et al., 2013). Conceptually, the different response systems might reflect the multi-faceted and complex nature of fear; Experimentally, the multiple response system presents a unique challenge to researchers as we do not yet know how these systems interact with each other as a whole. Clinically, however, the multiple response system of fear might be beneficial for patients as the modulation of one debilitating system may open up resources for managing other symptoms and ultimately improving quality of life and functioning.

#### *In search of the neural mechanism of reconsolidation*

The neural evidence for a reminder-specific pattern of extinction and re-extinction (Experiment 4) suggests a reduced dorsolateral prefrontal and hippocampus recruitment in the process of reconsolidation-extinction, which is an encouraging finding. This observation is consistent with the notion that reconsolidation may induce a fundamental change in the memory processes and reduce the inhibitory control from the prefrontal region (Schiller et

al., 2013a). The dorsolateral prefrontal region does not directly project to the amygdala, but it may indirectly control amygdala activity via projections to the ventromedial prefrontal cortex or lateral temporal cortex (Buhle et al., 2014; Ochsner & Gross, 2005). A recent study has shown that patients with dorsolateral prefrontal cortex (dlPFC) lesions are impaired in cognitively regulate their subjective fear compared to their normal controls, supporting the critical role of dlPFC in the cognitive regulation of subjective fear (Kroes et al., 2019). The question remains as to whether these patients with dlPFC lesions would benefit from the reminder-extinction procedure that targets the reconsolidation process to modulate their subjective fear responses.

Another potentially promising avenue for extending reconsolidation-based extinction is to harness the implicit nature of the physiological fear circuit (Taschereau-Dumouchel et al., 2018). Recently, Koizumi and colleagues (2017b) developed a novel approach to reducing conditioned fear by reinforcing the neural activation patterns representing the conditioned stimuli while bypassing participants' awareness of the content and the purpose of the procedure. Using fMRI neurofeedback, it may be plausible to quantify the magnitude of reactivation associated with a CS reminder cue. This manipulation would be similar to the implicit exposure to a CS reminder cue in Experiment 2 in theory, but gives more control to the experimenters or therapists for validating the reactivation procedure in a reconsolidation-extinction paradigm.

### **7.3 Clinical implications**

As reviewed in Chapter 1, there have been both success (Björkstrand et al., 2016; Soeter & Kindt, 2015b; Telch et al., 2017; Xue et al., 2012) and failures (Maples-Keller et al., 2017; Shiban et al., 2015a; van Schie, Veen, et al., 2017; Wood et al., 2015) in applying



reconsolidation-based treatments to the clinical populations. Although the prospect of engineering fear memory by offering brief interventions for anxiety or fear-related disorders with long-lasting effects is certainly exciting, there are considerable challenges to implement reconsolidation-based interventions in a clinic.

First, the circumstances in which memories are reactivated is sometimes difficult to achieve in a clinical setting. To reactivate a memory trace, the context of the reminder should ideally be identical to the encoding situation. One can imagine it is almost impossible to reproduce a past traumatic event involving complex memories in a therapy room. Second, complex trauma and older memories may be more resistant to modification (Kredlow et al., 2016), limiting the scope of the update mechanism in real-life situations as many traumatic experiences or memories may occur in childhood or adolescence, and patients may seek help at a later stage in life. Third, retrieving the memories per se is not enough to trigger the reconsolidation process; prediction errors are required to reactivate the memory trace for further destabilisation (Sevenster et al., 2013b). However, determining the optimal circumstances for a prediction error is clinically challenging and is harder to implement in a therapy session than in experimental settings.

Notwithstanding these problems, several experimental studies have shed light on these limitations and improved the relevance of the existing reconsolidation model for clinical interventions. For instance, Agren and colleagues (2017) introduced imaginal extinction, in which verbal instructions are given to encourage visualising the conditioned stimuli during extinction. They reported that both *in vivo* and imaginary extinction affect the reconsolidation of fear memories, and participants showed a reduction in their fear responses. Moreover, research into human and animal models of old and strong memories has highlighted a range of techniques that may prove useful for inducing memory reactivation,

such as extended reactivation trials, multiple treatment sessions, and pharmacological labialisation (Elseley & Kindt, 2017a).

Several innovative approaches utilising the reconsolidation-update mechanism have emerged recently, in addition to the traditional interference method including exposure or pharmacological intervention. James and colleagues (2015) investigated whether engaging in a visual-spatial task — playing Tetris during reconsolidation can interfere with the reconsolidation process and subsequently reduce the frequency of intrusive memories of an experimental trauma. They found that both memory reactivation and Tetris play, in combination, were effective in reducing subsequent intrusive memories. This idea was further tested by Iyadurai et al. (2018) by engaging patients who experienced a traumatic motor vehicle accident in the emergency department using the cognitive interference procedure. Patients who received a memory reminder of the traumatic event and Tetris gameplay reported a reduction of intrusive memories of their trauma by 62% relative to their control counterparts, who completed an activity log. These new creative studies are motivated by extending the application of the reconsolidation-update mechanisms to real-world clinics. Further research and results in this area are needed.

Before the field of memory reconsolidation and its translation to the clinical practice matures, clinicians may refer to an existing alternative to counteract the return of fear in the clinical scene, optimising inhibitory learning during extinction (Craske et al., 2014). These optimising strategies include the explicit violation of expectancy, deepened extinction, affect labelling, removal of safety signals, attention to the conditioned stimulus, and have been extensively tested in laboratory animals, healthy humans, and clinical samples (Craske et al., 2018).

#### **7.4 Limitation of fear conditioning/memory reconsolidation models**

It is hoped that the limitations associated with the experiments described in the current thesis have been sufficiently addressed in the respective chapters, and they are not repeated here. Rather, the general limitations of the Pavlovian conditioning paradigm and research in the return of fear are discussed.

Pavlovian conditioning is employed in the thesis because it is one of the oldest and most systematically studied paradigms in psychology and neuroscience. The neural circuitry of fear conditioning is well conserved across species (Fanselow & Wassum, 2016), and has a well-defined theoretical structure that allows investigations from the cellular to the behavioural levels of analysis to study the development, maintenance, and treatment of anxiety disorders. However, the ability to model a complex clinical phenomenon in a laboratory model using a reductionist approach inevitably has limitations. Understanding these limitations enables us to have realistic expectations about the research findings as well as the scope of translational work the model can offer. It also serves to improve experimental designs for optimising the applicability and practicality of the model in clinical settings.

First, the development of clinical anxiety is not always identified as a simple CS-US relationship acquired in a laboratory. Between 15% and 68% of clinical cases are unable to recall any cause of their fears, and only between 18% and 57.5% of individuals with specific phobias can recall a direct conditioning event leading to their phobia (Laborda & Miller, 2011). Although the absence of an identifiable traumatic event does not negate an instance of learning by conditioning, the high prevalence of the recall difficulty suggests that clinical anxiety is developed in a manner other than explicit, direct conditioning as modelled in the laboratory conditioning paradigm. Vicarious learning (Askew & Field, 2008), learning through verbal instructions (Olsson & Phelps, 2007) and second-order conditioning (Gewirtz

& Davis, 2000) might also contribute to the learning of pathological fear, but these forms of learning receive less attention in the field of conditioning research.

Second, the majority of the studies in fear conditioning research use arbitrary stimuli such as geometric figures or inanimate objects (e.g. pictures of houses or lamps) as conditioned stimuli (Lonsdorf et al., 2017), which lacks external validity and does not reflect the complexity of real-world experience. Fear-relevant stimuli (e.g. pictures of snakes and spiders or angry faces) are used in some studies, but they display a qualitatively different pattern of learning and extinction from fear-irrelevant stimuli. For instance, fear-relevant relative to fear-irrelevant conditioned stimuli lead to faster acquisition of conditioned fear (Ho & Lipp, 2014) and higher resistance to extinction (Mallan et al., 2013; Mineka & Öhman, 2002; but Åhs et al., 2018). Moreover, images of spiders and snakes may elicit emotions such as disgust in addition to fear. To overcome this limitation, some research groups have created animal-like, complex conditioned stimuli (Barry et al., 2014) or virtual reality paradigms (Kroes et al., 2017) that allow more control and generalisability to the real world.

Third, there is a lack of emphasis on the meaning of the CS-US association in human studies on fear conditioning. In clinical fears, individuals often reason why a traumatic event happen to them and ascribe a meaning to the event, such as shame, guilt or punishment, which is beyond a simple temporal relationship between a neutral stimulus and a dangerous outcome (Foa & Kozak, 1986). The presence of a meaningful conceptual relationship between the CS and US can strengthen the learning of the association, making it harder to be extinguished. For instance, faces followed by verbal insults led to a stronger conditioned response during acquisition, which was more sustained during extinction in individuals with social anxiety disorder compared to healthy controls (Blechert et al., 2015; Lissek et al., 2008). Interestingly, the cognitive appraisal could attenuate acquisition and facilitate

extinction (Bleichert et al., 2015). Along with mental imagery (Reddan et al., 2018) and perception of self-efficacy (Zlomuzica et al., 2015), these human-unique factors are rarely incorporated in conditioning research in humans despite being crucial for studying the process of fear learning and memories in humans.

To sum up, there are a number of limitations regarding the ways that fear conditioning research is conducted in humans, restricting its translational potentials. Nonetheless, there has also been significant progress using the Pavlovian conditioning paradigm for understanding the neural circuitry of conditioned defensive responses, as well as the development of a framework and strategies for improving extinction learning and its retention (Carpenter et al., 2019). Thankfully, there is consensus and progress on facilitating interaction between scientists and clinicians for improving the understanding and treatment of mental health disorders (Milton & Holmes, 2018). The science of extinction and reconsolidation will follow from the bench to the bedside, and back again.

## **7.5 Definitions of fear**

Throughout the current thesis, the terms fear conditioning/fear responses and threat conditioning/threat responses have been used interchangeably at times. It is important to note that Pavlov and his original followers did not use the term fear conditioning to describe associative learning in his laboratory animals. Rather, they named it defence conditioning, an empirically defined term based on observable behaviours. Over the years, the term ‘defence conditioned reflex’ came to describe what most neuroscientists know today as ‘learned fear’; fear and fear conditioning have been defined differently in diverse disciplines spanning from biology to philosophy, and even within the field of psychology and

neuroscience, fear has been conceptualised and defined broadly (Adolphs & Andler, 2018; LeDoux, 2014). Joseph LeDoux, one of the most influential researchers in the field of emotions, suggests that the terms ‘fear learning’ or ‘fear responses’ should be replaced by ‘threat’ and ‘defence responses’ for better demarcations between threat, responses to threat, and feelings of fear (LeDoux & Pine, 2016). Scientists, he argued, measure observable responses to threat in a Pavlovian conditioning experiment, which is different from a consciously felt state of fear.

There is disagreement on how best to define and investigate fear within the field of neuroscience. Some of the core unresolved issues in the debate include: is fear, or emotion more broadly, a conscious and subjective state? Can fear be studied in animals? Do we have a hardwired circuit of fear? Is fear constructed differently by different human brains? Consequently, there are different approaches for the investigation of the neuroscience of fear, which fall broadly into either the affective approach or the cognitive approach (Panksepp et al., 2017). In the following section, these two theoretical frameworks to understanding fear and their respective methodologies will be reviewed.

#### *The study of fear from Pankseppian’s affective neuroscience approach*

Jaak Panksepp first coined the term ‘affective neuroscience’ in the early 1990s to denote a unique research area that aims to understand the neural basis of human emotions using animal models (Panksepp, 1998). A few working hypotheses were also adopted to make the inferences from these cross-species studies translational and applicable to humans. First, primary emotions, including fear and attachment, are constituted at the level of the subcortical brain regions (Panksepp & Watt, 2011). For instance, rats whose cortex was removed surgically continued to display motivated and emotional behaviours with

subcortical deep brain stimulation (Huston & Ly, 1974). In humans, hydranencephalic children who are born without a complete cerebral cortex show appropriate emotional responses (Merker, 2007). Despite the evidence that the generation of emotions may not require the participation of the cortex, proponents of Pankseppian's affective approach suggest the emotional systems are hardwired subcortically, and this subcortical mechanism of generating emotions is conserved anatomically and physiologically across all mammals from an evolutionary perspective. Hence, the investigation of the neural substrates of emotions in animal models can be transferred to understanding the same system in humans.

