University of Glasgow | Institute of Neuroscience & Psychology

Studying the behavioural, physiological, and neural indices of associative learning in multi-trial paradigms: Methodological and analytical considerations.

Antonia Antony Jambazova

MA, MSc

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Psychology
Institute of Neuroscience and Psychology
College of Science and Engineering
University of Glasgow
62 Hillhead Street
Glasgow
G12 8QB

March 2021

Declaration of Originality Form

This form **must** be completed and signed and submitted with all assignments.

Please complete the information below (using BLOCK CAPITALS).

Name ANTONIA ANTONY JAMBAZOVA ....

Student Number ....

Course Name PHD PSYCHOLOGY – DOCTORAL THESIS..

Assignment Number/Name STUDYING THE BEHAVIOURAL, PHYSIOLOGICAL AND NEURAL INDICES OF ASSOCIATIVE LEARNING IN MULTI-TRIAL PARADIGMS: METHODOLOGICAL AND ANALYTICAL CONSIDERATIONS ....

An extract from the University's Statement on Plagiarism is provided overleaf. Please read carefully THEN read and sign the declaration below.

I confirm that this assignment is my own work and that I have:

Read and understood the guidance on plagiarism in the Student Handbook, including the University of Glasgow Statement on Plagiarism	☒
Clearly referenced, in both the text and the bibliography or references, **all sources** used in the work	☒
Fully referenced (including page numbers) and used inverted commas for **all text quoted** from books, journals, web etc. (Please check with the Department which referencing style is to be used)	☒
Provided the sources for all tables, figures, data etc. that are not my own work	☒
Not made use of the work of any other student(s) past or present without acknowledgement. This includes any of my own work, that has been previously, or concurrently, submitted for assessment, either at this or any other educational institution, including school (see overleaf at 31.2)	☒
Not sought or used the services of any professional agencies to produce this work	☒
In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations	☒

DECLARATION:

I am aware of and understand the University's policy on plagiarism and I certify that this assignment is my own work**,** except where indicated by referencing, and that I have followed the good academic practices noted above

Signed
....

Abstract

Since the pioneering work of Ivan Pavlov nearly a century ago, the empirical study of associative learning through classical conditioning has continued to grow. However, the high volume of classical conditioning investigations has resulted in an equal in magnitude methodological and analytical variation, which can often challenge cross-study comparisons, replicability and generalisability of findings (Haaker et al., 2019; Lonsdorf et al., 2017). Consequently, the field of conditioning has begun to focus on reducing excessive flexibility in data practices through increasing methodological rigour, consistency, and transparency. So far, research has concentrated on improving methods in areas such as the quantification of conditioned responding, analytical strategies, translational research and individual differences (Bach et al., 2018; Haaker et al., 2019; Korn et al., 2017; Lonsdorf et al., 2019; Lonsdorf & Merz, 2017; Ney et al., 2018; Sjouwerman & Lonsdorf, 2019). The aim of this thesis was to provide an additional contribution to recent methodological efforts in the field by focusing on an area that has not received as much empirical attention. Specifically, we discuss and examine the potential utility of multi-trial conditioning for studying psychophysiological indices of learning. In addition, throughout this thesis, we aimed to reinforce the value of transparent and robust data practises in aiding replicability and generalisability of conditioning research.

In Chapter 2, we report findings from an indirect behavioural replication of an established multi-trial task (i.e., Multi-CS Conditioning, Steinberg et al., 2013), accompanied by a discussion about the role of contingency awareness in conditioning. We also provide a re-analysis of a previous Multi-CS dataset (Rehbein et al., 2014) to highlight the value of robust and transparent data visualisation in guiding analytical decisions, and to illustrate how poor consideration of individual differences and underlying data distributions may explain the inconsistency in previous research using this task. Chapter 3 reports a novel visual blocked conditioning paradigm that delivers a high number of trials through attempting to elicit associative learning in multiple successive blocks. We investigated the potential utility of this task to overcome some of the technical and design challenges (e.g., detecting deep source activity, time-

frequency analysis) present in magnetoencephalography (MEG) research, studying the cortical and subcortical oscillatory dynamics of learning and extinction. The findings from this study suggested that the task does not reliably elicit conditioning in any of the outcome measures that we considered (MEG, pupil size, valence, and arousal ratings). Nevertheless, the reported results identified several design modifications that can aid future paradigm development. These were related to aspects such as trial duration, the type of CSs employed, and maintaining attention and contingency awareness. Chapter 4 reports findings from an auditory blocked conditioning task, modified based on the results from Chapter 3. The task was examined in the context of pupillary and subjective behavioural indices of conditioning, with a discussion of its application in future MEG designs. In addition, the study considers the potential of this multi-trial paradigm to offer better generalisability of findings when used in combination with robust analytical strategies (i.e., data-driven time window selection and design-appropriate mixed modelling). Finally, Chapter 5 discusses the implications of the findings reported in this thesis for future multi-trial conditioning research.

Table of Contents

List of Tables

List of Figures

List of Supplementary Tables

List of Supplementary Figures

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Christoph Scheepers for his support and guidance during the final year of my PhD. Thank you, for sharing all your data and mixed modelling wisdoms and for reviving my passion for research.

I would like to thank my second supervisor, Dr Nicola van Rijsbergen, for her continuous support throughout my Master's and PhD. You always provided me with alternative perspectives and constructive ideas, which challenged me to push myself forward.

I would also like to thank Dr Marc Recasens and Dr Tineke Grent-'t-Jong for sharing their technical and practical experience with MEG design and analyses. Thank you, to all participants who contributed their time and took part in these studies.

Finally, a big thank you to my friends and family for patiently supporting me throughout the PhD and to the best lab group of friends – Labless, for always being there through the ups and downs.

Abbreviations

(f)MRI	(Functional) magnetic resonance imaging
(k)Hz	(Kilo)hertz
(rm)ANOVA	(Repeated measures) analysis of variance
A/DCG	Anterior/Median cingulate gyri
AAL	Automated anatomical labelling
ACC	Anterior cingulate cortex
ANS	Autonomic nervous system
BLA	Basolateral amygdala
CA1	Cornu ammonis area of the hippocampus
CeA	Central Amygdala
CI	Confidence interval
CLM	Cumulative link mixed (models)
CR	Conditioned response
CS	Conditioned stimulus
CS-	Conditioned stimulus unpaired with an UCS
CS+	Conditioned stimulus paired with an UCS
dB	Decibels
E/MEG	Electro/magnetoencephalography
EDA	Electrodermal activity
EIFEL-ROF	Research Network for the European Interdisciplinary Study of Fear and Extinction Learning as well as the Return of Fear
EMG	Electromyography
ERF	Event-related field
FDR	False discovery rate
FFG	Fusiform gyrus
FFT	Fast fourier transformation
FIR	Finite impulse response
HPC	Hippocampus
IL	Infra-limbic cortex
ITI	Inter-trial interval
LA	Lateral amygdala
LCMV	Linearly constrained minimum variance

LME	Linear mixed effects (models)
LPP	Late positive potential
M/SOG	Middle/Superior occipital gyri
MNE	Minimum-norm estimation
MNI	Montreal Neurological Institute
mPFC	Medial pre-frontal cortex
OFC	Orbito-frontal cortex
OrbInf	Inferior frontal gyri
OrbMid	Middle frontal gyri
PL	Pre-limbic cortex
PSPM	Psychophysiological modelling
RMS	Root mean square
ROI	Region-of-interest
RSPM	Raven’s standard progressive matrices
SAM	Self-assessment Manikin scale
SCL	Symptom Checklist-90 Revised
SCR	Skin conductance response
SEM	Standard error of the mean
SNR	Signal-to-noise ratio
SSR	Steady-state response
STAI	State-trait anxiety inventory
TF	Time frequency
UCR	Unconditioned response
UCS	Unconditioned stimulus
vmPFC	Ventromedial pre-frontal cortex

List of publications

Chapter 4 is available as a pre-print manuscript, under a CC-By 4.0 licence:

Jambazova, A. A., Rijsbergen, N. van, & Scheepers, C. (2020). The more, the better: Auditory threat conditioning using multiple conditioned and unconditioned stimuli over many trials. 1–47. https://doi.org/10.31234/OSF.IO/P6Z5U

The content of Supplementary Materials 17 (Auditory control task) has been made available online under a CC-By 4.0 licence: https://osf.io/dehxa/

1 Chapter 1 – Introduction

1.1 Overview of associative learning

The ability to learn about predictive relationships between stimuli and motivationally significant outcomes (i.e., associative learning) and to use cues to anticipate future events, allows organisms to continuously adapt within their dynamic environment (Esber & Haselgrove, 2011; Le Pelley, 2004). This form of adaptive responding is shared among many animal species, from invertebrates to humans (Hawkins & Byrne, 2015; Morand-Ferron, 2017). For example, worms learn to avoid surfaces based on odour associations that predict the presence of pathogenic bacteria (Ardiel & Rankin, 2010). A lizard would learn to avoid toxic prey after ingesting a non-toxic amount of it (Morand-Ferron, 2017), while a heron may use sudden ripples in the water as cues for the potential location of prey (Esber & Haselgrove, 2011).

In humans, from an early age, emotions such as fear play an important role in supporting identification of motivationally significant threat-related events such as fearing separation from the primary caregiver in early childhood, or more socially relevant and abstract threats such as humiliation in adolescence (Shechner et al., 2014). Experiencing fear enables individuals to adaptively respond to and manage threat by forming and using associative memories of the relationship between threats and the cues that predict them. For instance, prior knowledge of an upcoming threat such as a radio announcement of a traffic accident blocking the road can allow a driver to remain in control of their vehicle (Goodman et al., 2018). However, fear can become maladaptive when physiological and behavioural responses to potential threat are exaggerated or when fear cannot be effectively regulated. This can cause a persisting fear response to stimuli that no longer signal danger or to an overgeneralisation to non-threatening situations (Dunsmoor & Paz, 2015). Such excessive or exaggerated fear reactions can have a detrimental impact on an individual’s wellbeing and lead to the development of anxiety disorders. Given its crucial role in both adaptive and maladaptive functioning, it is unsurprising that associative learning has been a popular topic of scientific investigation for more than a century, facilitating our understanding of how acquisition, expression and regulation of fear are acquired and of the treatment of anxiety disorders.

1.2 Historical foundations of associative learning research

The experimental study of learning began with the first associative learning theory proposed by Thorndike (1898), that was founded based on the idea that behavioural change occurs as a consequence of experience (Klein, 2019). In his work, Thorndike demonstrated that when a cat is placed in a puzzle box with food available outside, the cat gradually learns to engage in the behaviour that triggers the release mechanism of the box, allowing it to access the reward. The cat also learns to escape the box faster in subsequent trials, while other behaviours such as clawing and meowing that do not facilitate the release mechanism tend to decrease in frequency over time. Based on these observations, Thorndike suggested that learning occurs as a result of the formation of associations between a stimulus (e.g., the box) and a response (e.g., pressing the release mechanism). These stimulus-response associations were proposed to occur through the experience of trial and error and to be strengthened through the delivery of a reward (see Klein, 2019 for a detailed discussion).

Later work shifted attention away from stimulus-response associations and focused on the role of stimulus-stimulus contingencies in explaining associative learning processes. The influential research of Ivan Pavlov laid the solid theoretical and empirical foundations in associative learning that are still relevant to the present day. At the root of Pavlov's theory was his research in animal digestion and the discovery of reflexive responses, beginning in 1898. A detailed account of his work, however, was not widely available until 1927 when an English translation of his book (Pavlov, 1927) detailing the previous twenty-five years of his research was published (Boakes, 2003). In his work, Pavlov suggested that *unconditioned reflexes* are innate in both humans and animals and occur when an *unconditioned stimulus* (UCS) such as food, triggers an autonomic *unconditioned response* (UCR) such as salivation. He also suggested that reflexive responses can be learned through conditioning involving stimulus-

stimulus associations (i.e., a *conditioned reflex*). He demonstrated this type of learning by measuring saliva from a dog's salivary glands while presenting the animal with an initially neutral stimulus – the sound of a metronome. The sound served as a *conditioned stimulus* (CS) and was paired with meat powder acting as the UCS. While initially, the dog exhibited only an UCR by salivating in response to the UCS, the repeated CS-UCS pairings began to elicit a *conditioned response* (CR, i.e., salivation) to the CS which increased in magnitude over time (Klein, 2019). Another crucial discovery derived from Pavlov's comprehensive investigations of associative learning was that CRs acquired during conditioning can be weakened through the process of *extinction* learning. Pavlov suggested that repeatedly presenting the CS without the UCS following conditioning creates a new, inhibitory CS-UCS association that overrides the earlier associative memory and subsequently diminishes the magnitude of the CR (Wasserman & Miller, 1997).

These discoveries had a major impact on later contributions to learning research and the rise of behaviorism as pioneered by John Watson. Shifting the focus towards human research, he suggested that both adaptive and maladaptive behaviour could be learned (Klein, 2019). The infamous 'Little Albert' experiment provided empirical evidence for this by demonstrating that emotional responses such as fear, are also susceptible to conditioning (Watson & Rayner, 1920). In their study, a 9-month-old infant who initially experienced no fear of rats, was exposed to a white rat (CS). The child was presented with the strike of a hammer against a steel bar (UCS) every time they reached for the rat. Following several CS-UCS parings, the child began to exhibit fear in response to the rat, evidenced by crying and crawling away. This fear was also found to generalise to other similar objects (Watson & Rayner, 1920). The work of Watson and Rayner was followed by the 'Little Peter' experiment by Mary Cover Jones demonstrating the elimination of conditioned fear through *counterconditioning*. In this study (Jones, 1924), a 3-year-old boy with a fear of rabbits was conditioned to associate the animal with a pleasurable activity. The rabbit (CS) was moved closer in proximity to Peter while he was eating candy (UCS), until the child was able to interact with the rabbit by holding and touching it. Her discovery laid the foundations of our understanding of the aetiology of anxiety

disorders and aided the development of behavioural interventions for their treatment, such as systematic desensitisation (Fullana et al., 2020).
In later years, behaviourism focused on another form of learning (i.e., operant conditioning), which attempted to account for the impact of external stimuli on an organism's behaviour (Akpan, 2020). The theory of operant learning as defined by Skinner, was inspired by Thorndike's early work but incorporated the crucial role of reinforcement in determining conscious behaviour (Ruan & Wu, 2013). Skinner showed that the consequences of one's actions drive changes in behaviour (Zalta & Foa, 2012). For example, he demonstrated that a certain behaviour is more likely to occur if it causes a reduction in an unpleasant experience (i.e., negative reinforcement), (Zalta & Foa, 2012) or results in the delivery of a positive outcome (i.e., positive reinforcement), (Murphy & Lupfer, 2014). These principles of reinforcement have since been used to explain human behaviour and learning in a wide range of contexts, such as language acquisition, addiction, as well as the maintenance of anxiety (Akpan, 2020; Zalta & Foa, 2012).

1.3 Clinical applications of conditioning principles in the 20th century

Until the 1970s, behaviourism was the predominant approach contextualising pathological fear. Initially, based on early research (Jones, 1924; Watson & Rayner, 1920), anxiety was understood as the consequence of simple classical conditioning involving a traumatic experience (Lissek et al., 2005; Mineka & Zinbarg, 2006; Zalta & Foa, 2012). In 1947, the conceptualisation of anxiety was refined by Mowrer's two-factor theory to incorporate the influence of both classical and operant conditioning in the development and maintenance of anxiety (Mowrer, 1947). Mowrer suggested that in the context of fear, avoidance of the feared stimulus serves as a negative reinforcer by reducing physiological arousal. He further proposed that anxiety initially develops through classical conditioning (first factor), but it is subsequently maintained by operant conditioning (second factor) through avoidance of the feared situation. Specifically, avoidance was suggested to disrupt the development of extinction by preventing an individual from forming a safety associative memory that the CS does not signal danger (Krypotos, 2015). This discovery was fundamental for the development of exposure-based interventions, by postulating that the treatment of anxiety should not only focus on extinction of the feared response through repeatedly exposing a patient to the feared event but also, on eliminating avoidance through sustained exposure until anxiety has subsided (Krypotos, 2015; Zalta & Foa, 2012).

Later work by Wolpe (1968), focused on treatment strategies based on the principles of counterconditioning demonstrated by Jones (1924). According to his theory, anxiety could be reduced through the process of reciprocal inhibition, in which anxiety towards the feared stimulus (e.g., a rabbit) can be diminished through pairing the stimulus with a response that is incompatible with fear (e.g., eating candy). Extending Jones' work, Wolpe initially tested this hypothesis in cats who were conditioned to fear their cage through associating it with an electric shock. Later, their fear response was reduced through counterconditioning by providing the cats with food while they were in their cages. Based on these findings, Wolpe began implementing systematic desensitisation treatment for anxiety in humans (see Vinograd & Craske, 2020 for a discussion). The procedure involved patients alternating between completing a task that is physiologically incompatible with fear (i.e., deep muscle relaxation), and a gradual exposure to feared stimuli through imagery, beginning with stimuli that only induce mild fear (Vinograd & Craske, 2020). While this intervention was found to be successful in reducing anxiety at least for specific phobias, the clinical interest in it declined as a result of further work showing superior effectiveness of real exposure over mental imagery and in the absence of relaxation techniques (Zalta & Foa, 2012).

In the 1970s, criticisms of the behavioural approach to psychopathology emerged as it became apparent that conditioning alone could not account for factors such as individual differences. For instance, behavioural theories could not explain why anxiety disorders are not always triggered by conditioning or why traumatic conditioning does not always lead to the development of anxiety (Hofmann & Hay, 2018). The simple conditioning approach also failed to consider the impact of mental processes. With the rise of the "cognitive revolution", investigations began into the contributing role of higher order factors such as memory and attention in the development of psychopathology (Kindt, 2014). The clinical interest in behaviourism diminished with the shift towards cognitive theories explaining psychopathology as driven by disorder-specific cognitive biases (e.g., misinterpretation of physical sensations in panic disorder), (Hofmann, 2008). Nonetheless, exposure is still considered one of the most effective and critical aspects in the treatment of anxiety, with contemporary interventions combining both behavioural and cognitive approaches in the treatment of pathological fear (Kindt, 2014) .

1.4 The shift from behaviourism to neuroscience and psychophysiology

While the clinical interest in Pavlovian conditioning declined in the late 20th century, classical conditioning research continued to grow with a focus on the neurobiological mechanisms driving associative learning processes. Initial investigations aimed at establishing the biological basis of conditioning in animals (LeDoux, 2014), with early work focusing on understanding the role of stress hormones and a range of neurotransmitters on extinction processes (Milad & Quirk, 2012). Animal lesion research at the end of the 20th century began to provide insight into the functional role of the amygdala in the acquisition and expression of fear as well as into its anatomical connections with other brain regions (see Milad & Quirk, 2012 for a review). With technological advances in neuroimaging techniques (i.e., functional magnetic resonance imaging, fMRI), the interest in mapping the neural circuits of learning and extinction in humans began to grow substantially. (Fullana et al., 2020). The first neuroimaging evidence of amygdala involvement during human threat conditioning and extinction was provided by LaBar et al. (1998), corroborating neuropsychological findings showing that amygdala damage impairs the acquisition of conditioned responding in humans (e.g., Bechara et al., 1995). Using fMRI, they observed increased amygdala activity in response to stimuli paired (CS+) with an electric shock compared to unpaired stimuli (CS-). These findings provided a major contribution to associative learning research, confirming that the crucial role of the amygdala in threat learning is conserved across species (LaBar et al., 1998).

During the late 1980s, it was suggested that translating animal findings to studies in humans may also provide a means for gaining an insight into psychopathology. As such, classical conditioning became a popular animal model of anxiety disorders, with findings from animal studies informing human research in clinical populations (Milad & Quirk, 2012). These studies have demonstrated that persistently high fear responses during extinction in individuals who have experienced trauma, may be caused by hyper- activation in regions involved in the encoding and expression of threat (e.g., amygdala and the anterior cingulate cortex), as well as hypo-activation in regulatory regions such as the ventromedial pre-frontal cortex (vmPFC), (Milad et al., 2009; Shin et al., 1999).

These advances in animal and human conditioning research have allowed for the establishment of a comprehensive map of inter-related brain systems that detect and respond to threat-related information, often referred to as the 'fear circuit' (LeDoux, 2000; LeDoux & Pine, 2016). Today, it is widely agreed that this network of regions is translatable and preserved across mammals and involves the amygdala, hippocampus and the medial pre-frontal cortex (Fullana et al., 2020). For example, recent evidence from rodents has shown that information about the CS and UCS is initially encoded by the lateral amygdala (LA), (McCullough et al., 2016; Tovote et al., 2015) while behavioural and physiological fear reactions are triggered by the central amygdala (CeA), (LeDoux & Pine, 2016; McCullough et al., 2016). These defensive responses are mediated by cortical areas including the medial pre-frontal cortex (mPFC). In particular, the pre-limbic cortex (PL), part of the mPFC contains bi-directional projections to and from the amygdala, and is involved in fear expression, while the infra-limbic cortex (IL) projects to and downregulates basolateral amygdala (BLA) activity during the extinction of fear responses (McCullough et al., 2016; Tovote et al., 2015). The hippocampus (HPC) also plays an important role in the fear network as it is responsible for encoding of contextual and valence-specific information associated with the memory of the event (Lesting et al., 2011; McCullough et al., 2016).

Human fMRI research has established a similar network of circuits implicated in associative learning and extinction processes (Fullana et al., 2016, 2018; Sehlmeyer et al., 2009). In meta-analyses of fear conditioning and extinction studies, the anterior cingulate cortex (ACC), corresponding to the PL in rodents, and the anterior insula have been identified as the most reliably activated regions during threat processing and extinction (Fullana et al., 2016, 2018). When comparing responses to the CS+ relative to CS- during learning, a consistent deactivation of the vmPFC (corresponding to the IL in rodents), lateral orbitofrontal cortex (OFC) and the HPC is also observed.

Cross-species investigations, however, have not always yielded consistent findings. For instance, the functional role of the vmPFC in humans is still debated. The vmPFC has traditionally been seen as a region involved in the down-regulation of negative affect in a range of experimental situations including extinction learning (Delgado et al., 2008; Diekhof et al., 2011), that has been corroborated in animal research and human studies of emotion regulation (Diekhof et al., 2011; Gonzalez & Fanselow, 2020). However, there is accumulating evidence suggesting that activity in this region in humans may also be related to the processing of safety signals (CS- trials), (Fullana et al., 2016). Direct evidence for the role of the vmPFC in safety processing has been provided by Harrison et al. (2017), who demonstrated that vmPFC activity is positively correlated with CS- valence ratings, and activation for CS- trials persists even following adjustments for baseline activity, typically present during resting state imaging. Such functional distinction in the context of extinction paradigms may be linked to cross-species procedural differences whereby human conditioning studies heavily rely on the use of a control condition (i.e., CS-) with strong safety properties. This can also explain the difficulties in reliably detecting vmPFC activity in humans, as contrasting a safety stimulus (CS-) with a previously threatening stimulus (CS+) that quickly adopts a safety property can be analytically challenging. Specifically, computing CS+ > CS- contrasts during extinction learning creates a situation in which comparisons are made between two stimuli that may not exhibit substantial differences to allow for the detection of robust vmPFC involvement in humans (Fullana et al., 2018).

In addition, while animal studies have consistently demonstrated the crucial role of the amygdala during learning and extinction, human fMRI research has faced challenges in reliably detecting amygdala sources elicited from conditioning. Even though an early human meta-analysis revealed some evidence for detecting amygdala activity during learning (Mechias et al., 2010), later meta-analyses by Fullana et al. (2016, 2018) failed to detect a robust and consistent presence of amygdala activity during either threat learning or extinction. The poor translation in the context of amygdala involvement can be linked to the difficulty in localising source activity from the amygdala, especially from a task that is shown to trigger responses only in a small number of neurons (Fullana et al. 2020). For example, across the fMRI literature, both increases in activity in response to the CS+ and the CS- have been reported (Fullana et al., 2018). This pattern may not be entirely inconsistent with rodent data, as studies have shown that during threat learning, a similar number of neurons exhibit excitatory and inhibitory responses to the CS+ (Ciocchi et al., 2010). During extinction, different cell populations of the LA have also been shown to exhibit simultaneous increases and decreases in activity (Repa et al., 2001). Therefore, it is possible that different human fMRI studies tap onto activity from distinct neuronal populations and that the spatial resolution of fMRI may be insufficient to detect such fine-grained patterns. Finally, fMRI studies often rely on detecting differences using time-invariant CS+> CS- contrasts, by averaging activity over the trial duration. This approach prevents examinations of time-dependent neural differences in activity, that are likely to be present in the amygdala (Sehlmeyer et al., 2009). It is therefore possible that neural differences between CS+ and CS- are too subtle and rapidly extinguishing to be reliably and consistently detected with a method with very low temporal resolution such as fMRI (Lin et al., 2013).

Recent years have seen a growing interest in understanding the temporal dynamics of neural indices of learning, using electro- and magnetoencephalography (E/MEG) (see Miskovic & Keil, 2012; Trenado et al., 2018 for reviews). Yet, due to the inherent technical limitations of these measures (which restrict inferences primarily to the cortical surface) the focus of research has hitherto been predominantly on gaining insight into how conditioning is reflected in visual and auditory systems (Lonsdorf et al., 2017). In the past several years, however, efforts have been made to optimise these techniques (see Chapter 3) to allow inferences beyond the cortical surfaces, including structures deeper in the brain such as the amygdala (Attal et al., 2007; Balderston et al., 2013; Quraan et al., 2011a; Tzovara et al., 2019). A more detailed review of the E/MEG literature in learning and extinction will be provided in Chapter 3.

Another large body of research has focused on understanding the complex interplay between neural and autonomic activity and their relationship with behavioural, cognitive and affective processes. Consequently, a range of psychophysiological techniques assessing brain and autonomic nervous system (ANS) activity have been utilised in the study of classical conditioning (Gaffey & Wirth, 2014). Electrodermal activity (EDA) was the first measure used to index conditioning and to the present day, EDA measures such as the skin conductance response (SCR) have remained the most widely employed techniques (Lonsdorf et al., 2017). The startle eyeblink response derived from electromyography (EMG) and elicited through sudden sensory events, has been considered as the most reliable learning index in humans, and as such has also received a lot of empirical attention (Lonsdorf et al., 2017). Less commonly used physiological methods include pupillometry and heart rate changes, although pupil size has recently been employed more commonly and often in combination with SCR indices (Jentsch et al., 2020; Leuchs et al., 2019). An overview of the literature in SCR and pupil size as indices of learning and extinction will be discussed in Chapter 4.

Finally, since for ethical reasons contemporary conditioning paradigms are unlikely to elicit extreme behavioural responses, such as escape, behavioural measures of conditioning are rarely employed (Lonsdorf et al., 2017). Instead, subjective measures assessing individuals’ affective and cognitive states are more commonly used, including measures of valence and arousal, and reports of CS-UCS expectancy and contingency. Since different outcome measures are shown to tap onto different aspects of learning mechanisms (Leuchs et al., 2019), simultaneous recordings of multiple behavioural and psychophysiological outcome measures are beginning to be employed more frequently (Lonsdorf et al., 2017).

1.5 Variations in contemporary classical conditioning protocols

There is an abundance of classical conditioning protocols that have been used across the literature to study associative learning processes. These differ in terms of the type of stimuli that are being employed and procedurally in terms of reinforcement rate, timing, trial number and many other factors, some of which will be briefly reviewed below.

1.5.1 Type of stimuli

Various stimulus types have been used as CSs in the associative learning literature. These are often neutral, although emotional stimuli are occasionally used (Burkhouse et al., 2019; Pischek-Simpson et al., 2009; Rowles et al., 2012). Detecting a CR using emotional stimuli (e.g., fearful faces), however, can be problematic since they inherently elicit affective responses even prior to conditioning, and can mask conditioned responses (Lonsdorf et al., 2017). Across the literature, visual CSs are the most commonly employed including a wide range of categories such as gratings, geometric shapes, coloured lights, abstract images, faces, and animals (Sehlmeyer et al., 2009). When auditory stimuli are used, these typically involve tones or natural sounds (Bröckelmann et al., 2011; Fullana et al., 2016; Sehlmeyer et al., 2009). Although less common, olfactory, tactile, and taste CSs have been reported (Lonsdorf et al., 2017). In terms of UCSs, the administration of a painful stimulus such as an electric shock is the most commonly employed (Lonsdorf et al., 2017), however, for ethical reasons this is not always suitable for use in vulnerable populations such as clinical groups and children. Consequently, other unpleasant stimuli have been utilised. Most frequently, these have been auditory stimuli such as white noise and human screams (Glenn et al., 2012; Sperl, Panitz, & Hermann, 2016), but other highly arousing sounds have been used as well (Junghöfer et al., 2015b). Other stimuli such as air-puffs, olfactory and affective visual stimuli have also been reported (Lonsdorf et al., 2017; Steinberg et al., 2013).

1.5.2 Cued conditioning

Early conditioning studies (Jones, 1924; Watson & Rayner, 1920; Wolpe, 1968) typically used one CS (e.g. a rat, cage, or rabbit) that was paired with an UCS (e.g. loud noise, electric shock, or candy). More recently, such *single-cue* protocols have increasingly been replaced with *differential-cue* protocols, particularly for human studies, because the latter provide better statistical power and control for the presence of processes unrelated to associative learning (Lonsdorf et al., 2017). In these protocols, one neutral stimulus (CS+) is paired with the UCS while another stimulus (CS-) remains unpaired with CRs representing the difference between responses to the CS+ compared to the CS-. *Multiple-cue* procedures (Junghöfer et al., 2015b; Rehbein et al., 2014; e.g. Steinberg et al., 2013) involving multiple different CS+ and CS- items have also been used, although not as commonly as traditional differential-cue tasks. Since differentiating between a large number of stimuli is cognitively demanding in nature, these tasks have been employed to study neural activity in higher-order brain regions such as the pre-frontal cortex (Rehbein et al., 2014). The cognitive demand of such procedures, however, creates a situation in which awareness of the CS-UCS contingency may be difficult to establish, which may affect the development of a CR (Lonsdorf et al., 2017), (see Chapter 2 for a detailed discussion of Multi-CS conditioning).

In addition to variations in the number of CSs, procedures differ in relation to the timing between the presentation of the CS+ and the subsequent occurrence of the UCS. Specifically, in *delay conditioning* tasks the CS+ overlaps with or terminates with the onset of the UCS, while during *trace conditioning* a time interval in the range of 0.5 – 10 s separates the CS+ from the UCS, which is suggested to recruit working memory processes to a greater extent (Sehlmeyer et al., 2009). Furthermore, in both procedures, the UCS reinforcement rate (i.e., the probability of UCS occurrence) can be varied whereby in partial reinforcement protocols the CS+ is paired with the UCS only in portion of the trials. In 100% reinforcement procedures, the CS+ is always paired with the UCS. It is suggested that partial reinforcement produces a weaker CR during acquisition which takes longer to extinguish during extinction training (Lonsdorf et al., 2017). However, this procedure has its benefits as it can prevent UCS habituation and eliminate any potential confounding influence of the UCS by allowing for the measurement of CS+ unpaired responses (Lonsdorf et al., 2017). Conditioning tasks can also be preceded by a habituation phase in which the CSs are presented without the UCS. Using a familiarisation phase can be useful for establishing a baseline response to each stimulus that can later be used to account for potential baseline differences between conditions, while enabling participants to become familiar with the general task procedures (Lonsdorf et al., 2017).

1.5.3 Context conditioning

In contrast to cued protocols, in *context conditioning* the UCS is not predicted by a discrete cue but rather by the environment in which conditioning takes place (Marschner et al., 2008), with resulting CRs occurring in a more sustained fashion compared to cued CRs (Kroes et al., 2017). It has therefore been suggested that context conditioning models sustained states of anxiety to uncued threats, typical for generalised anxiety disorders (Grillon et al., 2006). In animal research the context is usually the conditioning chamber, while in humans, different details in the experimental task environment are modified (i.e., usually a background image or movies). For example, these may include presenting different scenes as contextual CSs (e.g. a bedroom and living room), one of which is paired with an UCS (Kroes et al., 2017). Recent technological advances have also allowed for utilizing virtual reality in context conditioning studies (Kroes et al., 2017). The distinction between context and cues, however, can sometimes be unclear as often, details within a context can serve as cues in which cases the context might serve as an occasion setter, modulating the conditioning rather than eliciting a CR alone (Kroes et al., 2017; Lonsdorf et al., 2017).