Fear is considered an aversive state of the nervous system accompanied by specific forms of autonomic and behavioural arousal to inform one of danger (Panksepp, 2000). Using electrical stimulation in both laboratory animals and humans, Panksepp mapped the amygdalo-hypothalamic-periaqueductal gray circuitry as the neural circuitry of fear in the brain (for a review, Davis et al., 2019). This pathway arises from the central amygdala and projects downward to the anterior and medial hypothalamus and to the PAG of the midbrain and other tegmental regions. The fear system controls autonomic and behavioural responses, as well as the phenomenological experience of fear. The neural substrates of this pathway have received some empirical supports from other groups studying fear circuitry in animals and humans (Fanselow, 2000; Mobbs et al., 2009; Schafe & LeDoux, 2008)

#### *The study of fear from the cognitive neuroscience approach*

While Panksepp's position is that subcortical activation is sufficient to generate emotions, some researchers with a cognitive neuroscience orientation hold that emotional and cognitive processes are inextricably linked. Higher cognitive processes such as appraisal and the role of consciousness are integral in the generation of emotions and their reactions

(Lazarus, 1991; Scherer, 1984). As such, the cognitive neuroscience approach to fear offers a different account of the neural basis of emotional experience in several ways.

First, human emotion systems differ from those in other animals in important ways because of evolution. While there is considerable phylogenetic continuity across mammalian species in the organisation of the neural systems underlying basic emotions and the threat processing network in particular (LeDoux, 2012), it is conceivable that the emotional subcortical system may have also evolved and served novel functions by natural selection through evolution. Consistent with this notion is the finding that neocortical size correlates with the size of social groups in human and non-human primates (Dunbar, 1993). Animal models inform the commonalities of emotion systems between humans and mammals, but not the differences between them.

Second, both the cortical and subcortical structures are implicated in the process of emotion generation (Davidson, 2003). While evidence for the survival circuit underlying threat is well defined in rodents and other mammalian species, higher cortical structures such as the ventromedial prefrontal cortex and the dorsal anterior cingulate cortex, are also implicated in appraising a situation, ascribing affective meaning to the situation and coordinating emotion-related autonomic reactions (Brosch & Sander, 2013; Roy et al., 2012). A meta-analysis study surveying 162 neuroimaging studies of emotion suggested that the dorsomedial prefrontal cortex is important for the cognitive generation of emotional states, and is closely associated with core limbic activation, forming a dmPFC-PAG-hypothalamic pathway (Kober et al., 2008). However, it is not yet clear that such a cortical emotion-relevant system exists in non-human animals and whether it can be studied appropriately in non-human animals.



Moreover, the same defence response may not evoke the same feelings, and affective labels are learned conceptual categories (Barrett et al., 2011; LeDoux, 2012). There is empirical evidence to suggest that emotional perceptions are influenced by the context in which they occur and emotion categories can be applied to different cognitive or bodily reactions in different contexts. In other words, even though there is a common survival circuit across animals, these circuits may not evoke the same feeling states that are described in humans.

Finally, an important aspect of emotion involves conscious experience (Davidson, 2003). Understanding in the neurobiological basis of emotions has advanced considerably in the past few decades, but most findings and discoveries have been based on outward, observable behaviours (e.g. freeze or flight). Current evidence suggests that emotion generation and emotion reactions can occur outside of conscious awareness (LeDoux, 2012). In the two-system model of fear, LeDoux has underscored the role of consciousness in the generation of fear (LeDoux, 2014; LeDoux & Pine, 2016). Specifically, threats elicit both a non-conscious defensive response via the defensive survival circuit and a conscious pathway that gives rise to the feeling of fear via some higher-order cognitive circuit (e.g., working memory). The defensive circuit is responsible for detecting and responding to threats with defensive behaviours and the associated physiological changes. However, according to LeDoux, the subjective feelings of fear are not products of subcortical circuits; rather they rely on a higher-order circuit that is responsible for cognitive processes such as attention and working memory (Brown et al., 2019). The higher-order circuit is mainly located in the cortical regions, and the neural correlates of these regions include the lateral and medial prefrontal cortex, as well as the parietal neocortex and the insula. The defensive circuit and the high order circuit interact but do not share the same pathways in the brain's fear system.

Essentially, the conscious feeling of fear arises through the cognitive processing of raw neural materials, which are provided by the activation of the survival circuits. Therefore, studies of defensive behaviours in animals are important for understanding the survival circuit but are of limited value in understanding the feeling of fear in humans.

In sum, the affective and cognitive perspectives mentioned above represent some disputes in how to investigate and formulate fear by a subset of influential researchers in the field of neuroscience. The two contrasting views highlight the importance of a unified, clear definition of fear, and better differentiation of the subcomponents within this construct. Only a clear definition of fear will consolidate the knowledge developed in the understanding of fear and fear learning in the past decades and advance the progress of research in the coming decades to realise their implications for the psychopathology and treatments of fear-related and anxiety disorders in humans.

## **7.6 Future directions**

Decades of research in emotions and memories have generated valuable knowledge about fear learning and its extinction. Currently, basic and behavioural neuroscience approaches, including *in vitro* and *in vivo* animal models, are effective in decoding the biochemical or neural mechanisms of fear. However, this knowledge can only be appraised, and its clinical utility can only be evaluated, in combination with human studies. The fear conditioning paradigm allows such a translational approach, and dialogues between basic and clinical neuroscience researchers should continue to facilitate in the development of translational models for mental health. For instance, combining ideas from animal behavioural science and human psychology may aid in designing new interventions for use in humans. One such example is illustrated by Holmes et al. who tested the use of a computer task after

memory reactivation to reduce post-accident intrusive traumatic memories in physically stable patients after an emergency admission.

Another promising avenue is to harness the unconscious processes in the defensive circuit and its interaction with the higher-order circuit in order to modulate the conscious feeling of fear. The idea of targeting unconscious processes in psychological interventions is not new. It began in the Freudian era, but the emphasis in later behavioural psychology and cognitive psychology has hampered the development of targeting the non-conscious or implicit process in psychological interventions. Recent advance in cognitive neuroscience and the scientific studies of consciousness have enabled new scientific tools to target unconscious processes in psychological interventions. For instance, using modern neuroscience techniques such as fMRI real-time neurofeedback, participants' neural patterns representing a specific object or content can be obtained and manipulated by the reinforcement of the neural pattern. At least two proof-of-concept studies to-date have demonstrated this neural reinforcement method for modulating conditioned physiological fear responses, and the intervention can proceed outside of the participants' awareness (Koizumi et al., 2017a; Taschereau-Dumouchel et al., 2018). Other than fMRI, there is evidence suggesting that exposure to fear objects utilising a continuous flash suppression technique or backward masking can attenuate conditioned physiological fear responses (Oyarzún et al., 2019; Siegel et al., 2018). Similar to neural reinforcement, these techniques target the unconscious defensive circuit and change the information that feeds-forward to the conscious, subjective feeling of fear. Most importantly, these techniques are inexpensive and can be implemented readily in a therapy room. Nevertheless, more rigorous experimentation is needed to delineate further how, when, and for whom these manipulations work.

## 7.7 Closing remark

The four experiments presented in this thesis were developed and implemented to optimise treatment models of anxiety and fear-related disorders and bridge the gap between laboratory-based treatment research and clinical practice. The findings highlight the subtleties of memory reconsolidation in humans, and the results contribute to a broader understanding of the neural, physiological and behavioural mechanism underlying reconsolidation-extinction and the recovery of fear.

*"We were inspired by the observation that the progress of psychological treatment is lagging behind theoretical advances and promising findings in the laboratory."*

- Scheveneels, Boddez, Vervliet, & Hermans, 2016

As a scientist and a practitioner, I could not agree more with Scheveneels and other senior researchers' view in the field. Nevertheless, the momentum for integrating psychological treatment and neuroscience finding is strong, and my search for new interventions for reprogramming maladaptive fear and translating them to the clinical scene has just begun.

## Chapter 8 References

- Adolphs, R., & Andler, D. (2018). Investigating Emotions as Functional States Distinct From Feelings. *Emotion Review*, *10*(3), 191–201. <https://doi.org/10.1177/1754073918765662>
- Agren, T., Engman, J., Frick, A., Bjorkstrand, J., Larsson, E.-M., Furmark, T., & Fredrikson, M. (2012). Disruption of Reconsolidation Erases a Fear Memory Trace in the Human Amygdala. *Science*, *337*(6101), 1550–1552. <https://doi.org/10.1126/science.1223006>
- Agren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E. M., Furmark, T., & Fredrikson, M. (2012). Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science*, *337*(6101), 1550–1552. <https://doi.org/10.1126/science.1223006>
- Agustina López, M., Jimena Santos, M., Cortasa, S., Fernández, R. S., Carbó Tano, M., & Pedreira, M. E. (2016). Different dimensions of the prediction error as a decisive factor for the triggering of the reconsolidation process. *Neurobiology of Learning and Memory*, *136*, 210–219. <https://doi.org/10.1016/j.nlm.2016.10.016>
- Åhs, F., Rosén, J., Kastrati, G., Fredrikson, M., Agren, T., & Lundström, J. N. (2018). Biological preparedness and resistance to extinction of skin conductance responses conditioned to fear relevant animal pictures: A systematic review. *Neuroscience and Biobehavioral Reviews*, *95*(October), 430–437. <https://doi.org/10.1016/j.neubiorev.2018.10.017>
- Alberini, C. M. (2008). The role of protein synthesis during the labile phases of memory: Revisiting the skepticism. *Neurobiology of Learning and Memory*, *89*(3), 234–246. <https://doi.org/10.1016/j.nlm.2007.08.007>
- Alfei, J. M., Monti, R. I. F., Molina, V. A., Bueno, A. M., & Urcelay, G. P. (2015). Prediction error and trace dominance determine the fate of fear memories after post-training manipulations. *Learning and Memory*, *22*(8), 385–400. <https://doi.org/10.1101/lm.038513.115>
- Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., Lucock, M., Lutz, W., & Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy*, *94*, 1–8. <https://doi.org/10.1016/j.brat.2017.04.006>
- Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review*, *101*(3), 414–445. <https://doi.org/10.1037/0033-295X.101.3.414>
- Andreatta, M., Glotzbach-Schoon, E., Mühlberger, A., Schulz, S. M., Wiemer, J., & Pauli, P. (2015). Initial and sustained brain responses to contextual conditioned anxiety in humans. *Cortex*, *63*, 352–363. <https://doi.org/10.1016/j.cortex.2014.09.014>
- Askew, C., & Field, A. P. (2008). The vicarious learning pathway to fear 40 years on. *Clinical Psychology Review*, *28*(7), 1249–1265. <https://doi.org/10.1016/j.cpr.2008.05.003>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bach, D. R., Weiskopf, N., & Dolan, R. J. (2011). A stable sparse fear memory trace in human amygdala. *Journal of Neuroscience*, *31*(25), 9383–9389. <https://doi.org/10.1523/JNEUROSCI.1524-11.2011>
- Ball, T. M., Knapp, S. E., Paulus, M. P., & Stein, M. B. (2017). Brain activation during fear extinction predicts exposure success. *Depression and Anxiety*, *34*(3), 257–266. <https://doi.org/10.1002/da.22583>
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. <https://doi.org/10.1177/0963721411422522>
- Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the fribbles: Novel stimuli for

- use within behavioural research. *Frontiers in Psychology*, 5(FEB), 1–8. <https://doi.org/10.3389/fpsyg.2014.00103>
- Bisaz, R., Travaglia, A., & Alberini, C. M. (2014). The neurobiological bases of memory formation: from physiological conditions to psychopathology. *Psychopathology*, 47(6), 347–356. <https://doi.org/10.1159/000363702>
- Björkstrand, J., Agren, T., Åhs, F., Frick, A., Larsson, E. M., Hjorth, O., Furmark, T., & Fredrikson, M. (2016). Disrupting Reconsolidation Attenuates Long-Term Fear Memory in the Human Amygdala and Facilitates Approach Behavior. *Current Biology*, 26(19), 2690–2695. <https://doi.org/10.1016/j.cub.2016.08.022>
- Björkstrand, J., Agren, T., Frick, A., Engman, J., Larsson, E.-M., Furmark, T., & Fredrikson, M. (2015). Disruption of Memory Reconsolidation Erases a Fear Memory Trace in the Human Amygdala: An 18-Month Follow-Up. *PloS One*, 10(7), e0129393. <https://doi.org/10.1371/journal.pone.0129393>
- Blechert, J., Wilhelm, F. H., Williams, H., Braams, B. R., Jou, J., & Gross, J. J. (2015). Reappraisal facilitates extinction in healthy and socially anxious individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, 46, 141–150. <https://doi.org/10.1016/j.jbtep.2014.10.001>
- Bos, M. G. N., Beckers, T., & Kindt, M. (2014). Noradrenergic blockade of memory reconsolidation: A failure to reduce conditioned fear responding. *Frontiers in Behavioral Neuroscience*, 8(NOV), 1–8. <https://doi.org/10.3389/fnbeh.2014.00412>
- Bouton, M. E. (2002). Context, Ambiguity, and Unlearning: Sources of Relapse after Behavioural Extinction. *Biological Psychiatry*, 52(10), 976–986. [http://www.abs.gov.au/ausstats/subscriber.nsf/log?openagent&4364055001do009\\_20142015.xls&4364.0.55.001&Data=Cubes&06977CCBBDA622B4CA257F150009FA28&0&2014-15&08.12.2015&Latest](http://www.abs.gov.au/ausstats/subscriber.nsf/log?openagent&4364055001do009_20142015.xls&4364.0.55.001&Data=Cubes&06977CCBBDA622B4CA257F150009FA28&0&2014-15&08.12.2015&Latest)
- Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 11(5), 485–494. <https://doi.org/10.1101/lm.78804>
- Bouton, M. E. (2017). Extinction: Behavioral Mechanisms and Their Implications ☆. In *Learning and Memory: A Comprehensive Reference* (Second Edi, Vol. 1, Issue August 2016). Elsevier. <https://doi.org/10.1016/b978-0-12-809324-5.21006-7>
- Bouton, M. E., Mineka, S., & Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorder. *Psychological Review*, 108(1), 4–32. <https://doi.org/10.1037/0033-295X.108.1.4>
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607. <https://doi.org/10.1111/j.1469-8986.2008.00654.x>
- Brascamp, J. W., & Naber, M. (2016). Eye tracking under dichoptic viewing conditions: a practical solution. *Behavior Research Methods*, 1–7. <https://doi.org/10.3758/s13428-016-0805-2>
- Brockelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghofer, M. (2011a). Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing: Magnetoencephalographic Correlates. *Journal of Neuroscience*, 31(21), 7801–7810. <https://doi.org/10.1523/JNEUROSCI.6236-10.2011>
- Brockelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghofer, M. (2011b). Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing: Magnetoencephalographic Correlates. *Journal of Neuroscience*, 31(21), 7801–7810. <https://doi.org/10.1523/JNEUROSCI.6236-10.2011>
- Brosch, T., & Sander, D. (2013). Comment: The appraising brain: Towards a neuro-cognitive model of appraisal processes in emotion. *Emotion Review*, 5(2), 163–168. <https://doi.org/10.1177/1754073912468298>
- Brown, R., Lau, H., & Ledoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, 23(9), 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>
- Buhle, J. T., Silvers, J. A., Wage, T. D., Lopez, R., Onyemekwu, C., Kober, H., Webe, J., & Ochsner,

- K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex*, *24*(11), 2981–2990. <https://doi.org/10.1093/cercor/bht154>
- Carpenter, J. K., Pinaire, M., & Hofmann, S. G. (2019). From extinction learning to anxiety treatment: Mind the gap. *Brain Sciences*, *9*(7). <https://doi.org/10.3390/brainsci9070164>
- Cassini, L. F., Flavell, C. R., Amaral, O. B., & Lee, J. L. C. (2017). On the transition from reconsolidation to extinction of contextual fear memories. *Learning and Memory*, *24*(9), 392–399. <https://doi.org/10.1101/lm.045724.117>
- Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences*, *6*(12), 524–531.
- Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1742). <https://doi.org/10.1098/rstb.2017.0025>
- Craske, M. G., & Mystkowski, J. L. (2006). Exposure Therapy and Extinction: Clinical Studies. In *Fear and learning: From basic processes to clinical implications*. (pp. 217–233). American Psychological Association. <https://doi.org/10.1037/11474-011>
- Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H. U. (2017). Anxiety disorders. *Nature Reviews Disease Primers*, *3*(May). <https://doi.org/10.1038/nrdp.2017.24>
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. <https://doi.org/10.1016/j.brat.2014.04.006>
- Critchley, H. D. (2002). Book Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, *8*(2), 132–142. <https://doi.org/10.1177/107385840200800209>
- Culver, N. C., Vervliet, B., & Craske, M. G. (2015). Compound Extinction: Using the Rescorla–Wagner Model to Maximize Exposure Therapy Effects for Anxiety Disorders. *Clinical Psychological Science*, *3*(3), 335–348. <https://doi.org/10.1177/2167702614542103>
- Das, R. K., Lawn, W., & Kamboj, S. K. (2015). Rewriting the valuation and salience of alcohol-related stimuli via memory reconsolidation. *Translational Psychiatry*, *5*, e645. <https://doi.org/10.1038/tp.2015.132>
- Davidson, R. J. (2003). Seven sins in the study of emotion: Correctives from affective neuroscience. *Brain and Cognition*, *52*(1), 129–132. [https://doi.org/10.1016/S0278-2626\(03\)00015-0](https://doi.org/10.1016/S0278-2626(03)00015-0)
- Davis, K. L., Montag, C., & Davis, K. L. (2019). *Selected Principles of Pankseppian Affective Neuroscience*. *12*(January), 1–11. <https://doi.org/10.3389/fnins.2018.01025>
- Dawson, M. E., Catania, J. J., Schell, A. M., & Grings, W. W. (1979). Autonomic classical conditioning as a function of awareness of stimulus contingencies. *Biological Psychology*, *9*(1), 23–40. [https://doi.org/10.1016/0301-0511\(79\)90020-6](https://doi.org/10.1016/0301-0511(79)90020-6)
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*(6), 853–869. <https://doi.org/10.1037/0033-2909.127.6.853>
- Dębiec, J., Bush, D. E. A., & LeDoux, J. E. (2011). Noradrenergic enhancement of reconsolidation in the amygdala impairs extinction of conditioned fear in rats—a possible mechanism for the persistence of traumatic memories in PTSD. *Depression and Anxiety*, *28*(3), 186–193. <https://doi.org/10.1002/da.20803>
- Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala response to emotional stimuli without awareness: Facts and interpretations. *Frontiers in Psychology*, *7*(JAN), 1–13. <https://doi.org/10.3389/fpsyg.2016.02029>
- Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning and Memory*, *11*(5), 549–554. <https://doi.org/10.1101/lm.78004>
- Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2007). Reinstatement of conditioned responses in human differential fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(3), 237–251. <https://doi.org/10.1016/j.jbtep.2006.04.001>
- Dolman, C. P. (1919). Tests for Determining the Sighting Eye. *American Journal of Ophthalmology*,

- 2(12), 867. [https://doi.org/10.1016/S0002-9394\(19\)90258-3](https://doi.org/10.1016/S0002-9394(19)90258-3)
- Doyère, V., Dèbiec, J., Monfils, M. H., Schafe, G. E., & LeDoux, J. E. (2007). Synapse-specific reconsolidation of distinct fear memories in the lateral amygdala. *Nature Neuroscience*, *10*(4), 414–416. <https://doi.org/10.1038/nn1871>
- Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? *Annual Review of Psychology*, *55*(1), 51–86. <https://doi.org/10.1146/annurev.psych.55.090902.142050>
- Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. <https://doi.org/10.1017/S0140525X00032325>
- Dunsmoor, J. E., Kroes, M. C. W., Li, J., Daw, N. D., Simpson, H. B., & Phelps, E. A. (2019). Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. *Journal of Neuroscience*, *39*(17), 3264–3276. <https://doi.org/10.1523/JNEUROSCI.2713-18.2019>
- Dunsmoor, J. E., Niv, Y., Daw, N., & Phelps, E. A. (2015). Rethinking Extinction. *Neuron*, *88*(1), 47–63. <https://doi.org/10.1016/j.neuron.2015.09.028>
- Eelen, P., & Vervliet, B. (2007). Fear Conditioning and Clinical Implications: What Can We Learn From the Past? In *Fear and learning: From basic processes to clinical implications*. (pp. 17–35). American Psychological Association. <https://doi.org/10.1037/11474-001>
- Elsley, J. W. B., & Kindt, M. (2017a). Breaking boundaries: Optimizing reconsolidation-based interventions for strong and old memories. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *24*(9), 472–479. <https://doi.org/10.1101/lm.044156.116>
- Elsley, J. W. B., & Kindt, M. (2017b). Tackling maladaptive memories through reconsolidation: From neural to clinical science. *Neurobiology of Learning and Memory*, *142*, 108–117. <https://doi.org/10.1016/j.nlm.2017.03.007>
- Elsley, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin*, *144*(8), 797–848. <https://doi.org/10.1037/bul0000152>
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*, *26*(8), 3508–3526. <https://doi.org/10.1093/cercor/bhw157>
- Fang, Z., Li, H., Chen, G., & Yang, J. (2016). *Unconscious Processing of Negative Animals and Objects: Role of the Amygdala Revealed by fMRI*. *10*(April), 1–12. <https://doi.org/10.3389/fnhum.2016.00146>
- Fanselow, M. S. (2000). Contextual fear, gestalt memories, and the hippocampus. *Behavioural Brain Research*, *110*(1–2), 73–81. [https://doi.org/10.1016/S0166-4328\(99\)00186-2](https://doi.org/10.1016/S0166-4328(99)00186-2)
- Fanselow, M. S., & Wassum, K. M. (2016). The origins and organization of vertebrate pavlovian conditioning. *Cold Spring Harbor Perspectives in Biology*, *8*(1), 1–28. <https://doi.org/10.1101/cshperspect.a021717>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Feinstein, J., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, *21*(1), 34–38. <https://doi.org/10.1016/j.cub.2010.11.042>
- Feng, P., Zheng, Y., & Feng, T. (2015). Spontaneous brain activity following fear reminder of fear conditioning by using resting-state functional MRI. *Scientific Reports*, *5*(1), 16701. <https://doi.org/10.1038/srep16701>
- Flykt, A., Esteves, F., & Öhman, A. (2007). Skin conductance responses to masked conditioned stimuli: Phylogenetic/ontogenetic factors versus direction of threat? *Biological Psychology*, *74*(3), 328–336. <https://doi.org/10.1016/j.biopsycho.2006.08.004>
- Foa, E. B., & Kozak, M. J. (1986). Emotional Processing of Fear. Exposure to Corrective Information. *Psychological Bulletin*, *99*(1), 20–35. <https://doi.org/10.1037/0033-2909.99.1.20>
- Fricchione, J., Greenberg, M. S., Spring, J., Wood, N., Mueller-Pfeiffer, C., Milad, M. R., Pitman, R. K., & Orr, S. P. (2016). Delayed extinction fails to reduce skin conductance reactivity to fear-



- conditioned stimuli. *Psychophysiology*, 53(9), 1343–1351. <https://doi.org/10.1111/psyp.12687>
- Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews*, 88(December 2017), 16–25. <https://doi.org/10.1016/j.neubiorev.2018.03.002>
- Fullana, M. A., Harrison, B., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, 21, 500–508. <https://doi.org/10.1038/mp.2015.88>
- Gaal, S. Van, Ridderinkhof, K. R., Scholte, H. S., & Lamme, V. A. F. (2010). *Unconscious Activation of the Prefrontal No-Go Network*. 30(11), 4143–4150. <https://doi.org/10.1523/JNEUROSCI.2992-09.2010>
- Garfinkel, S. N., Abelson, J. L., King, A. P., Sripada, R. K., Wang, X., Gaines, L. M., & Liberzon, I. (2014). Impaired contextual modulation of memories in PTSD: An fMRI and psychophysiological study of extinction retention and fear renewal. *Journal of Neuroscience*, 34(40), 13435–13443. <https://doi.org/10.1523/JNEUROSCI.4287-13.2014>
- Gayet, S., Paffen, C. L. E., Belopolsky, A. V., Theeuwes, J., & Van der Stigchel, S. (2016). Visual input signaling threat gains preferential access to awareness in a breaking continuous flash suppression paradigm. *Cognition*, 149, 77–83. <https://doi.org/10.1016/j.cognition.2016.01.009>
- Germeroth, L. J., Carpenter, M. J., Baker, N. L., Froeliger, B., LaRowe, S. D., & Saladin, M. E. (2017). Effect of a brief memory updating intervention on smoking behavior: A randomized clinical trial. *JAMA Psychiatry*, 74(3), 214–223. <https://doi.org/10.1001/jamapsychiatry.2016.3148>
- Gershman, S. J., Monfils, M.-H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *ELife*, 6, 1–41. <https://doi.org/10.7554/eLife.23763>
- Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning and Behavior*, 40, 255–268. <https://doi.org/10.3758/s13420-012-0080-8>
- Gewirtz, J. C., & Davis, M. (2000). Using Pavlovian higher-order conditioning paradigms to investigate the neural substrates of emotional learning and memory. *Learning and Memory*, 7(5), 257–266. <https://doi.org/10.1101/lm.35200>
- Gilmartin, M. R., Balderston, N. L., & Helmstetter, F. J. (2014). Prefrontal cortical regulation of fear learning. *Trends in Neurosciences*, 37(8), 445–464. <https://doi.org/10.1016/j.tins.2014.05.004>
- Golkar, A., Bellander, M., Olsson, A., & Öhman, A. (2012). Are fear memories erasable?—reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Frontiers in Behavioral Neuroscience*, 6(November), 1–10. <https://doi.org/10.3389/fnbeh.2012.00080>
- Golkar, A., & Öhman, A. (2012). Fear extinction in humans: Effects of acquisition-extinction delay and masked stimulus presentations. *Biological Psychology*, 91(2), 292–301. <https://doi.org/10.1016/j.biopsycho.2012.07.007>
- Gomes, N., Silva, S., Silva, C. F., & Soares, S. C. (2017). Beware the serpent: the advantage of ecologically-relevant stimuli in accessing visual awareness. *Evolution and Human Behavior*, 38(2), 227–234. <https://doi.org/10.1016/j.evolhumbehav.2016.10.004>
- Greco, J. A., & Liberzon, I. (2016). Neuroimaging of Fear-Associated Learning. In *Neuropsychopharmacology* (Vol. 41, Issue 1, pp. 320–334). Nature Publishing Group. <https://doi.org/10.1038/npp.2015.255>
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.
- Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014a). A review on human reinstatement studies: an overview and methodological challenges. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 21, 424–440.
- Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014b). A review on human reinstatement studies: an overview and methodological challenges. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 21(9), 424–440. <https://doi.org/10.1101/lm.036053.114>
- Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology*, 57(1), 5–14. <https://doi.org/10.1016/j.ijpsycho.2005.01.006>