1.5.4 Generalisation

In addition to assessing how individuals acquire CRs during threat acquisition, the past decade has seen growing interest in gaining insights into how an acquired threat or fear towards one stimulus generalises to other similar stimuli. When *threat generalisation* occurs, the effects elicited by threat learning (e.g., fear) extend to other similar events that individuals associate with the initial threat experience (Dunsmoor et al., 2009). This process is adaptive in nature as it allows individuals to recognise and adequately respond to potential threat of a novel stimulus through prior experience (Dunsmoor et al., 2009). Threat generalisation, however, can become maladaptive when overgeneralisation to non-threatening stimuli elicits a fear response. This overgeneralisation process has been suggested to be a crucial aspect underlying anxiety disorders (Dymond et al., 2015) and can be linked to perceptual similarities such as general physical properties (e.g., all dogs), to specific features that are perceived as threatening (e.g., sharp teeth), or to conceptual ones (e.g., a fear of all stimuli or situations that may be perceived as potentially life-threatening), (Bennett et al., 2015; Dymond et al., 2015). Experimentally, threat generalisation protocols based on perceptual similarities are the most commonly employed as these are easily quantifiable (Dunsmoor & Paz, 2015). In these tasks, following conditioning, responses to a series of generalisation stimuli (GSs) resembling the CS are measured. This form of generalisation has been tested using a range of stimuli varying in colour, shape or size (Dunsmoor & Paz, 2015). Conceptual forms of generalisation include, for example, using words as CSs and their synonyms as GSs, with semantically related stimuli eliciting a threat generalisation response similar to that elicited by the CS (Boyle et al., 2016).

1.5.5 Extinction and return of fear

During extinction training, the CSs are presented again without being paired with the UCS, allowing for a new memory trace to form which signals the newly acquired safety of the CS+. In recent years, *immediate extinction* following acquisition within the same experimental session is the most frequently employed procedure in studies on humans (Lonsdorf et al., 2017). *Delayed extinction*, in contrast is more common in the animal literature whereby extinction training is delivered at a later point in time, usually 24 hours after conditioning (Lonsdorf et al., 2017). A recent review of the animal and human literature suggested that while successfully reducing the CR, immediate extinction often fails to secure long-term retention, causing spontaneous CR recovery within 24 hours of extinction training (Maren, 2014). Maren (2014) suggested that these extinction deficits occur since brain systems involved in the acquisition of threat are still active immediately after conditioning and therefore, inhibit activity in extinction-related regions. This observation may suggest that immediate trauma interventions could be ineffective and that longer consolidation periods may be required for maintaining long-term extinction (Maren, 2014). Yet, immediate extinction procedures can still be informative for immediate treatment intervention research and can offer insights into the effects of consolidation interruption (Lonsdorf et al., 2017). Furthermore, such procedures can provide time and cost-effective means for studying the learning processes underlying extinction development and for developing new paradigms.

The return of fear is experimentally manipulated in investigations attempting to model clinical relapse (Lonsdorf et al., 2017), which is a common problem in the treatment of anxiety (de Jong et al., 2019). Similar to the study of threat acquisition, a wide range of protocols are available for this purpose. *Spontaneous recovery* is typically studied in the absence of any experimental manipulations by re-exposing participants to the CSs at least 24 hours following extinction training (Lonsdorf et al., 2017). The return of fear can also be elicited through *reinstatement* protocols in which participants are re-exposed to the UCS or to a non-extinguished CS (Haaker et al., 2014; Halladay et al., 2012). From a clinical perspective these procedures can offer insights into the processes that drive aggravation of symptomatology following re-exposure to a traumatic event (Norrholm et al., 2006). Finally, *fear renewal* protocols provide contextual manipulations. For example, a CR acquired and extinguished in context A can be renewed in a new context B, or a CR elicited in context A can be extinguished in context B, but subsequently renewed in context A (Lonsdorf et al., 2017).

1.5.6 Awareness

Standard classical conditioning tasks rely on eliciting a CR through establishing a contingency awareness of the relationship between the CSs and the UCS. A perpetuating debate across the conditioning literature is whether contingency awareness is a necessary component in conditioning, and whether and under what conditions associative learning can occur in the absence of awareness (see Mertens & Engelhard, 2020 for a review and Chapter 2 for more details). Consequently, a number of tasks have been designed to determine the role of awareness in conditioning. These frequently focus on diverting attention away from the contingency by implementing a secondary task that is cognitively demanding, such as requiring participants to discriminate between a series of tones (Dawson et al., 2007) or numbers (Tabbert et al., 2011) . Other procedures that do not require the implementation and potential confounding influence of a secondary task involve manipulations in the discriminability of CSs (Schultz & Helmstetter, 2010). In addition, subliminal conditioning is commonly implemented, which typically involves the presentation of CSs below the perceptual threshold using masking procedures (Balderston et al., 2014b; Raes & Raedt, 2011). Subsequently, a range of tasks can be utilised to assess the extent to which awareness has been established including CS discrimination tasks, online expectancy ratings and post-experimental questionnaires. However, the accuracy and sensitivity of some of these measures have been heavily criticised due to issues such as low power as a result of a small number of trials and prolonged delays between conditioning and the assessment of awareness. (Mertens & Engelhard, 2020).

1.6 The current state of the art

Research interest in classical conditioning has continued to grow over time. To date, it is fair to say that the paradigm has become one of the most common approaches for studying the underlying mechanisms of associative learning; for instance, a Google Scholar search for the term ‘fear conditioning’ reveals over 15,000 results for the year 2020 alone. Areas of investigation involve both human and animal studies examining a wide range of topics including, but not limited to, development, psychopathology, pharmacology, neurobiology, and psychophysiology. This dramatic increase in classical conditioning studies has also resulted in a great level of design and methodological variation (Lonsdorf et al., 2017), some of which was reviewed in the previous section. Furthermore, differences in methodology are often accompanied by equally varied analytical strategies. This high degree of methodological and analytical heterogeneity creates difficulty in comparing findings across studies in both human (Lonsdorf et al., 2017) and cross-species research (Haaker et al., 2019).

With the rise in awareness of the *replicability crisis* in psychology (Aarts et al., 2015), a considerable amount of attention has been given to increasing reproducibility of psychological research through improving transparency and research practice. In the context of conditioning research, this inspired the formation of a multi-disciplinary research group *(Research Network for the European Interdisciplinary Study of Fear and Extinction Learning as well as the Return of Fear,* i.e., EIFEL-ROF*)* which has been making substantial efforts to improve the robustness of methods for studying fear and anxiety through the provision of methodological guidelines, reviews, meta-analyses and cross-laboratory replications (*European Meeting of Human Fear Conditioning*, n.d.). One of the first papers (Lonsdorf et al., 2017) derived from this collaboration focused on providing a detailed review and a set of guidelines for novices for the design and statistical analysis of classical conditioning experiments, along with an extensive review of potential outcome measures for indexing conditioning including behavioural, psychophysiological and neural read-outs. Later work focused on improving methodology in relation to issues such as individual differences, exclusion criteria, SCR quantification, and translational research (Haaker et al., 2019; Lonsdorf et al., 2019; Lonsdorf & Merz, 2017; Sjouwerman & Lonsdorf, 2019). Outside of this research collaboration, there has been work focusing on improving analytical tools in classical conditioning research (Bach et al., 2015; Bach & Friston, 2013; Korn et al., 2017; Ney et al., 2018). Each of these issues will be briefly discussed below.

1.6.1 Individual differences

Individual differences in anxiety, which are also present when modelling threat responding experimentally, constitute a prominent topic in clinical practice and conditioning research. Despite the enormous efforts to gain insights into the development and maintenance of anxiety disorders in the previous century, there are still significant gaps in our understanding of why exposure to a traumatic event does not always lead to the development of anxiety and why the effectiveness of interventions varies across individuals. In the context of clinical models of anxiety such as classical conditioning, a similar pattern is observed where, even when identical procedures are employed, individual differences in the magnitude of conditioned responding are common (Lonsdorf &

Merz, 2017). To contextualise this issue, Lonsdorf and Merz (2017) provided a detailed review of the role of biological, genetic, psychological, procedural and analytical variation that may contribute to the likelihood of observing individual differences in threat learning (Lonsdorf & Merz, 2017).

In terms of methodology, Lonsdorf and Merz (2017) suggested that a special consideration should be paid in relation to factors that can mediate the manifestation of individual differences such as the strength of the experimental manipulation, baseline response differences across populations, sample characteristics and exclusion criteria. In addition, they argued that since different read-outs my tap onto distinct aspects of learning, capturing response variability may be facilitated by indexing conditioning using multiple outcome measures. Furthermore, the authors encouraged the use of adequate statistical tools suitable for inferences about individual differences such as ensuring direct between-group comparisons, including potential covariates in the analyses, avoiding artificial dichotomisation of variables through procedures such as median-splits, and being aware of the risks of selection bias through arbitrary data exclusion.

1.6.2 Exclusion criteria

More recently, Lonsdorf et al. (2019) provided an empirical illustration of the major impact that data exclusion practices and researcher degrees of freedom have on the inferences and conclusions that are drawn from conditioning data. Since it is believed that a considerably large and robust CR is a prerequisite for studying learning and extinction processes, data exclusion of participants who have failed to develop a CR or were non-responsive to the experimental stimuli (i.e., non-learners and non-responders respectively) is common (Lonsdorf et al., 2019). These practices, however, are often arbitrary and highly variable across the literature. Specifically, Lonsdorf et al. (2019) showed that 22% of the reviewed literature adopted performance-based exclusion of individuals who did not exhibit a CR, and each of these studies adopted a different definition of non-learners. Similarly, 32% of the literature employed data exclusion of non-responders, with a similar degree of variability in the definitions.

In addition, through re-analyses of example datasets from previous conditioning research, the paper pointed out several issues arising from heterogeneous definitions and analytical practices. First, inferences were shown to differ greatly depending on the adopted definition of non-learners. In addition, it was found that if one outcome variable failed to exhibit a CR, then this was not necessarily paralleled in other outcome variables. It was further demonstrated that arbitrary data exclusion criteria may impact on statistical inferences by creating sampling bias. For instance, in one of the reported example datasets, individuals with high trait anxiety exhibited lower differential CRs. Exclusion of those participants would thus, introduce a sampling bias towards a population with low trait anxiety. The considerable variation in the conclusions that can be drawn from the same dataset was suggested to pose a significant risk to the replicability and generalisability of findings to different samples, but also to clinical translation. Consequently, Lonsdorf et al. (2019) offered comprehensive guidelines and solutions to these problems, by encouraging transparent reporting of exclusion criteria through open science practices and adequate data visualisation tools that capture all of the available data. They further argued that data exclusion should be justified theoretically as well as practically through manipulation checks ensuring that participants truly failed to learn the CS-UCS contingency.

1.6.3 SCR Quantification

Another source of significant methodological and analytical variation is the definition and quantification of the most commonly used index of conditioned responding, the SCR (Sjouwerman & Lonsdorf, 2019). In addition, common, current practices used to define stimulus-induced SCR latencies rely on early empirical work characterising SCR response patterns that today may be seen as outdated due to recent technological advances allowing for more precise data acquisition and temporal resolution (Sjouwerman & Lonsdorf, 2019). Adding a further contribution to recent aims of improving the robustness of conditioning research, Sjouwerman and Lonsdorf (2019) provided up-to-date recommendations for SCR quantification. In this study, they examined the temporal trajectory of SCR responses across different modalities and the modulating role of additional factors. They demonstrated that SCR latencies are modulated by cognitive factors such as CS-UCS contingency awareness as well as by individual characterstics such as sex but not personality traits. Consistent with earlier recommendations, they showed that the typical latency of the SCR is best captured between 1 – 4 s post stimulus onset, however response latencies were found to vary according to the stimulus modality. For example, tactile stimuli were found to have the shortest latencies while visual stimuli elicited the largest latencies and audotiry stimuli exhibited mid-range latencies. Based on these data, the authors proposed a refined set of modality-specific guidelines for SCR quantification aimed at increasing analytical sensitivity.

1.6.4 Analytical tools

The choice of analytical tools in the study of classical conditioning, specifically in relation to psychophysiological outcome measures is as heterogeneous as that of methodological and procedural aspects, without the availability of a universally accepted approach. Currently, a wide range of procedures are employed to reduce data from physiological measures such as peak scoring and area under the curve in pre-defined time windows (Korn et al., 2017). These methods have several disadvantages (see Chapter 4 for a discussion) and finding an optimal balance between sensitivity and specificity of the time window of interest can be extremely difficult (Bach et al., 2018). Consequently, in parallel with studies by the *EIFEL-ROF* network, several investigations have focused on improving estimates of psychophysiological outcome measures. For example, the Dominik Bach’s laboratory has offered an alternative to conventional ways of making inferences about unobservable psychological constructs (e.g., threat anticipation) from measurable physiological responses (e.g. SCR), (Bach et al., 2018). The approach relies on psychophysiological modelling (PSPM) to estimate the values of psychological constructs (e.g. anticipation) given the observed physiological signal, while also providing a goodness-of-fit measure through estimating how well a psychological construct can be predicted from a given physiological measure (i.e., retroactive validity analysis), (Bach & Melinscak, 2020). The PSPM approach has been applied to a number of physiological responses including SCR, pupil size, and heart rate and has been shown to often outperform conventional methods for psychophysiological analysis (Korn et al., 2017; Ojala & Bach, 2020). Nonetheless, the PSPM method is still relatively novel and as stated by the authors, requires further investigations and across a wider range of experimental settings (Ojala & Bach, 2020).

In contrast, Ney at al. (2018) offered a different set of recommendations that may be more suitable to researchers with less extensive mathematical background that also aims at increasing replicability and improving inferences of physiological data derived from conditioning research. In their review, Ney at al. (2018) identified a number of major issues in the analysis of psychophysiological data, including lack of power, researcher degrees of freedom in relation to analytical choices, post-hoc selection of analytical tools and removal of data, lack of transparency in reporting, and poor estimation of individual variability due to data reduction. The proposed solution to these problems was a transition towards the use of analytical tools such as time-series analysis, predictive and multi-level modelling accompanied by more liberal multiple comparisons correction techniques that boost power (e.g., FDR). Furthermore, it was suggested that increasing the number of trials may increase the reliability and accuracy of findings while transparent data reporting, including making all data available was argued to provide a solution to the problem of arbitrary data exclusion.

1.6.5 Translational research

Classical conditioning has long been used as a translational tool for bridging the gap in our understanding of anxiety disorders, by allowing the translation of underlying mechanisms observed in animals to more complex processes in humans (Haaker et al., 2019). Cross-species differences in methodology, however, can introduce significant problems when comparing animal and human data. Consequently, the methodological review by Haaker et al. (2019) provides a comprehensive account of factors complicating cross-species comparisons, that require consideration when drawing conclusions from the existing literature, or when designing new experiments aimed at measuring comparable cross-species processes. These include procedural and paradigm differences, variation in the outcome measures employed as well as challenges associated with the cross-species translation of individual differences.

Regarding procedural variation, several factors that have the potential of triggering slightly different or additional underlying processes were considered. One of the most notable differences identified was that of the control conditions used to dissociate associative and non-associative processes. In human research, control conditions are established using within-subject differential protocols in which one CS is paired with the UCS while another is not. In contrast, animal studies rely either on single-cue between-subject protocols or when differential procedures are employed, presentation of the CSs is conducted on different days. This is a crucial difference since unlike single-cue protocols, differential procedures in which a CS- is presented, deploy both threat and safety learning processes (Haaker et al., 2019). It was also suggested that attention should be paid in respect to the UCS. While electric shocks have been shown to produce comparable sensory effects, it was highlighted that the potential effects of certain procedural differences such as UCS intensity should not be ignored. As a result, Haaker et al. (2019) recommended that due to the uncommon use of high intensity UCS in humans, cross-species comparisons should only be made based on animal studies using moderate UCS intensity.

Another major difference suggested to pose a translational challenge is that procedural instructions are only administered in human studies. The presence of instructions subsequently changes the underlying process indexed by the CR, with instructed and uninstructed protocols reflecting fear expression and learning respectively. Therefore, it was recommended that cross-species comparisons should be limited to human studies that used a minimal amount of instruction. In terms of measurement, Haaker et al. (2019) argued for a careful consideration of the degree to which different outcome measures used in animal and human studies tap onto similar learning mechanisms, especially since different indices of learning in humans alone do not necessarily converge.

A final point discussed by Haaker et al. (2019) was that of the translational challenges of individual differences research and of the associated cross-species differences in the factors that typically influence the magnitude of CRs. For example, they raised the importance of considering sex sample variation when comparing studies, specifically in relation to cross-species differences in the temporal dynamics of sex hormone concentrations and the tendency for animal studies to report data based on male animals and for human data to be based on mixed-sex samples. Recommendations of additional factors to be considered also included the difficulty in mapping developmental differences between species, ethical constraints in relation to the study of acute stressors, and the significant difficulty in translating human personality traits to animal species. In highlighting the practical and methodological constraints of translational research while also providing potential solutions for maximising cross-species comparability, this extensive review offers an important perspective of translational research that is rarely considered.

1.6.6 Outstanding issues

One design feature that has received considerably less attention is that of trial number. The majority of conditioning tasks rely on a relatively small number of items and trials per condition to index learning and extinction, in order to prevent habituation of response measures or to the UCS (Lonsdorf et al., 2017). This design constraint, however, can pose significant limitations on the mechanisms that can be investigated and the tools that can be used to study conditioning in humans. For example, the measurement of psychophysiological signal is often accompanied by high level of noise elicited through a range of environmental and random factors (Ney et al., 2018). Consequently, increasing the number of trials per condition can offer potential means for improving estimations in these methods. The extent to which a large number of trials is required for dealing with measurement noise is not uniform across measures and is also dependent on the analysis of choice. Certain techniques such as electro- and magnetoencephalography and the demands of certain analytical tools such as time-frequency analysis or the detection of deep brain sources require a large number of trials as a prerequisite for establishing a sufficient signal-to-noise ratio (Quraan et al., 2011a; Steinberg et al., 2013; Tzovara et al., 2019). In contrast, for other psychophysiological measures the use of a great amount of trials is not a standard practice but it has been suggested as a possible strategy for increasing reliability and accuracy (Ney et al., 2018).

In the context of conditioning, there are two approaches that can be taken to achieve a high number of trials - increasing the number of repetitions of a single

CS+ and CS- items or increasing the number of unique items per condition. The risks associated with greater repetitions is habituation of the CR or to the UCS, although studies have successfully implemented this type of design previously in the context of E/MEG (Dolan et al., 2006; Kluge et al., 2011; Moses et al., 2005, 2007; Tesche et al., 2007; Tzovara et al., 2019). In the latter case, the main risk with introducing a great variation of CSs is hampering with the development of contingency awareness which may prevent CR acquisition or reduce its magnitude (Lonsdorf et al., 2017). Nonetheless, previous research in E/MEG has shown that employing such multi-cs conditioning tasks allow for the detection of conditioned responding in the absence of awareness (Junghöfer et al., 2017; Rehbein et al., 2014, 2015; Steinberg et al., 2013). The main benefit of such an approach is that it can offer potential means for improving the generalisability of inferences as it allows for the implementation of analytical tools that simultaneously model random variability not only among participants but also across items. In contrast to traditional conditioning protocols, both of these approaches have so far been used in a limited number of experimental contexts and have provided variable support for their utility.

1.7 Aims of the thesis

Since the development of the classical conditioning paradigm in the 1920s, an extensive body of research has examined the physiological, behavioural, neural and neurobiological mechanisms underlying a range of associative learning-related processes. This has provided an enormous contribution to our ongoing understanding of how humans perceive and regulate threat, of the potential causes for the development and maintenance of anxiety disorders, and of their treatment. The growing amount of methodological work in the field has also highlighted various procedural and analytical aspects that may constrain the inferences we draw about learning and extinction, while also offering means for improving the rigour and replicability of conditioning research.

The aim of this thesis is to contribute to recent methodological efforts and theoretical debates in the field of conditioning in several ways. First, the thesis examines the utility of three potential approaches for increasing trial number in conditioning tasks that may facilitate future studies utilising noisy psychophysiological measures. Within these investigations, this work also attempts to add a contribution to two ongoing debates in the literature related to the role of contingency awareness in conditioning, and the feasibility in detecting deep structure activity using MEG and a conditioning task with a large number of trials. Finally, throughout the different chapters, this thesis also takes into account recent recommendations for the use of analytical strategies in conditioning, aimed at increasing transparency and replicability (Lonsdorf et al., 2019; Lonsdorf & Merz, 2017; Ney et al., 2018) as well as for improving generalisability of findings in psychological research in general (Barr et al., 2013; Yarkoni, 2020).

Specifically, Chapter 2 will examine an already established multi-trial paradigm (i.e., Multi-CS Conditioning, see Steinberg et al., 2013) that relies on acquiring a large number of trials through the use of a high number of unique items per condition. In this chapter, we provide data from an indirect behavioural replication of the Multi-CS conditioning task that has previously provided inconsistent evidence for its ability to elicit subjective behavioural CRs (i.e., valence and arousal ratings), despite detecting differential neural activation (Bröckelmann et al., 2011, 2013; Rehbein et al., 2014; Steinberg et al., 2012).

We discuss our findings in the context of the ongoing debate about the role of contingency awareness in associative learning, since the use of a great number of items in this task prevents individuals from learning the relationship between the CSs and the UCS. In addition, we employed statistical tools designed to improve generalisability of inferences. Through a re-analysis of a published Multi-CS dataset (Rehbein et al., 2014), we also demonstrate how the use of transparent data visualisation strategies may facilitate our understanding of the potential factors that may explain our results as well as the inconsistency in previous findings. Finally, we offer a set of recommendations that may improve inferences in future Multi-CS conditioning research, including increasing clarity when defining constructs and their measurement, encouraging the use of generalisable statistical tools that model by-item and by-subject random variability when using a large number of items, and of transparent data visualisation, and improving the outcome measures used to assess valence and arousal.

Chapter 3 focuses on a novel *visual blocked conditioning task*. The aim was to utilise a high number of trials while at the same time establishing a balance between risk of habituation (due to item repetition) and poor contingency awareness (due to many unique items). This task was tested in the context of MEG, as this is one of the approaches for which the implications of multi-trial conditioning are potentially considered to be the greatest. In particular, the study detailed in this chapter investigates a niche in the field that suffers from significant constraints both technically and from a design perspective. One of those constraints is linked to recent efforts in attempting to detect deep structure brain sources (e.g. the amygdala) using MEG (Dumas et al., 2011; Quraan et al., 2011b; Tzovara et al., 2019). Additionally, however, the human conditioning literature so far has struggled to provide a comprehensive account of the role of neural oscillations, specifically at the theta range in threat learning and extinction. Such investigations have potentially been limited by experimental design constraints as in order to understand the temporal dynamics of oscillatory activity and analyse effects at the time and frequency domain simultaneously, a large number of trials is typically required. As such, this chapter attempts to examine the feasibility in studying the role of theta oscillations in cortical and subcortical structures in humans through maximising the SNR with a high number of trials and employing analytical techniques that have previously been shown to detect deep structure activity. In addition, we assessed conditioned responding in several other outcome variables, including subjective valence and arousal ratings, as well as pupil size.

The findings reported in Chapter 3 provided poor evidence for the reliable detection of conditioned responding in any of the investigated outcome measures and thus, limited the desired quantification of the oscillatory mechanisms driving learning and extinction. However, the results from this study provided a degree of confidence that the blocked nature of the task was not the primary cause for the poor CR elicitation and highlighted a number of potential design factors that may increase the likelihood of detecting a CR in future work employing a blocked conditioning task. These included aspects such as trial duration, the use of less complex CSs, and the importance of maintaining participants’ attention.

In light of these findings, the study detailed in Chapter 4 aimed to re-design the blocked conditioning task in the auditory domain. In part due to technical and time constraints, the work moved away from MEG measurement and towards discussing the utility of multi-trial paradigms in psychophysiological measures in general, with a particular focus on pupil size, while also measuring valence and arousal conditioned responding. The study also discusses the potential of this task in combination with suitable analytical techniques for improving inferences, by expanding on the issue of generalisability discussed in earlier chapters. In particular, the analyses of this study avoided potential issues arising from arbitrary time window selection by utilising robust data-driven time window selection methods. Furthermore, statistical inferences were based on design-appropriate mixed effects models that considered by-subject and by-item variability of effects simultaneously, thus ensuring optimum generalisability of results. While this approach is less sophisticated than the PSPM method discussed earlier in this introductory chapter, it can offer certain benefits in terms of analytical consistency across measures, ease of use, flexibility in relation to data pre-processing and suitability for users with less extensive mathematical background.

Finally, chapter 5 provides a summary of the findings from the experimental chapters with a discussion of their implications for future work in multi-trial conditioning research.

2 Chapter 2 - Can associative learning without awareness be elicited using Multi-CS conditioning? An indirect behavioural paradigm replication.

2.1 Introduction

A widely debated issue in the conditioning literature is whether associative learning processes and related conditioned responding are dependent on the conscious awareness of the CS-UCS contingency. Theoretically, this debate has been informed by the single and dual process accounts of learning. The dual-process framework suggests that implicit (affective) and explicit (expectancy) learning are dissociated and can occur independently of one another, and as such, this model proposes that conditioning can occur in the absence of CS-UCS contingency awareness (Schultz & Helmstetter, 2010). In contrast, the single-process account suggests that affective and expectancy learning are driven by the same underlying mechanism and consequently, experimental manipulations should elicit qualitatively similar implicit and explicit responses (Lipp & Purkis, 2005). Indeed, this model assumes that contingency awareness is necessary for establishing a CR (Lovibond & Shanks, 2002).

Early evidence and a commonly used example to support the dual-process model originates from lesion studies showing that hippocampal damage disrupts the memory of the CS-UCS relationship but not autonomic responses such as the SCR. On the other hand, amygdala damage prevents the acquisition of an autonomic CR in the presence of contingency awareness (Schultz & Helmstetter, 2010 but see Bechara et al., 1995). Later empirical work has focused on examining the effects of awareness using a masking procedure on the CSs or a distraction task that interfere with the development of contingency awareness (Mertens & Engelhard, 2020; Schultz & Helmstetter, 2010). Such studies often report the acquisition of CRs in the absence of explicit awareness, at least in some outcome measures and conditioning protocols (Schultz & Helmstetter, 2010; Sevenster et al., 2014; Weike et al., 2006), although findings to the contrary and in support of the single-process model are also available (Dawson et al., 2007; Lipp & Purkis, 2005; Weidemann et al., 2016).

Some insight into the factors that may drive the mixed evidence in support of the dual-process account come from a recent systematic review and meta-analysis of unaware conditioning in physiological outcome measures (Mertens & Engelhard, 2020). The review found that the majority of findings in the literature were derived from tasks that suffer from significant methodological limitations, where the likelihood of observing contingency-unaware conditioning decreases with greater methodological quality (Mertens & Engelhard, 2020). Specifically, it was shown that poor masking procedures and contingency awareness measures, publication bias, and researcher degrees of freedom are commonly present across studies, limiting the conclusions that can be drawn from them (Mertens & Engelhard, 2020).

In addition, trial-order effects were found to be largely uncontrolled for in studies reporting unaware CRs. (Mertens & Engelhard, 2020). A common practice in conditioning studies is to use a pseudo-random presentation order with the common restriction of presenting the same stimulus no more than twice in succession. Such a procedure ensures that the CS+-UCS pairings are not presented in multiple successive trials. This approach, however, can facilitate expectancy especially when events appear to occur randomly, leading individuals to seek predictability (Singh et al., 2013). As a result, even if participants are unaware of the CSs and their relationship with the UCS, the mere presentation of one stimulus type (e.g. a CS-) increases the probability that the next trial will contain the other stimulus type (e.g. a CS+), which subsequently affects UCS expectancy (Mertens & Engelhard, 2020). This expectancy, driven by trial-order contingency is shown to induce a CR without awareness of the CS-UCS relationship. (Singh et al., 2013; Wiens et al., 2003), although some evidence that not all outcome measures are susceptible to such effects has also been reported (Sevenster et al., 2014).

Disentangling the role of behaviour in unaware conditioning has also been challenging. From a theoretical standpoint, the dual-process model of learning suggests that subjective feelings about a stimulus are directly driven by conscious, cognitive processes, but that affective responses can also arise automatically outside of conscious awareness when non-consciously acquired associative memories trigger defensive responses such as physiological arousal. These defensive responses are suggested to indirectly trigger subjective experiences (LeDoux & Pine, 2016). Therefore, a subjective experience of threat should occur regardless of whether a threat-related associative memory is acquired consciously or non-consciously. For example, a phobia of dogs would result in a subjectively perceived threat and fear of dogs irrespective of whether the person has experienced a conscious negative event with a dog (e.g. being bitten by a dog as a child) or an non-conscious association between a dog and the threat of being bitten. By this logic, explicit online and offline subjective behavioural measures of conditioning should still elicit a differential CR.

However, similar to findings in physiological measures, the evidence supporting unaware conditioning in offline behavioural outcome measures such as valence and arousal judgements is mixed. Studies examining the behavioural effects of standard differential conditioning paradigms using masking or attention-distracting procedures to prevent awareness are limited, potentially because it is suggested that online physiological measures can be more sensitive in detecting unaware conditioning than offline subjective behavioural measures (Corneille & Mertens, 2020). Results from such studies have shown that while unaware participants exhibit differential activity in the fear network in the brain (e.g. the amygdala, hippocampus, orbitofrontal and anterior cingulate cortices), they show no conditioned responding in offline valence and arousal ratings or SCR (Klucken et al., 2009; Tabbert et al., 2011). These findings contradict the dual-process account suggesting that due to their automatic and defensive nature, physiological responses should occur in the absence of awareness and trigger subjective experiences of threat (LeDoux & Pine, 2016). Research in evaluative conditioning (i.e., changes in the likeness of CSs as a result of pairing them with affective UCSs) has shown some support for unaware conditioning (Hütter et al., 2012), although evidence to the contrary is also available (Kattner, 2012). In addition, it has been demonstrated that contingency awareness is crucial for evaluative conditioning to occur, regardless of whether implicit or explicit measures are used (Pleyers et al., 2007). A detailed review of the evaluative conditioning literature also concluded that there is little empirical support for eliciting changes in subjective evaluative responses in the absence of awareness, and that findings supporting the dual-process account of learning are limited to specific experimental procedures (Corneille & Stahl, 2019).

This large body of conflicting findings across physiological and behavioural measures reinforces recent observations from reviews of the available evidence (Corneille & Stahl, 2019; Mertens & Engelhard, 2020), suggesting that methodological and procedural issues perpetuate the unresolved debate regarding the dissociative nature of conscious and non-conscious learning processes. Methodological problems, however, are not specific to this debate, but an ongoing problem identified in threat conditioning research in general, which has received a considerable amount of attention in recent years. Specifically, efforts have been placed in improving methodological practices that can maximise replicability and reliable inferences, such as reducing researcher degrees of freedom through pre-registrations, improving research and reporting transparency and eliminating excessive data analysis flexibility (Lonsdorf et al., 2017, 2019; Ney et al., 2018). These lie at the core of improving inferences drawn from unaware conditioning as well, combined with the necessity for systematically investigating the conditions and outcome measures under which such an effect may be observed.

Consequently, the present chapter aims to contribute to the ongoing dual-process model debate and recent efforts in improving inferences in threat learning. In this chapter, we provide an indirect behavioural replication of an established multi-trial paradigm (i.e., Multi-CS conditioning), proposed to elicit conditioned responding in the absence of contingency awareness (Steinberg et al., 2013). Unlike more traditional methods for studying unaware conditioning in which masking, secondary attention-distracting or cognitively demanding tasks are used to interfere with the perception of the CSs, in Multi-CS conditioning a large number of unmasked CS+ and CS- stimuli are used to elicit threat learning. In this task, each stimulus is presented only a limited number of times and for a relatively short duration (20-800 ms). This creates a highly demanding environment in which multiple complex and perceptually similar stimuli have to be differentiated, which prevents the acquisition of contingency awareness

(Steinberg et al., 2013). The use of a large number of CSs that are repeated a very limited number of times was implemented to allow for the investigation of rapid cortical activation in noisy electrophysiological measures which require a large number of trials to achieve an acceptable signal-to-noise ratio, while also reducing the likelihood of observing CR habituation due to high number of stimulus repetitions (Steinberg et al., 2013).