- Hartley, C. A., & Phelps, E. A. (2010). Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 35(1), 136–146. <https://doi.org/10.1038/npp.2009.121>
- Hedger, N., Gray, K. L. H., Garner, M., & Adams, W. J. (2016). Are visual threats prioritized without awareness? A critical review and meta-analysis involving 3 behavioral paradigms and 2696 observers. *Psychological Bulletin*, 142(9), 934–968. <https://doi.org/10.1037/bul0000054>
- Helpman, L., Marin, M. F., Papini, S., Zhu, X., Sullivan, G. M., Schneier, F., Neria, M., Shvil, E., Malaga Aragon, M. J., Markowitz, J. C., Lindquist, M. A., Wager, T., Milad, M., & Neria, Y. (2016). Neural changes in extinction recall following prolonged exposure treatment for PTSD: A longitudinal fMRI study. *NeuroImage: Clinical*, 12, 715–723. <https://doi.org/10.1016/j.nicl.2016.10.007>
- Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. *Biological Psychiatry*, 60(4), 361–368. <https://doi.org/10.1016/j.biopsych.2005.10.006>
- Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van Den Bergh, O., & Eelen, P. (2005a). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy*, 43(4), 533–551. <https://doi.org/10.1016/j.brat.2004.03.013>
- Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van Den Bergh, O., & Eelen, P. (2005b). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy*, 43(4), 533–551. <https://doi.org/10.1016/j.brat.2004.03.013>
- Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, 40(3), 217–234. [https://doi.org/10.1016/S0005-7967\(01\)00006-7](https://doi.org/10.1016/S0005-7967(01)00006-7)
- Ho, Y., & Lipp, O. V. (2014). Faster acquisition of conditioned fear to fear-relevant than to nonfear-relevant conditional stimuli. *Psychophysiology*, 51(8), 810–813. <https://doi.org/10.1111/psyp.12223>
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative Conditioning in Humans: A Meta-Analysis. *Psychological Bulletin*, 136(3), 390–421. <https://doi.org/10.1037/a0018916>
- Huang, L., Yang, T., & Li, Z. (2003). Applicability of the Positive and Negative Affect Scale in Chinese. *Chinese Mental Health Journal*, 17(1), 54–56.
- Huston, J. P., & Ly, A. A. B. (1974). *The Thalamic Rat: General Behavior, Operant Learning with Rewarding Hypothalamic Stimulation, and Effects of Amphetamine*. 12(3), 433–448.
- Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., Nobre, A. C., & Holmes, E. A. (2018). Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: A proof-of-concept randomized controlled trial. *Molecular Psychiatry*, 23(3), 674–682. <https://doi.org/10.1038/mp.2017.23>
- James, E. L., Bonsall, M. B., Hoppitt, L., Tunbridge, E. M., Geddes, J. R., Milton, A. L., & Holmes, E. A. (2015). Computer game play reduces intrusive memories of experimental trauma via reconsolidation-update mechanisms. *Psychological Science*, 1–15. <https://doi.org/10.1177/0956797615583071>
- James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., ... Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159), 1789–1858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jones, M. C. (1924). A Laboratory Study of Fear: The Case Of Peter. *The Pedagogical Seminary and*

- Journal of Genetic Psychology*, 31(4), 308–315.  
<https://doi.org/10.1080/08856559.1924.9944851>
- Junghöfer, M., Rehbein, M. A., Maitzen, J., Schindler, S., & Kissler, J. (2017). An evil face? Verbal evaluative multi-CS conditioning enhances face-evoked mid-latency magnetoencephalographic responses. *Social Cognitive and Affective Neuroscience*, 12(4), 695–705.  
<https://doi.org/10.1093/scan/nsw179>
- Junghofer, M., Winker, C., Rehbein, M. A., & Sabatinelli, D. (2017). Noninvasive Stimulation of the Ventromedial Prefrontal Cortex Enhances Pleasant Scene Processing. *Cerebral Cortex*, 27(6), 3449–3456. <https://doi.org/10.1093/cercor/bhx073>
- Kalisch, R., & Gerlicher, A. M. V. (2014). Making a mountain out of a molehill: On the role of the rostral dorsal anterior cingulate and dorsomedial prefrontal cortex in conscious threat appraisal, catastrophizing, and worrying. *Neuroscience and Biobehavioral Reviews*, 42, 1–8.  
<https://doi.org/10.1016/j.neubiorev.2014.02.002>
- Kalisch, R., Korenfeld, E., Stephan, K. E., Weiskopf, N., Seymour, B., & Dolan, R. J. (2006). Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *Journal of Neuroscience*, 26(37), 9503–9511.  
<https://doi.org/10.1523/JNEUROSCI.2021-06.2006>
- Kar, S. K., & Sarkar, S. (2016). Neuro-stimulation techniques for the management of anxiety disorders: An update. *Clinical Psychopharmacology and Neuroscience*, 14(4), 330–337.  
<https://doi.org/10.9758/cpn.2016.14.4.330>
- Karpinski, A., Briggs, J. C., & Yale, M. (2019). A direct replication: Unconscious arithmetic processing. *European Journal of Social Psychology*, 49, 637–644.  
<https://doi.org/10.1002/ejsp.2390>
- Katkin, E. S., Wiens, S., & Öhman, A. (2001). Nonconscious fear conditioning, visceral perception, and the development of gut feelings. *Psychological Science*, 12(5), 366–370.  
<https://doi.org/10.1111/1467-9280.00368>
- Kattoor, J., Thürling, M., Gizewski, E. R., Forsting, M., Timmann, D., & Elsenbruch, S. (2014). Cerebellar contributions to different phases of visceral aversive extinction learning. *Cerebellum*, 13(1), 1–8. <https://doi.org/10.1007/s12311-013-0512-9>
- Kindt, M. (2018). The surprising subtleties of changing fear memory: A challenge for translational science. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1742).  
<https://doi.org/10.1098/rstb.2017.0033>
- Kindt, M., & Soeter, M. (2013a). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology*, 92(1), 43–50.  
<https://doi.org/10.1016/j.biopsycho.2011.09.016>
- Kindt, M., & Soeter, M. (2013b). Reconsolidation in a human fear conditioning study: a test of extinction as updating mechanism. *Biological Psychology*, 92(1), 43–50.  
<https://doi.org/10.1016/j.biopsycho.2011.09.016>
- Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nature Neuroscience*, 12(3), 256–258.  
<https://doi.org/10.1038/nn.2271>
- Klucken, T., Kruse, O., Schweckendiek, J., Kuepper, Y., Mueller, E. M., Hennig, J., & Stark, R. (2016). No evidence for blocking the return of fear by disrupting reconsolidation prior to extinction learning. *Cortex*, 79, 112–122. <https://doi.org/10.1016/j.cortex.2016.03.015>
- Kober, H., Barrett, F., Joseph, J., Bliss-moreau, E., Lindquist, K., & Wager, T. D. (2008). *Functional grouping and cortical – subcortical interactions in emotion: A meta-analysis of neuroimaging studies*. 42, 998–1031. <https://doi.org/10.1016/j.neuroimage.2008.03.059>
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, 11(1), 16–22. <https://doi.org/10.1016/j.tics.2006.10.012>
- Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., & Lau, H. (2017a). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behaviour*, 1(1), 1–7. <https://doi.org/10.1038/s41562-016-0006>

- Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., & Lau, H. (2017b). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behaviour*, *1*(1), 1–7. <https://doi.org/10.1038/s41562-016-0006>
- Korn, C. W., Staib, M., Tzovara, A., Castegnetti, G., & Bach, D. R. (2017). A pupil size response model to assess fear learning. *Psychophysiology*, *54*(3), 330–343. <https://doi.org/10.1111/psyp.12801>
- Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 857–875. <https://doi.org/10.1098/rstb.2007.2093>
- Kouider, S., Eger, E., Dolan, R., & Henson, R. N. (2009). Activity in face-responsive brain regions is modulated by invisible, attended faces: Evidence from masked priming. *Cerebral Cortex*, *19*(1), 13–23. <https://doi.org/10.1093/cercor/bhn048>
- Kredlow, M. A., Unger, L. D., & Otto, M. W. (2016). Harnessing reconsolidation to weaken fear and appetitive memories: a meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, *142*(3), Online First Publication. <https://doi.org/10.1037/bul0000034>
- Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336–1342. <https://doi.org/10.3758/s13428-018-1075-y>
- Kroes, M. C. W., Dunsmoor, J. E., Hakimi, M., Oosterwaal, S., Meager, M. R., & Phelps, E. A. (2019). Patients with dorsolateral prefrontal cortex lesions are capable of discriminatory threat learning but appear impaired in cognitive regulation of subjective fear. *Social Cognitive and Affective Neuroscience*, *14*(6), 601–612. <https://doi.org/10.1093/scan/nsz039>
- Kroes, M. C. W., Dunsmoor, J. E., Mackey, W. E., McClay, M., & Phelps, E. A. (2017). Context conditioning in humans using commercially available immersive Virtual Reality. *Scientific Reports*, *7*(1), 1–14. <https://doi.org/10.1038/s41598-017-08184-7>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). <https://doi.org/10.18637/jss.v082.i13>
- Kyu, H. H., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, M., Abebe, Z., Abil, O. Z., Aboyans, V., Abrham, A. R., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., ... Murray, C. J. L. (2018). Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1859–1922. [https://doi.org/10.1016/S0140-6736\(18\)32335-3](https://doi.org/10.1016/S0140-6736(18)32335-3)
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. <https://doi.org/10.1038/nrn1825>
- Labar, K. S., Gatenby, J. C., Gore, J. C., Ledoux, J. E., & Phelps, E. A. (1998). Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study to investigate amygdala function in human populations have produced inconsistent results across techniques. Whereas neuropsychological studies have repo. *Neuron*, *20*, 937–945.
- Laborda, M. A., & Miller, R. R. (2011). S-R Associations, Their Extinction, and Recovery in an Animal Model of Anxiety: A New Associative Account of Phobias Without Recall of Original Trauma. *Behavior Therapy*, *42*(2), 153–169. <https://doi.org/10.1016/j.beth.2010.06.002>
- Lang, P. J., Davis, M., & Öhman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of Affective Disorders*, *61*(3), 137–159. [https://doi.org/10.1016/S0165-0327\(00\)00343-8](https://doi.org/10.1016/S0165-0327(00)00343-8)
- Lau, H. C., & Passingham, R. E. (2007). Unconscious Activation of the Cognitive Control System in the Human Prefrontal Cortex. *Journal of Neuroscience*, *27*(21), 5805–5811. <https://doi.org/10.1523/JNEUROSCI.4335-06.2007>
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- LeDoux, J. (2012). Rethinking the Emotional Brain. *Neuron*, *73*(4), 653–676.

- <https://doi.org/10.1016/j.neuron.2012.02.004>
- Ledoux, J., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, *19*(5), 269–282. <https://doi.org/10.1038/nrn.2018.22>
- Ledoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, *111*(4), 2871–2878. <https://doi.org/10.1073/pnas.1400335111>
- LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(8), 2871–2878. <https://doi.org/10.1073/pnas.1400335111>
- Ledoux, J. E., & Muller, J. (1997). Emotional memory and psychopathology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1362), 1719–1726. <https://doi.org/10.1098/rstb.1997.0154>
- LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *American Journal of Psychiatry*, *173*(11), 1083–1093. <https://doi.org/10.1176/appi.ajp.2016.16030353>
- Lee, J. L. C. (2009). Reconsolidation: maintaining memory relevance. *Trends in Neurosciences*, *32*(8), 413–420. <https://doi.org/10.1016/j.tins.2009.05.002>
- Lee, J. L. C., Nader, K., & Schiller, D. (2017). An Update on Memory Reconsolidation Updating. *Trends in Cognitive Sciences*, *21*(7), 531–545. <https://doi.org/10.1016/j.tics.2017.04.006>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package “emmeans.” *Mran.Microsoft.Com*, *34*(1), 126. <https://doi.org/10.1080/00031305.1980.10483031>>License
- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1). <https://doi.org/10.18637/jss.v069.i01>
- Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017a). Neural correlates of pupil dilation during human fear learning. *NeuroImage*, *147*(August 2016), 186–197. <https://doi.org/10.1016/j.neuroimage.2016.11.072>
- Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017b). Neural correlates of pupil dilation during human fear learning. *NeuroImage*, *147*(August 2016), 186–197. <https://doi.org/10.1016/j.neuroimage.2016.11.072>
- Leuchs, L., Schneider, M., & Spoormaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, *56*(1), 1–16. <https://doi.org/10.1111/psyp.13283>
- Leung, H. T., Reeks, L. M., & Westbrook, R. F. (2012). Two ways to deepen extinction and the difference between them. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(4), 394–406. <https://doi.org/10.1037/a0030201>
- Li, S., & Graham, B. M. (2016). Estradiol is associated with altered cognitive and physiological responses during fear conditioning and extinction in healthy and spider phobic women. *Behavioral Neuroscience*, *130*(6), 614–623. <https://doi.org/10.1037/bne0000166>
- Lipp, O. V., Kempnich, C., Jee, S. H., & Arnold, D. H. (2014). Fear conditioning to subliminal fear relevant and non fear relevant stimuli. *PLoS ONE*, *9*(9). <https://doi.org/10.1371/journal.pone.0099332>
- Lissek, S., Baas, J. M. P., Pine, D. S., Orme, K., Dvir, S., Nugent, M., Rosenberger, E., Rawson, E., & Grillon, C. (2005). Airpuff startle probes: An efficacious and less aversive alternative to white-noise. *Biological Psychology*, *68*(3), 283–297. <https://doi.org/10.1016/j.biopsycho.2004.07.007>
- Lissek, S., Levenson, J., Biggs, A. L., Johnson, L. L., Ameli, R., Pine, D. S., & Grillon, C. (2008). Elevated fear conditioning to socially relevant unconditioned stimuli in social anxiety disorder. *American Journal of Psychiatry*, *165*(1), 124–132. <https://doi.org/10.1176/appi.ajp.2007.06091513>
- Liu, J., Zhao, L., Xue, Y., Shi, J., Suo, L., Luo, Y., Chai, B., Yang, C., Fang, Q., Zhang, Y., Bao, Y., Pickens, C. L., & Lu, L. (2014). An Unconditioned Stimulus Retrieval Extinction Procedure to Prevent the Return of Fear Memory. *Biological Psychiatry*, *76*(11), 895–901. <https://doi.org/10.1016/j.biopsycho.2014.03.027>
- Lonergan, M. H., Olivera-Figueroa, L. A., Pitman, R. K., & Brunet, A. (2013). Propranolol’s effects

- on the consolidation and reconsolidation of long-term emotional memory in healthy participants: a meta-analysis. *Journal of Psychiatry & Neuroscience: JPN*, 38(4), 222–231. <https://doi.org/10.1503/jpn.120111>
- Lonsdorf, T. B., Haaker, J., Fadai, T., & Kalisch, R. (2014). No evidence for enhanced extinction memory consolidation through noradrenergic reuptake inhibition - Delayed memory test and reinstatement in human fMRI. *Psychopharmacology*, 231(9), 1949–1962. <https://doi.org/10.1007/s00213-013-3338-8>
- Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., ... 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 and Biobehavioral Reviews*, 77, 247–285. <https://doi.org/10.1016/j.neubiorev.2017.02.026>
- Lonsdorf, T. B., Merz, C. J., & Fullana, M. A. (2019). Fear Extinction Retention: Is It What We Think It Is? *Biological Psychiatry*, 85(12), 1074–1082. <https://doi.org/10.1016/j.biopsych.2019.02.011>
- Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28(1), 3–26. <https://doi.org/10.1037/0097-7403.28.1.3>
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF)*. CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9. ISBN 91-630-7164-9
- Luo, Y. X., Xue, Y. X., Liu, J. F., Shi, H. S., Jian, M., Han, Y., Zhu, W. L., Bao, Y. P., Wu, P., Ding, Z. B., Shen, H. W., Shi, J., Shaham, Y., & Lu, L. (2015). A novel UCS memory retrieval-extinction procedure to inhibit relapse to drug seeking. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms8675>
- Mallan, K. M., Lipp, O. V., & Cochrane, B. (2013). Slithering snakes, angry men and out-group members: What and whom are we evolved to fear? *Cognition and Emotion*, 27(7), 1168–1180. <https://doi.org/10.1080/02699931.2013.778195>
- Manber Ball, T., Ramsawh, H. J., Campbell-Sills, L., Paulus, M. P., & Stein, M. B. (2013). Prefrontal dysfunction during emotion regulation in generalized anxiety and panic disorders. *Psychological Medicine*, 43(7), 1475–1486. <https://doi.org/10.1017/S0033291712002383>
- Maples-Keller, J. L., Price, M., Jovanovic, T., Norrholm, S. D., Odenat, L., Post, L., Zwiebach, L., Breazeale, K., Gross, R., Kim, S. J., & Rothbaum, B. O. (2017). Targeting memory reconsolidation to prevent the return of fear in patients with fear of flying. *Depression and Anxiety*, 34(7), 610–620. <https://doi.org/10.1002/da.22626>
- Maren, S., & Holmes, A. (2016). Stress and fear extinction. In *Neuropsychopharmacology* (Vol. 41, Issue 1, pp. 58–79). Nature Publishing Group. <https://doi.org/10.1038/npp.2015.180>
- Maren, S., Phan, K. L., & Liberzon, I. (2013a). The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, 14(June), 417–428. <https://doi.org/10.1038/nrn3492>
- Maren, S., Phan, K. L., & Liberzon, I. (2013b). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, 14(6), 417–428. <https://doi.org/10.1038/nrn3492>
- Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, 1(1), 1–23. <https://doi.org/10.5334/joc.18>
- Mcgaugh, J. L., & Mcgaugh, J. L. (2012). *Memory — a Century of Consolidation*. 248(2000), 248–252. <https://doi.org/10.1126/science.287.5451.248>
- Meir Drexler, S., Merz, C. J., Hamacher-Dang, T. C., Marquardt, V., Fritsch, N., Otto, T., & Wolf, O. T. (2014). Effects of postretrieval-extinction learning on return of contextually controlled cued fear. *Behavioral Neuroscience*, 128(4), 474–481. <https://doi.org/10.1037/a0036688>
- Meir Drexler, S., Merz, C. J., Hamacher-Dang, T. C., & Wolf, O. T. (2016). Cortisol effects on fear memory reconsolidation in women. *Psychopharmacology*, 233(14), 2687–2697.

- <https://doi.org/10.1007/s00213-016-4314-x>
- Meissner, C. A., & Brigham, J. C. (2001). Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review. *Psychology, Public Policy, and Law*, 7(1), 3–35. <https://doi.org/10.1037/1076-8971.7.1.3>
- Meneguzzo, P., Tsakiris, M., Schioth, H. B., Stein, D. J., & Brooks, S. J. (2014). Subliminal versus supraliminal stimuli activate neural responses in anterior cingulate cortex, fusiform gyrus and insula: a meta-analysis of fMRI studies. *BMC Psychology*, 2(1). <https://doi.org/10.1186/s40359-014-0052-1>
- Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30(1), 63–81. <https://doi.org/10.1017/S0140525X07000891>
- Merlo, E., Milton, A. L., Goozee, Z. Y., Theobald, D. E., & Everitt, B. J. (2014). Reconsolidation and extinction are dissociable and mutually exclusive processes: Behavioral and molecular evidence. *The Journal of Neuroscience*, 34(7), 2422–2431. <https://doi.org/10.1523/JNEUROSCI.4001-13.2014>
- Mertens, G., & De Houwer, J. (2016). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. *Biological Psychology*, 113, 91–99. <https://doi.org/10.1016/j.biopsycho.2015.11.014>
- Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neuroscience & Biobehavioral Reviews*, 108, 254–268. <https://doi.org/10.1016/j.neubiorev.2019.11.012>
- Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, 42(4), 456–464. <https://doi.org/10.1111/j.1469-8986.2005.00302.x>
- 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. <https://doi.org/10.1146/annurev.psych.121208.131631>
- Milad, M. R., Rosenbaum, B. L., & Simon, N. M. (2014). Neuroscience of fear extinction: Implications for assessment and treatment of fear-based and anxiety related disorders. *Behaviour Research and Therapy*, 62, 17–23. <https://doi.org/10.1016/j.brat.2014.08.006>
- Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of Fear Extinction in Humans Activates the Ventromedial Prefrontal Cortex and Hippocampus in Concert. *Biological Psychiatry*, 62(5), 446–454. <https://doi.org/10.1016/j.biopsych.2006.10.011>
- Milligan-Saville, J. S., & Graham, B. M. (2016). Mothers do it differently: reproductive experience alters fear extinction in female rats and women. *Translational Psychiatry*, 6(10), e928. <https://doi.org/10.1038/tp.2016.193>
- Milton, A. L., & Holmes, E. A. (2018). Of mice and mental health: Facilitating dialogue and seeing further. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1742), 1–6. <https://doi.org/10.1098/rstb.2017.0022>
- Mineka, S., & Öhman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry*, 52(10), 927–937. [https://doi.org/10.1016/S0006-3223\(02\)01669-4](https://doi.org/10.1016/S0006-3223(02)01669-4)
- Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde Amnesia Produced by Electroconvulsive Shock after Reactivation of a Consolidated Memory Trace. *Science*, 160(3827), 554–555. <https://doi.org/10.1126/science.160.3827.554>
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2), 183–246. <https://doi.org/10.1017/S0140525X09000855>
- Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., Petrovic, P., Dolan, R. J., & Frith, C. D. (2009). From threat to fear: The neural organization of defensive fear systems in humans. *Journal of Neuroscience*, 29(39), 12236–12243. <https://doi.org/10.1523/JNEUROSCI.2378-09.2009>