Findings from Multi-CS conditioning studies have indeed shown some fast differential activation patterns in response to visual and auditory CSs paired with a range of UCSs (e.g. electric shock, unpleasant sounds, white noise), in pre-frontal and sensory regions following as well as during conditioning (Bröckelmann et al., 2011, 2013; Rehbein et al., 2014, 2015; Steinberg et al., 2013). However, there are concerns about the degree to which the findings from some of these investigations (Bröckelmann et al., 2011, 2013; Steinberg et al., 2012, 2013) indeed reflect the acquisition of a CR. Specifically, the results from these studies are compromised by a technical discrepancy with standard conditioning research in that they failed to examine the typical signatures of conditioned responding, i.e., the differences between CS+ and CS- conditions during learning. Instead, statistical comparisons were based on neural activity during non-reinforced presentations typically used to study the process of extinction. A re-analysis (Rehbein et al., 2014) of one of the datasets in the study by Steinberg et al. (2013), however, argued against the likelihood of this activity reflecting extinction, as they observed similar pre-frontal activity during (reinforced) and after conditioning (non-reinforced). Nonetheless, these results may not be sufficient to support this claim, as recent meta-analyses examining the networks involved in threat learning and extinction have demonstrated that pre-frontal regions are engaged in both processes (Fullana et al., 2016, 2018).

The issue of what underlying process is measured by this task is even more prominent with respect to the derived behavioural CR indices in all of the Multi-CS conditioning studies. Since offline rating data were not collected immediately after conditioning, statistical comparisons were only made between ratings obtained following extinction and prior to habituation. Therefore, these subjective ratings of valence and arousal in fact reflect the extinction of CRs rather than their acquisition. Considering that some conditioning studies report resistance to extinction in subjective evaluative judgements (Gawronski & Mitchell, 2014; Sehlmeyer et al., 2011; Winn et al., 2018), it is likely that any reported significant subjective behaviour effects from these investigations are linked to a CR resistant to extinction. Furthermore, the derived behavioural results from these comparisons are not as consistent as the reported neural findings. For example, for auditory CSs presented for 20 ms, valence CRs were observable only in an implicit affective priming task (Bröckelmann et al., 2013) but not in explicit subjective measures (Bröckelmann et al., 2011; Junghöfer et al., 2015b). For neutral face CSs, Rehbein et al. (2014) reported a change in subjective valence for CS- trials from habituation to post-conditioning (extinction) only (although this effect was only approaching significance), and no subjective arousal or implicit behavioural effects. Steinberg et al.(2012) showed only a valence but not an arousal effect while Roesmann et al. (2020) demonstrated both valence and arousal threat generalisation effects after conditioning (extinction). In addition, the study by Rehbein et al. (2015) showed that conditioning effects may partly depend on individual's perceived contingency rather than the actual contingency. They found that CS trials that were perceived as CS- were rated as more pleasant after extinction than habituation, while both perceived and actual CS- trials were rated as less arousing after extinction than habituation. Such inconsistency in behavioural effects raises concerns about the inferences and interpretations that can be drawn about neural activity in isolation, and in the absence of a behavioural or physiological measure that offers some degree of corroboration, confirming the likelihood that this activity truly reflects unaware conditioning.

The discrepancy between neural and behavioural measures cannot be explained by some of the above-mentioned methodological problems such as poor masking or procedural differences, since findings from these studies are based on tasks with very similar experimental procedures involving unmasked stimuli. Trial order effects are also unlikely to explain the pattern of results since some of the studies reported findings based on neural measurements that were derived from non-reinforced CS presentations during/after extinction (Bröckelmann et al., 2011, 2013; Steinberg et al., 2012, 2013). Trial sequencing cannot drive the behavioural effects from Multi-CS conditioning studies either, since these were based on offline ratings collected following extinction which would not be susceptible to trial order effects. Since these experimental parameters cannot explain the inconsistency in behavioural findings, other factors must be mediating the likelihood of observing a conditioning effect from this task.

For example, analytical decisions may be equally important and can have dramatic impact on the inferences we make and their reliability and generalisability. The reported behavioural findings from Multi-CS conditioning research were derived using repeated measures ANOVA (rm ANOVA) tests on data averaged at the participant level. Similarly, the results regarding the reported differential neural activity were also based on rm ANOVAs on aggregated data. This approach is shown to inflate test statistics and lead to anticonservative inferences (Judd et al., 2012; Westfall et al., 2017; Yarkoni, 2020) as it only considers by-subject variability, but fails to account for by-item variability (or conflates the latter with trial-level measurement error, respectively). Assuming that all items would be affected by the experimental manipulation to an equal degree can be problematic, even more so when eliciting effects using a large number of unique, complex, highly dimensional stimuli such as faces, that show high degree of heterogeneity both within and between individuals (Jenkins et al., 2011). While the results reported from Multi-CS conditioning with faces (Rehbein et al., 2014, 2015; Roesmann et al., 2020) relied on the same face databases, it is unclear whether the same visual stimuli were used across experiments. Even if the same stimuli were employed (e.g. Bröckelmann et al., 2011, 2013), it is possible that by-subject and by-item interactions can cause varying sensitivity to the experimental conditions, which can lead to poor effect replicability between studies. Therefore, to reduce the risk of Type I error and ensure that results not only generalise to a new population of participants but also to new sets of stimuli of the same type, it is important to simultaneously consider both by-subject and by-item random variability and their relationships within the experimental manipulation (Barr et al., 2013; Judd et al., 2017).

A related problem is the common adoption of data visualisation of condition means using tools such as bar or line graphs, that accompany the statistical results. Since this approach does not represent the raw data underlying the presented measure of central tendency, it can lead to gross over or under-estimation of the magnitude of effects. Summarising data in this manner can also obscure important patterns in the data, differences in the distributions and in the case of within-subjects designs, it fails to provide information regarding the consistency of patterns across individuals (Rousselet et al., 2017; Weissgerber et al., 2015). In the context of Multi-CS conditioning, we cannot be certain whether the variability in effects is driven by for example, a subset of individuals or if the analysis of means adequately captures the underlying distribution. To increase transparency and improve inferences, recent efforts have focused on improving the use of robust data visualisation tools that can reveal important information about individual differences, outliers and other interesting or unexpected data patterns (Allen et al., 2018). For example, the use of rainclouds can be extremely informative as along with central tendency summary measures, they provide information about the underlying distribution as well as individual data points (Allen et al., 2018).

2.1.1 The present study

The indirect behavioural replication of the Multi-CS conditioning paradigm aimed to gain a better insight of the factors that drive the inconsistent behavioural findings derived from Multi-CS conditioning, with a particular focus on providing more generalisable inferences. We made only minimal and necessary modifications to the original task while closely following the design presented by Rehbein et al. (2015), specifically in relation to parameters such as CS and UCS inter-stimulus and inter-trial intervals. We used white noise as the UCS, presented at approximately 80-85 dB, which was slightly lower than that used by Rehbein et al. (2015), and with a 50 ms longer duration. In addition, since attention is shown to play an important role in facilitating and enabling CR acquisition (Field & Moore, 2005), and passive viewing may risk participants' attention drifting away, we included a few catch trials per block to maintain task engagement.

In order to disentangle the role of extinction processes in previous Multi-CS conditioning studies, we recorded ratings following habituation, acquisition and extinction phases, unlike previous studies (Rehbein et al., 2014, 2015) which measured valence and arousal before habituation (pre-learning) and after extinction (post-learning). Since the behavioural ratings are measured offline after each experimental phase, it appears safe to assume that any valence and arousal CRs are not affected by trial sequencing and, if present, will reflect a genuine change in the perception of the stimuli and not a simple UCS expectancy response.

The most significant deviation from the original task was in the response scale we used. The research employing Multi-CS conditioning has relied on the Self-Assessment-Manikin (SAM) scale (Bradley & Lang, 1994). The scale was used either in its original 9-point discrete format (Bröckelmann et al., 2011; Junghöfer et al., 2015b; Rehbein et al., 2015; Roesmann et al., 2020) or as a modified interval scale (Rehbein et al., 2014, ranging from -300 to 300), although it is unclear how a continuous transformation was applied to a discrete 9 point pictorial scale. While the SAM scale is still widely used in psychological research, it is difficult to implement in computerised tasks and it is relatively outdated, particularly in the graphics domain in which the pictorial affective states representations can today be perceived as unintuitive and ambiguous (Betella & Verschure, 2016). As a result, the present study used a standard 7-point Likert scale to measure subjective valence and arousal which we believe should be less ambiguous but qualitatively similar to the discrete SAM scales used previously, as both measure emotionality on an ordinal scale.

In addition, we used a different set of neutral faces serving as the CSs and employed design-appropriate mixed modelling that simultaneously accounts for by-item and by-subject random variability, combined with transparent data visualisation. We conducted two sets of analyses. Our main analysis focused on modelling changes in valence and arousal by first accounting for potential baseline differences during habituation, by subtracting habituation ratings from acquisition and extinction ratings. Since this baselining procedure distorts the ordinal nature of the data, we modelled the data using linear mixed effects (LME) models. In our secondary analyses, we aimed to corroborate our main results by considering the three-phased data in its original ordinal format and applying cumulative-link mixed models which take into account the ordinal nature of the data. We also conducted several secondary analyses. First, similar to Rehbein et al. (2015), we examined the impact of perceived contingency on conditioned responding. In addition, we re-analysed one of the Multi-CS conditioning datasets (Rehbein et al., 2014) to demonstrate how the use of transparent data visualisation can improve and guide our understanding of experimental effects.

2.2 Methods

2.2.1 Participants

Twenty-three participants (16 females) aged 19-29 (mean = 21.3, SD = 2.6) took part in the study. Three participants were excluded from the contingency awareness task as they gave the same response on all trials, however, all 23 participants were included in the valence and arousal analyses. Participants were recruited by undergraduate students at the Psychology Department, as part of a group research project, and received £6 per hour for their participation. The study was approved by the College of Science and Engineering ethics committee (300170261). Written informed consent was obtained from each participant.

2.2.2 Stimuli

Conditioned stimuli (CS) were 104 neutral, frontal view faces (52 females) of White background, obtained from the Chicago Face Database (Ma, Correll & Wiitenbrink, 2015). Stimuli from the database are normed both in terms of physical and subjective properties of each facial identity (see Ma et al., 2015). The stimulus selection process ensured that faces were similar in luminance or levels of attractiveness, trustworthiness, and emotionality. Stimuli were colour photographs scaled to 510 x 510 pixels. For each subject, 52 of the faces were randomly selected and assigned as CS+ while the remaining 52 served as CS-.

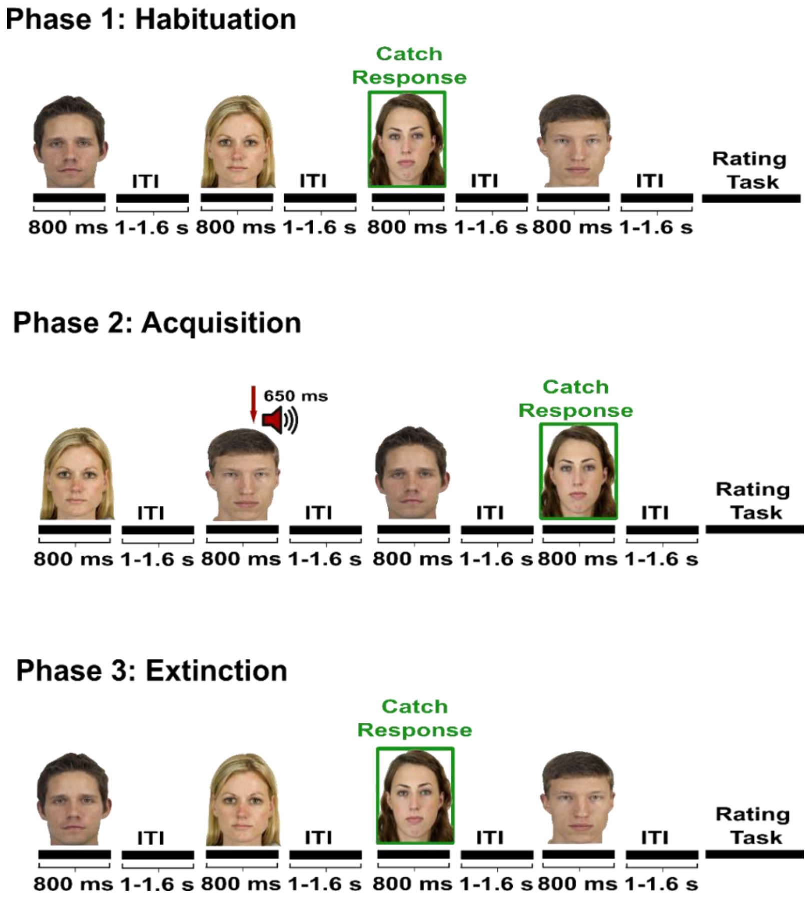
A 150 ms white noise of approximately 80-85 dB was used as the Unconditioned Stimulus (UCS), since it is shown to produce a stronger and more reliable affective learning and extinction in experimental designs containing high number of trials (Sperl et al., 2016). This was also the UCS type used by Rehbein et al. (2015).

2.2.3 Procedure

The task contained three experimental phases – Habituation, Acquisition, and Extinction (see Figure 1). During each phase, a total of 156 CS+ and 156 CS- trials were presented in three separate blocks (52 CS+ and 52 CS- per block). Each CS was presented together with a black fixation cross, positioned at the centre of the image (the nose), for 800 ms. The inter-trial interval (ITI) had a duration of 1300 ms ±300 ms and was accompanied by a black fixation cross in the centre of the screen. Participants were asked to maintain fixation at the centre of the screen at all times. Trial order was randomised across participants with the restrictions that the first trial was always a CS- and no more than three trials of the same stimulus type (e.g., a CS+) could occur consecutively. To maintain subjects' attention on the task, 2-3 trials were randomly selected in each block and presented twice in succession at a random time point in each block. Participants were instructed that they will be presented with a series of faces that they have to view while maintaining fixation at the centre of the screen and respond using a button press when the same face occurs twice in a row. As such, participants were not informed of the CS-UCS contingency. During the Acquisition phase, CS+ trials were paired with the UCS, which occurred 650 ms post CS+ onset, while CS- trials were never paired with the UCS. During Habituation and Extinction, the CSs were presented alone, without the UCS. At the end of each phase, participants completed a face rating task where they were asked to rate each CS on valence and arousal using a 7-point Likert scale (1 not at all pleasant/arousing to 7 extremely pleasant/arousing). The task scripts can be found at https://osf.io/c6uhy/.

At the end of the experiment, participants completed a surprise contingency awareness task, consisting of 24 trials. For each subject, 12 CS+ (6 female faces) and 12 CS- (6 female faces) stimuli were randomly selected from the total number of stimuli. On each trial, a fixation cross was presented for 600 ms followed by the CS, which was presented for 800 ms. Trial order was randomised across participants. On each trial, participants were asked to indicate whether the face they were presented with was paired with a sound during Phase 2 of the experiment. In addition, participants were asked to indicate how confident they were in their judgement using a 7-point Likert scale (1 not at all confident to 7 extremely confident).

Figure 1
Paradigm and trial structure.


Phase 1: Habituation
Catch Response
ITI
ITI
ITI
ITI
Rating Task
800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s
Phase 2: Acquisition
650 ms
Catch Response
ITI
ITI
ITI
ITI
Rating Task
800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s
Phase 3: Extinction
Catch Response
ITI
ITI
ITI
ITI
Rating Task
800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s 800 ms 1-1.6 s

Note. Permission to re-use the sample of images from the CFD (Ma et al., 2015) was obtained from the copyright holder, the University of Chicago, Center for Decision Research.

2.3 Results

All analyses were performed in R. Analysis scripts are available at https://osf.io/c6uhy/. Data are only available upon request, due to a section of the participant's information sheet, restricting public data sharing.

2.3.1 Valence and Arousal Ratings

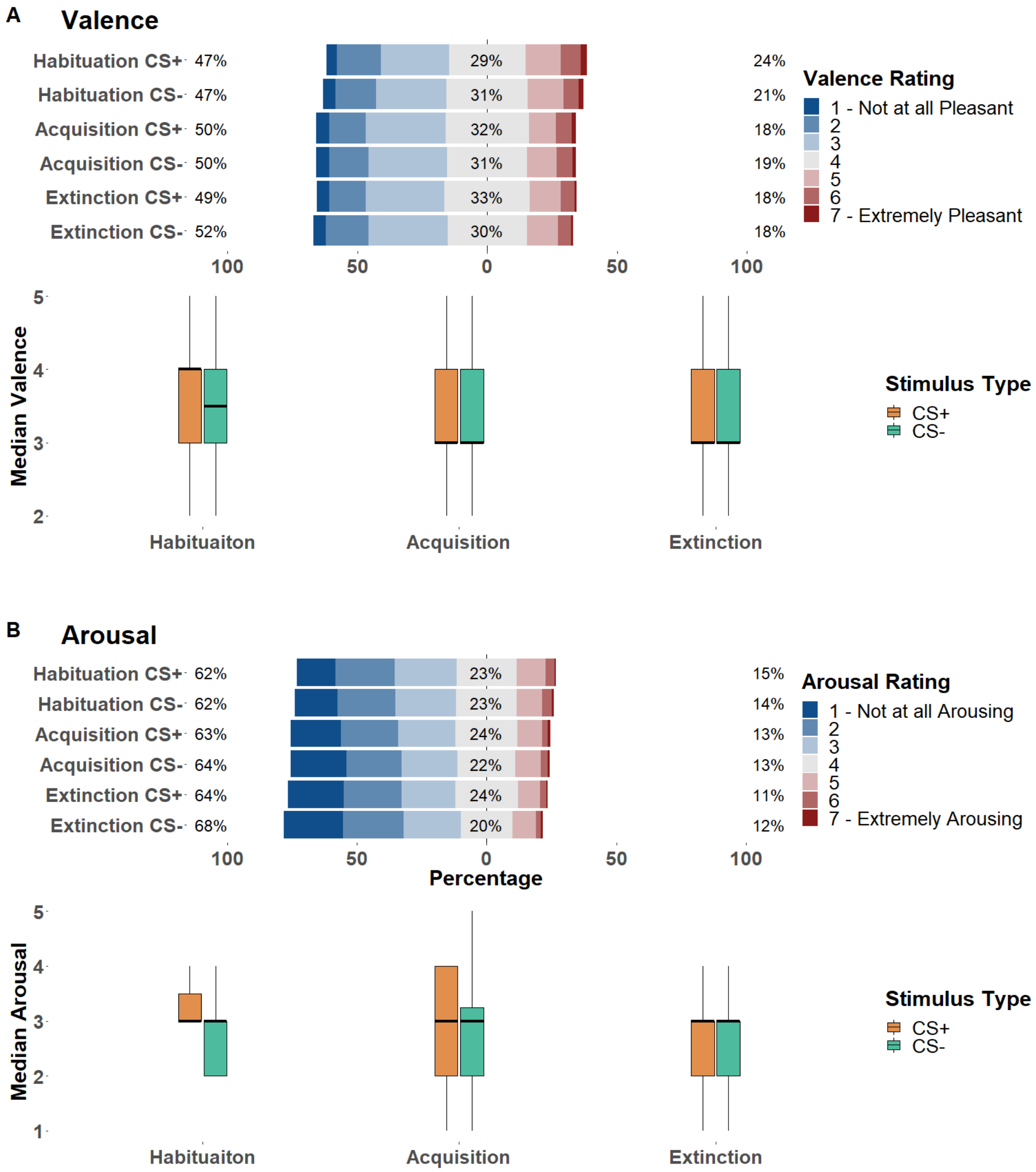
Figure 2 shows the percentage of responses belonging to each Likert scale point (1-7) as well as the median score across the experimental phases and stimulus type. As seen in Figure 2, for valence ratings, around 30% of responses had a rating of 4 and 50% of responses had a rating below 4. During Habituation and regardless of Stimulus Type, valence ratings were 2-6% higher than those during Acquisition and Extinction. For arousal ratings, approximately 20% of responses had a rating of 4 and ~60 % had ratings below 4, irrespective of Stimulus Type or Experimental Phase. Ratings above 4 in response to both CS+ and CS- stimuli were also 1- 4% more prevalent during Habituation than during Acquisition and Extinction. This descriptive summary suggests that overall, stimuli were predominantly perceived as neutral in valence and low in arousal. This pattern was also reflected in the median ratings.

Prior to quantifying any potential differences between conditions, we baselined the data with respect to the Habituation phase to account for any potential baseline differences. The habituation-adjusted data was obtained by subtracting each item's rating during Habituation from the same item's rating during the Acquisition and Extinction. Next, we used linear-mixed effects (LME) modelling to predict valence and arousal ratings. For each rating type, the model consisted of a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS+ vs CS-) fixed effects design. We used mean-centred contrasts for the two categorical fixed effect predictors. The models included by-subject and by-item random intercepts, and by-subject and by-item random slopes for both main effects and the interaction (see Supplementary Materials 1 for random effects summary). We assessed main effects and interactions using type III Wald chi-square tests (see Table 1). The tests revealed no significant effects for valence ratings, and a significant main effect of *Experimental Phase* for arousal ratings. Note, however, that the p-value of this main effect is large ($p < 0.049$) and approaching 0.05. Post-hoc marginal mean contrasts (package *emmeans*) were performed after averaging over the levels of Stimulus Type. These showed that faces were perceived as more arousing during Acquisition than Extinction, however, this effect was not statistically significant ($p = 0.065$), (see Table 2). This is not surprising considering that the confidence intervals and distributions of ratings during Acquisition and Extinction overlap significantly (see Figure 4), and the strength of association (effect size) for all main effects and interactions is approaching 0 (see Table 2).

Since the non-baselined data were measured on an ordinal scale, we validated our LME findings in a secondary analysis on the three-phased dataset using cumulative link mixed (CLM) models (package *ordinal*), (see Supplementary Materials 2). The CLM results are comparable to the LME modelling findings.

Figure 2

Distribution of valence and arousal ratings across conditions



Note. Percentage of ratings (Top) belonging to each Likert point (1-7) and median rating (Bottom) for CS+ and CS- stimuli across the experimental phases for A) Valence and B) Arousal ratings.

Table 1

Type III Wald Chi-square tests *and R-squared values for the valence and arousal model and each of the fixed effects.*

	Chisq	Df	P-value	R^2 Fixed (CI)
Valence				
Full Model				0 (0 – 0.03)
Experimental Phase	0.78	1.000	0.37	0 (0 – 0.000)
Stimulus Type	0.03	1.000	0.85	0 (0 – 0.002)
Experimental Phase X Stimulus Type	0.39	1.000	0.53	0 (0 - 0.001)
Arousal				
Full Model				0.001 (0-0.004)
Experimental Phase	3.86	1.000	0.049	0.001 (0 – 0.004)
Stimulus Type	0.01	1.000	0.92	0 (0– 0.001)
Experimental Phase X Stimulus Type	0.16	1.000	0.68	0 (0 – 0.001)

Table 2

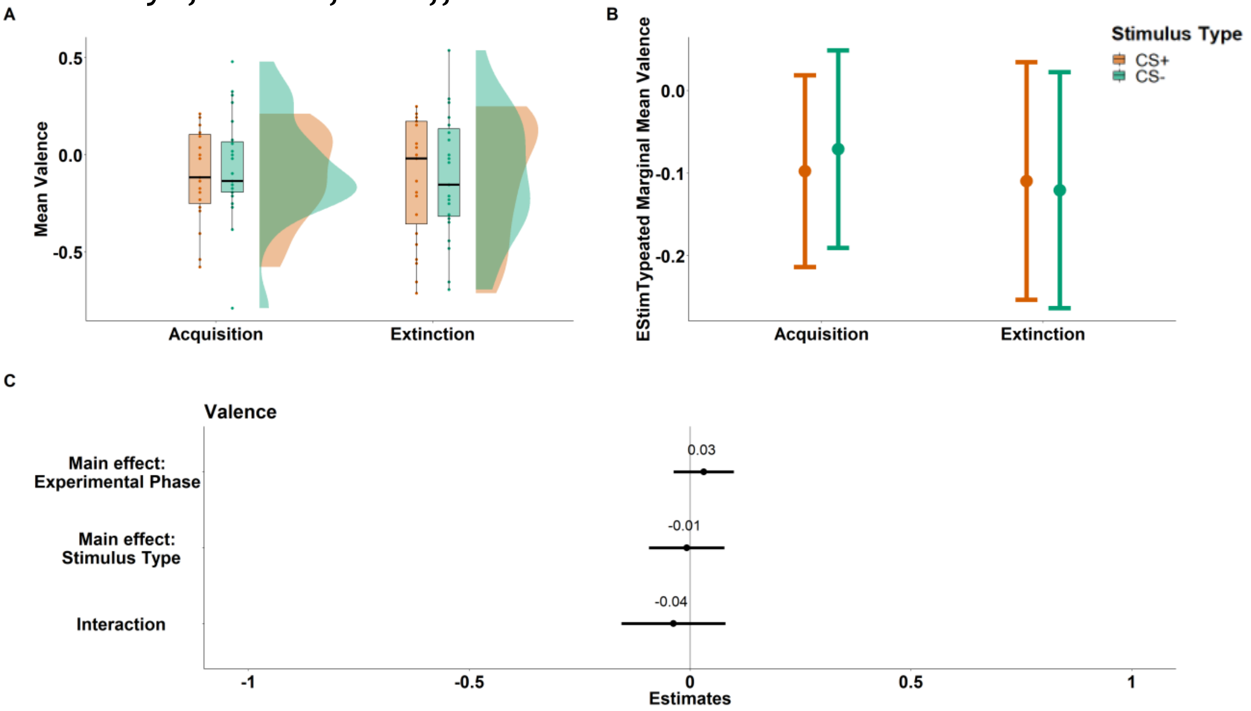
Estimated marginal means and related contrasts derived for the arousal model.

Experimental Phase Estimated Marginal Means						
Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	
Acquisition	0.13	0.10	23.79	-0.34	0.07	
Extinction	-0.22	0.09	23.64	-0.42	-0.02	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
Acquisition – Extinction	0.09	0.04	19.90	-0.01	0.18	0.06

Note. Contrasts were computed using Kenward-Roger method for degrees of freedom approximation.

Figure 3

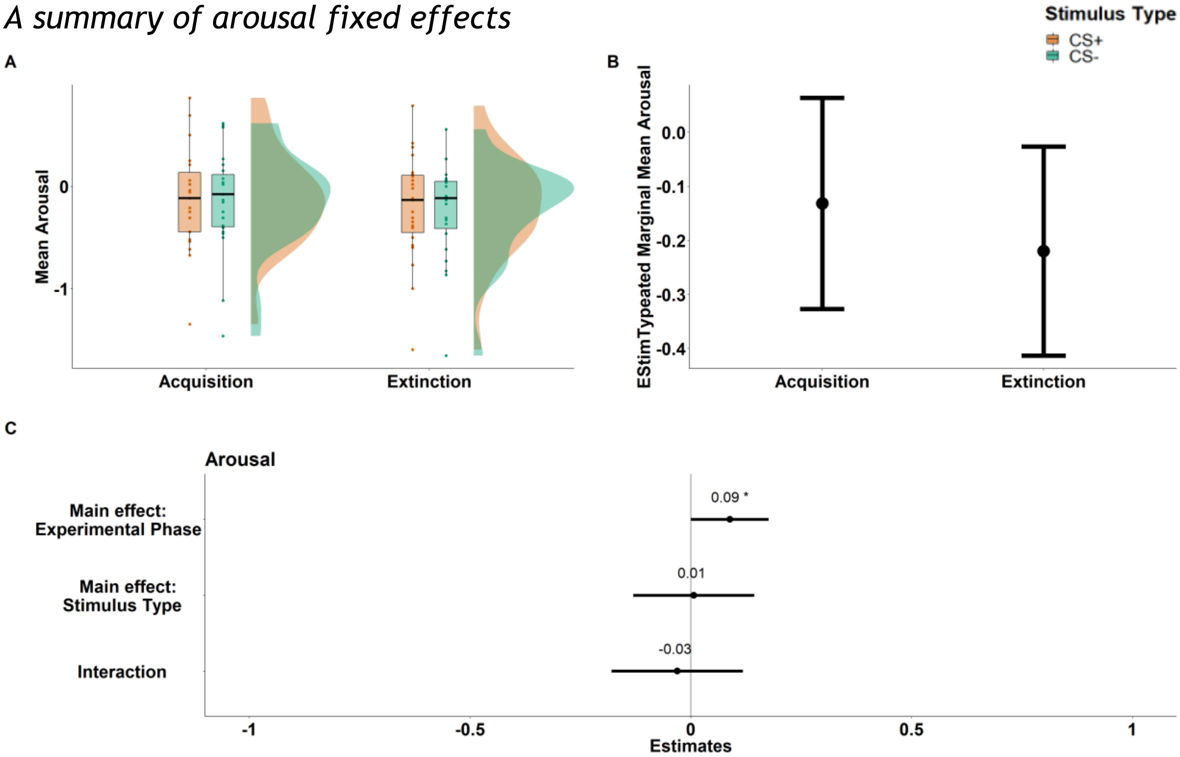
A summary of valence fixed effects



Note. A) Distribution of mean valence ratings of Habituation-baselined data. B) Estimated marginal means per condition derived from the linear mixed effects model of valence ratings (error bars represent 95% CIs for the means conditioned on the random effects). C) Fixed effect estimates (labelled dots) derived from the linear mixed effects model of valence ratings; bars represent 95% CIs for the estimates.

Figure 4

A summary of arousal fixed effects



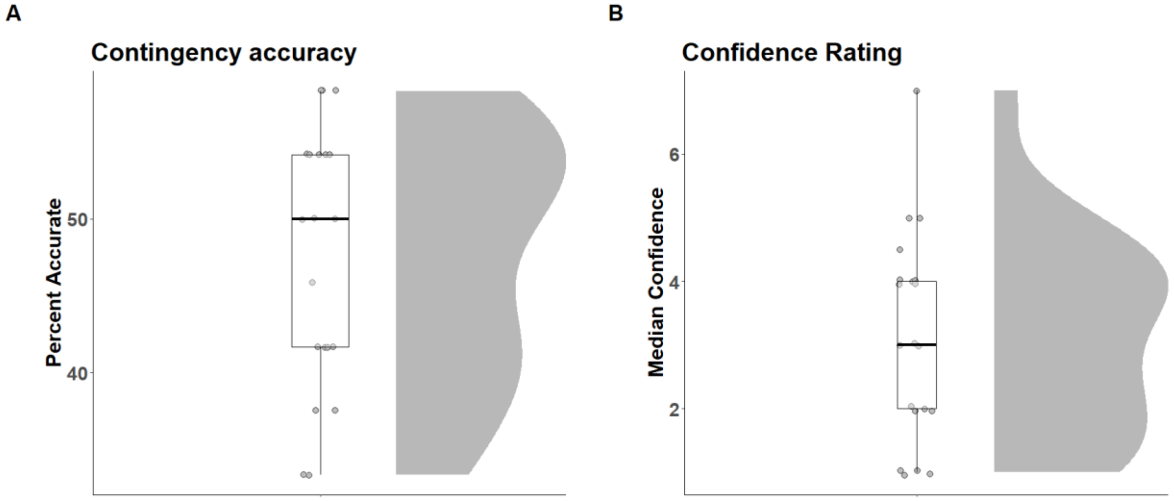
Note. A) Distribution of mean valence ratings of Habituation-baselined data. Estimated marginal means from the linear mixed effects model of arousal ratings for B) Experimental Phase and C) Stimulus Type (error bars represent 95% CIs for the means conditioned on the random effects). D) Fixed effect estimates (labelled dots) derived from the linear mixed effects model of arousal ratings; bars represent 95% CIs for the estimates.