- Monfils, M., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-Reconsolidation Boundaries: Key to Persistent Attenuation of Fear Memories. *Science*, 324(5929), 951–955. <https://doi.org/10.1126/science.1167975>
- Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*, 5(12), 1032–1042. [https://doi.org/10.1016/S2215-0366\(18\)30270-0](https://doi.org/10.1016/S2215-0366(18)30270-0)
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5), 1936–1947. <https://doi.org/10.1523/jneurosci.16-05-01936.1996>
- Morís, J., Barberia, I., Vadillo, M. A., Andrades, A., & López, F. J. (2017). Slower reacquisition after partial extinction in human contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(1), 81–93. <https://doi.org/10.1037/xlm0000282>
- Morris, J. S., Ohrnan, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, 393(6684), 467–470. <https://doi.org/10.1038/30976>
- Morris, R. G. M., Inglis, J., Ainge, J. A., Olverman, H. J., Tulloch, J., Dudai, Y., & Kelly, P. A. T. (2006). Memory Reconsolidation: Sensitivity of Spatial Memory to Inhibition of Protein Synthesis in Dorsal Hippocampus during Encoding and Retrieval. *Neuron*, 50(3), 479–489. <https://doi.org/10.1016/j.neuron.2006.04.012>
- Mungee, A., Kazzer, P., Feeser, M., Nitsche, M. A., Schiller, D., & Bajbouj, M. (2013). Transcranial direct current stimulation of the prefrontal cortex. *NeuroReport*, 25(7), 1. <https://doi.org/10.1097/WNR.000000000000119>
- Nader, K., Schafe, G. E., & LeDoux, J. E. (2000). The labile nature of consolidation theory. *Nature Reviews. Neuroscience*, 1(3), 216–219. <https://doi.org/10.1038/35044580>
- Nader, Karim. (2015). Reconsolidation and the Dynamic Nature of Memory. *Cold Spring Harbor Perspectives in Biology*, 7(10), a021782. <https://doi.org/10.1101/cshperspect.a021782>
- Nader, Karim, Schafe, G. E., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, 406(6797), 722–726. <https://doi.org/10.1038/35021052>
- Nuutinen, M., Mustonen, T., & Häkkinen, J. (2017). CFS MATLAB toolbox: An experiment builder for continuous flash suppression (CFS) task. *Behavior Research Methods*, 9. <https://doi.org/10.3758/s13428-017-0961-z>
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–249. <https://doi.org/10.1016/j.tics.2005.03.010>
- Öhman, A., & Soares, J. J. F. (1994). “Unconscious anxiety”: Phobic responses to masked stimuli. *Journal of Abnormal Psychology*, 103(2), 231–240. <https://doi.org/10.1037/0021-843X.103.2.231>
- Öhman, A., & Soares, J. J. F. (1998). Emotional Conditioning to Masked Stimuli: Expectancies for Aversive Outcomes Following Nonrecognized Fear-Relevant Stimuli. *Journal of Experimental Psychology: General*, 127(1), 69–82. <https://doi.org/10.1037/0096-3445.127.1.69>
- Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after pavlovian, observational, and instructed fear. *Psychological Science*, 15(12), 822–828. <https://doi.org/10.1111/j.0956-7976.2004.00762.x>
- Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, 10(9), 1095–1102. <https://doi.org/10.1038/nn1968>
- Orsini, C. a., Kim, J. H., Knapska, E., & Maren, S. (2011). Hippocampal and Prefrontal Projections to the Basal Amygdala Mediate Contextual Regulation of Fear after Extinction. *Journal of Neuroscience*, 31(47), 17269–17277. <https://doi.org/10.1523/JNEUROSCI.4095-11.2011>
- Oyarzún, J. P., Càmara, E., Kouider, S., Fuentemilla, L., & de Diego-Balaguer, R. (2019). Implicit but not explicit extinction to threat-conditioned stimulus prevents spontaneous recovery of threat-potentiated startle responses in humans. *Brain and Behavior*, 9(1), 1–13. <https://doi.org/10.1002/brb3.1157>
- Oyarzún, J. P., Lopez-Barroso, D., Fuentemilla, L., Cucurell, D., Pedraza, C., Rodriguez-Fornells, A., & de Diego-Balaguer, R. (2012). Updating fearful memories with extinction training during



- reconsolidation: A human study using auditory aversive stimuli. *PLoS ONE*, 7(6). <https://doi.org/10.1371/journal.pone.0038849>
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.
- Panksepp, J., & Watt, D. (2011). What is Basic about Basic Emotions? Lasting Lessons from Affective Neuroscience. *Emotion Review*, 3(4), 387–396. <https://doi.org/10.1177/1754073911410741>
- Panksepp, Jaak. (2000). Fear and Anxiety Mechanisms of the Brain : clinical implications. *Biological Psychiatry*, 155–177.
- Panksepp, Jaak, Lane, R. D., Solms, M., & Smith, R. (2017). Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. *Neuroscience and Biobehavioral Reviews*, 76, 187–215. <https://doi.org/10.1016/j.neubiorev.2016.09.010>
- Parkinson, J., & Haggard, P. (2014). Subliminal priming of intentional inhibition. *Cognition*, 130(2), 255–265. <https://doi.org/10.1016/j.cognition.2013.11.005>
- Pavlov, I. P. (1927). Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex. In *Oxford University Press* (Vol. 2). Oxford University Press.
- Pedreira, M. E. (2004). Mismatch Between What Is Expected and What Actually Occurs Triggers Memory Reconsolidation or Extinction. *Learning & Memory*, 11(5), 579–585. <https://doi.org/10.1101/lm.76904>
- Pedreira, María Eugenia, & Maldonado, H. (2003). Protein Synthesis Subverts Reconsolidation or Extinction Depending on Reminder Duration. *Neuron*, 38(6), 863–869. [https://doi.org/10.1016/S0896-6273\(03\)00352-0](https://doi.org/10.1016/S0896-6273(03)00352-0)
- Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, 14(2), 198–202. <https://doi.org/10.1016/j.conb.2004.03.015>
- Przybylski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, 84(1–2), 241–246. [https://doi.org/10.1016/S0166-4328\(96\)00153-2](https://doi.org/10.1016/S0166-4328(96)00153-2)
- Quirk, G. J. (2002). Memory for Extinction of Conditioned Fear Is Long-lasting and Persists Following Spontaneous Recovery. *Learning & Memory*, 9(6), 402–407. <https://doi.org/10.1101/lm.49602>
- Rachman, S. (1979). The return of fear. *Behaviour Research and Therapy*, 17(2), 164–166. [https://doi.org/10.1016/0005-7967\(79\)90028-7](https://doi.org/10.1016/0005-7967(79)90028-7)
- Raes, A. K., & Raedt, R. De. (2011). Interoceptive awareness and unaware fear conditioning: Are subliminal conditioning effects influenced by the manipulation of visceral self-perception? *Consciousness and Cognition*, 20(4), 1393–1402. <https://doi.org/10.1016/j.concog.2011.05.009>
- Raio, C. M., Carmel, D., Carrasco, M., & Phelps, E. A. (2012a). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology*, 22(12). <https://doi.org/10.1016/j.cub.2012.04.023>
- Raio, C. M., Carmel, D., Carrasco, M., & Phelps, E. A. (2012b). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology*, 22(12), R477–R479. <https://doi.org/10.1016/j.cub.2012.04.023>
- Rajbhandari, A. K., Tribble, J. E., & Fanselow, M. S. (2017a). Neurobiology of Fear Memory ☆. In *Learning and Memory: A Comprehensive Reference* (Second Edi, Vol. 4, Issue October 2016). Elsevier. <https://doi.org/10.1016/b978-0-12-809324-5.21100-0>
- Rajbhandari, A. K., Tribble, J. E., & Fanselow, M. S. (2017b). Neurobiology of Fear Memory ☆. In *Learning and Memory: A Comprehensive Reference* (Second Edi, Vol. 4, Issue October 2016, pp. 487–503). Elsevier. <https://doi.org/10.1016/B978-0-12-809324-5.21100-0>
- Reddan, M. C., Wager, T. D., & Schiller, D. (2018). Attenuating Neural Threat Expression with Imagination. *Neuron*, 100(4), 994–1005.e4. <https://doi.org/10.1016/j.neuron.2018.10.047>
- Rehbein, M. A., Steinberg, C., Wessing, I., Pastor, M. C., Zwitterlood, P., Keuper, K., & Junghöfer, M. (2014). Rapid plasticity in the prefrontal cortex during affective associative learning. *PLoS ONE*, 9(10). <https://doi.org/10.1371/journal.pone.0110720>

- Reinhardt, I., Jansen, A., Kellermann, T., Schüppen, A., Kohn, N., Gerlach, A. L., & Kircher, T. (2010). Neural correlates of aversive conditioning: Development of a functional imaging paradigm for the investigation of anxiety disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *260*(6), 443–453. <https://doi.org/10.1007/s00406-010-0099-9>
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151–160. <https://doi.org/10.1037/0003-066X.43.3.151>
- Rescorla, R. A. (2006). Deepened extinction from compound stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 135–144. <https://doi.org/10.1037/0097-7403.32.2.135>
- Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement BT - Classical conditioning II: current research and theory. In A. H. Black; & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). Appleton-Century-Crofts. <http://jshd.pubs.asha.org/Article.aspx?articleid=1775379%5Cnpapers3://publication/uuid/1A852E2C-BD69-44DE-BAE6-3DFafa705330>
- Roesmann, K., Wiens, N., Winker, C., Rehbein, M. A., Wessing, I., & Junghoefer, M. (2020). Fear generalization of implicit conditioned facial features – Behavioral and magnetoencephalographic correlates. *NeuroImage*, *205*(March 2019), 116302. <https://doi.org/10.1016/j.neuroimage.2019.116302>
- Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156. <https://doi.org/10.1016/j.tics.2012.01.005>
- Sandrini, M., Censor, N., Mishoe, J., & Cohen, L. G. (2013). Causal Role of Prefrontal Cortex in Strengthening of Episodic Memories through Reconsolidation. *Current Biology*, *23*(21), 2181–2184. <https://doi.org/10.1016/j.cub.2013.08.045>
- Sara, S. J. (2008). Reconsolidation: Historical Perspective and Theoretical Aspects. In *Learning and Memory: A Comprehensive Reference* (Second Edi, Vol. 1, Issue December 2016). Elsevier. <https://doi.org/10.1016/b978-012370509-9.00061-9>
- Schafe, G.E., & LeDoux, J. E. (2008). Neural and Molecular Mechanisms of Fear Memory. *Learning and Memory: A Comprehensive Reference*, 157–192. <https://doi.org/10.1016/b978-012370509-9.00045-0>
- Schafe, Glenn E., & LeDoux, J. E. (2000). Memory Consolidation of Auditory Pavlovian Fear Conditioning Requires Protein Synthesis and Protein Kinase A in the Amygdala. *The Journal of Neuroscience*, *20*(18), RC96–RC96. <https://doi.org/10.1523/JNEUROSCI.20-18-j0003.2000>
- Schiller, D., Cain, C. K., Curley, N. G., Schwartz, J. S., Stern, S. A., Ledoux, J. E., & Phelps, E. A. (2008). Evidence for recovery of fear following immediate extinction in rats and humans. 394–402. <https://doi.org/10.1101/lm.909208.4>
- Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M.-H., & Phelps, E. A. (2013a). Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20040–20045. <https://doi.org/10.1073/pnas.1320322110>
- Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M.-H., & Phelps, E. a. (2013b). Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50). <https://doi.org/10.1073/pnas.1320322110>
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., Ledoux, J. E., & Phelps, E. A. (2010a). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53. <https://doi.org/10.1038/nature08637>
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., Ledoux, J. E., & Phelps, E. a. (2010b). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53. <https://doi.org/10.1038/nature08637>
- Schiller, D., & Phelps, E. a. (2011). Does reconsolidation occur in humans? *Frontiers in Behavioral Neuroscience*, *5*(October), 24. <https://doi.org/10.3389/fnbeh.2011.00074>

- Schroyens, N., Beckers, T., & Kindt, M. (2017). In search for boundary conditions of reconsolidation: A failure of fear memory interference. *Frontiers in Behavioral Neuroscience*, *11*(April), 1–13. <https://doi.org/10.3389/fnbeh.2017.00065>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. <https://doi.org/10.1126/science.275.5306.1593>
- Schweckendiek, J., Klucken, T., Merz, C. J., Tabbert, K., Walter, B., Ambach, W., Vaitl, D., & Stark, R. (2011). Weaving the (neuronal) web: Fear learning in spider phobia. *NeuroImage*, *54*(1), 681–688. <https://doi.org/10.1016/j.neuroimage.2010.07.049>
- Sevenster, D., Beckers, T., & Kindt, M. (2012). Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiology of Learning and Memory*, *97*(3), 338–345. <https://doi.org/10.1016/j.nlm.2012.01.009>
- Sevenster, D., Beckers, T., & Kindt, M. (2013a). Prediction error governs pharmacologically induced amnesia for learned fear. *Science*, *339*(February), 830–833. <https://doi.org/10.1126/science.1177170>
- Sevenster, D., Beckers, T., & Kindt, M. (2013b). Prediction error governs pharmacologically induced amnesia for learned fear. *Science (New York, N.Y.)*, *339*(6121), 830–833. <https://doi.org/10.1126/science.1231357>
- Sevenster, D., Beckers, T., & Kindt, M. (2014). Fear conditioning of SCR but not the startle reflex requires conscious discrimination of threat and safety. *Frontiers in Behavioral Neuroscience*, *8*(FEB), 32. <https://doi.org/10.3389/fnbeh.2014.00032>
- Shiban, Y., Brütting, J., Pauli, P., & Mühlberger, A. (2015a). Fear reactivation prior to exposure therapy: Does it facilitate the effects of VR exposure in a randomized clinical sample? *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 133–140. <https://doi.org/10.1016/j.jbtep.2014.09.009>
- Shiban, Y., Brütting, J., Pauli, P., & Mühlberger, A. (2015b). Fear reactivation prior to exposure therapy: does it facilitate the effects of VR exposure in a randomized clinical sample? *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 133–140. <https://doi.org/10.1016/j.jbtep.2014.09.009>
- Shiozawa, P., & Sato, I. A. (2016). Transcranial Magnetic Stimulation for Anxiety Symptoms: An Updated Systematic Review and Meta-Analysis. *Abnormal and Behavioural Psychology* Trevizol et al. *Abnorm Behav Psychol*, *2*(January), 1. <https://doi.org/10.4172/abp.1000108>
- Siegel, P., & Warren, R. (2013a). Less is still more: Maintenance of the very brief exposure effect 1 year later. *Emotion*, *13*(2), 338–344. <https://doi.org/10.1037/a0030833>
- Siegel, P., & Warren, R. (2013b). The effect of very brief exposure on experienced fear after in vivo exposure. *Cognition and Emotion*, *27*(6), 1013–1022. <https://doi.org/10.1080/02699931.2012.756803>
- Siegel, P., Warren, R., Jacobson, G., & Merritt, E. (2018). Masking exposure to phobic stimuli reduces fear without inducing electrodermal activity. *Psychophysiology*, *55*(5), 1–14. <https://doi.org/10.1111/psyp.13045>
- Siegel, P., Warren, R., Wang, Z., Yang, J., Cohen, D., Anderson, J. F., Murray, L., & Peterson, B. S. (2017). Less is more: Neural activity during very brief and clearly visible exposure to phobic stimuli. *Human Brain Mapping*, *38*(5), 2466–2481. <https://doi.org/10.1002/hbm.23533>
- Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. <https://doi.org/10.1002/wcs.1323>
- Sjouwerman, R., Niehaus, J., Kuhn, M., & Lonsdorf, T. B. (2016). Don't startle me??? Interference of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology*, *53*(12), 1889–1899. <https://doi.org/10.1111/psyp.12761>
- Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, *109*(48), 19614–19619. <https://doi.org/10.1073/pnas.1211645109>
- Soeter, Marieke, & Kindt, M. (2012). Erasing fear for an imagined threat event. *Psychoneuroendocrinology*, *37*(11), 1769–1779. <https://doi.org/10.1016/j.psyneuen.2012.03.011>