2.3.2 Contingency awareness

Figure 5 shows the percentage accuracy in correctly identifying faces that were paired with the UCS and associated confidence ratings. As seen in the figure, 60% of participants had accuracy below 50% and the maximum accuracy rate was 58%. The low accuracy rate is also reflected in the relatively low confidence ratings. Contingency awareness was also evaluated by computing d-prime (d') estimates of participants' ability to correctly discriminate between CS+ and CS- faces (package *sdt.rmcs*). To calculate d-prime, participants' responses were divided into hits – correctly identifying a CS+ face, correct rejections – correctly identifying a CS- face, false alarms – incorrectly classifying a CS- face as CS+, and misses, incorrectly classifying a CS+ face as CS-. A sensitivity index d' was calculated (see Figure 6) by taking the difference of the z values for the hits and misses. To assess whether participants had a bias towards preferentially responding to either one of the stimuli, a bias index c was also computed. This represented the number of standard deviations from the midpoint of the difference between the z values of the hits and misses. Finally, a one-sample t-test was performed to assess whether d-prime and bias estimates were significantly greater than 0. Confirming the descriptive results, these revealed that participants did not perform above chance on the contingency awareness task (t (19) = -1.4, p < 0.17) and there was no significant bias in responding (t (19) = 0.57, p < 0.57). Note, that these analyses excluded three subjects who gave the same contingency awareness response in all trials. Overall, these findings suggest that participants failed to acquire an awareness of the contingency between the CSs and UCS. Furthermore, our exploratory analyses suggest that the perceived CS-UCS contingency did not influence conditioned responding (see Supplementary Materials 3).

Figure 5

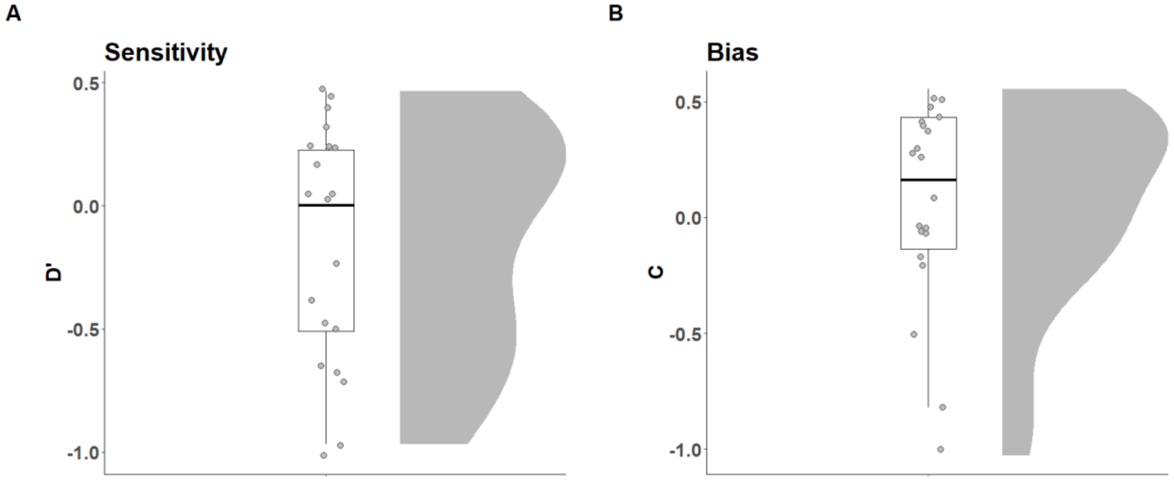
Contingency accuracy and associated confidence rating.



Note. Percentage accuracy was calculated by summing the number of correct responses and dividing by the overall number of trials * 100.

Figure 6

Distribution of sensitivity and bias estimates.



2.3.3 Re-analysis of the Rehbein et al. (2014) dataset

In this secondary analysis, we used the valence rating data provided by Rehbein et al. (2014) since they reported a small in magnitude conditioned response in valence ratings. These analyses aimed to demonstrate the utility of robust and transparent graphical tools in providing valuable information that can be complementary to but also guide statistical inferences.

Figure 7A shows a traditional plot that often accompanies result sections across the literature, similar to that employed by Rehbein et al. (2014) to support their findings in relation to the valence CR. The figure shows the average valence rating per condition. In contrast, Figure 7B represents the same mean ratings per condition, accompanied by the individual data points, boxplots including the median valence, as well as split-violin plots. As seen in the figure, multiple patterns can be seen in the data. First, even providing confidence intervals around the mean can be informative in relation to the magnitude of the observed effects. For example, while the mean difference between CS+ and CS- post-extinction are larger than those obtained pre-habituation, the uncertainty around each condition is very similar and overlapping. Visualising the full range of data points further emphasises that the magnitude of effects is very small, the range in ratings is very large, and that there are a few outliers in the data.

Another interesting pattern that emerges is that the mean and median ratings change the direction of condition differences. For example, when looking at the mean during post-extinction, we can see that CS- faces (mean =2.92) were rated as more pleasant than CS+ (mean = -1.08) faces, similar to what is depicted in Figure 7A. However, when looking at the median which is a measure of central tendency that is more robust to outliers (Rousselet et al., 2017), we can see that CS+ faces (median = 2.14) are rated as more pleasant than CS- faces (median = -1.37). The magnitude of the mean and median differences is also very similar. Finally, looking at the boxplot and split-violin plots, we can see that the distributions overlap substantially between conditions, further confirming the small effects.

Visualising the data in this transparent manner suggests various patterns that if considered, can guide analytical strategies and lead to more robust inferences, even if traditional analytical strategies that do not consider by-item variability are employed. First, observing the full data distribution and the presence of outliers suggests that the mean may not be suitable in providing robust estimate of condition differences, since it is highly sensitive to small changes in distributions and the presence of outliers (Rousselet et al., 2017). In this case, if an rm ANOVA or a t-test is performed to test for differences between conditions, 20% trimming of the mean would be a more reliable alternative as it provides more power in the presence of outliers (Wilcox, 2017). To illustrate, Table 3 provides a comparison between the results of a standard paired t-test and a robust t-test using 20% trimming of the mean. Here, we compared the CS- ratings between pre-habituation and post-extinction, since this was the main effect reported by Rehbein et al. (2014). As seen in the table, the results from the paired t-test give identical results from those provided in the original paper, with a p-value ‘approaching’ the significance level of 0.05. However, the results from the robust t-test suggest that there are no differences between the two conditions. These findings confirm that the presence of outliers can shift the pattern of results and dramatically impact the inferences we make and clearly demonstrates the value of appropriate data visualisation tools in guiding robust inferences.

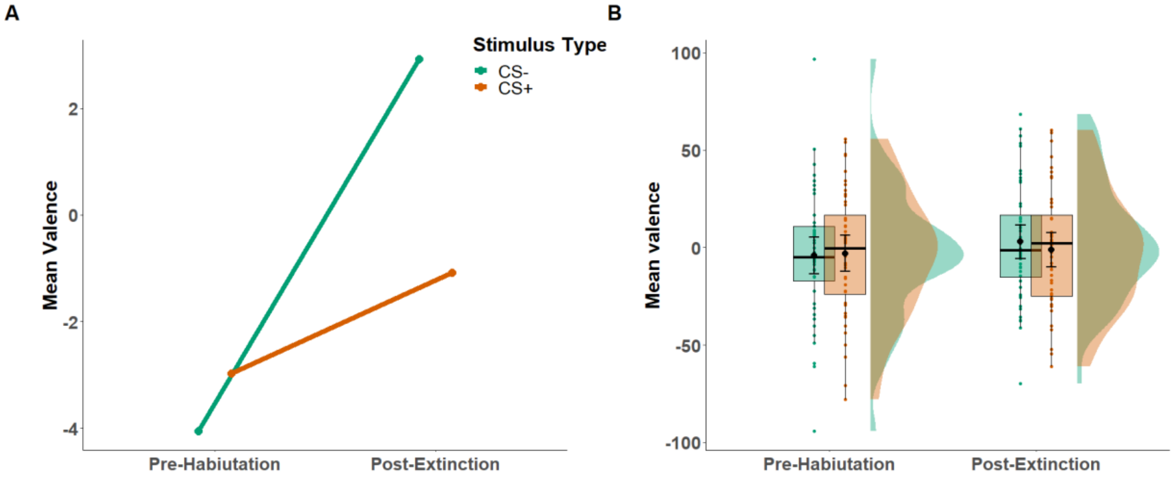
Table 3

Comparison between the outputs derived from a t-test and a robust t-test on trimmed means.

method	estimate	statistic	p-value	df	Lower CI	Upper CI
t-test	6.98	-1.98	0.054	47	-14.8	0.11
robust t-test	-4.79	-1.21	0.236	29	-12.9	3.11

Figure 7

Comparison of graphical tools.



Note. A) Mean valence ratings per condition. B) The coloured dots represent the average valence rating for each individual, the black dot and associated error bars depict the mean and 95% confidence intervals, and the boxplot and split-violin plots convey information about the underlying data distribution.

The large variability in responses also raises the question of whether individual differences can explain the small effects observed from this dataset. This question can be easily answered by visualising the effects at a within-subject level (see Figures 8 and 9). When looking at the largest effect from this dataset (see Figure 8), i.e., the lower mean valence for CS- faces in pre-habituation compared to post-extinction, we can indeed see that this is the case for a subset of participants. However, it is also visible that other participants show an effect in the opposite direction or no effect at all. This is the case for CS+ trials as well. This pattern is not surprising since individual differences in conditioned responding have been well established, with the issue of non-responders having been extensively discussed (Lonsdorf et al., 2019; Lonsdorf & Merz, 2017). When visualising paired differences between conditions within each experimental phase (see Figure 9), another pattern emerges. The great variance in responses suggests that the stimuli were not simply neutral and at least in some individuals, there are baseline valence differences between CS+ and CS- trials during pre-habituation. This can cause additional uncertainty when making inferences regarding changes in valence elicited by the conditioning manipulation when prior differences may already be present. Overall, the visualisation of paired differences across individuals confirms that the response to the experimental manipulation differed across participants, which further highlights the necessity of modelling random by-subject and by-item variability and their dependencies within the experimental conditions.

Figure 8

Within-subject differences when comparing the differences between pre-habituation and post-extinction for each stimulus type.

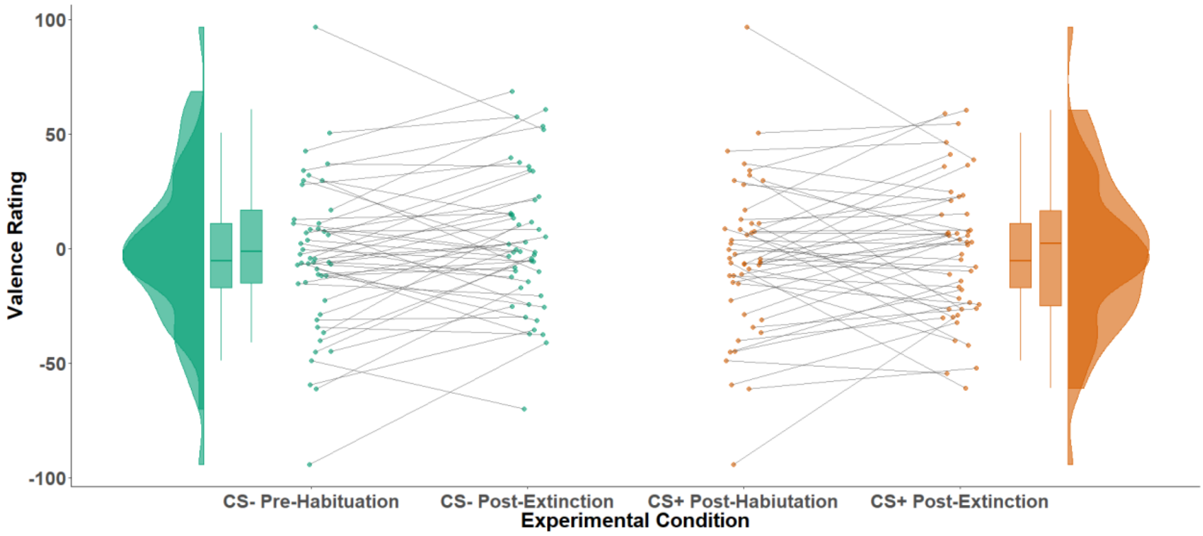
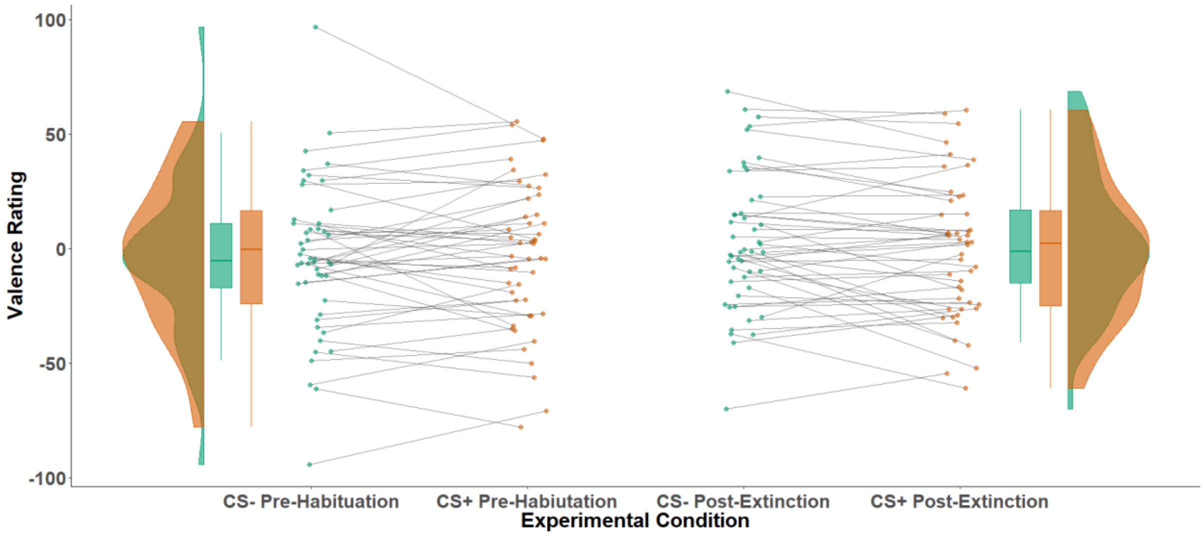


Figure 9

Within-subject differences when comparing the differences between CS+ and CS- trials within each experimental phase.



2.4 Discussion

The goal of this study was to contribute to the ongoing efforts in understanding whether threat learning can develop in the absence of contingency awareness, through a replication of a Multi-CS conditioning paradigm that so far, has provided conflicting evidence regarding the feasibility of eliciting subjective behavioural CRs (Bröckelmann et al., 2011; Rehbein et al., 2014, 2015). Specifically, we aimed to gain an insight into the factors that may contribute to these inconsistent reports by utilising robust and transparent analytical and data visualisation strategies. An additional complication affecting the interpretability of these findings was the presence of technical differences in defining what constitutes a measure of a CR, as the results from Multi-CS conditioning tasks were derived from statistical comparisons following extinction, instead of from differences between conditions following threat acquisition. Any changes in evaluative judgements observed from these comparisons, however, were interpreted as behavioural indices of learning rather than indices of a non-extinguished CR following extinction. While technically, observing a non-extinguished CR suggests that conditioned responding/associative learning has occurred at some point in time, it does not provide direct evidence for it. To provide clarification of this issue, the replication focused on disentangling the extent to which the paradigm can measure the acquisition of threat related CRs *as well as* their extinction. To do so, we measured ratings of valence and arousal following habituation, threat acquisition and extinction.

Like some of the previous Multi-CS conditioning investigations (Bröckelmann et al., 2011; Rehbein et al., 2014, 2015) our analyses revealed that participants had no subjective awareness of the CS-UCS contingency. Our descriptive analyses suggested limited evidence for unaware conditioning, driven by high response variability across conditions in both valence and arousal ratings. Furthermore, we used linear mixed effects modelling to provide a generalisable quantification of the presence of learning and extinction indices, but we found no evidence for conditioned valence or arousal responses during either acquisition or extinction. This was the case in our main analysis which controlled for potential baseline habituation differences as well as when ordinal modelling was performed on the raw data (see Supplementary Materials 2). In addition, unlike Rehbein et al. (2015), we demonstrated that the perceived contingency had no influence on valence and arousal ratings (see Supplementary Materials 3). These results, provide some support for the single process account of learning (Mertens & Engelhard, 2020) which argues that contingency awareness is a prerequisite for the development of a conditioned response. Nonetheless, it is also possible that our contingency awareness measure was not sensitive enough to detect a relationship between awareness and the rating data.

Although our findings are consistent with evidence supporting the role of contingency awareness in threat learning (Dawson et al., 2007; Klucken et al., 2009; Lipp & Purkis, 2005; Tabbert et al., 2011; Weidemann et al., 2016), it is still important to establish why at least some of the previous Multi-CS conditioning studies report differential CRs (Bröckelmann et al., 2013; Rehbein et al., 2015; Steinberg et al., 2012). These studies have observed significant CRs in experimental contexts that typically challenge the successful detection of conditioning, such as utilising a cognitively demanding task and employing offline behavioural measures that are generally considered to be less sensitive (Corneille & Mertens, 2020). In addition, these studies have reported effects following extinction training which is prone to rapid CR habituation (Dunsmoor et al., 2019; Leuchs et al., 2019). This suggests that some form of implicit processing effect that is large enough to be detectable is taking place. Consistent with previous research (Luck & Lipp, 2015a, 2015b; Wendt et al., 2020), this implicit process appears to create evaluative CRs that are resistant to extinction.

A potential reason for failing to detect differential CRs in the present study, as well as for the inconsistency in previous behavioural results from Multi-CS conditioning, may be the issue of generalisability and differences in the analytical strategies that drive inferences. It is highly probable that previous results about valence and arousal CRs, based on conventional analyses using data averaged up to participant level, simply do not generalise to a different set of stimuli of the same type. The p-values, test statistics and any subsequently drawn conclusions from such analyses can only apply to the stimulus set employed in the study (Westfall et al., 2017). Consequently, utilising a different set of similar stimuli can produce an effect in the opposite direction that is driven by different levels of random item variability rather than by the experimental manipulation (Yarkoni, 2020). To account for this issue, our results were derived from a new set of face stimuli and design-appropriate, simultaneous modelling of the random variability observed at the item as well as the subject levels. Employing this approach increases the confidence that our findings can generalise to other populations of people and stimuli of the same type, and that the chance of observing a false positive is significantly reduced.

It is also likely that reports of behavioural CRs following extinction training, derived from at least some Multi-CS conditioning studies are driven by a small proportion of participants. Through re-visualisation of the effects observed by Rehbein et al. (2014) (see Supplementary Materials 4) we indeed showed that participants were influenced by the experimental manipulation to a different degree, and sometimes in a divergent manner. This is in line with other investigations in standard conditioning protocols (Lonsdorf & Merz, 2017) showing that CRs are highly susceptible to individual differences. Our secondary analysis also revealed several other factors that can affect inferences and conclusions. We found that interpreting condition differences resulting from the experimental manipulation can be complicated by the presence of baseline differences between CS+ and CS- faces, at least in some individuals. We also showed that the choice of central tendency measure for assessing condition differences can be crucial in driving the direction of results. Furthermore, we demonstrated that visualising the data beyond the mean and including the full range of responses can contextualise the magnitude of effects but also guide analytical strategies. For example, comparing mean differences using a traditional t-test in the presence of outliers in the dataset by Rehbein et al. (2014), was found to produce potential condition differences. However, the presence of significant effects driven by outliers and slight changes in distributions were diminished when using a robust alternative, even with a tool that does not account for by-item variability (i.e., a t-test on trimmed means), (see Rousselet et al., 2017). These results highlight the utility of robust graphical tools in improving inferences through facilitating our understanding of underlying data patterns.

An additional factor that can affect replicability and should be considered is the measurement itself. Previous Multi-CS conditioning studies relied on interval or discrete SAM scales. The present study used a 7-point Likert scale to assess subjective valence and arousal. This was done to avoid the use of a measure that can be less intuitive to participants (we assume that most participants are used to Likert scales these days, even in non-academic contexts). While there are no systematic investigations examining the comparative reliability and validity of SAM and standard Likert scales, it is reasonable to assume that no drastic discrepancies should be present since both measure ordinal responses. The differences in the scale ranges also should not have posed a significant problem since previous research has demonstrated minimal changes in reliability between 7 and 9-point scales, at least in Likert measures (Preston & Colman, 2000). Likert scale responses, however, are typically subject to participant response bias patterns (i.e., a tendency to use neutral values or extremes), (Sung & Wu, 2018). This was potentially the case in our data, as we observed that a large proportion of responses belonged to the *neutral* response category for valence ratings and to the *not arousing* response categories for arousal ratings, even after the conditioning manipulation was applied. Response biases of this type, therefore, may disrupt the detection of potential effects. In contrast, interval measures, such as visual analogue scales, may offer a more sensitive alternative, as they are suggested to increase the likelihood of obtaining a more exact measure and to capture a greater variability in responses (Reips & Funke, 2008; Sung & Wu, 2018). Evidence for such improved sensitivity in the context of Multi-CS conditioning is unconvincing, since the only study using an interval scale found no significant condition differences (although a *marginally significant* valence effect was reported), (Rehbein et al., 2014). Furthermore, in this study, the original ordinal SAM scale was transformed to an analogous scale. However, since no details were available regarding how the – 300 to 300 continuous values were modified within a 9-point pictorial representation scale, it is unclear whether the SAM interval measure was truly interval. Nonetheless, determining whether response variability can be increased using an interval scale, and whether this would facilitate the detection of CRs using Multi-CS conditioning would be beneficial for understanding the role of measurement in unaware conditioning.

2.4.1 Future recommendations

Based on our findings, we provide several recommendations focusing on technical, methodological, and analytical factors that can improve inferences from Multi-CS conditioning. Our first recommendation is related to increasing clarity when defining concepts and their measurement. In the general conditioning literature, the use of measures that may not assess the underlying construct that they intended to measure fuels and perpetuates the debate regarding the role of contingency awareness in conditioning. In the context of Multi-CS conditioning, this problem is aggravated by the poor differentiation between the processes of learning and extinction. It is therefore crucial for future studies to clearly distinguish between measures of associative learning and acquisition of a CR, and of extinction and the extinguishment of the CR. Consequently, conclusions based on inferences drawn from statistical tests should reflect as closely as possible the construct that was intended to be measured. This would ease comparisons between studies and facilitate quantification of other factors that may be relevant when attempting to elicit conditioning without awareness.

Second, efforts for improving replicability of conditioning effects in the absence of awareness using multiple CSs should focus on ensuring generalisability of inferences beyond single studies and minimising the occurrence of false positives. This can be achieved by moving away from analytical strategies relying on subject-level aggregated data and towards employing design-appropriate modelling of random variability across subjects as well as items. In addition, whenever possible findings should be accompanied by transparent data visualisation at the participant rather than average level, as this can facilitate understanding of the underlying distribution and response variability and guide analytical decisions and interpretation.

Finally, we recommend that future studies focus on establishing a reliable and sensitive offline measure of behavioural CRs, to ensure that the processes elicited through the experimental manipulation are measured as accurately as possible. We encourage the use of measures that reflect more contemporary graphical interfaces that are intuitive to participants over the use of outdated measures of emotional constructs, although it is worth noting that this is not a problem specific to this paradigm. In fact, the original SAM scale paper (Bradley & Lang, 1994) was cited nearly 5000 times in the past 5 years alone (Google Scholar search), suggesting its continuous and wide-spread use. Nonetheless, there are no recent investigations examining its validity and reliability. Consequently, relying on measures due to their popularity rather than validity, can prove detrimental for measuring rapidly habituating processes and small effects. Since Likert scales can pose an additional set of problems relating to response biases and poor response variability, future work should focus on examining the utility and sensitivity of continuous measures, such as visual analogue scales in measuring associative learning processes using Multi-CS conditioning.

2.5 Chapter 2 Supplementary Materials

2.5.1 Supplementary Materials 1: Random effects summaries derived from mixed models

Valence

Supplementary Table 1

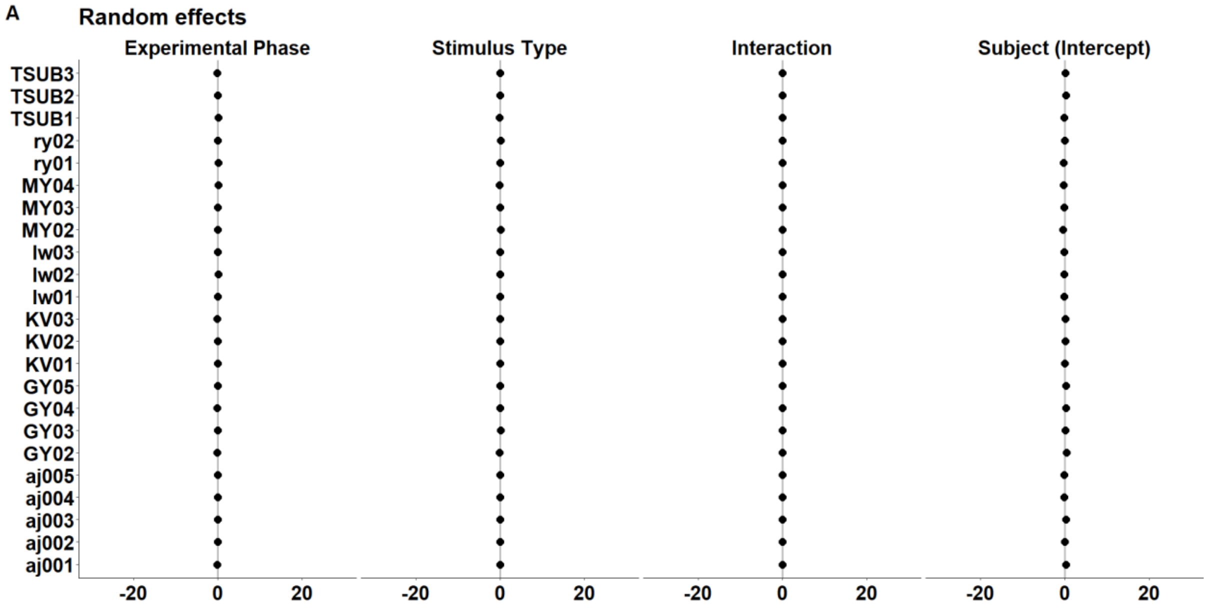
Summary of fixed estimates and random effect variance for the valence model.

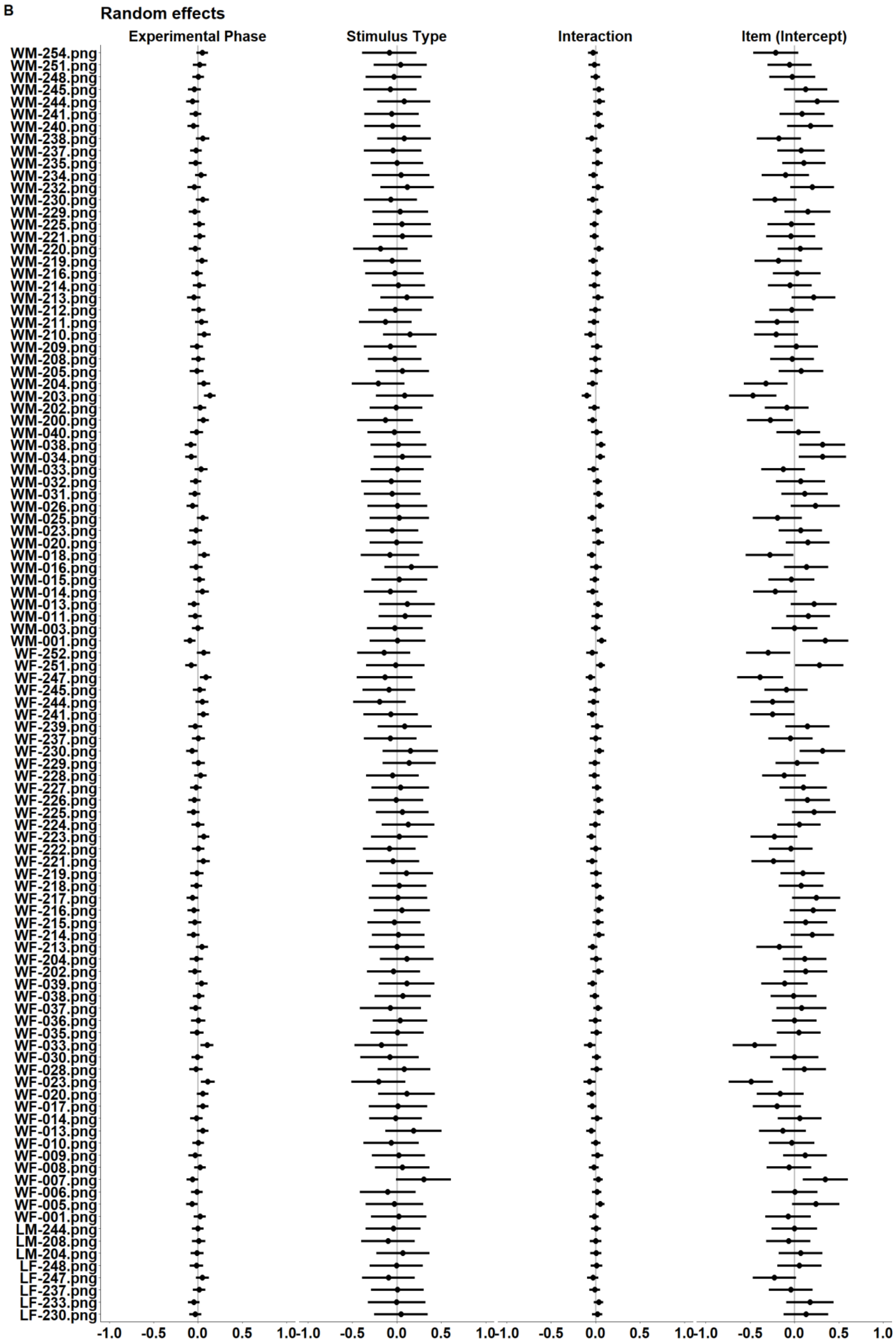
	Mean Valence		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.10	-0.22 – 0.02	0.091
Experimental Phase	0.03	-0.04 – 0.10	0.375
Stimulus Type	-0.01	-0.09 – 0.08	0.857
Interaction	-0.04	-0.16 – 0.08	0.530
Random Effects			
σ^2	1.0667		
τ_{00} Subject	0.0506		
τ_{00} Item	0.0631		
τ_{11} Subject: Phase	0.0033		
τ_{11} Subject: Stimulus Type	0.0328		
τ_{11} Subject: Interaction	0.0019		
τ_{11} Item: Phase	0.0065		
τ_{11} Item: Stimulus Type	0.0098		
τ_{11} Item: Interaction	0.0001		
N Subject	23		
N Item	104		
Observations	4784		
Marginal R^2	0.000		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 1

By-subject and by-item random coefficients and intercept for the valence model.




B
Random effects
Experimental Phase
Stimulus Type
Interaction
Item (Intercept)
WM-254.png
WM-251.png
WM-248.png
WM-245.png
WM-244.png
WM-241.png
WM-240.png
WM-238.png
WM-237.png
WM-235.png
WM-234.png
WM-232.png
WM-230.png
WM-229.png
WM-225.png
WM-221.png
WM-220.png
WM-219.png
WM-216.png
WM-214.png
WM-213.png
WM-212.png
WM-211.png
WM-210.png
WM-209.png
WM-208.png
WM-205.png
WM-204.png
WM-203.png
WM-202.png
WM-200.png
WM-040.png
WM-038.png
WM-034.png
WM-033.png
WM-032.png
WM-031.png
WM-026.png
WM-025.png
WM-023.png
WM-020.png
WM-018.png
WM-016.png
WM-015.png
WM-014.png
WM-013.png
WM-011.png
WM-003.png
WM-001.png
WF-252.png
WF-251.png
WF-247.png
WF-245.png
WF-244.png
WF-241.png
WF-239.png
WF-237.png
WF-230.png
WF-229.png
WF-228.png
WF-227.png
WF-226.png
WF-225.png
WF-224.png
WF-223.png
WF-222.png
WF-221.png
WF-219.png
WF-218.png
WF-217.png
WF-216.png
WF-215.png
WF-214.png
WF-213.png
WF-204.png
WF-202.png
WF-039.png
WF-038.png
WF-037.png
WF-036.png
WF-035.png
WF-033.png
WF-030.png
WF-028.png
WF-023.png
WF-020.png
WF-017.png
WF-014.png
WF-013.png
WF-010.png
WF-009.png
WF-008.png
WF-007.png
WF-006.png
WF-005.png
WF-001.png
LM-244.png
LM-208.png
LM-204.png
LF-248.png
LF-247.png
LF-237.png
LF-233.png
LF-230.png
-1.0
-0.5
0.0
0.5
1.0-1.0
-0.5
0.0
0.5
1.0-1.0
-0.5
0.0
0.5
1.0-1.0
-0.5
0.0
0.5
1.0

Arousal

Supplementary Table 2

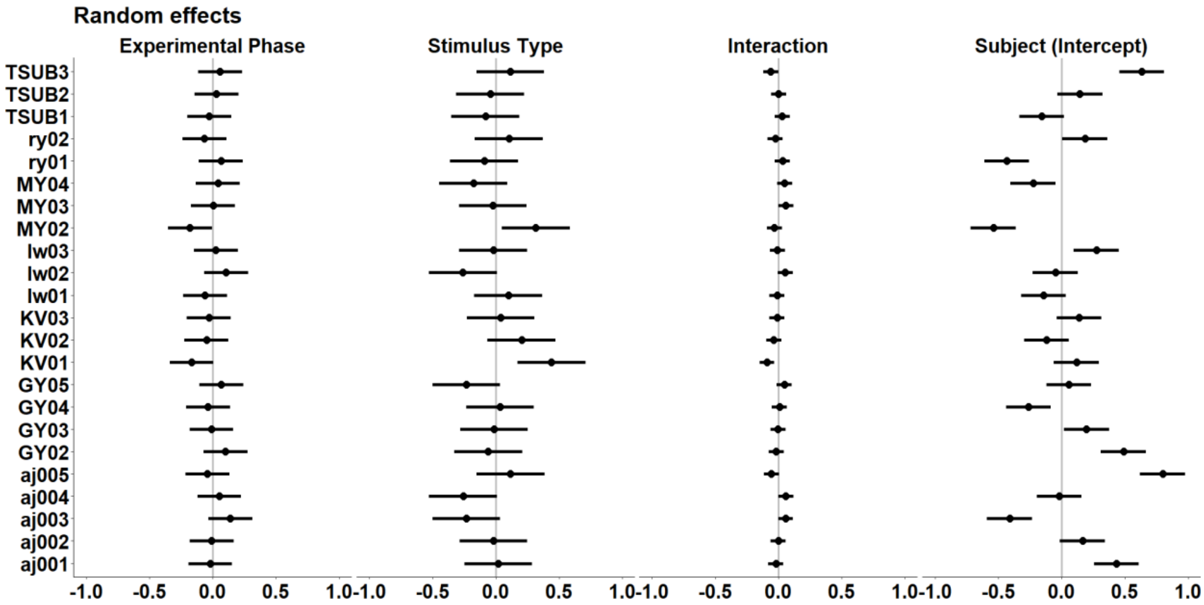
Summary of fixed estimates and random effect variance for the arousal model.

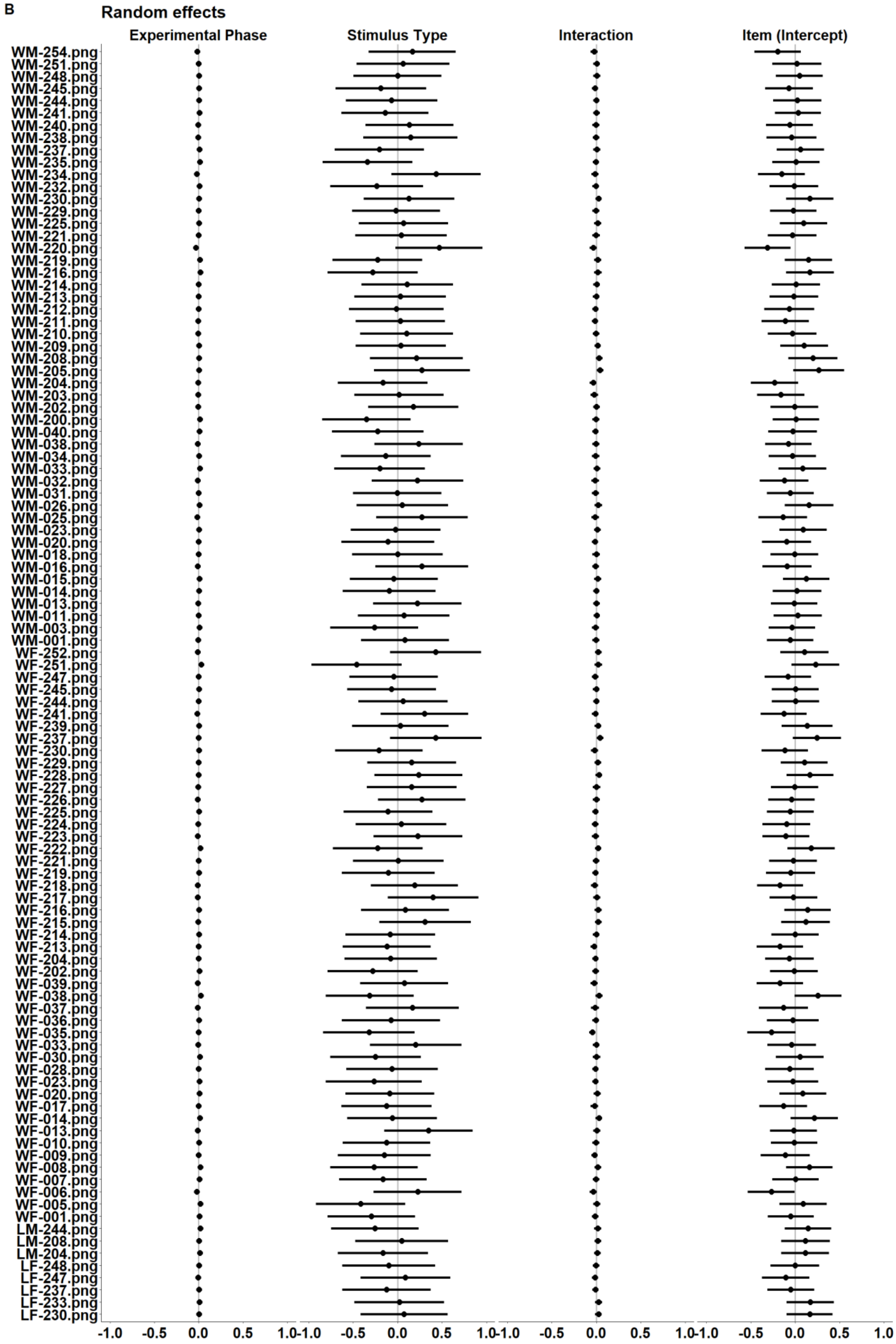
	Mean Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.18	-0.36 – 0.01	0.069
Experimental Phase	0.09	0.00 – 0.18	**0.049**
Stimulus Type	0.01	-0.13 – 0.14	0.926
Interaction	-0.03	-0.18 – 0.12	0.685
Random Effects			
σ^2	1.6833		
τ_{00} Subject	0.0331		
τ_{00} Item	0.1977		
τ_{11} Subject: Phase	0.0002		
τ_{11} Subject: Stimulus Type	0.1091		
τ_{11} Subject: Interaction	0.0007		
τ_{11} Item: Phase	0.0139		
τ_{11} Item: Stimulus Type	0.0494		
τ_{11} Item: Interaction	0.0026		
N Subject	23		
N Item	104		
Observations	4784		
Marginal R^2	0.001		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 2

By-subject and by-item random coefficients and intercept for the arousal model.




B
Random effects
Experimental Phase
Stimulus Type
Interaction
Item (Intercept)
WM-254.png
WM-251.png
WM-248.png
WM-245.png
WM-244.png
WM-241.png
WM-240.png
WM-238.png
WM-237.png
WM-235.png
WM-234.png
WM-232.png
WM-230.png
WM-229.png
WM-225.png
WM-221.png
WM-220.png
WM-219.png
WM-216.png
WM-214.png
WM-213.png
WM-212.png
WM-211.png
WM-210.png
WM-209.png
WM-208.png
WM-205.png
WM-204.png
WM-203.png
WM-202.png
WM-200.png
WM-040.png
WM-038.png
WM-034.png
WM-033.png
WM-032.png
WM-031.png
WM-026.png
WM-025.png
WM-023.png
WM-020.png
WM-018.png
WM-016.png
WM-015.png
WM-014.png
WM-013.png
WM-011.png
WM-003.png
WM-001.png
WF-252.png
WF-251.png
WF-247.png
WF-245.png
WF-244.png
WF-241.png
WF-239.png
WF-237.png
WF-230.png
WF-229.png
WF-228.png
WF-227.png
WF-226.png
WF-225.png
WF-224.png
WF-223.png
WF-222.png
WF-221.png
WF-219.png
WF-218.png
WF-217.png
WF-216.png
WF-215.png
WF-214.png
WF-213.png
WF-204.png
WF-202.png
WF-039.png
WF-038.png
WF-037.png
WF-036.png
WF-035.png
WF-033.png
WF-030.png
WF-028.png
WF-023.png
WF-020.png
WF-017.png
WF-014.png
WF-013.png
WF-010.png
WF-009.png
WF-008.png
WF-007.png
WF-006.png
WF-005.png
WF-001.png
LM-244.png
LM-208.png
LM-204.png
LF-248.png
LF-247.png
LF-237.png
LF-233.png
LF-230.png
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0

2.5.2 Supplementary Materials 2: CLMM Modelling of the three-phased rating data

Due to the ordinal nature of the data, the rating data were analysed again using cumulative link mixed models (package *ordinal*). We examined the interaction between *Stimulus Type* (CS+/CS-) and *Experimental Phase* (Habituation, Acquisition, Extinction) predicts arousal and valence ratings. The models included arousal/valence ratings as the outcome variable, *Stimulus Type* and *Experimental Phase* and the interaction between them as the fixed effects. The random effect structure included *Subjects* and *Items* as random intercepts and by-subject and by-item random slopes for the *Stimulus Type* and *Experimental Phase* interaction. In each model, mean-centred (deviation coding) contrasts were used for the two categorical fixed effects. For the *Experimental Phase* fixed effect, the Habituation phase was used as the baseline level. The*"nlminb"* optimizer was used to maximise the marginal likelihood function. Main effects and interactions were assessed using type II Likelihood-ratio test (package *RVAideMemoire*).

The likelihood-ratio tests did not reveal any significant effects at the level of 0.05 for either valence or arousal ratings (see Supplementary Table 3 and Supplementary Figure 3 and 4). In terms of arousal, these results suggest that the differences between Acquisition and Extinction observed using the LME modelling on habituation-baselined data are much smaller when and ordinal model is fitted to the data without baselining in respect to the Habituation phase. Therefore, it is possible that some minor baseline condition differences may exist during Habituation or that accounting for the ordinal nature of the data reduces the magnitude of observed differences.

The predicted probabilities for each ratings category for valence and arousal across experimental phases and stimulus type can be visualised in Supplementary Figure 3A and 4A. For both CS+ and CS- stimuli and across experimental phases the predicted response probability for valence ratings was highest for ratings of 3 and 4, suggesting that the stimuli were largely perceived as neutral in valence. For arousal, ratings below 4 had the highest probability, suggesting that the stimuli were largely perceived as not arousing.

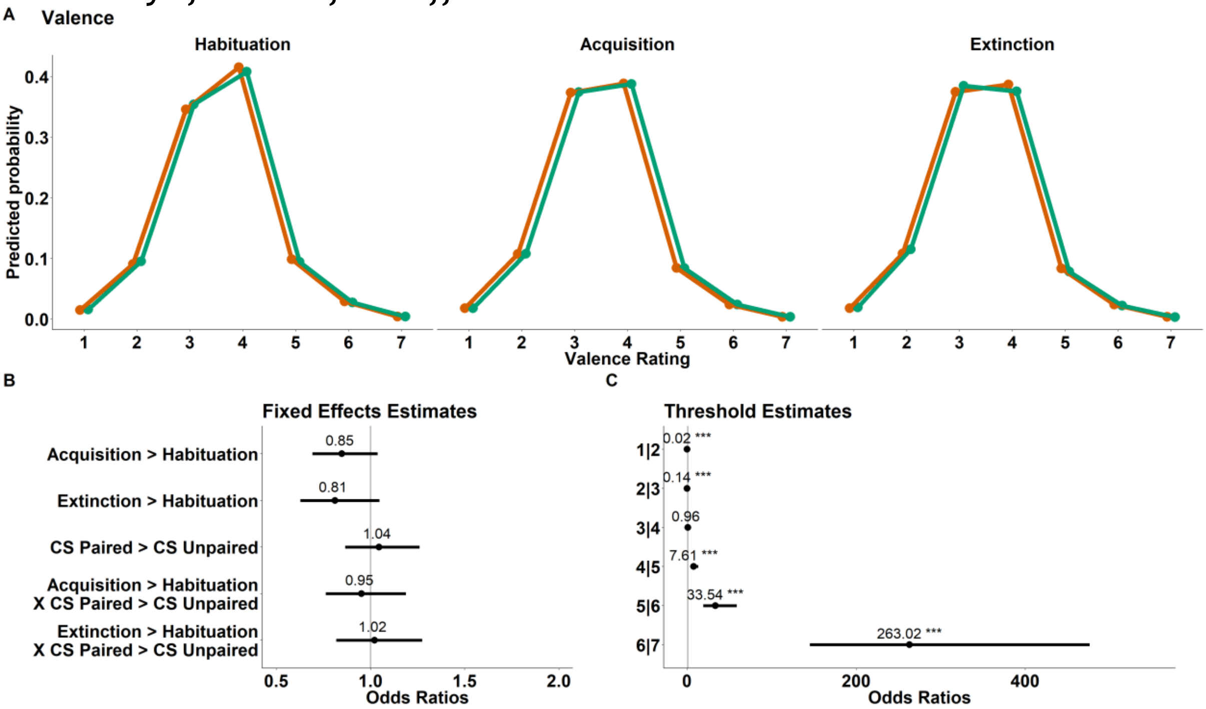
Supplementary Table 3

Type II Likelihood-ratio test of main effects and interactions.

	LR Chisq	Df	P-value	McFadenn Pseudo R^2	Negelkerke Pseudo R^2
Valence Ratings				0.0001	0.0005
Experimental Phase	2.68	2.000	0.27		
Stimulus Type	0.19	1.000	0.65		
Experimental Phase X Stimulus Type	0.44	2.000	0.80		
Arousal Ratings				0.0003	0.001
Experimental Phase	5.19	2.000	0.07		
Stimulus Type	1.38	1.000	0.24		
Experimental Phase X Stimulus Type	0.72	2.000	0.69		

Supplementary Figure 3

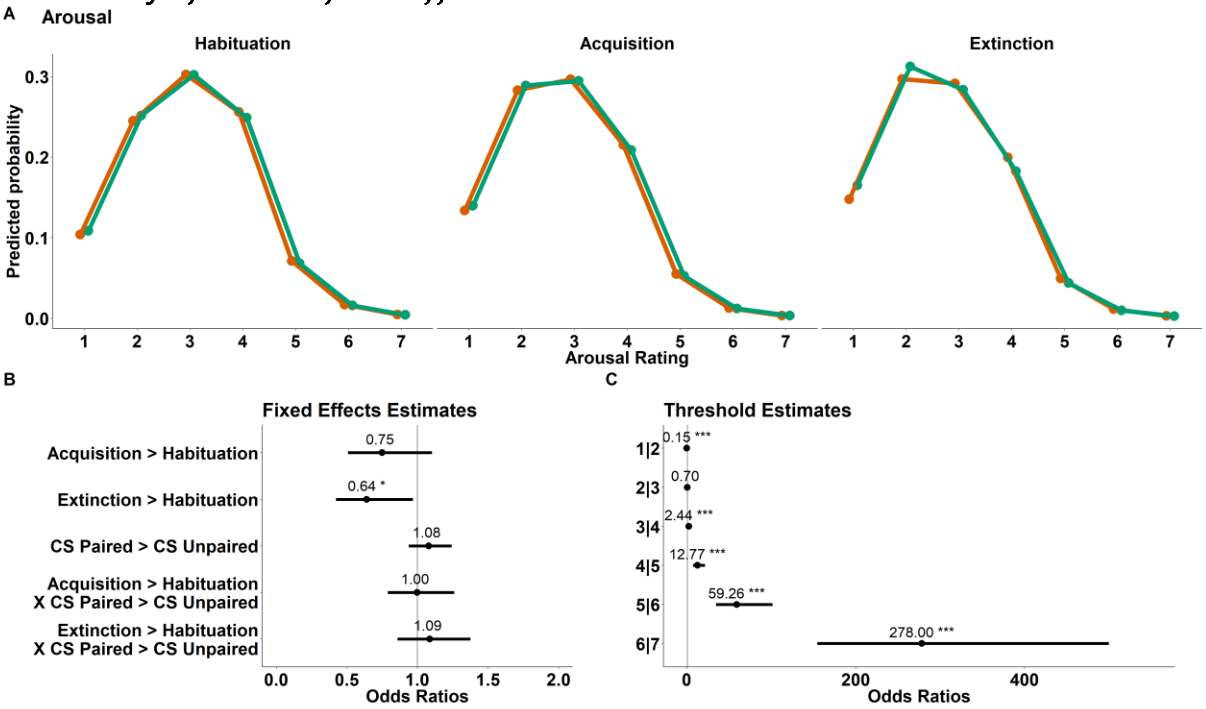
A summary of valence fixed effects.



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of valence ratings B) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates. C) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates.

Supplementary Figure 4

A summary of arousal fixed effects.



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of arousal ratings B) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates. C) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates.

2.5.3 Supplementary Materials 3: Mediating role of perceived contingency on conditioned responding

To determine whether conditioned responding is influenced by participants' subjective perception of the CS-UCS contingency, a second set of LME modelling was performed for both valence and arousal, in which participant's subjective report of the CS-UCS contingency was added as an interacting factor, similar to previous reports by Rehbein et al. (2015). Since the contingency awareness task was performed on a random subset of 24 CSs per participants, the modelling was performed by using only habituation-baselined rating data for these stimuli. Specifically, the models included a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS+ vs CS-) by 2 *Perceived Contingency* (CS+ vs CS-) fixed effects design, with mean-centred contrasts for the two categorical fixed effect predictors. The models included by-subject and by-item random intercepts, together with by-subject and by-item random slopes for all main effects and interactions. Main effects and interactions were assessed using Type III Wald Chi Square tests. Those revealed no significant effects for either valence or arousal ratings (see Supplementary Table 4).

Supplementary Table 4

Type III Wald Chi-square tests *and R-squared values for the valence and arousal models and each of the fixed effects.*

	Chisq	Df	P-value
Valence			
Experimental Phase	0.220	1.000	0.639
Stimulus Type	0.006	1.000	0.937
Contingency Report	0.124	1.000	0.725
Experimental Phase X Stimulus Type	0.008	1.000	0.928
Experimental Phase X Contingency Report	0.216	1.000	0.642
Contingency Report X Stimulus Type	0.005	1.000	0.942
Experimental Phase X Stimulus Type X Contingency Report	0.145	1.000	0.703
Arousal			
Experimental Phase	0.157	1.000	0.692
Stimulus Type	0.565	1.000	0.452
Contingency Report	0.620	1.000	0.431
Experimental Phase X Stimulus Type	0.378	1.000	0.539
Experimental Phase X Contingency Report	0.126	1.000	0.723
Contingency Report X Stimulus Type	0.001	1.000	0.981
Experimental Phase X Stimulus Type X Contingency Report	0.271	1.000	0.603

3 Chapter 3 - Oscillatory, behavioural and pupillary signatures of associative learning and extinction derived using visual blocked conditioning

3.1 Introduction

There has been an increasing amount of evidence suggesting that synchronised neural oscillations at different frequency bands may play an important role in supporting information processing and integration, by facilitating communication between brain regions (Keil & Senkowski, 2018). In the context of associative learning and extinction, rodent studies have demonstrated that oscillations within the theta frequency band (4-8 Hz) are important for coordinating activity across the fear network (Karalis et al., 2016; Lesting et al., 2011; McCullough et al., 2016). Local field potentials studies have revealed that theta synchrony between the basal amygdala (BA) and the medial pre-frontal cortex (mPFC) are linked to CS+>CS- discrimination during learning, while that between the lateral amygdala (LA) and the hippocampus (HPC) are associated with fear expression (Likhtik et al., 2014; Seidenbecher et al., 2003). Furthermore, theta coherence within and across the CA1 part of the HPC, the infralimbic cortex (IL) and lateral amygdala (LA) correlate with behavioural CRs, i.e., conditioned freezing in rodents (Lesting et al., 2013). During extinction, theta synchronisation between the CA1, LA and IL, is shown to decrease, with this pattern of decreased oscillatory activity being driven by top-down influences of the mPFC (Lesting et al., 2011).

The literature on oscillatory dynamics of learning and extinction in humans, however, is relatively sparse due to a number of technical and design constraints that can limit the range of mechanisms that can be studied. For instance, while detecting signal originating from the amygdala has been a challenge in fMRI conditioning research (Fullana et al., 2016, 2018), detection of subcortical activity with MEG can be even more problematic, for several reasons. First, reliably localising amygdala sources can be difficult due to its small size and deep location (Tzovara et al., 2019), especially since the spatial resolution of MEG decreases considerably with increasing distance from the sensors (Meyer, Rossiter, et al., 2017). A common pitfall that can further complicate localisation of activity from weak sources in deep structures using MEG is leakage from other regions. For example, visual paradigms tend to elicit strong, evoked occipital responses that can leak into other sources, thereby masking their signal (Mills et al., 2012). In addition, simulation studies have shown that detecting deep structure activity using MEG requires a very large number of trials to achieve an adequate SNR, (Quraan et al., 2011a; Steinberg et al., 2013; Tzovara et al., 2019). This is also the case when investigating oscillatory dynamics, which requires the simultaneous examination of trial-level source activity along the time and frequency domains (this will be discussed in more detail below). The use of many trials in conditioning studies, however, is uncommon since conditioned responding habituates rapidly over repeated CS presentations (Lonsdorf et al., 2017; Ojala & Bach, 2020).

Due to these technical and design limitations, only a handful of MEG conditioning investigations examine or report activity in subcortical structures or investigate the oscillatory signatures of learning (Balderston et al., 2014b; Lithari et al., 2015, 2016; Moses et al., 2007; Tzovara et al., 2019). Instead, the majority of E/MEG research has focused on the learning and extinction indices in sensory regions reflected by evoked brain activity (Bröckelmann et al., 2013; Dolan et al., 2006; Kluge et al., 2011; Lithari et al., 2015, 2016; Moratti et al., 2006, 2017; Moratti & Keil, 2005; Moses et al., 2005; Tesche et al., 2007). This has left a significant gap in our understanding of the temporal and oscillatory dynamics of learning within the fear network. As such, the goal of this study was to design a multi-trial task and test its utility for examining the cortical and subcortical oscillatory dynamics of associative learning. The design of the paradigm was informed by the existing body of MEG investigations attempting to localise deep structure activity during emotion processing, as well as by the current E/MEG evidence of the neural correlates of learning and extinction. These are reviewed below.

In terms of deep source localisation, an accumulating body of simulation and empirical research has focused on the optimisation of methodological and analytical parameters that may allow for the reliable detection of activity deep in the brain using MEG. For instance, it has been shown that coupled with a sufficient number of trials, source reconstruction tools such as minimum-norm estimation (MNE) and beamforming can successfully detect activity originating from the amygdala and the HPC (Attal & Schwartz, 2013; Mills et al., 2012). More recent research has transitioned towards a modelling approach of source reconstruction, which attempts to estimate whether models including deep sources explain the data better than a purely neocortical model (i.e., generative models), (Meyer, Rossiter, et al., 2017; Tzovara et al., 2019). In addition, technological advances are now beginning to allow for more precise spatial measurements, using high-precision approaches such as individualised 3D head casts and wearable on-scalp systems (Meyer, Bonaiuto, et al., 2017; Tierney et al., 2020).

Only recently, a high-precision MEG has been implemented (Tzovara et al., 2019) in the study of the neural correlates of conditioning. The remainder of the existing literature has so far mostly focused on delivering a high number of trials and/or on the use of beamforming, MCE, and MNE estimation tools for detecting deep brain sources (Balderston et al., 2014b; Lithari et al., 2016; Moses et al., 2007). One of the first studies reporting amygdala activity used a partial reinforcement protocol (Moses et al., 2007). They showed that specifically during acquisition, the amygdala exhibited a stronger peak amplitude around 300 ms for the unpaired CS+ than the CS-. However, these findings could not be replicated in a paradigm employing a shorter CS presentation (Tesche et al., 2007), suggesting that trial duration may be another factor influencing the detection of amygdala activation. Furthermore, Moses et al. (2007) demonstrated that amygdala activity for the unpaired CS+ was linked to both the onset and offset of the CS+, supporting the role of the amygdala in the encoding of affective, contingency information as well as in anticipatory processes. Such evoked bi-phasic amygdala activity in response to CS+ compared to CS- has been elicited during threat acquisition without awareness as well (Balderston et al., 2014b). Neural activity in other subcortical structures has been reported using fearful faces as CSs (Lithari et al., 2015). Directed functional connectivity showed that conditioning mediated thalamic connections to the fusiform and parahipocampal gyrus. However, information flow to the amygdala from the occipital areas, fusiform and parahipocampal gyri was not driven by threat acquisition as this connectivity pattern was observed across all phases and thus, was potentially driven by processing of the fearful stimuli.

The temporal dynamics of other well-established regions in the fear network are not yet well understood as the limited number of MEG conditioning investigations (N ~ 20) do not report consistent patterns of activation. For example, Tesche et al. (2007) reported greater activation in the mPFC during acquisition in latencies starting around 350 ms to 550 ms for unpaired CS+ compared to CS-. This activation was suggested to be linked to anticipation of the UCS. Activation was also greater for unpaired CS+ compared to paired CS+ at latencies following UCS onset, suggesting that this activity may reflect an omission response. In addition, differential activity in the left orbito-frontal cortex has been observed but only for individuals who exhibited declarative heart rate changes to the CS+ during acquisition (Moratti & Keil, 2005). Studies using multiple CSs to elicit conditioning have reported an increased activation for CS+ during learning compared to habituation in the right inferior frontal PFC, around 50-80 ms post CS onset. However, this effect was non-significant when comparing CS+ and CS- trials during acquisition (Rehbein et al., 2014). A right dorsolateral PFC and pre-motor cortex activation at a latency between 87-118 ms post CS onset, has also been reported in response to CS+ compared to CS- during learning, but only for individuals with high and not low trait anxiety (Rehbein et al., 2015). Evidence regarding the temporal dynamics of extinction processes is even sparser, as only a small number of MEG studies included an extinction phase. In studies that directly compare CS+ and CS- during extinction, a differential vmPFC in evoked steady state response (SSR) activity to CS+ has been observed (Moratti et al., 2017), although reports of no differential source activity during extinction are also available (Lithari et al., 2015; Rehbein et al., 2015).

More consistent evidence of differential activation during threat acquisition have been reported in sensory regions. For visual paradigms, studies have reported greater activation post CS onset for CS+ than CS- in the occipital areas and the fusiform gyrus, across SSR (Lithari et al., 2015; Moratti et al., 2006, 2017; Moratti & Keil, 2005), and evoked activity paradigms (Dolan et al., 2006). The onset of these effects, however, has been dependent on the type of protocol employed. In SSR studies using gratings as CSs presented with long presentation times (~ 4 -13 seconds), differential activation has been reported at time windows starting after 1.3 seconds post-CS onset (Moratti et al., 2006; Moratti & Keil, 2005). In contrast, evoked activity paradigms using faces, presented for a short duration (800 ms) elicit visual responses much earlier, around 150 ms (Balderston et al., 2014b; Dolan et al., 2006; Rehbein et al., 2015). This activation pattern is consistent with the N/M170 response suggested to underlie face perception (Liu et al., 2002). However, the detection of M170 has not been consistently associated with condition differences, with studies reporting both the presence (Dolan et al., 2006; Rehbein et al., 2015) and the absence (Balderston et al., 2014b) of an M170 CR.

Similarly, EEG evoked activity studies examining the N170 component have found no evidence of condition differences (Stolz et al., 2019), while others have reported both decreased (Sperl et al., 2021) as well as increased (Camfield et al., 2016) activity in response to the CS+. Such inconsistency may be explained by procedural differences, awareness, and the type of CSs that were used. More consistent findings of conditioned responding in the EEG literature have been observed during later time windows. Specifically, enhanced evoked activity towards CS+ has been reported (Ferreira de Sá et al., 2019; Pastor et al., 2015; Sperl et al., 2021) in time windows associated with the late positive potential (LPP), suggested to reflect high arousal and stimulus salience (Sperl et al., 2021).

In the auditory domain, MEG conditioning studies have reported enhanced activity in regions such as Heschl’s gyrus in response to CS+ than CS- stimuli. In Multi-CS conditioning tasks, these CRs have been observed in both in early (20-50 ms) and mid (100 – 150 ms) latencies (Bröckelmann et al., 2011, 2013). In standard conditioning protocols effects have been reported at mid (85- 115 ms) and late (180 – 270 ms) latencies, although an early component (30-50 ms) that exhibited an amplitude reduction in response to the CS+ has also been observed. (Kluge et al., 2011). Furthermore, an enhanced differential activation has been observed in partial reinforcement protocols, in response to the UCS during the unpaired CS+ presentation (Moses et al., 2005; Tesche et al., 2007). However, it is likely that the latter reflects an omission response like that observed in auditory studies where an unexpected sound omission elicits post-omission auditory activity (Raij et al., 1997).

The literature on the oscillatory signatures underlying threat learning and extinction using MEG is even more limited. Lithari et al. (2016) reported increased alpha and beta power for CS+ compared to CS- during learning in the somatosensory cortex and the insula. More recently, Tzovara et al. (2019) provided direct evidence for the feasibility in detecting deep structure sources using high precision MEG. Their generative modelling suggested that models including subcortical sources explained the data better than those including cortical sources only. Their analysis demonstrated reduced theta power around 130 ms post CS onset in the amygdala for CS+ compared to CS- trials during acquisition. In addition, they found no significant differences during extinction. Neural synchrony between the amygdala and the HPC was also found to increase during learning. When examining average oscillatory power, theta but not gamma power predicted threat in the form of lower theta for CS+ than CS- during learning, in both the amygdala and the HPC. However, this pattern is in the opposite direction of that reported in the rodent literature. One explanation for this cross-species discrepancy is that theta oscillations may have a divergent functional role in humans. According to the authors, however, it is also possible that the divergent results may be driven by procedural differences. Specifically, they argued that the greater theta power in response to CS- might have been elicited by the absence of an UCS, serving as a reward.

The sparse number of studies investigating the oscillatory mechanisms of conditioning in humans can be partly explained by the methodological constraints of the conditioning paradigm, that are at odds with the analytical demands of time-frequency decomposition. As mentioned earlier, standard conditioning tasks employ a small number of trials to elicit learning and extinction, due to amplitude reductions resulting from repeated stimulus presentation (Lonsdorf et al., 2017; Ojala & Bach, 2020). Yet, electrophysiological measures often benefit from a larger number of trials in reducing the SNR. For example, in the analysis of event-related fields, an average over trials is typically computed, averaging out random noise and activity that is not phased-locked to the stimulus onset (Herrmann et al., 2014). With increasing number of trials, such random activity approaches zero, retaining only the evoked response (Roach & Mathalon, 2008). Examining activity over time and at different frequencies, that varies on a trial-by-trial basis (i.e., induced activity not time-locked to the stimulus onset), (Herrmann et al., 2014) requires trial-level time-frequency decomposition. It would therefore be expected that even a larger number of trials than that needed for ERF analysis is required to detect a reliable oscillatory signal from trial-level computations. For example, Cohen (2016) recommended utilising 100 trials per condition for time-frequency investigations. More importantly, trial number is crucial for accurate source localisation of deep structure activity. For example, simulation studies have shown that localisation accuracy of hippocampal activity is relatively poor (~11 mm away from the source) even with 50 trials but substantially increases when 100 trials are used, with 150 trials providing the optimal SNR (Quraan et al., 2011a)

In most MEG studies employing a large number of trials to study conditioning mechanisms, researchers often opt for repeating the same single CS+ and CS- stimuli many times (>=100 trials), (Dolan et al., 2006; Kluge et al., 2011; Moses et al., 2005, 2007; Tesche et al., 2007; Tzovara et al., 2019). This method, however, can pose a significant risk for habituation of the CR. An alternative approach that attempts to resolve this issue is Multi-CS conditioning (Junghöfer et al., 2017; Rehbein et al., 2014, 2015; Steinberg et al., 2013). In this task, conditioning is elicited using many different CSs (40-60 per CS type) in an attempt to prevent habituation to the stimuli by limiting repetitions and increasing the number of unique CSs. As such, this method relies on establishing a CR without awareness. Yet, there is little agreement in the literature regarding the feasibility of eliciting CRs in the absence of awareness. For example, a recent systematic and meta-analytic review concluded that there is little evidence for unaware conditioning since methodological quality and a number of factors including trial order effects and publication bias question the reliability of evidence on unaware conditioning (Mertens & Engelhard, 2020). A potential solution to the limitations of the above-mentioned approaches would be to obtain a better balance between stimulus repetitions and the number of unique stimuli. This may minimise the negative effects of habituation and poor contingency awareness and enhance the likelihood of detecting electrophysiological signatures of threat learning.