- Soeter, Marieke, & Kindt, M. (2010). Dissociating response systems: erasing fear from memory. *Neurobiology of Learning and Memory*, 94(1), 30–41. <https://doi.org/10.1016/j.nlm.2010.03.004>
- Soeter, Marieke, & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. *Learning & Memory*, 18(6), 357–366. <https://doi.org/10.1101/lm.2148511>
- Soeter, Marieke, & Kindt, M. (2012a). Stimulation of the noradrenergic system during memory formation impairs extinction learning but not the disruption of reconsolidation. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 37(5), 1204–1215. <https://doi.org/10.1038/npp.2011.307>
- Soeter, Marieke, & Kindt, M. (2012b). Erasing fear for an imagined threat event. *Psychoneuroendocrinology*, 37(11), 1769–1779. <https://doi.org/10.1016/j.psyneuen.2012.03.011>
- Soeter, Marieke, & Kindt, M. (2015a). Retrieval cues that trigger reconsolidation of associative fear memory are not necessarily an exact replica of the original learning experience. *Frontiers in Behavioral Neuroscience*, 9(May), 1–10. <https://doi.org/10.3389/fnbeh.2015.00122>
- Soeter, Marieke, & Kindt, M. (2015b). An Abrupt Transformation of Phobic Behavior After a Post-Retrieval Amnesic Agent. *Biological Psychiatry*, 78(12), 880–886. <https://doi.org/10.1016/j.biopsych.2015.04.006>
- Sparta, D. R., Smithuis, J., Stamatakis, A. M., Jennings, J. H., Kantak, P. A., Ung, R. L., & Stuber, G. D. (2014). Inhibition of projections from the basolateral amygdala to the entorhinal cortex disrupts the acquisition of contextual fear. *Frontiers in Behavioral Neuroscience*, 8(MAY), 6–11. <https://doi.org/10.3389/fnbeh.2014.00129>
- Sperandio, I., Bond, N., & Binda, P. (2018). Pupil Size as a Gateway Into Conscious Interpretation of Brightness. *Frontiers in Neurology*, 9(December), 1–9. <https://doi.org/10.3389/fneur.2018.01070>
- Spering, M., & Carrasco, M. (2015). Acting without seeing: Eye movements reveal visual processing without awareness. *Trends in Neurosciences*, 38(4), 224–258. <https://doi.org/10.1016/j.tins.2015.02.002>
- Spring, J. D., Wood, N. E., Mueller-Pfeiffer, C., Milad, M. R., Pitman, R. K., & Orr, S. P. (2015). Prereactivation propranolol fails to reduce skin conductance reactivity to prepared fear-conditioned stimuli. *Psychophysiology*, 52(3), 407–415. <https://doi.org/10.1111/psyp.12326>
- Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory Consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8), a021766. <https://doi.org/10.1101/cshperspect.a021766>
- Stein, T., Utz, V., & van Opstal, F. (2020). Unconscious semantic priming from pictures under backward masking and continuous flash suppression. *Consciousness and Cognition*, 78(January 2020), 102864. <https://doi.org/10.1016/j.concog.2019.102864>
- Steinberg, C., Bröckelmann, A.-K., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2013). Preferential responses to extinguished face stimuli are preserved in frontal and occipito-temporal cortex at initial but not later stages of processing. *Psychophysiology*, 50(3), 230–239. <https://doi.org/10.1111/psyp.12005>
- Steinberg, C., Bröckelmann, A.-K., Rehbein, M., Dobel, C., & Junghöfer, M. (2013). Rapid and highly resolving associative affective learning: Convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological Psychology*, 92(3), 526–540. <https://doi.org/10.1016/j.biopsycho.2012.02.009>
- Steinberg, C., Dobel, C., Schupp, H. T., Kissler, J., Elling, L., Pantev, C., & Junghöfer, M. (2012a). Rapid and highly resolving: Affective evaluation of olfactorily conditioned faces. *Journal of Cognitive Neuroscience*, 24(1), 17–27. [https://doi.org/10.1162/jocn\\_a\\_00067](https://doi.org/10.1162/jocn_a_00067)
- Steinberg, C., Dobel, C., Schupp, H. T., Kissler, J., Elling, L., Pantev, C., & Junghöfer, M. (2012b). Rapid and highly resolving: affective evaluation of olfactorily conditioned faces. *Journal of Cognitive Neuroscience*, 24(1), 17–27.
- Steinfurth, E. C. K., Kanen, J. W., Raio, C. M., Clem, R. L., Haganir, R. L., & Phelps, E. A. (2014a). Young and old Pavlovian fear memories can be modified with extinction training during reconsolidation in humans. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 21(7), 338–341.

- <https://doi.org/10.1101/lm.033589.113>
- Steinfurth, E. C. K., Kanen, J. W., Raio, C. M., Clem, R. L., Huganir, R. L., & Phelps, E. a. (2014b). Young and old Pavlovian fear memories can be modified with extinction training during reconsolidation in humans. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *21*(7), 338–341. <https://doi.org/10.1101/lm.033589.113>
- Sylvers, P., Lilienfeld, S. O., & LaPrairie, J. L. (2011). Differences between trait fear and trait anxiety: Implications for psychopathology. *Clinical Psychology Review*, *31*(1), 122–137. <https://doi.org/10.1016/j.cpr.2010.08.004>
- Tabbert, K., Stark, R., Kirsch, P., & Vaitl, D. (2005). Hemodynamic responses of the amygdala, the orbitofrontal cortex and the visual cortex during a fear conditioning paradigm. *International Journal of Psychophysiology*, *57*(1), 15–23. <https://doi.org/10.1016/j.ijpsycho.2005.01.007>
- Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., Kawato, M., & Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3470–3475. <https://doi.org/10.1073/pnas.1721572115>
- Taschereau-Dumouchel, V., Liu, K., & Lau, H. (2018). Unconscious psychological treatments for physiological survival circuits. *Current Opinion in Behavioral Sciences*, *24*, 62–68. <https://doi.org/10.1016/j.cobeha.2018.04.010>
- Telch, M. J., York, J., Lancaster, C. L., & Monfils, M. H. (2017). Use of a Brief Fear Memory Reactivation Procedure for Enhancing Exposure Therapy. *Clinical Psychological Science*, *5*(2), 367–378. <https://doi.org/10.1177/2167702617690151>
- Terry, L., & Holliday, J. H. (1972). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace: A replication. *Psychonomic Science*, *29*(3), 137–138. <https://doi.org/10.3758/BF03342570>
- Thome, J., Koppe, G., Hauschild, S., Liebke, L., Schmahl, C., Lis, S., & Bohus, M. (2016). Modification of fear memory by pharmacological and behavioural interventions during reconsolidation. *PLoS ONE*, *11*(8), 1–20. <https://doi.org/10.1371/journal.pone.0161044>
- Thompson, A., & Lipp, O. V. (2017). Extinction during reconsolidation eliminates recovery of fear conditioned to fear-irrelevant and fear-relevant stimuli. *Behaviour Research and Therapy*, *92*, 1–10. <https://doi.org/10.1016/j.brat.2017.01.017>
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>
- Treanor, M., Brown, L. A., Rissman, J., & Craske, M. G. (2017). Can Memories of Traumatic Experiences or Addiction Be Erased or Modified? A Critical Review of Research on the Disruption of Memory Reconsolidation and Its Applications. *Perspectives on Psychological Science*, *12*(2), 290–305. <https://doi.org/10.1177/1745691616664725>
- Troiani, V., Price, E. T., & Schultz, R. T. (2014). *Unseen fearful faces promote amygdala guidance of attention*. <https://doi.org/10.1093/scan/nss116>
- Troiani, V., & Schultz, R. T. (2013). Amygdala, pulvinar, and inferior parietal cortex contribute to early processing of faces without awareness. *Frontiers in Human Neuroscience*, *7*(June), 1–12. <https://doi.org/10.3389/fnhum.2013.00241>
- Trouche, S., Sasaki, J. M., Tu, T., & Reijmers, L. G. (2013). Fear Extinction Causes Target-Specific Remodeling of Perisomatic Inhibitory Synapses. *Neuron*, *80*(4), 1054–1065. <https://doi.org/10.1016/j.neuron.2013.07.047>
- Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. <https://doi.org/10.1038/nn1500>
- van Schie, K., van Veen, S. C., Hendriks, Y. R., van den Hout, M. A., & Engelhard, I. M. (2017). Intervention strength does not differentially affect memory reconsolidation of strong memories. *Neurobiology of Learning and Memory*, *144*, 174–185. <https://doi.org/10.1016/j.nlm.2017.07.011>
- van Schie, K., Veen, S. C. van, van den Hout, M. A., & Engelhard, I. M. (2017). Modification of

- episodic memories by novel learning: a failed replication study. *European Journal of Psychotraumatology*, 8(sup1), 1315291. <https://doi.org/10.1080/20008198.2017.1315291>
- Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalization, and return of fear: A critical review of renewal research in humans. *Biological Psychology*, 92(1). <https://doi.org/10.1016/j.biopsycho.2012.01.006>
- Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology*, 9(1), 215–248. <https://doi.org/10.1146/annurev-clinpsy-050212-185542>
- Vieira, J. B., Wen, S., Oliver, L. D., & Mitchell, D. G. V. (2017). Enhanced conscious processing and blindsight-like detection of fear-conditioned stimuli under continuous flash suppression. *Experimental Brain Research*, 235(11), 3333–3344. <https://doi.org/10.1007/s00221-017-5064-7>
- Visser, R. M., Lau-Zhu, A., Henson, R. N., & Holmes, E. A. (2018). Multiple memory systems, multiple time points: How science can inform treatment to control the expression of unwanted emotional memories. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1742). <https://doi.org/10.1098/rstb.2017.0209>
- Vlassova, A., Donkin, C., & Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), 16214–16218. <https://doi.org/10.1073/pnas.1403619111>
- Wang, S.-H., & Morris, R. G. M. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual Review of Psychology*, 61, 49–79, C1-4. <https://doi.org/10.1146/annurev.psych.093008.100523>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, 3(1), 1–14. <https://doi.org/10.1037/h0069608>
- Weike, A. I., Schupp, H. T., & Hamm, A. O. (2007). Fear acquisition requires awareness in trace but not delay conditioning. *Psychophysiology*, 44(1), 170–180. <https://doi.org/10.1111/j.1469-8986.2006.00469.x>
- Weinberger, J., Siegel, P., Siefert, C., & Drwal, J. (2011). What you cannot see can help you: The effect of exposure to unreportable stimuli on approach behavior. *Consciousness and Cognition*, 20(2), 173–180. <https://doi.org/10.1016/j.concog.2011.01.003>
- Whalen, P. J., Rauch, S. L., Etkoff, N. L., McInerney, S. C., Lee Michael, B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, 18(1), 411–418. <https://doi.org/10.1523/jneurosci.18-01-00411.1998>
- Wiemer, J., Mühlberger, A., & Pauli, P. (2014). Illusory correlations between neutral and aversive stimuli can be induced by outcome aversiveness. *Cognition and Emotion*, 28(2), 193–207. <https://doi.org/10.1080/02699931.2013.809699>
- Wiens, S., & Öhman, A. (2002). Unawareness is more than a chance event: Comment on Lovibond and Shanks (2002). *Journal of Experimental Psychology: Animal Behavior Processes*, 28(1), 27–31. <https://doi.org/10.1037/0097-7403.28.1.27>
- Winters, B. D., Tucci, M. C., & DaCosta-Furtado, M. (2009). Older and stronger object memories are selectively destabilized by reactivation in the presence of new information. *Learning and Memory*, 16(9), 545–553. <https://doi.org/10.1101/lm.1509909>
- Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21(9), 655–679. <https://doi.org/10.1016/j.euroneuro.2011.07.018>
- Wolpe, J. (1981). Behavior therapy versus psychoanalysis: Therapeutic and social implications. *American Psychologist*, 36(2), 159–164. <https://doi.org/10.1037/0003-066X.36.2.159>