The aim of this study was to test the utility of a novel paradigm in allowing for the characterisation of the role of theta band oscillations in cortical and subcortical regions previously identified as underlying threat learning and extinction (Karalis et al., 2016; Lesting et al., 2011; McCullough et al., 2016). We focused on three key analytical and design components that can optimise the likelihood of detecting reliable signal from deep sources. First, we employed a visual blocked conditioning task in which learning and extinction were established several times, in 9 consecutive blocks, each containing the three standard conditioning phases (Habituation, Acquisition and Extinction) but a different set of neutral faces serving as CSs. Therefore, our paradigm attempts to maximise SNR through a large number of trials, minimise repetition-related habituation effects through a lower number of repetitions per unique CS, while increasing the likelihood of establishing awareness of the CS-UCS contingency through a small number of unique CSs per block. Employing a larger number of trials through using a greater range of CSs, however, also requires adequate consideration of item-related random variability, to ensure that any effects are generalisable not only to populations of subjects but also items of the same type. As such, following recommendations by Barr et al. (2013), our analyses used design-appropriate mixed modelling and wherever possible, accounted for both by-subject and by-item random variability.

Second, based on previous recommendations suggesting the use of subtraction of control conditions to reduce the impact of leakage from other regions with stronger activity patterns (Mills et al., 2012; Quraan et al., 2011a), we subtracted source activity during Habituation from that during Acquisition and Extinction. Finally, similar to previous research demonstrating the feasibility in detecting subcortical activity using MEG (Attal et al., 2007; Attal & Schwartz, 2013; Dumas et al., 2013), we maximised the accurate estimation of neural currents in deep sources, by using anatomical segmentation of limbic structures for each participant. Furthermore, we used depth weighted MNE for source reconstruction. The benefit of MNE is that unlike beamforming approaches which assume that sources are not temporally correlated, this method is not affected by correlations between sources (Sánchez & Halliday, 2013). In addition, unlike standard MNE which suffers from biases towards superficial sources, depth-weighted MNE is suggested to increase sensitivity in detecting deep brain activity (Attal et al., 2012).

We measured conditioned responses using multiple outcome measures, including MEG (sensor level ERFs and source level time frequency decomposition), pupil size and offline behavioural ratings (valence and arousal). Consistent with the animal literature, we expected to observe an increase in theta power for CS+ relative to CS- trials during learning, across the fear network.[1] We focused on regions-of-interest (ROIs) including the amygdala, HPC, thalamus, the rostral and caudal ACC and insula. Since human electrophysiological studies (Dolan et al., 2006; Lithari et al., 2015; Moratti et al., 2006, 2017; Moratti & Keil, 2005; Rehbein et al., 2014) have also identified sensory regions as differentially responding during learning, we expected to see an increase in theta power in the lateral occipital cortex and the fusiform gyrus. Based on previous E/MEG findings demonstrating the potential sensitivity of the M170 component to the encoding of face stimuli during conditioning (Camfield et al., 2016; Rehbein et al., 2015; Sperl et al., 2021), we expected to see evidence for such effects around 200 ms post-CS onset. In line with recent evidence suggesting that the mPFC in human conditioning may be involved in safety processing (Harrison et al., 2017), we predicted an increased theta power in the mPFC (lateral and medial OFC) in response to safety cues (CS- compared to CS+) during learning, and a diminishing difference during extinction. We also expected a decrease in differential CS+>CS- activation in the rest of the fear network, during extinction. Finally, since there is some evidence to suggest that human conditioning involves oscillatory activity within other frequencies (Lithari et al., 2016), we performed a set of exploratory analyses focusing on understanding the potential additional contributions of other frequency bands (i.e., alpha, beta, and gamma) across the whole brain. In terms of behavioural and pupillary signatures, we expected to observe a larger pupil size, lower valence and an increase in subjective arousal when comparing CS+ to CS- trials during learning and a reduction in this

[1] This hypothesis was generated prior to the findings of Tzovara et al. (2019) who reported the opposite activation pattern.

difference during extinction. Table 4 provides a summary of the main research questions the analyses performed to test our hypothesis and the main conclusions drawn from these.

Table 4

A summary of main research questions and related hypotheses and analyses.

Research question(s)	Hypotheses	Outcome Measure(s)	Analysis type	Main conclusion
Is the M170 component sensitive to conditioning using face CSs? Is there evidence for sensor-level conditioning across sensors and over time?	Differences between CS+ and CS- conditions/ peak ERFs will be evidenced around 200 ms post CS onset	MEG sensor level: ERFs 0.01 – 20 Hz	Cluster-based Monte- Carlo permutation across time and sensors to: - examine differential M170 responses in peak occipital sensors - confirm that the paradigm successfully elicits M170 due to the use of face stimuli - examine differential responses over time and across sensors	The task successfully elicits an M170 component, but this is not sensitive to the conditioning manipulation. No condition differences in any sensors or time points
Is there evidence for theta power differential activity during conditioning and	- Increased theta power across the fear network and occipital	MEG: - Source level	- Monte- Carlo permutation across time, theta	- No evidence for conditioning in ROIs in any of

extinction across the fear network and sensory regions? Is there evidence for theta power differential activity during conditioning and extinction across the fear network and sensory regions?	regions for CS+ >CS- during learning. - Increased theta power in the lateral and medial OFC in response to CS- >CS+ during learning - Reduction of CS+>CS- differences during extinction in ROIs	time frequency decomposition in ROIs - Source level mean theta power across time in ROIs	frequencies in ROIs to detect potential time and frequencies that may be sensitive to learning and extinction. - LME modelling of mean theta power across time and frequencies to further examine potential condition differences. Mean power over time was computed since the permutation test did not show any significant time points or frequencies.	the frequencies or time points - No evidence for conditioning in ROIs in mean theta power
Is there evidence for differential power activity during conditioning and extinction in other brain regions and frequency bands?	Exploratory analyses with no concrete predictions	MEG source level:	Monte- Carlo permutation across time, frequencies, and brain regions to explore potential condition differences across the brain and frequencies	No evidence for conditioning across the whole brain and other frequency bands

Is there evidence for conditioning and extinction in pupil size?	Larger pupil size for CS+>CS- during learning, reduction during extinction	Pupil size	LME modelling to examine potential condition differences.	Overall pupillary constriction, potentially reflecting the pupillary light reflex. Greater constriction during acquisition than extinction
Is there evidence for conditioning and extinction in valence and arousal?	Lower valence and higher arousal for CS+>CS- during learning, reduction during extinction	Valence and arousal ratings	LME modelling to examine potential condition differences in habituation-baselined data.	Baseline valence differences that diminish following baseline correction, No valence differences during learning and extinction
			CLMM modelling on raw data to account for the ordinal nature of the data	Some evidence of conditioning in arousal ratings, reflected by a greater arousal during acquisition than extinction and for CS+ compared to CS- faces, but no interaction

3.2 Methods

3.2.1 Participants

Twenty English native speakers aged between 18 and 30 took part in the study (see Table 5 for demographic information). All participants had normal or corrected-to-normal vision, normal hearing, no metal on their body and no diagnosis of psychological or neurodevelopmental disorders. Pupil data from one participant is missing due to a technical problem during data saving. Participants were recruited from the University of Glasgow's Subject Pool. They provided a written informed consent to take part in the study and received £6 per hour for their time. The study was approved by the College of Science and Engineering ethics committee (300170261).

Table 5

Demographic information.

Sex	N (N Pupil*)	Mean Age	Age Range	Mean STAI-Trait (SD)	Mean STAI-State (SD)	Mean ERQ Reappraisal (SD)	Mean ERQ Suppression (SD)	Mean RSPM (SD)
Females	11 (10)	22	18-30	44.2 (7.5)	37 (9.3)	5.7 (0.7)	3.6 (0.9)	59.8 (18.4)
Males	9 (9)	21.2	18-26	37.7 (7.3)	29.3 (3.6)	5.2 (0.9)	3.4 (1.08)	75.9 (14.3)

3.2.2 Psychological Assessment

Participants were asked to complete a range of self-report measures of psychological functioning. These included, 1) The Emotion Regulation Questionnaire (ERQ),(Gross & John, 2003), 2) the Spielberger State Trait Anxiety Inventory (STAI), (Spielberger et al., 1983) 3) the Raven's Standard Progressive Matrices (RSPM), (Raven, 1941) providing a measure of non-verbal cognitive ability and 4) the Symptom Checklist-90 Revised (SCL), (Vaurio, 2011) which provides a measure of general psychopathology. These were included in a secondary analysis examining the potential mediating role of psychological factors on conditioned responding, although we found limited evidence for this (see Supplementary Materials 4). In addition, participants reported their demographic information, such as age sex and years of education.

3.2.3 Stimuli

Conditioned stimuli (CSs) were 36 neutral, frontal view faces (18 females) of White background, obtained from the Chicago Face Database (Ma, Correll & Wiitenbrink, 2015). Stimuli from the database are normed both in terms of physical and subjective properties of each facial identity (see Ma et al., 2015). The stimulus selection process ensured that female and male faces were similar in levels of attractiveness or emotionality (happiness, anger and fear). Stimuli were colour photographs scaled to 340 x 340 pixels. For each subject, 18 of the faces were randomly selected and assigned as CS+ while the remaining 18 served as CS-. The assignment of faces as CS+ and CS- was counterbalanced across participants. Specifically, for each block a different set of 4 CSs were randomly selected from the total of 36 faces. Two of these were randomly assigned to the CS+, and the remaining two – to CS- condition. This procedure was repeated 10 times, creating 10 stimulus sets for each block. An additional 10 sets were created by swapping the CS+/CS- assignment of the original 10 sets. Stimuli were presented on a mirrored projection screen with width of 80 cm and height of 65 cm, at a distance of 160 cm.

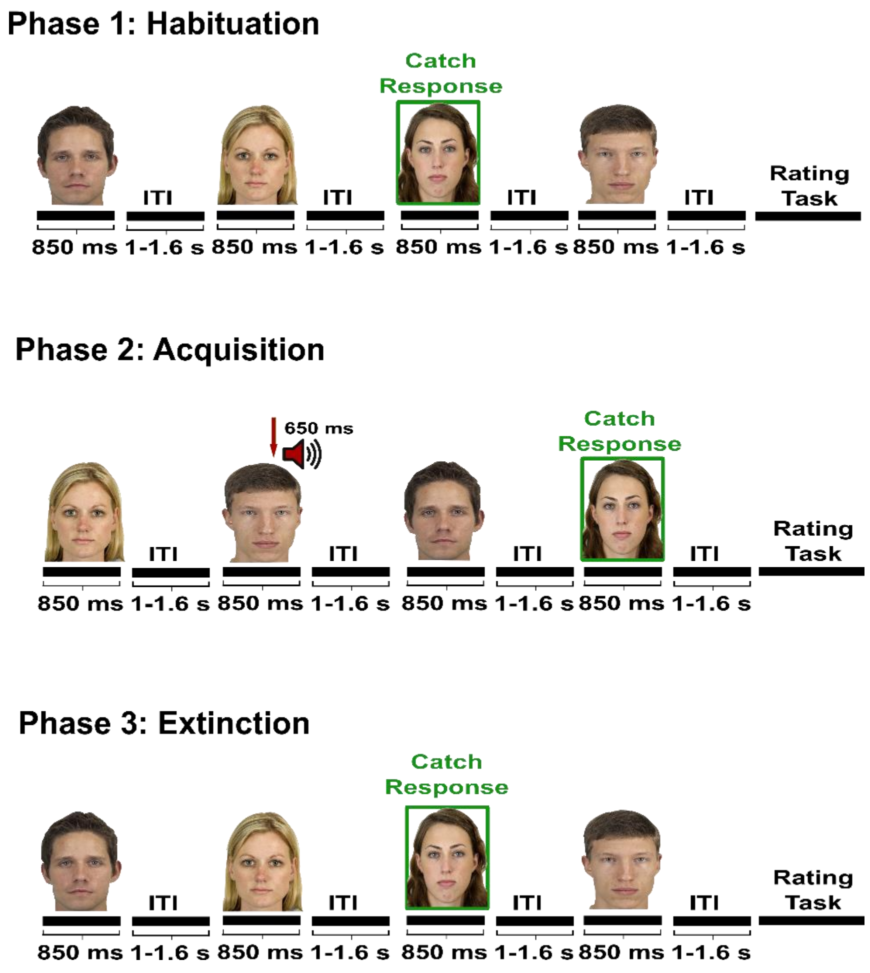
A 200 ms alarm sound of approximately 85-90 dB was used as the Unconditioned Stimulus (UCS). The UCS was selected based on a separate auditory norming study (N = 14, see Supplementary Materials 5 and https://osf.io/u6qza/ for results, stimuli, and task and analysis code). The UCS was delivered through 4m plastic tubes and earpieces with band pass frequency of 4 kHz.

3.2.4 Procedure

The experiment was conducted within two separate sessions. In session A, participants completed the psychological assessment measures and underwent T1-weighted MRI scan if they did not already have one available for access. In session B, participants completed the blocked conditioning task while MEG and eye tracking data were recorded. Session A was conducted at least two days before or following Session B, to avoid MRI-induced magnetic noise (J. Gross et al., 2013). The task was comprised of 9 blocks, each containing three experimental phases – Habituation, Acquisition, and Extinction (see Figure 10). Each block contained a different subset of 2 CS+ and 2 CS- faces, randomly selected from the total randomised set. Each CS was presented 9 times (a total of 162 trials per condition across blocks) on a gray screen, together with a black fixation cross, positioned at the centre of the image (the nose) for 850 ms. The inter-trial interval (ITI) had a duration of 1300 ms ±300 ms and was accompanied by a black fixation cross. To minimise ocular artefacts, participants were asked to maintain fixation at the centre of the screen at all times. Trial order was randomised across participants with the restrictions that the first trial was always a CS- and no more than two trials of the same stimulus type (e.g., a CS+) could occur consecutively. To maintain subjects' attention during each block, participants were required to respond to two catch trials by pressing a button on the response pad. Catch trials were two additional faces surrounded by a green frame, randomly selected from a total of 8. Participants were instructed that they will be presented with a series of faces that they have to view while maintaining fixation at the centre of the screen and respond to faces surrounded by a green frame. During the acquisition phase, CS+ trials were paired with the UCS, which occurred 650 ms post CS+ onset for 200 ms, while CS- trials were never paired with the UCS. During the habituation and extinction phases, CSs were presented alone without the UCS. At the end of each phase, participants completed a face rating task where they were asked to rate each CS on valence and arousal using an 8-point Likert scale (1 not at all pleasant/arousing to 8 extremely pleasant/arousing). At the end of the Acquisition phase, participants also rated the UCS on valence and arousal. Note that for arousal, participants were instructed to only rate stimuli in terms of the negative aspects of arousal associated with feelings of fear or unpleasant experience. Task scripts are available at https://osf.io/nxt68/.

Figure 10

Visual Blocked Conditioning block example.



Note. Permission to re-use the sample of four images from the CFD (Ma et al., 2015) was obtained from the copyright holder, the University of Chicago, Center for Decision Research.

3.2.5 Data acquisition

Pupil response was recorded using EyeLink 1000 long-range eye tracker. Data were recorded continuously during each trial presentation, with initial sampling rate of 1000 Hz. MEG data were acquired in a magnetically shielded room using a whole-head, 248-channel system (MAGNES® 3600WH, 4D-Neuroimaging, CA, USA). Prior to the MEG recording, five coils were attached to the participant’s head. These coils were then used to digitise the head shape (FASTRACK, Polhemus Inc., VT, USA) of each subject. This was done to allow for co-registration with participants' T1-weighted MRI (3D MPRAGE) as well as to monitor head position before and after each block.

3.2.6 Pre-processing

3.2.6.1 Pupil

Initial pre-processing was performed in Matlab 2017a using functions provided by Urai et al. (2017) in combination with the Fieldtrip Toolbox (Oostenveld et al., 2011). Raw .edf files were converted to .asc format (*edf2asc*) and data were reduced to trials with a length of -100 to 1700 ms with respect to CS onset. EyeLink-identified and peak-detected blinks (>3SD from mean pupil size) were padded with 200 ms on either side and linearly interpolated (*blink_interpolate*). Data were exported to R for further pre-processing. First, data were down sampled to 100 Hz by taking every 100 ms of data and discarding remaining data points. Any remaining missing data points were linearly interpolated using package *imputeTS*. Next, pupil data were log10 transformed and multiple linear regression was performed for each participant (log pupil size as outcome, and X and Y eye position as predictors) in order to remove small eye movement-related artifacts. The residual pupil size during each trial and time point was extracted from the regressions and used in subsequent analyses. Log10 pupil size change from baseline (mean pupil size between -0.1 and 0 s) was then calculated for each trial and time point using the formula *change* = *pupil size-baseline*. Finally, to obtain a measure of proportional change from baseline we calculated the inverse of the baselined pupil data. The resulting data were averaged across CS+ and CS- trials. To examine effects in relation to the conditioned response (CR), for each subject, data were averaged across the entire time window before UCS onset (0-0.6s)

3.2.6.2 MEG

Raw data were pre-processed using the Fieldtrip Toolbox (Oostenveld et al., 2011) in Matlab 2017a. Power line noise at 50 Hz was removed using a discrete Fourier transform. Environmental noise was reduced by performing principal component analysis (PCA) on the MEG reference channels using *ft_denoise_pca*. For each phase, CS+ and CS- epochs of – 0.9 to 0.9 ms with respect to CS onset were extracted from the continuous data. Data were then down sampled to 500 Hz. For each subject, eight channels producing excessive noise were removed. Trials containing a maximum amplitude above +/- 4pT were rejected following which PCA was performed to detect and remove subject-specific noisy channels and system-related artifacts. Cardiac and ocular artifacts were projected out (N= 2:5) of the data using a 50-step independent component analysis (ICA, *runica*). A final visual inspection was performed and trials containing a maximum amplitude of +/- 3pT were removed. Sensors that were discarded during any of the pre-processing stages were repaired using spline interpolation (*ft_channel_repair*).

3.2.6.3 Source estimation

MEG-MRI co-registration was performed using the subject's digitised head shape and landmark information (nasion and peri-auricular points). Gray and white matter, and deep brain structures (the amygdala, HPC and thalamus) were automatically segmented using Freesurfer (Dale et al., 1999). Source estimation was conducted using the Brainstorm Toolbox (Tadel et al., 2011). A mixed surface (~18,000 vertices) was created for each individual by merging the segmented cortical and subcortical structures. The segmented cortical and subcortical structures were downsampled (~15,000 and ~3,000 vertices respectively) and merged into a mixed surface (~18,000 vertices). To compute the head (forward) model, a whole-brain volume was then created using an overlapping sphere model in which a local sphere is fitted for each sensor. We used a constrained approach for both cortical and subcortical structures. The inverse model was calculated for each trial using depth-weighted Minimum Norm Estimation (MNE). To avoid contamination caused by slow shifts in the data, the noise covariance matrix was computed for each block separately from the baseline window of -900 ms to -2 ms. An exploratory source estimation using LCMV beamforming was also performed to ensure the consistency of results across source reconstruction methods (see Supplementary Materials 9). The results from the LCMV beamforming revealed comparable findings to those derived from the MNE.

3.2.6.4 Sensor level event-related fields analysis

Since the paradigm employed neutral faces as CSs, we expected to observe task-related activity at sensor level, in channels reflecting visual activity. To examine this, we computed sensor level Event-related fields (ERFs) between 0.01 and 20 Hz. We chose 20 Hz as the upper limit as this frequency limit has previously been employed in MEG conditioning studies using face CSs (Balderston et al., 2014b). For each subject, we computed an average across trials in each condition. Since band-pass filtering typically results in signal smearing and onsets of effects can be distorted (see Rousselet, 2012; VanRullen, 2011 but also Supplementary Materials 6), data were trimmed to a time window that does not include the UCS (-0.64 to 0.64 s). This was done to ensure that any observed effects are not contaminated by signal associated with the sound. Averaged data were band-pass filtered (0.01-20 Hz) using a FIR causal filter. The filter was selected based on an exploratory analysis of filtering artifacts (see Supplementary Materials 6). To ease interpretation, planar transformation was performed prior to visualisation and statistical analysis.

3.2.6.5 Source-level time-frequency analyses

Trial-level sources derived from the Desikan-Killiany atlas were exported from Brainstorm to Fieldtrip. Prior to computing the time-frequency (TF) maps, trial-level source activity within each region was averaged over the two hemispheres (results were comparable in analyses without hemisphere averaging). Since accurate detection of low frequency oscillations requires several cycles within the analysis time window (i.e., ~1.5 – 2 seconds), and the trial length in our dataset is 1.2 s prior to computing the TF maps, each trial was zero-padded with 2 second on each side. TF analysis was performed for each trial, in each ROI. Similar to previous research examining subcortical activity, we used Morlet wavelets (Tzovara et al., 2019). Specifically, we used 5-cycle wavelets during a time window of -2.64 to 2.64 s with a 3 ms resolution and in frequencies between 1 and 120 Hz with a 1 Hz resolution. Similar to previous studies examining oscillations in subcortical structures in humans (Khemka et al., 2017; Tzovara et al., 2019), our confirmatory analyses focused on differences in theta power, specifically in the range between 1 and 8 Hz as this range has been shown to functionally correspond to the 4-10 Hz responses observed in rodents (Jacobs, 2014). In addition, we performed a set of exploratory analyses that examined the potential contributions of frequencies above 8 Hz.

3.3 Results

All scripts necessary to reproduce the analyses and results outputs are available at https://osf.io/nxt68/. Data are only available upon request, due to a section of the participant's information sheet, restricting public data sharing.

3.3.1 Sensor Level Event-related fields 0.01 – 20 Hz

Since there is evidence to suggest that conditioning in humans can be manifested within sensory regions (Dolan et al., 2006; Lithari et al., 2015; Moratti et al., 2006, 2017; Moratti & Keil, 2005; Rehbein et al., 2014), we examined potential condition differences in occipital sensor-level ERFs. Due to employing faces as CSs, we expected to also observe evidence for the M170 component, suggested to drive the encoding of facial stimuli (Liu et al., 2002).

Figure 11 shows the topography of the average ERFs across conditions, demonstrating clear activity in occipital sensors. Before examining potential condition differences, we baselined the data by subtracting the acquisition and extinction ERFs from those during habituation. Next, to quantify potential differences between CS+ and CS- trials, for each experimental phase, we conducted a two-tailed cluster-based Monte-Carlo permutation test across all sensors and during 0 – 0.65 s post CS onset. We used 2000 permutations and an alpha level of 0.025, and cluster alpha of 0.05. FDR multiple comparisons correction was applied in both the time and sensor domain. The tests revealed no significant differences (see Figure 12). The analysis was also repeated on the data prior to habituation baselining, and again no significant condition differences were observed during either acquisition or extinction.

Figure 11

Grand average ERFs across all conditions.

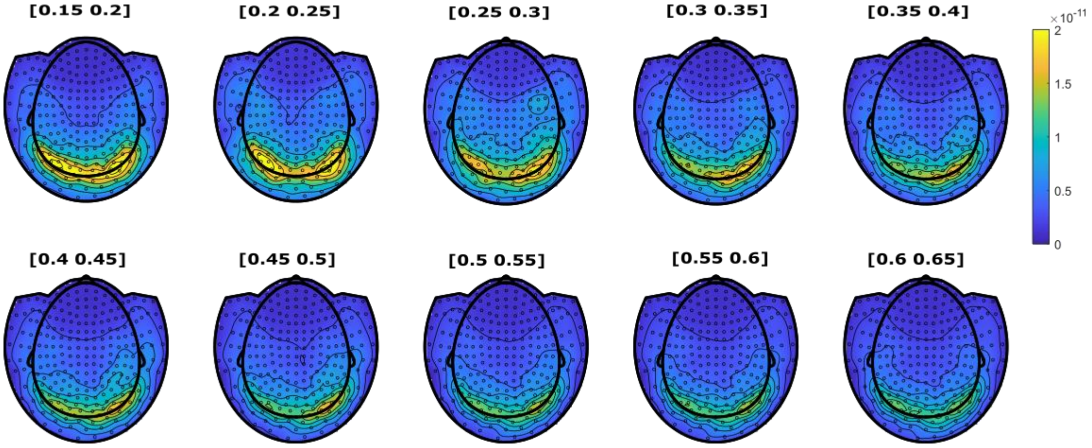
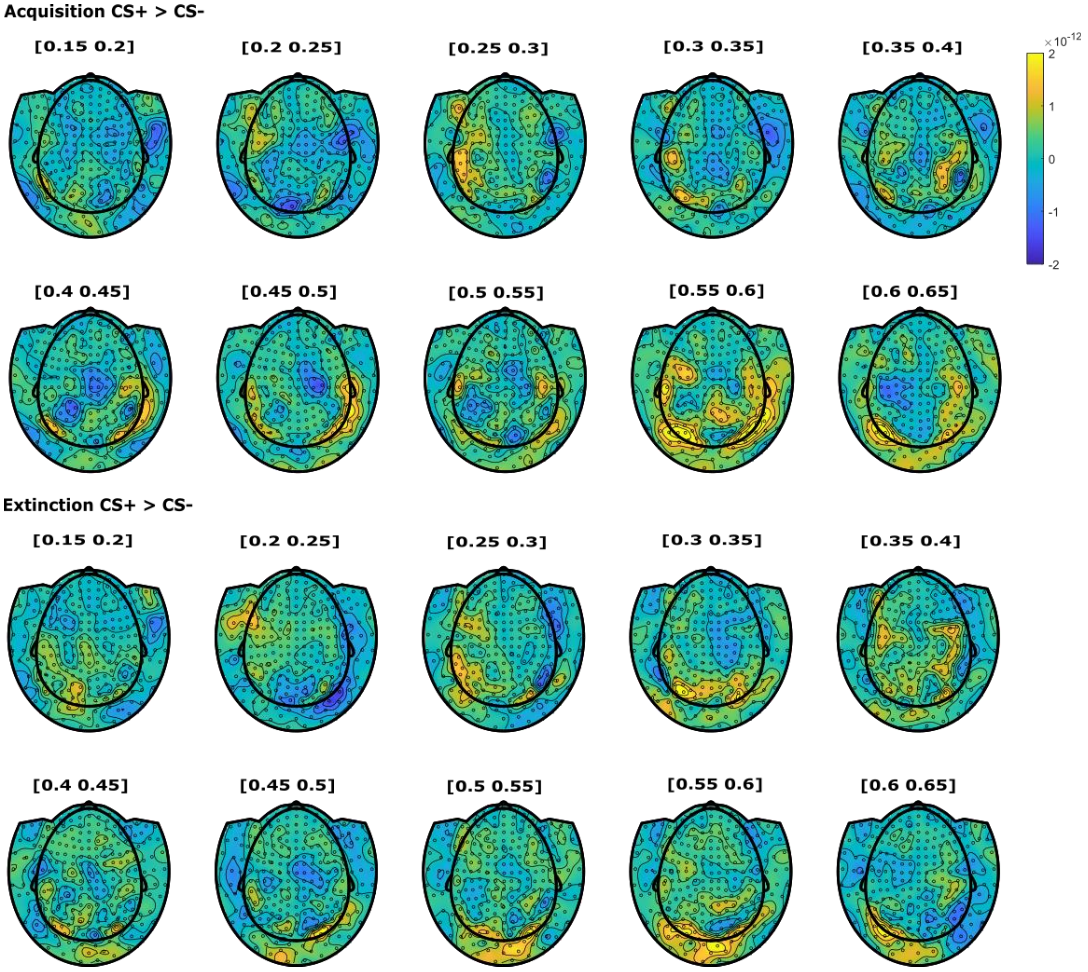


Figure 12

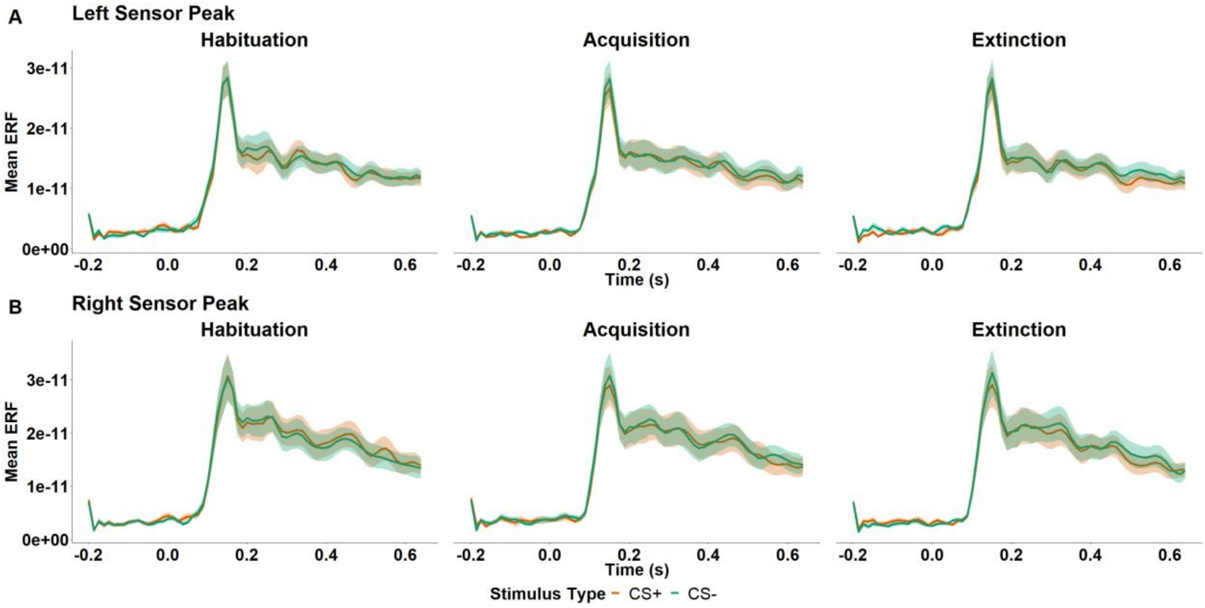
Habituation-baselined grand average ERFs, contrasting the difference between CS+ and CS- conditions during Acquisition and Extinction.



As seen in Figure 11, the peak sensor level activation is observed at around 200 ms. To examine this further, we extracted the ERFs from sensors that exhibited the highest activation in the grand average ERFs during the time window between 0.2 and 0.25 post CS onset. Across both the left (A162) and right (A188) hemispheres, these were located in the posterior region, likely reflecting occipital activity. As seen in Figure 13, across conditions and in both left and right hemispheres, the ERFs clearly peak just before 200 ms. This activation pattern is likely to reflect the M170 component. However, as seen in the figure, there are no indications of differences between conditions, evidenced by the clear overlap in ERFs and their CIs. This was confirmed in paired two-tailed permutation t-tests in latencies between 140 and 200 ms (2000 permutations, FDR correction in time domain, alpha = 0.025). This time window was selected as the M170 is suggested to occur within these latencies (Lueschow et al., 2015). The permutation tests were performed separately for Acquisition and Extinction for the left and right peak sensors.

Figure 13

Peak sensor grand average ERFs across conditions.



Note. The shaded area indicates the standard error of the mean.

These analyses suggest that there are no significant condition differences within the M170 component but also within any sensor or time point during the CS presentation. Nonetheless, we still pursued the source level analyses since reducing the dimensionality of the data by examining differential activity only within theta and specific ROIs, may increase the SNR sufficiently to detect conditioning.

3.3.2 MEG time-frequency analyses

We performed time frequency decomposition to examine the role of theta power in conditioning and extinction in ROIs. TF maps were averaged over trials, within conditions and similar to the ERF analysis, were baselined against the Habituation phase (see Figure 17, time frequency maps for each condition can be seen in Supplementary Materials 7). As seen in the figure, the average power within the theta range (2-8 Hz) in the caudal ACC is greater for CS+ than CS- trials during Acquisition and Extinction. The opposite pattern of lower theta power for CS+ trials is observed for most other regions, with the amygdala exhibiting most pronounced difference in theta power. A pattern of increased mean power in frequencies between 10 and 20 Hz is also observed for CS+ trials compared to CS- trials, in most ROIs.

3.3.2.1 Cluster-based permutation

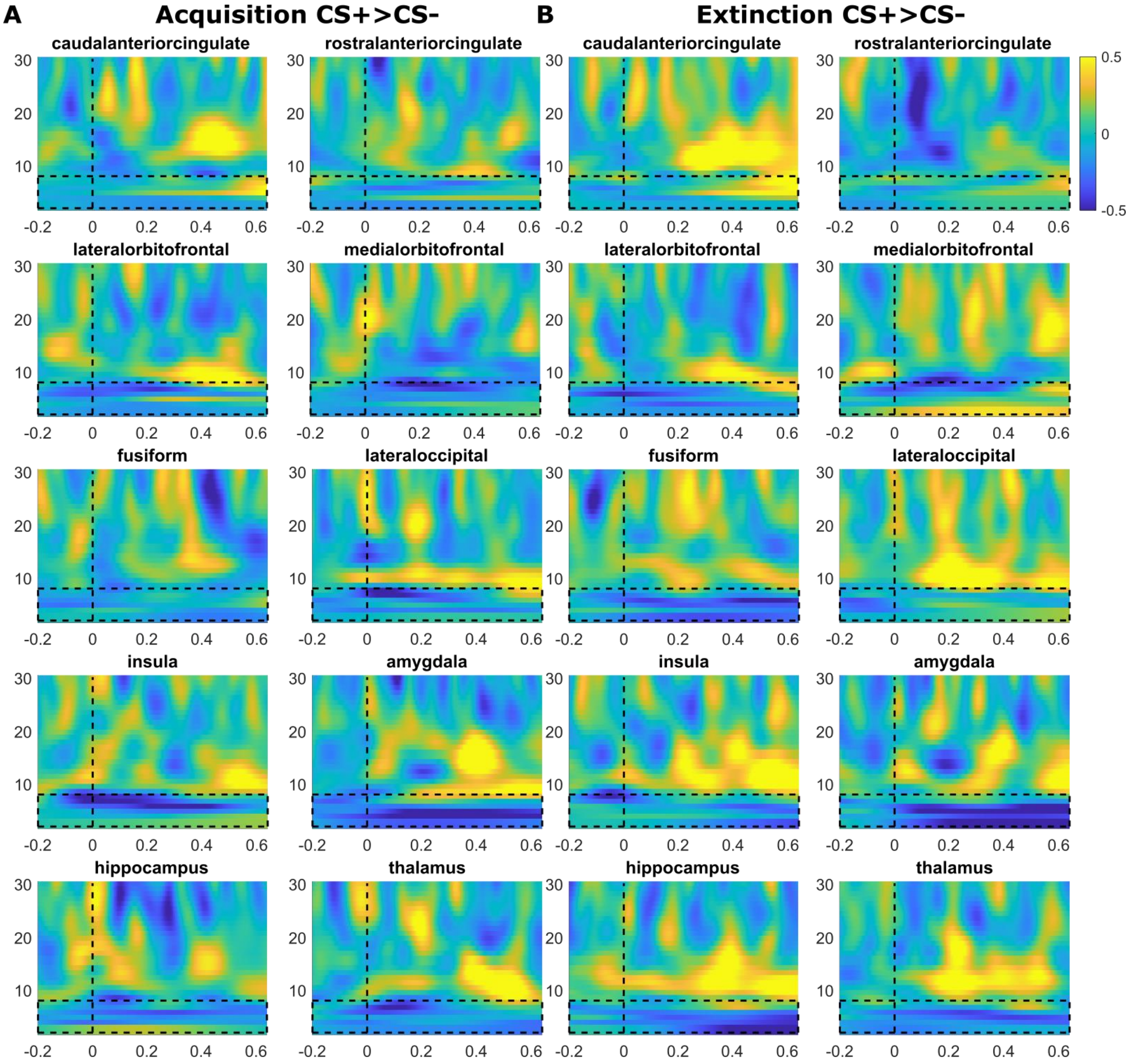
To identify potential time points and frequencies within theta that may be sensitive to the conditioning manipulation, we performed a confirmatory two-tailed Monte-Carlo permutation t-test in the ROIs, on the differences between CS+ and CS- conditions within Acquisition and Extinction. We looked at frequencies between 2 and 8 Hz during the time window of 0 to 0.64 s post CS onset. We used 2000 permutations, alpha level of 0.025 and applied FDR correction in the time, frequency, and signal domain to correct for multiple comparisons. The test revealed no statistical differences (see https://osf.io/nxt68/ for permutation outputs). Similar findings were obtained when computing permutation tests examining differences between CS+ and CS- using separate t-tests for acquisition and extinction and when analysing the TF data before habituation-baselining (see https://osf.io/nxt68/ for permutation outputs).

To examine potential differences in other frequency bands, our exploratory analysis focused on frequencies above theta (9 to 120 Hz), during the time window of 0 to 0.64 s. Again, we computed a two-tailed Monte-Carlo permutation t-test in the ROIs, on the differences between CS+ and CS- trials within Acquisition and Extinction. Similar, to the confirmatory analysis, no statistical differences were observed (see Supplementary Materials 8 for TF maps in frequencies above 30 Hz). We obtained similar results when computing TF maps on LCMV source reconstructed data (see Supplementary Materials 9).

Since the upper frequency bands of theta can sometimes overlap with the lower bands of alpha, we examined this further by computing a Fast Fourier Transformation in frequencies up to 20 Hz. This allowed us to gain a more refined understanding of the specific frequencies that may potentially be more sensitive to condition differences (see Supplementary Materials 10). Nonetheless, none of these analyses revealed any significant differences between conditions.

Figure 14

Habituation-baselined grand average time frequency maps, contrasting the difference between CS+ and CS- conditions during Acquisition and Extinction.



3.3.2.2 LME modelling of mean theta power

We performed linear mixed-effects (LME) modelling in R (package *lme4*) to further quantify any potential differences between conditions. We performed a separate model for each ROI, with mean theta power as the outcome variable. Initially, the goal of this analysis was to use the time windows and frequencies identified by the permutation test performed in section 2.8.2.1. However, since the permutation test did not identify any potential time windows of interest, we averaged power in each trial across all theta frequencies (2-8 Hz) and the entire trial time window (0 – 0.64 s) in ROIs. For each ROI, the models comprised of a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS Positive vs CS Negative) fixed effects design. The random effect structure included *Subjects* as random intercepts and by-subject random slopes for each main effect and the three-way interaction. During the pre-processing of the MEG data, some trials were manually and automatically removed due to excessive noise. However, during this process only information regarding the condition was retained. Since information about trial number and items was not retained in this dataset, we were unable to consider by-item random variability in our model. Main effects and interactions were assessed using a Type III Wald chi-square test. There were no significant effects at the level of 0.05 (see Table 6 and Figure 18B). Considering that even in its simpler form, the model revealed no significant differences, it is unlikely that including item random variability would have changed this pattern of findings. As seen in Figure 18A, consistent with the results from the permutation and the LME modelling, there is a significant overlap in the distributions between conditions within each experimental phase and ROIs.

Table 6

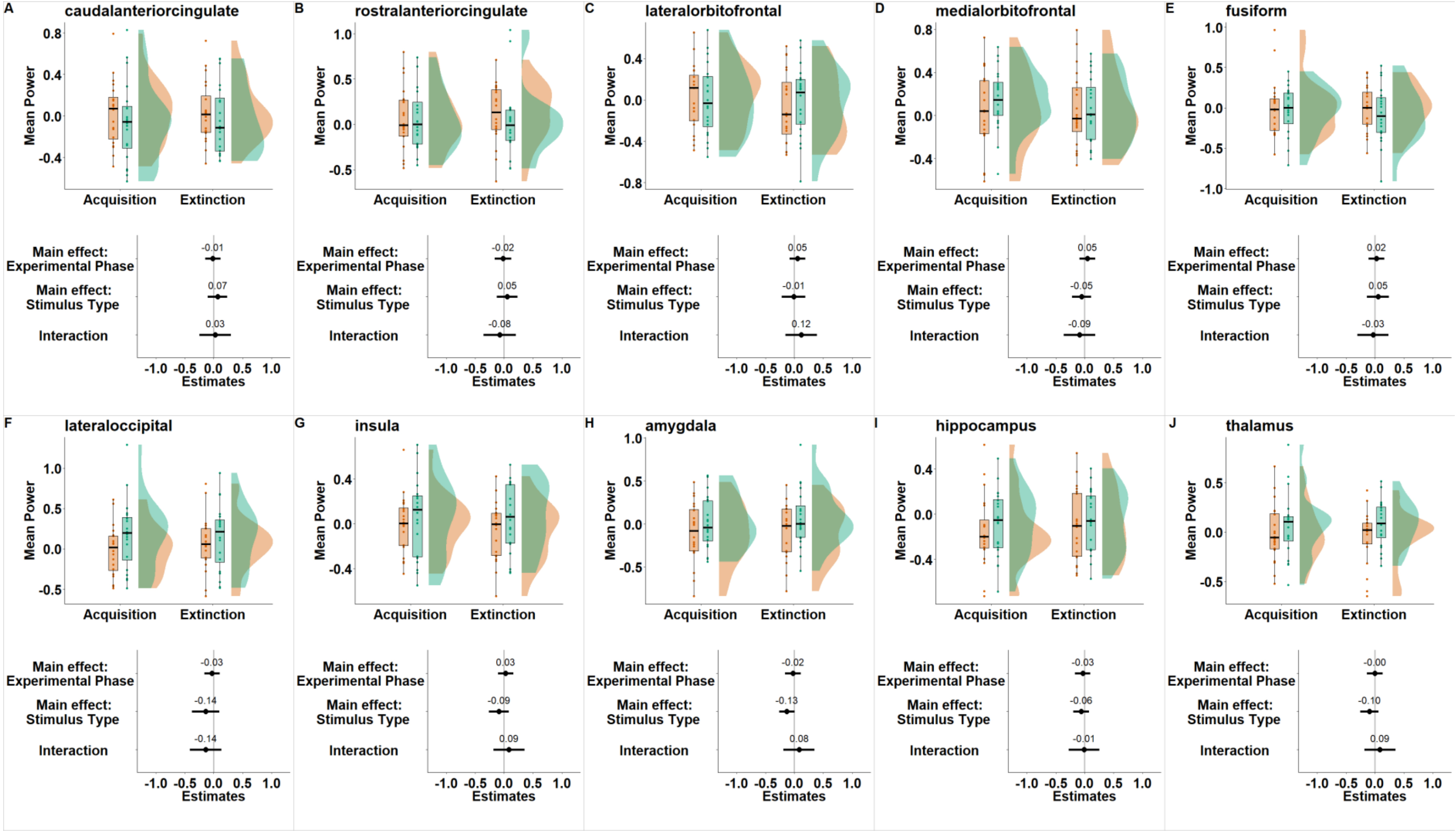
Type III Chi-square test for each of the fixed effects derived from the mean theta power LME mode within each ROI.

	Chisq	Df	P-value	R^2 Fixed
caudalanteriorcingulate				0.0001
Experimental Phase	0.044	1.000	0.834	
Stimulus Type	0.630	1.000	0.427	
Experimental Phase X Stimulus Type	0.035	1.000	0.851	
rostralanteriorcingulate				0.0001
Experimental Phase	0.073	1.000	0.788	
Stimulus Type	0.356	1.000	0.551	
Experimental Phase X Stimulus Type	0.283	1.000	0.595	
lateralorbitofrontal				0.0001
Experimental Phase	0.548	1.000	0.459	
Stimulus Type	0.014	1.000	0.907	
Experimental Phase X Stimulus Type	0.704	1.000	0.402	
medialorbitofrontal				0.0001
Experimental Phase	0.466	1.000	0.495	
Stimulus Type	0.417	1.000	0.518	
Experimental Phase X Stimulus Type	0.411	1.000	0.522	
fusiform				0.0001
Experimental Phase	0.127	1.000	0.722	
Stimulus Type	0.279	1.000	0.597	
Experimental Phase X Stimulus Type	0.058	1.000	0.810	
lateraloccipital				0.0004
Experimental Phase	0.170	1.000	0.680	
Stimulus Type	1.320	1.000	0.251	
Experimental Phase X Stimulus Type	0.992	1.000	0.319	
insula				0.0002
Experimental Phase	0.161	1.000	0.688	
Stimulus Type	1.049	1.000	0.306	
Experimental Phase X Stimulus Type	0.395	1.000	0.530	
amygdala				0.0003
Experimental Phase	0.132	1.000	0.716	
Stimulus Type	3.542	1.000	0.060	
Experimental Phase X Stimulus Type	0.341	1.000	0.559	
HPC				0.0001
Experimental Phase	0.261	1.000	0.610	
Stimulus Type	0.705	1.000	0.401	
Experimental Phase X Stimulus Type	0.006	1.000	0.938	

	Chisq	Df	P-value	
thalamus				0.0002
Experimental Phase	0.000	1.000	0.992	
Stimulus Type	1.407	1.000	0.236	
Experimental Phase X Stimulus Type	0.382	1.000	0.537	

Figure 15

Distribution of mean theta power of Habituation-baselined data and fixed effect estimates derived from the multiple regression models in ROIs.



3.3.3 Summary of MEG results

The results from the MEG analyses we performed revealed no indication of neural conditioning or extinction effects. This was the case for our primary analyses focusing on sensor-level ERFs and theta power within ROIs as well as for our secondary analyses exploring potential source level differences across the brain and frequencies above theta. These findings raise questions of whether this task can successfully induce conditioning, at least that is detectable at neural level.

3.3.4 Pupil size

To account for potential baseline differences during Habituation, the pupil size data during Habituation was subtracted from that Acquisition and Extinction. First, a mean pupil size across time and trials was calculated during Habituation. This was performed for each subject, block, and item separately. This baseline Habituation pupil size was then subtracted from each time point during Acquisition and Extinction. The average proportional pupil size change from baseline during the three experimental phases as well as in the habituation-baselined data can be seen in Figure 19. As seen in the figure, a clear UCR is seen shortly following UCS offset (~0.85 s) in the form of a larger mean pupil size in response to CS+ than CS- trials. It can also be seen that the pupil time course across the three-phased data for the duration of the CR (0-0.65 s) is characterised by constriction rather than dilation. However, minimal condition differences in pupil size are observed during Acquisition and Extinction across the three-phased as well as habituation-baselined data.

The development and extinction of the CR in the habituation-baselined data was examined inferentially in an LME model (package *lme4*). The model comprised of a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS+ vs CS) fixed effects design. The outcome variable was the mean pupil size over the duration of the CR (0 – 0.6 s post CS onset, see Figure 20A). This was computed for each subject, block and trial. The model included mean-centred contrasts (deviation coding) for the two categorical fixed effects. In order to account for random variation between subjects and items a maximal model was fitted as per recommendations provided by Barr et al. (2013). Specifically, *Subjects* and *Items* were added as random intercepts. Due to the within-items and within-subjects experimental manipulations, by-subject and by-item random slopes for each main effect and the interaction were also included (see Supplementary Materials 11 for random effects summary). A type III Wald chi-square test performed on the model (see Table 7 and Figure 20C for model estimates) revealed a significant main effect of *Experimental Phase* at the level of $p < 0.05$. Post-hoc analysis of the main effect performed using estimated marginal means contrasts (see Table 8) and Kenward-Roger method for degrees of freedom estimation (package *emmeans*) revealed a larger mean pupil size during Extinction than during Acquisition, significant at the level of 0.05 (t (19.8) = -2.65, p = 0.016).

Table 7

Type III Wald chi-square tests and R-squared values for the pupil dilation model and each of the fixed effects.

	Chisq	Df	P-value	R^2 Fixed (CI)
				0.001 (0 – 0.003)
Experimental Phase	8.34	1.000	0.004	0 (0 – 0.001)
Stimulus Type	0.08	1.000	0.78	0 (0 – 0.001)
Experimental Phase X Stimulus Type	0.41	1.000	0.52	0 (0 – 0.003)

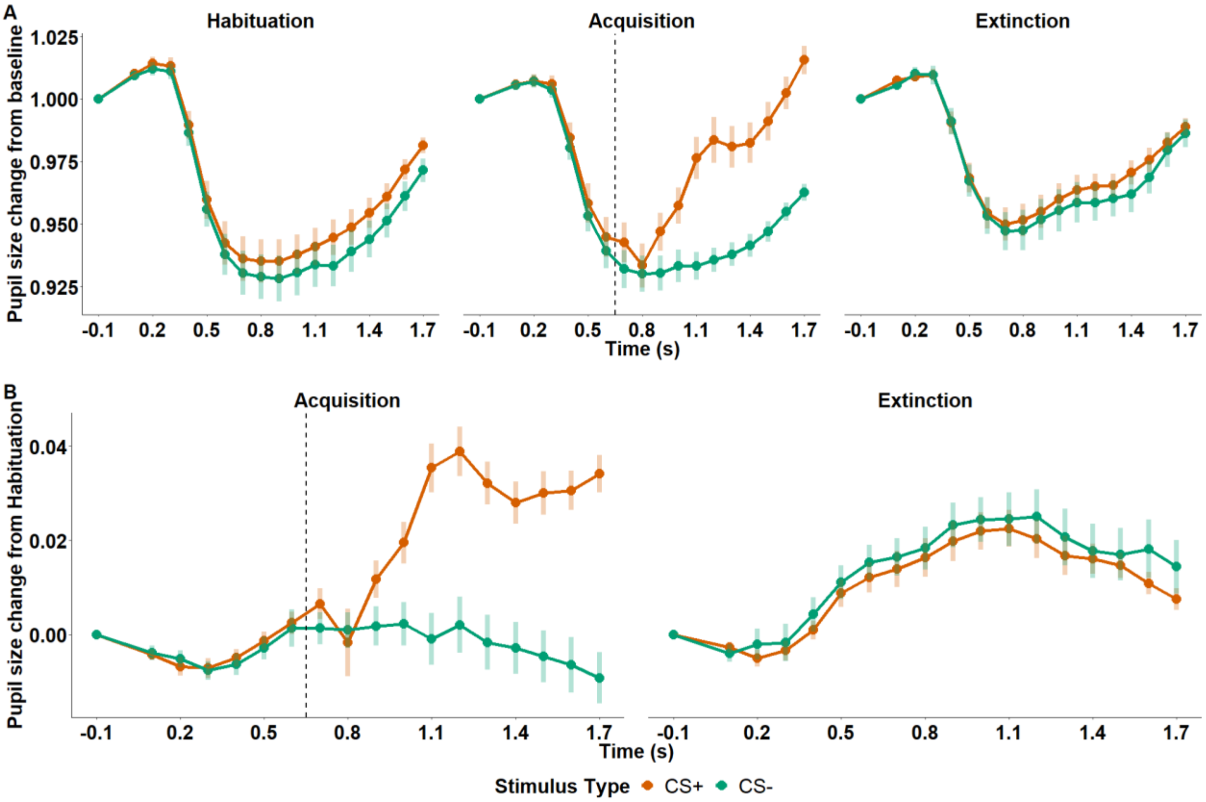
Table 8

Estimated marginal means and related contrasts derived for the pupil model.

Estimated Marginal Means						
Experimental Phase	Estimate	SE	df	Lower CI	Upper CI	
Acquisition	-0.005	0.002	13.7	-0.009	-0.001	
Extinction	0.001	0.002	13.9	-0.003	0.004	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
Acquisition - Extinction	-0.006	0.002	19.8	-0.01	-0.001	0.016

Figure 16

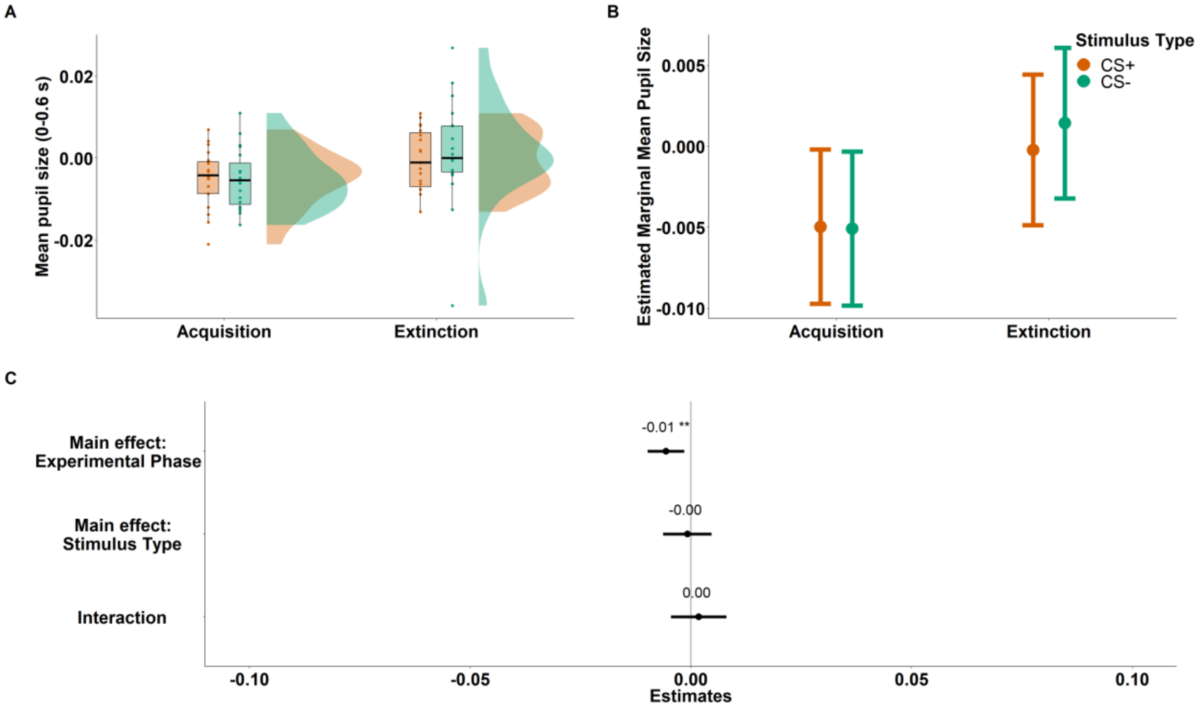
Proportional mean pupil size over time.



Note. The vertical bars indicate the standard error of the mean. The vertical dashed line indicates the US onset. A) Change in mean pupil size from baseline across Habituation, Acquisition and Extinction. B) Habituation - baselined mean pupil size. Acquisition and Extinction time courses reflect the difference between Acquisition and Habituation and Extinction and Habituation, respectively.

Figure 17

A summary of pupil size fixed effects



Note. A) Distribution of mean pupil size between 0 and 0.6 s post CS onset of Habituation-baselined data. B) Estimated marginal mean pupil size derived from the pupil model. C) Fixed effect estimates derived from the pupil model.

3.3.4.1 Summary of pupil size results

The findings from the pupil size analyses provided poor evidence for conditioning or extinction at a pupil level. When visualising the three-phased data, we observed an overall pupil constriction across conditions during the first second post-CS onset. When examining the data following habituation baselining, the constriction remained evident only during acquisition. This was also reflected in the results from the LME modelling, which showed an overall larger (and constricted) mean pupil size during extinction compared to acquisition. The observed constriction pattern may be suggestive that the trial duration was not sufficiently long to allow for the pupil to dilate.

3.3.5 Valence and arousal ratings

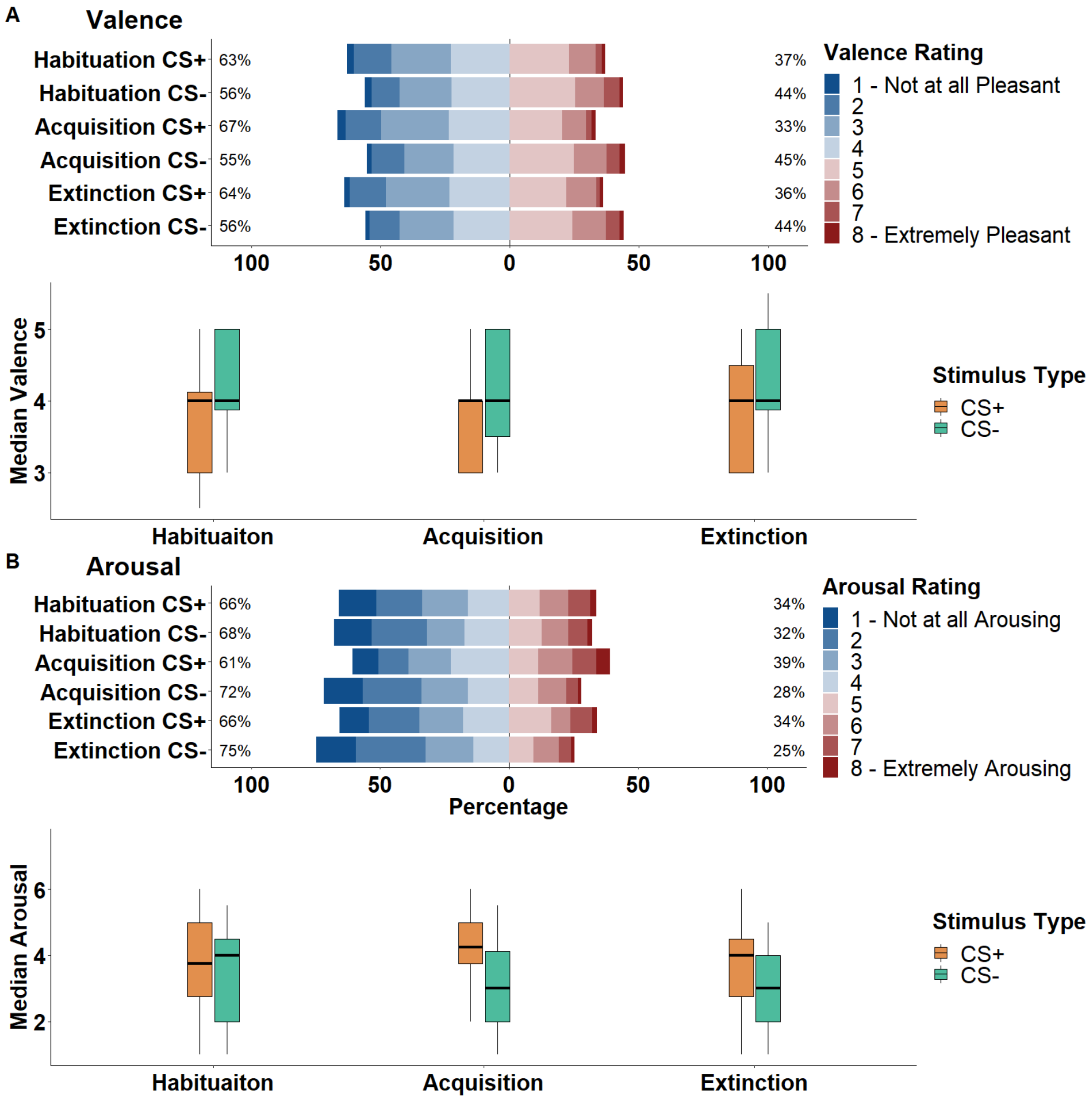
To examine potential valence and arousal related CRs, we fit two sets of mixed models to the rating data. Similar to the pupil size data, the analyses reported here was performed on the Habituation-baselined data. To derive this baselined data, each item's rating during Habituation was subtracted from that item's rating during the Acquisition and Extinction. Valence and arousal ratings were predicted using an LME model, consisting of a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS+ vs CS-) fixed effects design. Mean-centred contrasts were used for the two categorical fixed effect predictors. By-subject and by-item random intercepts were added, together with by-subject and by-item random slopes for both main effects and the interaction (see Supplementary Materials 11 for random effects summary). Main effects and interactions were assessed using Type III Wald chi-square tests. Since the valence and arousal data were measured on an ordinal scale (1-8), a second set of analyses were performed on the three-phased dataset using cumulative link mixed (CLMM) models (package *ordinal*), (see Supplementary Materials 12). These models revealed findings comparable to LME modelling findings.

When descriptively examining valence ratings, a potential indication of baseline differences can be seen, with 44% of ratings in response to CS- faces belonging to scores of 4 and above, in comparison to 37% in response to CS+ faces (see Figure 21). In other words, during Habituation, at least descriptively, CS+ faces were perceived as less pleasant than CS- faces. This pattern is maintained across Acquisition and Extinction. When condition differences were assessed inferentially using LME modelling on the Habituation-baselined data, there were no fixed effects significant at the level of 0.05 (see Table 9 and Figure 22). However, differences were present in the CLM modelling of the three-phased data in the form of a main effect of stimulus type, confirming the presence of baseline valence effects (see Supplementary Materials 12). The potential source of these baseline differences was explored in Supplementary Materials 13. These analyses revealed that factors such as stimulus sex as well as normative ratings of attractiveness and perceived anger of the neutral faces influenced valence ratings, however, these did not mediate the condition differences.

In terms of arousal ratings, compared to Habituation there was a slight increase in the percentage of responses belonging to ratings of 4 and above (39%) for CS+ compared to CS- faces during Acquisition. The difference between CS+ and CS- faces during Acquisition (11%) is also larger than that during Habituation (4%) and slightly larger than that during Extinction (9%), (see Figure 18). The results from the LME model (see Table 9 and Figure 23), however, showed that only the main effects of *Experimental Phase* and *Stimulus Type* reached statistical significance at the level of 0.05. Post-hoc contrasts revealed that CS+ faces were rated as more arousing than CS- faces and that faces were rated as more arousing during Acquisition compared to Extinction (see Table 10). A similar pattern of results was observed in the CLM modelling (see Supplementary Materials 12).

Figure 18

Valence and arousal ratings.



Note. A) Valence and B) Arousal ratings. Top: Percentage of responses belonging to each of the 8 response categories. Bottom: Median ratings across subjects.

Table 9

Type III Wald Chi-square tests and R-squared values for the complete valence and arousal model and each of the fixed effects.

	Chisq	Df	P-value	R^2 Fixed (CI)
Valence				
Full Model				0.001 (0 – 0.01)
Experimental Phase	0.38	1.000	0.95	0 (0 – 0.005)
Stimulus Type	0.39	1.000	0.53	0.001 (0 – 0.006)
Experimental Phase X Stimulus Type	0.55	1.000	0.46	0 (0 - 0.005)
Arousal				
Full Model				0.05 (0.03-0.07)
Experimental Phase	8.83	1.000	0.003	0.01 (0 .003– 0.025)
Stimulus Type	6.48	1.000	0.01	0.035 (0.02 – 0.06)
Experimental Phase X Stimulus Type	2.54	1.000	0.11	0.003 (0 – 0.01)

Table 10

Estimated marginal means and related contrasts derived for the arousal model.