- Wood, N. E., Rosasco, M. L., Suris, A. M., Spring, J. D., Marin, M.-F., Lasko, N. B., Goetz, J. M., Fischer, A. M., Orr, S. P., & Pitman, R. K. (2015). Pharmacological blockade of memory reconsolidation in posttraumatic stress disorder: three negative psychophysiological studies. *Psychiatry Research*, 225(1–2), 31–39. <https://doi.org/10.1016/j.psychres.2014.09.005>
- Xue, Y., Luo, Y., Wu, P., Shi, H., Xue, L., Chen, C., Zhu, W., Ding, Z., Bao, Y., Shi, J., Epstein, D. H., Shaham, Y., & Lu, L. (2012). A Memory Retrieval-Extinction Craving and Relapse. *Science*, 336(April), 241–245. <https://doi.org/10.1126/science.1215070>
- Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion (Washington, D.C.)*, 7(4), 882–886. <https://doi.org/10.1037/1528-3542.7.4.882>
- Ye, X., He, S., Hu, Y., Yu, Y. Q., & Wang, K. (2014). Interference between conscious and unconscious facial expression information. *PLoS ONE*, 9(8), 1–6. <https://doi.org/10.1371/journal.pone.0105156>
- Zbozinek, T. D., Hermans, D., Prenoveau, J. M., Liao, B., & Craske, M. G. (2015). Post-extinction conditional stimulus valence predicts reinstatement fear: Relevance for long-term outcomes of exposure therapy. *Cognition and Emotion*, 29(4), 654–667. <https://doi.org/10.1080/02699931.2014.930421>
- Zimmermann, J., & Bach, D. R. (2020). Impact of a reminder/extinction procedure on threat-conditioned pupil size and skin conductance responses. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 27(4), 164–172. <https://doi.org/10.1101/lm.050211.119>
- Zlomuzica, A., Preusser, F., Schneider, S., & Margraf, J. (2015). Increased perceived self-efficacy facilitates the extinction of fear in healthy participants. *Frontiers in Behavioral Neuroscience*, 9(OCTOBER), 1–12. <https://doi.org/10.3389/fnbeh.2015.00270>
- Zuccolo, P. F., & Hunziker, M. H. L. (2019). A review of boundary conditions and variables involved in the prevention of return of fear after post-retrieval extinction. *Behavioural Processes*, 162(September 2018), 39–54. <https://doi.org/10.1016/j.beproc.2019.01.011>

## Appendix A.

### Supplemental analyses

Table 8-1 Supplemental analyses for Experiment 1 (a) Result summary: Coefficient estimates, Standard Error,  $t$  statistics, and significance levels  $p$  for all predictors in the acquisition and re-extinction phase. Significant beta values suggest that pupil responses of the corresponding CS type were significantly different compared to those of the implicit CS+ (as the intercept). (N = 59). (b) Estimated marginal means of the pupil responses and unpleasantness rating of CSs, their standard errors/deviations and confidence intervals in each experimental phase (N = 59)

(a)

Outcome	Phase	Parameters	$\beta$	SE	95% CI		$t$	$p$
<i>Pupil response</i>	Acquisition	intercept	0.59	0.08	0.44	0.76	7.48	<.001
		CS <sub>exp+</sub>	0.09	0.10	-0.11	0.29	0.89	0.373
		CS -	0.10	0.10	-0.10	0.30	0.96	0.333
	Re-extinction	intercept	0.59	0.07	0.45	0.73	8.30	<.001
		CS <sub>exp+</sub>	-0.11	0.09	-0.29	0.07	-1.71	0.242
		CS -	-0.04	0.09	-0.22	0.14	-0.44	0.660

(b)

Outcome	Phase	Type	EMMS	SE	95% CI	
<i>Pupil responses</i>	Acquisition	CS <sub>imp+</sub>	0.60	0.08	0.44	0.76
		CS <sub>exp+</sub>	0.69	0.08	0.53	0.85
		CS -	0.70	0.08	0.54	0.86
	Re-extinction	CS <sub>imp+</sub>	0.59	0.07	0.45	0.73
		CS <sub>exp+</sub>	0.48	0.07	0.34	0.62
		CS -	0.55	0.07	0.41	0.69
<i>CS unpleasantness rating</i>	Acquisition		<u>Mean</u>	<u>SD</u>		
		CS <sub>imp+</sub>	2.97	1.39		
		CS <sub>exp+</sub>	3.1	1.41		
	Extinction	CS -	1.83	0.89		
		CS <sub>imp+</sub>	1.98	1.22		
		CS <sub>exp+</sub>	2.05	1.25		
	Re-extinction	CS -	1.67	1.06		
		CS <sub>imp+</sub>	2.03	1.15		
		CS <sub>exp+</sub>	1.93	1.12		
		CS -	1.47	0.78		



Table 8-2. Supplemental Analyses for Experiment 2: Demographic Characteristics of the Sample (N = 59)

	Explicit group ( $n = 26$ )	Implicit group ( $n = 33$ )	$t$ - test ( $p$ )
Age	23.46 (6.05)	21.21 (3.75)	1.66 (.105)
Gender: male (female)	5 (21)	10 (23)	--
Education Level	14.58 (3.37)	14.52 (2.72)	0.07 (.940)
STAI-T	46.30 (10.27)	46.58 (9.28)	-0.10 (.918)

*Note.* STAI-T = State-Trait Anxiety Inventory - Trait

Table 8-3. Experiment 2 result summary: Coefficient estimates (beta), Standard Error,  $t$  statistics, and significance level  $p$  for each predictor in estimating pupillary responses.

*Note:* Implicit reactivation group ( $n = 33$ ), Explicit reactivation group ( $n = 26$ )

Group	Session	Parameters	$\beta$	SE	$t$	$p$
<i>Implicit reactivation</i>	Acquisition	CS -	0.52	0.09	5.92	<0.001
		Reminded CS+	0.44	0.06	6.98	<0.001
		Non-reminded CS+	0.51	0.06	8.10	<0.001
<i>Explicit reactivation</i>	Acquisition	CS -	0.65	0.09	7.11	<0.001
		Reminded CS+	0.32	0.08	4.11	<0.001
		Non-reminded CS+	0.43	0.08	5.51	<0.001

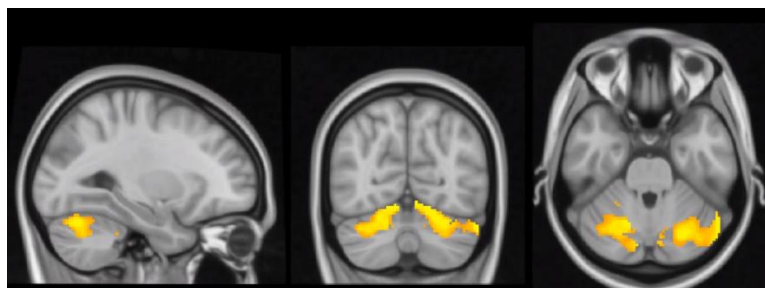
Table 8-4. Pearson correlation coefficients between unpleasantness ratings (post-extinction) and pupil responses (post-reinstatement)

	Pupil responses
<i>Implicit Reactivation group</i>	
Unpleasantness rating – reminded CS	-0.12 ( $p = 0.524$ )
Unpleasantness rating – non-reminded CS	0.02 ( $p = 0.901$ )
<i>Explicit Reactivation group</i>	
Unpleasantness rating – reminded CS	-0.16 ( $p = 0.418$ )
Unpleasantness rating – non-reminded CS	-0.15 ( $p = 0.451$ )

Table 8-5. Supplemental analyses for Experiment 3: Estimated means differences and standard errors for pupil responses for participants with low detectability of CS-US associations (Day 1) ( $n = 18$ )

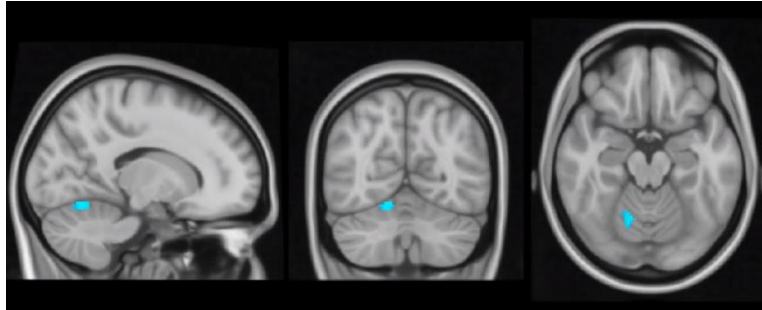
<b>Day: session</b>	<b>contrast</b>	<b>Parameter estimate differences</b>	<b>SE</b>	<b>df</b>	<b><i>t</i>-value</b>	<b><i>p</i></b>
Day 1: Acquisition	rCS vs CS-	0.24	0.13	92	1.95	<b>.055</b>
	nrCS vs CS-	0.30	0.13	92	2.40	<b>.019</b>
	rCS+ vs nrCS+	-0.06	0.13	92	-0.45	.653

Table 8-6. Supplemental analyses for Experiment 4: Localization and statistics for cerebellar ROI-analyses on Day 1 (Acquisition). The figure illustrates the bilateral cerebellar lobules VI and Crus 1 activation during early acquisition on Day 1 ( $p_{\text{FWE-corr}} < .05$ )



	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{\text{FWE-corr}}$
Acquisition (Run 1 & 2)	CS+ vs CS-	VI	L	579	-22	-70	-28	3.54	<0.001
					-16	-76	-20	3.54	
					-8	-72	-14	3.54	
					-10	-76	-16	3.35	
					-8	-72	-28	2.62	
		VI	R	489	-24	-56	-32	2.59	0.004
					20	-62	-30	3.54	
					28	-62	-28	3.54	
					20	-70	-26	3.54	
					26	-68	-28	3.35	
		Crus 1	L	606	10	-78	-20	2.71	0.003
					-54	-54	-34	3.54	
					-52	-68	-32	3.54	
					-54	-60	-32	3.54	
					-22	-72	-28	3.54	
Crus 1	R	347	-46	-72	-24	3.54	0.001		
			-16	-78	-20	3.54			
			36	-68	-30	3.54			
			10	-80	-26	3.54			
			20	-72	-26	3.54			
Acquisition (Run 3 & 4)	CS+ vs CS-	VI	L	107	30	-68	-34	3.35	0.034
					18	-80	-32	3.24	
					14	-78	-30	3.24	
					-20	-54	-24	3.54	
					-16	-58	-20	3.35	
					-12	-70	-14	2.69	

Table 8-7. Localization and statistics for cerebellar ROI-analyses on Day 3 (Re-Extinction). The figure illustrates the right cerebellar lobules VI deactivation in the rCS+ > nrCS+ contrast during re-extinction on Day 3. (16, -64, -18,  $p_{\text{FEW-corr}} = .031$ )



	Contrast	Structure	Side	Size (voxels)	x	y	z	Zmax	$p_{\text{FEW-corr}}$
Re-extinction (run 1)	rCS+ vs nrCS+	VI	R	50	16	-64	-18	3.54	0.031

Remark: The results indicate deactivation of the brain region in the contrast comparing rCS+ and nrCS+