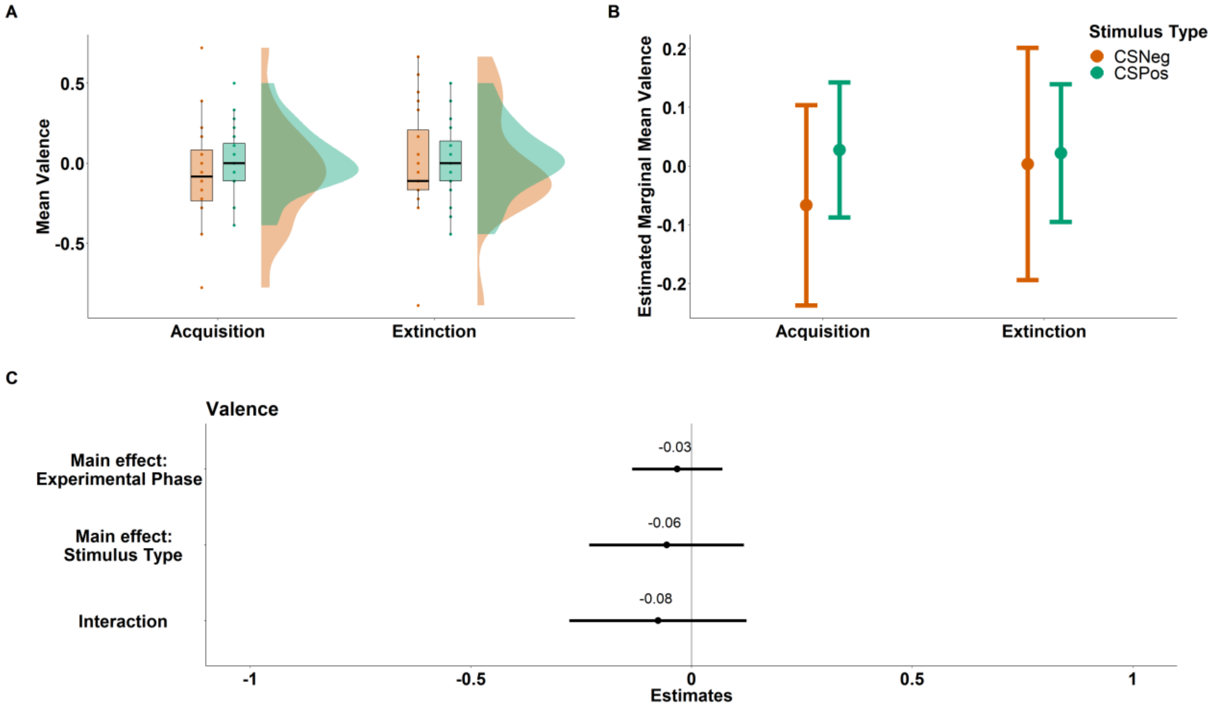
Stimulus Type Estimated Marginal Means						
Stimulus Type	Emmean	SE	df	Lower CI	Upper CI	
CS+	0.23	0.16	21.3	-0.11	0.57	
CS-	-0.26	0.08	26.6	-0.43	-0.09	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
CS+ - CS-	0.49	0.19	21.4	0.09	0.88	0.02

Experimental Phase Estimated Marginal Means						
Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	
Acquisition	0.12	0.10	21.71	-0.08	0.33	
Extinction	-0.15	0.09	23.03	-0.35	0.05	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
Acquisition – Extinction	0.28	0.09	17.44	0.08	0.47	0.01

Note. Contrasts were computed using Kenward-Roger method for degrees of freedom approximation.

Figure 19

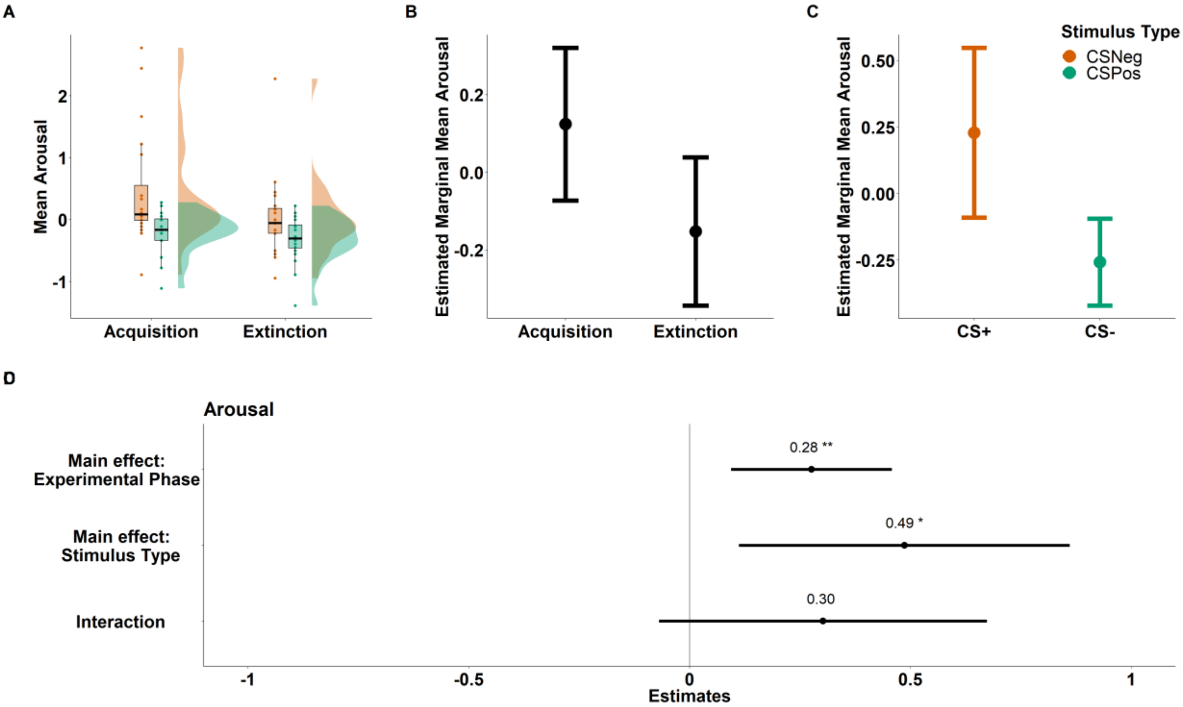
A summary of valence fixed effects.



Note. A) Distribution of mean valence ratings of Habituation-baselined data. B) Estimated marginal means per condition derived from the linear mixed effects model of valence ratings (error bars represent 95% CIs for the means conditioned on the random effects). C) Fixed effect estimates (labelled dots) derived from the linear mixed effects model of valence ratings; bars represent 95% CIs for the estimates.

Figure 20

A summary of arousal fixed effects



Note. A) Distribution of mean valence ratings of Habituation-baselined data. Estimated marginal means from the linear mixed effects model of arousal ratings for B) Experimental Phase and C) Stimulus Type (error bars represent 95% CIs for the means conditioned on the random effects). D) Fixed effect estimates (labelled dots) derived from the linear mixed effects model of arousal ratings; bars represent 95% CIs for the estimates.

3.3.5.1 Summary of rating data results

The visualisation and analysis of valence effects in the three-phased data (see Supplementary Materials 12) indicated baseline differences during habituation that perpetuated during acquisition and extinction. When habituation-baselining was performed, we observed no significant valence differences. We found some evidence for conditioning in the arousal ratings, evidenced by an increase in arousal for CS+ than CS- faces and during acquisition compared to extinction. However, we observed no significant interaction between Experimental Phase and Stimulus Type.

3.4 Discussion

In this study, we report findings from a novel conditioning paradigm that aimed to maximise the reliable detection of oscillatory dynamics in cortical and subcortical structures underlying threat learning and extinction. In addition, we measured conditioned responses not only at neural level but physiologically (pupil size) and behaviourally (subjective valence and arousal ratings). Our findings indicate that at its current form, the paradigm does not evoke reliable signatures of conditioned responding. Nonetheless, the results from the present study are informative in highlighting several factors that can potentially enhance the detection of CRs when using tasks with a large number of trials.

3.4.1 MEG

In our initial analyses, we investigated the extent to which the visual blocked conditioning can elicit differential activity at sensor level, by performing cluster-based permutation across time and sensors. This analysis revealed no significant condition differences. In addition, we aimed to confirm that the paradigm can reliably elicit the M170 component that underlies face encoding. As expected and similar to other E/MEG studies employing face CSs (Balderston et al., 2014b; Dolan et al., 2006; Rehbein et al., 2015; Sperl et al., 2021; Stolz et al., 2019), we observed a clear M170 in peak occipital sensors around 200 ms post-CS onset across conditions. However, the M170 was not sensitive to the experimental manipulation. Since across the literature, the M170 has not been consistently shown to index a CR, it is possible that detecting a CR in this component is dependent on specific experimental and procedural parameters. Yet, it is also highly likely that conditioning in this task was simply not established considering that none of our MEG analyses detected any evidence for a CR.

Our source level analyses aimed to determine the role of theta oscillations in learning and extinction by performing trial-level time frequency decomposition. To quantify potential condition differences, we employed cluster-based permutation to identify time windows that may be sensitive to the experimental manipulation, in ROIs implicated in threat processing. In addition, we examined the average theta power differences during the entire trial duration using LME modelling. The results from these analyses, however, did not reveal any differences in ROIs. While Yet, descriptively, we observed a similar activation patterns in deep structures to that reported recently in the amygdala and the HPC (Tzovara et al., 2019). For instance, similar to Tzovara et al. (2019) our secondary analyses (see Supplementary Materials 10) examining the power distribution within frequencies below 20 Hz, showed that oscillatory power was highest in the lowest frequencies, between 2 and 4 Hz. However, in our data this pattern was maintained across conditions without any substantial differences. In addition, we observed a small peak in power in frequencies between 8 and 12 Hz across all conditions, although these findings cannot be compared against the data provided by Tzovara et al. (2019) since their mean power analysis only focused on frequency bands up to 8 Hz.

Furthermore, in line with recent findings (Tzovara et al., 2019), at a descriptive level we found that the average theta oscillatory power was lower in response to the CS+ compared to the CS- in deep structures. Such pattern contrasts the consistently observed increase in theta power to the CS+ in the rodent literature (Karalis et al., 2016; Lesting et al., 2011; McCullough et al., 2016). This discrepancy can partly be explained by procedural differences in human and animal investigations. For example, rodent studies use a wide range of protocols to provide a control condition to the CS+, including between and within-subject designs, fixed and pseudo-random CS order presentations, and differential cue protocols that typically present the CS+ and CS- on separate days (Haaker et al., 2019). Ultimately, however, these procedures rarely elicit multi-process competition. In contrast, human conditioning research largely relies on differential cue protocols that create an environment in which aversive and safety learning processes may compete. While still highly speculative, observing a higher theta power to the CS- in deep sources may be qualitatively similar to the patterns reported in fMRI studies in which activation in the mPFC is greater in response to the CS-, reflecting safety learning (Harrison et al., 2017). The presence of these competing processes can engage different learning mechanisms to those observed in rodent research, which in turn can complicate the direct comparison between animal and human findings (Haaker et al., 2019). Considering that human conditioning paradigms are also potentially far less aversive than those used in rodents for ethical reasons, it is possible that variations in the employed task parameters can lead to distinct patterns of theta oscillations in humans and animals. Another possibility is that the functional properties of oscillatory activity within the theta range during learning may differ in humans and rodents. Yet, this may be difficult to determine without first disentangling the impact of cross-species procedural differences.

Procedural variation can also complicate the direct comparison between the results from the present study and that by Tzovara et al. (2019), for a number of reasons. First, the results provided by Tzovara et al. (2019) were based on analyses of data that were not baselined with respect to a habituation phase. In contrast, we only observed lower theta power to the CS+ in deep structures in the habituation-baselined data, which was performed to account for potential leakage from other sources and for baseline differences. When only considering, non-habituation baselined TF activity, the opposite pattern was observed, at least in the amygdala and the HPC (see Supplementary Materials 7). These findings suggest that the direction of results can shift substantially in the presence of a baseline habituation procedure, although it is unclear whether this change may be driven by baseline differences or due to the leakage correction. Furthermore, our analyses included a substantially higher number of ROIs than Tzovara et al. (2019) and a less precise deep source estimation procedure, both of which would reduce the likelihood of detecting significant differential activation.

3.4.2 Physiological and behavioural measures

With respect to the pupillary signatures of conditioning, when examining the three-phase data it was evident that the pupil size time course across conditions was characterised by constriction rather than dilation. This was likely driven by the short trial duration (0.65 s) in the present design which would have prevented the full resolution of the pupil (baseline - constriction – dilation - baseline). In standard pupillometry studies, the pupil size typically peaks at around 1 s after stimulus onset. In cases where multiple stimuli are presented in close succession, peak latencies are even larger (van Rij et al., 2019). Therefore, it is likely that the observed pupil constriction reflects only the initial stage of the pupil resolution cycle, specifically the pupillary light reflex (i.e., the rapid pupil constriction in response to light (Becket Ebitz & Moore, 2017). Following habituation baselining this overall constriction was only observed during acquisition and evidenced by a main effect of experimental phase whereby pupil size was larger during extinction than acquisition. This is an opposite pattern to the initially expected and typically observed increase in pupil size during acquisition (Jentsch et al., 2020; Kluge et al., 2011; Korn et al., 2017; Leuchs et al., 2019; Tzovara et al., 2018). Considering the identified issue of trial duration, this effect is difficult to interpret.

Partly in line with our predictions, the analysis of arousal ratings, revealed an overall higher arousal for CS+ than CS- ratings as well as higher arousal during acquisition than extinction. However, the interaction between the experimental phase and stimulus type was non-significant. These findings provide some evidence that conditioning may have taken place. This was evidenced by the presence of a CR that potentially did not fully habituate during extinction and by the acquisition phase itself being perceived as more arousing. In terms of valence, there were no significant effects during acquisition or extinction when ratings were baselined with respect to the habituation. There were, however, indications of baseline differences leaking into the experimental phases when descriptively examining the data before baselining. These differences were corroborated by the CLM modelling (see Supplementary Materials 12) but diminished in the habituation-baselined analysis. A secondary analysis attempted to better understand the sources of variability in valence during Habituation (see Supplementary Materials 13). We found limited evidence that the counterbalancing procedure influenced the baseline effects. However, we showed that the ratings of attractiveness and perceived anger of faces provided in the normative database of the stimuli, correlated with the valence ratings obtained in the present study. Specifically, there was a positive correlation between valence and attractiveness and a negative correlation between anger and valence. Female faces were also rated as more pleasant than male faces. However, these factors did not moderate the difference in valence between CS+ and CS- faces. According to these findings, the faces used in the present study were not perceived as completely neutral which could explain the variation in valence ratings. Therefore, despite controlling for baseline effects, it is still possible that the initial variation in the perceived valence of the stimuli influenced the formation of the CR across outcome measures.

3.4.3 Methodological considerations

This study aimed at increasing the likelihood of detecting conditioned responding in deep sources by paying special consideration to both design and analytical aspects. Our study employed a blocked conditioning task in which learning and extinction were established in multiple consecutive blocks. This aimed to maximise the SNR by obtaining the large number of trials required for deep structure detection (Quraan et al., 2011a), while increasing the likelihood of establishing contingency awareness by presenting a small number of unique CSs in each block. In terms of source reconstruction, we used a combination of techniques that have been shown to be successful at detecting subcortical activity using MEG (Balderston et al., 2014b; Dumas et al., 2013). We derived realistic anatomical information of limbic structures, through using anatomical segmentation from each participant's MRI, and combined those with the participant's cortical surface. Next, we used depth-weighted MNE to reconstruct sources in our ROIs which has the advantages over both MNE and beamforming in that it does not assume lack of temporal correlation between sources and increases sensitivity in detecting subcortical activity (Attal et al., 2012). Finally, we reduced the impact of leakage from other sources through subtracting baseline activity during habituation from that during acquisition and extinction. As such, we can be relatively confident that we have maximised the accurate detection of subcortical activity to a reasonable level and provided some control over source leakage.

Our analyses also offered a more generalised account of conditioning effects, especially at a behavioural and pupil level where we modelled both by-subject and by-item random variability. Unfortunately, however, we were unable to model random by-item variation in our MEG average theta power analysis, since item information during data pre-processing was not retained. Considering that the non-maximal model we used failed to reveal any statistically significant effects, it is unlikely that a more complex random structure including by-item random variability would have caused a substantial change in the findings. Consequently, due to the time-consuming nature of the pre-processing pipeline, we opted against the re-analysis of the raw MEG data.

While our MEG and pupil results showed little evidence for the acquisition of a CR, a clear but small increase in self-reported arousal was observed for the CS+ compared to the CS-. This suggests that the paradigm may indeed be capable of inducing conditioning but that any effects were too small to be reliably detected in noisy psychophysiological measures. Since this is the first time a blocked conditioning paradigm has been used to investigate fear conditioning, it is important to consider the factors that may drive the observed small effects. For example, it is possible that the magnitude of the CR differed over blocks. This may be the result of a CR habituation over blocks due to the repetitive exposure to the conditioning task. Alternatively, although less likely, the repetitive elicitation of conditioning could result in a learning effect whereby CRs become stronger over blocks. In either case, the overall CR we observed would be significantly reduced by the presence of blocks with minimal condition differences. Our supplementary analyses (see Supplementary Materials 14) examining the pupillary and behavioural patterns for each block, however, revealed limited evidence for the presence of learning or habituation effects.

We also found no evidence to suggest that the poor CR acquisition was driven by low aversiveness of or habituation to the UCS. In particular, unlike previous studies using a UCS that was not validated against the perceived aversiveness of other stimuli, the choice of an aversive stimulus in the present study was informed by a separate control study (see Supplementary Materials 5). Furthermore, participants in the present study showed no evidence of response habituation to the UCS as they consistently rated the UCS as unpleasant across blocks (see Supplementary Materials 15). This was also corroborated when examining the pupil time course at a block level, where the UCR was visible in all blocks and remained relatively stable (see Supplementary Materials 14).

Another candidate for explaining the lack of a CR at least at a pupil level is the trial duration. Studies employing classical conditioning in fMRI or psychophysiologically (pupil size or SCR), typically involve long trial duration (3-8

s) and ITI (7-17 s), (e.g. Jentsch et al., 2020; Korn et al., 2017; Tzovara et al., 2018) that enable the pupil to reach its peak dilation. However, to allow for the desirable large number of trials, the trial duration was reduced to 0.85 s in order to maintain an acceptable experiment duration. While trial duration can enhance the detection of a pupillary CR, previous studies using Multi-CS conditioning (Junghöfer et al., 2017; Rehbein et al., 2015; Steinberg et al., 2013) have demonstrated conditioning effects at both behaviourally (although not consistently) and in MEG, with a similar number of trials and trial duration. Similarly, the use of a short trial duration in EEG conditioning studies is not uncommon (Camfield et al., 2016; Ferreira de Sá et al., 2019; Pastor et al., 2015). Therefore, it is likely that other factors have contributed to the observed effects in the present study.

For example, it is possible, that inducing conditioning effects using a social stimulus such as a face is more difficult than when basic stimuli such as shapes and simple sounds are employed. Faces are highly complex and multi-dimensional stimuli and as such, involve the processing of multiple socially relevant components including identity, sex, age, emotion, attractiveness and gaze (Leopold & Rhodes, 2010; Rossion, 2014). As it has been shown (Carter et al., 2003), higher-level cognitive processes during conditioning can have a significant impact on the establishment of a CR, with greater levels of cognitive demand interfering with CR elicitation. While the present study did not involve additional cognitive or attention distracting tasks, the mere requirement for face discrimination in the presence of a weak effect may have been sufficient to hamper its detection.

A related contributing factor is that of attention and contingency awareness. As previously discussed, contingency awareness has been shown to be crucial for the process of associative learning (Mertens & Engelhard, 2020). An essential component facilitating the acquisition of contingency awareness is that of attention, with poor attention to the relationship between conditioned and unconditioned stimuli interfering with CR development (Weidemann et al., 2016). To reduce the impact of additional cognitive demands from a secondary attention-demanding task, the present paradigm employed a low cognitive load component to sustain participant’s attention during the experiment. In each block we implemented two *catch* trials in which participants responded in the presence of a face surrounded by a green frame. However, this manipulation was aimed at sustaining overall attention, and thus did not guarantee that participant attended to the relationship between stimuli. In addition, the present study relied on the assumption that presenting a limited number of unique CSs in each block will ensure contingency awareness, but our task did not explicitly measure participants' acquisition of contingency. Therefore, it is likely that the poor differentiation between CS+ and CS- stimuli in most of our outcome measures was driven by low contingency awareness and/or poor attention.

3.4.4 Conclusions

The present study lays the foundations for the development of a classical conditioning paradigm that can successfully and reliably measure psychophysiological and neural signatures of associative learning and extinction. Since these measures are inherently noisy and the detection of subcortical structures driving these processes in non-invasive human imaging is challenging, we paid particular attention to maximising source localisation through experimental design (large number of trials, validated aversive UCS) and analytical strategies (baseline subtraction, individual anatomical segmentation and source reconstruction suitable for deep structure detection).

While the paradigm did not elicit reliable neural and pupillary conditioning signatures, the establishment of a CR through subjective arousal ratings suggests that the blocked design of the task does indeed have the potential to elicit associative learning, but that any effects were too negligible to be detected via noisy physiological outcome measures. The results from the present study suggest that the blocked design was not a primary cause for the observed small effects and highlight a number of other design parameters that require further consideration in order to enhance the detection of CRs when using a large number of trials. Future work should focus on optimising the balance between number of trials and trial duration, if a pupillary CR is to be studied. The implementation of simpler CSs such as basic shapes or tones could potentially enhance CR development by eliminating the confounding influence of complex social cognition processes. The use of simpler CSs can also reduce the likelihood of observing baseline subjective behaviour effects by reducing variability in multiple dimensions (i.e., emotionality and appearance). Alternatively, if face stimuli are utilised, these should be first normed independently to ensure that they elicit similar ratings of attractiveness and emotionality.

More importantly, a consideration should be paid in measuring and ensuring contingency awareness through sustaining participants' attention to the relationship between conditioned and unconditioned stimuli. A potential solution would be to include a low cognitive load task following each trial, where participants make simple perceptual judgements. Combined with the analytical and design strategies implemented in the current study, addressing these methodological issues should theoretically be sufficient to allow for the reliable measurement of the neural and psychophysiological signatures of conditioning.

3.5 Chapter 3 supplementary materials

3.5.1 Supplementary Materials 4: Relationship between conditioning measures and psychological self-report measures

We assessed potential relationship between measures of conditioning as well as with self -reported measures of psychological functioning. Specifically, we separately examined the variables that may moderate the interaction between experimental phase and stimulus type in predicting pupil size as well mean theta power.

The relationship between pupil size, valence and arousal during acquisition and extinction and self-reported anxiety, emotion regulation, general psychopathology and non-verbal ability can be seen in Supplementary Figure 5 and 6 in the form of scatterplots accompanied by Pearson's r correlations by stimulus type where applicable. Note that these correlations are provided for descriptive purposes only and are not accompanied by significance testing. As seen in the figures, during Acquisition, there seems to be a moderate, positive correlation between pupil size and arousal ratings for CS- trials. In terms of self-report measures, there is a moderate negative correlation between pupil size and trait anxiety for CS+ trials, which is even larger for state anxiety. Both, expressive suppression, and cognitive reappraisal exhibit a negative correlation with pupil size for CS- trials, while general psychopathology is positively correlated with pupil size for CS+ trials. Finally, a negative and positive correlations for CS+ and CS- trials respectively are observed between non-verbal ability and pupil size. During Extinction, there was a moderate negative correlation between pupil size and arousal for CS+ trials, and between pupil size trait, state anxiety, expressive suppression, and cognitive reappraisal for both CS+ and CS- trials.

To examine these differences inferentially we used a linear mixed effect model in which pupil size was added as the outcome variable. The model included the fixed effects and interaction between *Experimental Phase* and *Stimulus Type* and the fixed effects of valence, arousal ratings, state and trait anxiety, cognitive reappraisal, expressive suppression, and non-verbal ability. Data from the general psychopathology measure and state anxiety was not included since these were highly correlated with the trait measure of anxiety. The model also included all two and three-way interactions between the covariates and *Experimental Phase* and *Stimulus Type*. Type III Wald chi-square tests revealed no significant main effects or interactions (see Supplementary Table 5).

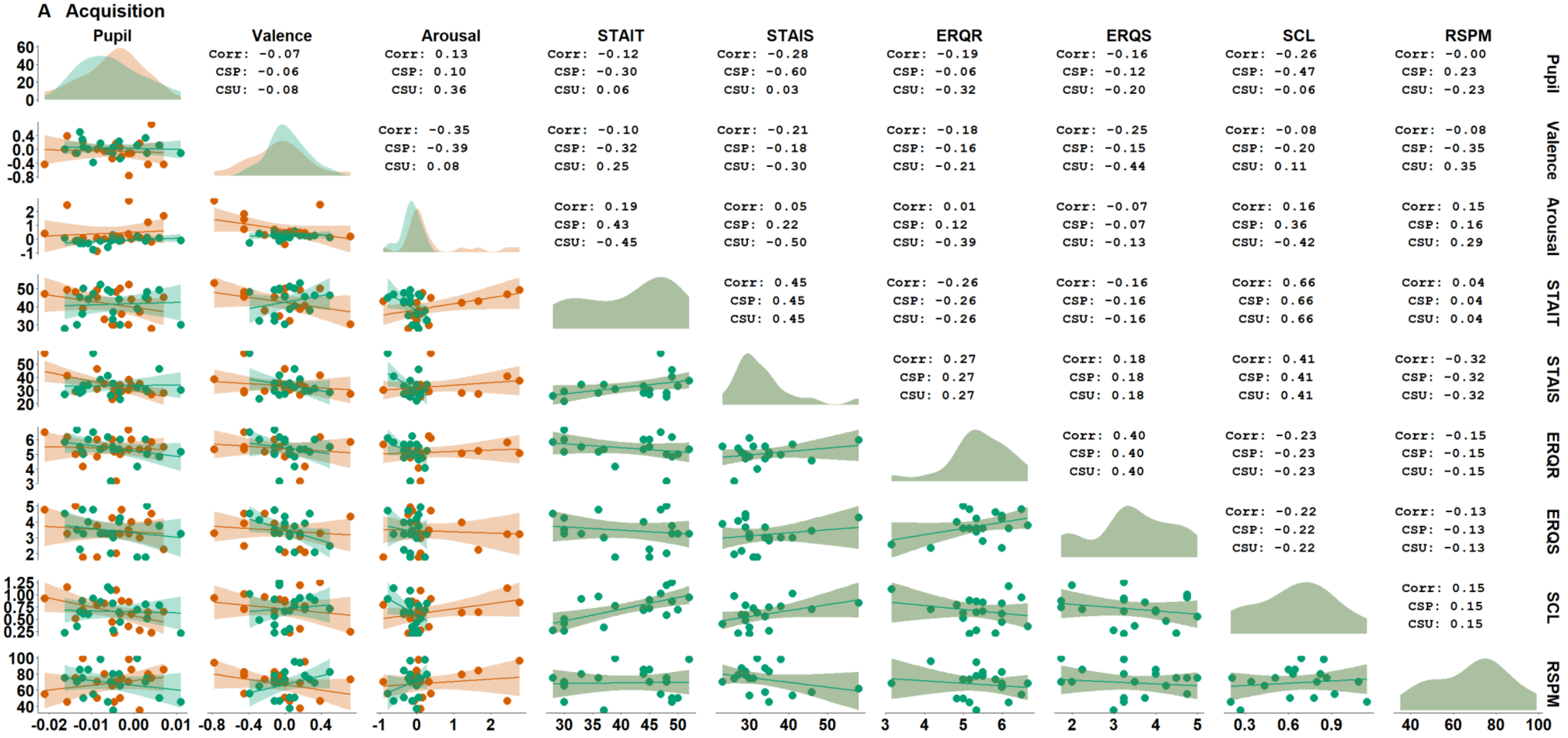
Supplementary Table 5

Type III Wald Chi-square tests for each of the fixed effects derived from the LME pupil model.

	Chisq	Df	P-value	R^2 Fixed
Full Model				0.017
Experimental Phase	0.231	1.000	0.631	
Stimulus Type	0.284	1.000	0.594	
Valence	1.599	1.000	0.206	
Arousal	0.206	1.000	0.650	
STAIT	0.097	1.000	0.756	
ERQS	0.757	1.000	0.384	
ERQR	1.044	1.000	0.307	
RSPM	0.223	1.000	0.637	
Experimental Phase X Stimulus Type	0.424	1.000	0.515	
Experimental Phase X Valence	0.377	1.000	0.539	
Experimental Phase X Arousal	0.919	1.000	0.338	
Experimental Phase X STAIT	0.417	1.000	0.519	
Experimental Phase X ERQS	0.209	1.000	0.647	
Experimental Phase X ERQR	0.024	1.000	0.876	
Experimental Phase X RSPM	0.000	1.000	0.993	
Stimulus Type X Valence	0.000	1.000	0.998	
Stimulus Type X Arousal	0.354	1.000	0.552	
Stimulus Type X STAIT	2.378	1.000	0.123	
Stimulus Type X ERQS	0.036	1.000	0.850	
Stimulus Type X ERQR	0.047	1.000	0.829	
Stimulus Type X RSPM	0.199	1.000	0.656	
Experimental Phase X Stimulus Type X Valence	1.278	1.000	0.258	
Experimental Phase X Stimulus Type X Arousal	0.001	1.000	0.972	
Experimental Phase X Stimulus Type X STAIT	0.180	1.000	0.672	
Experimental Phase X Stimulus Type X ERQS	0.065	1.000	0.799	
Experimental Phase X Stimulus Type X ERQR	0.043	1.000	0.835	
Experimental Phase X Stimulus Type X RSPM	0.586	1.000	0.444	

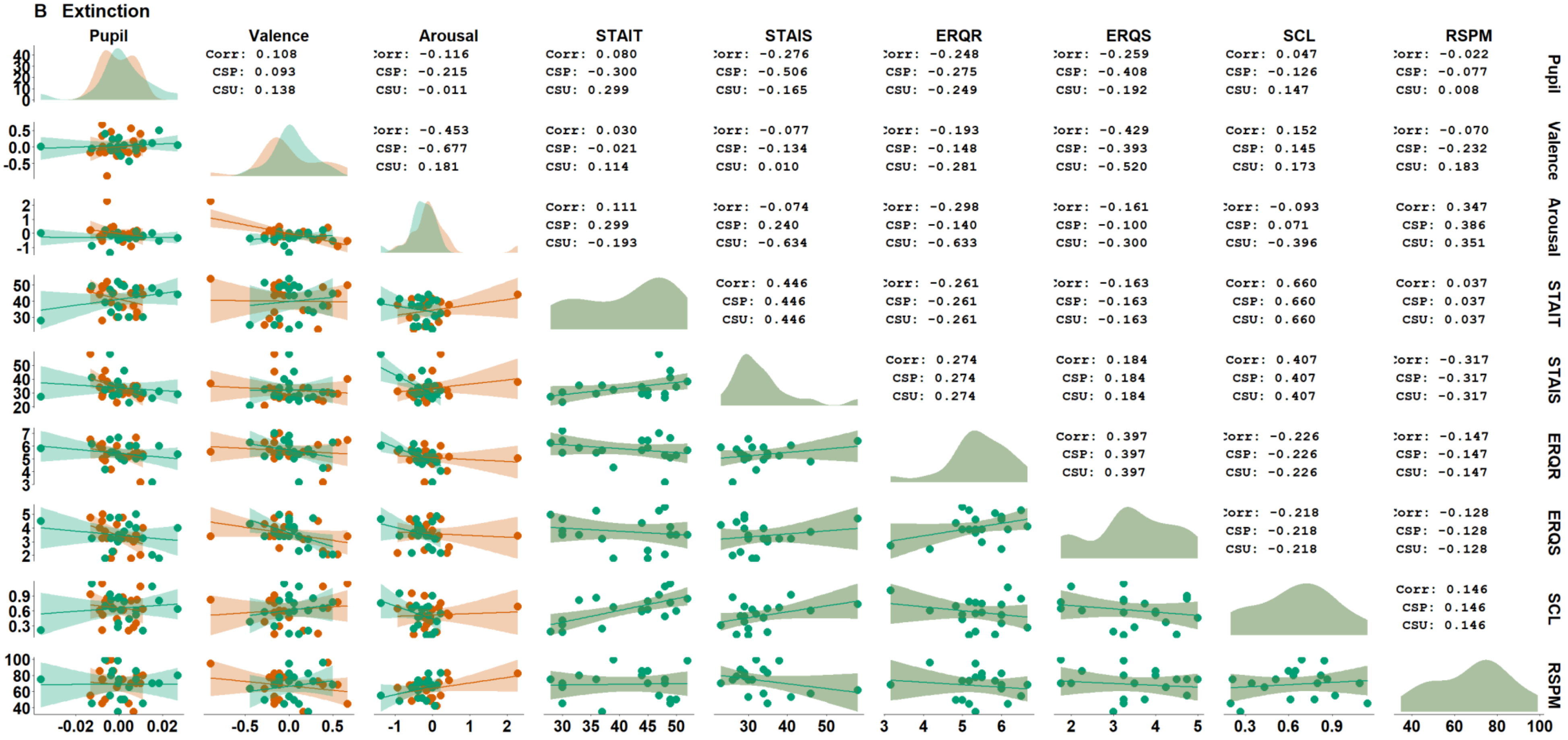
Supplementary Figure 5

Relationship between conditioning effects and psychological self-report measures during Acquisition.



Supplementary Figure 6

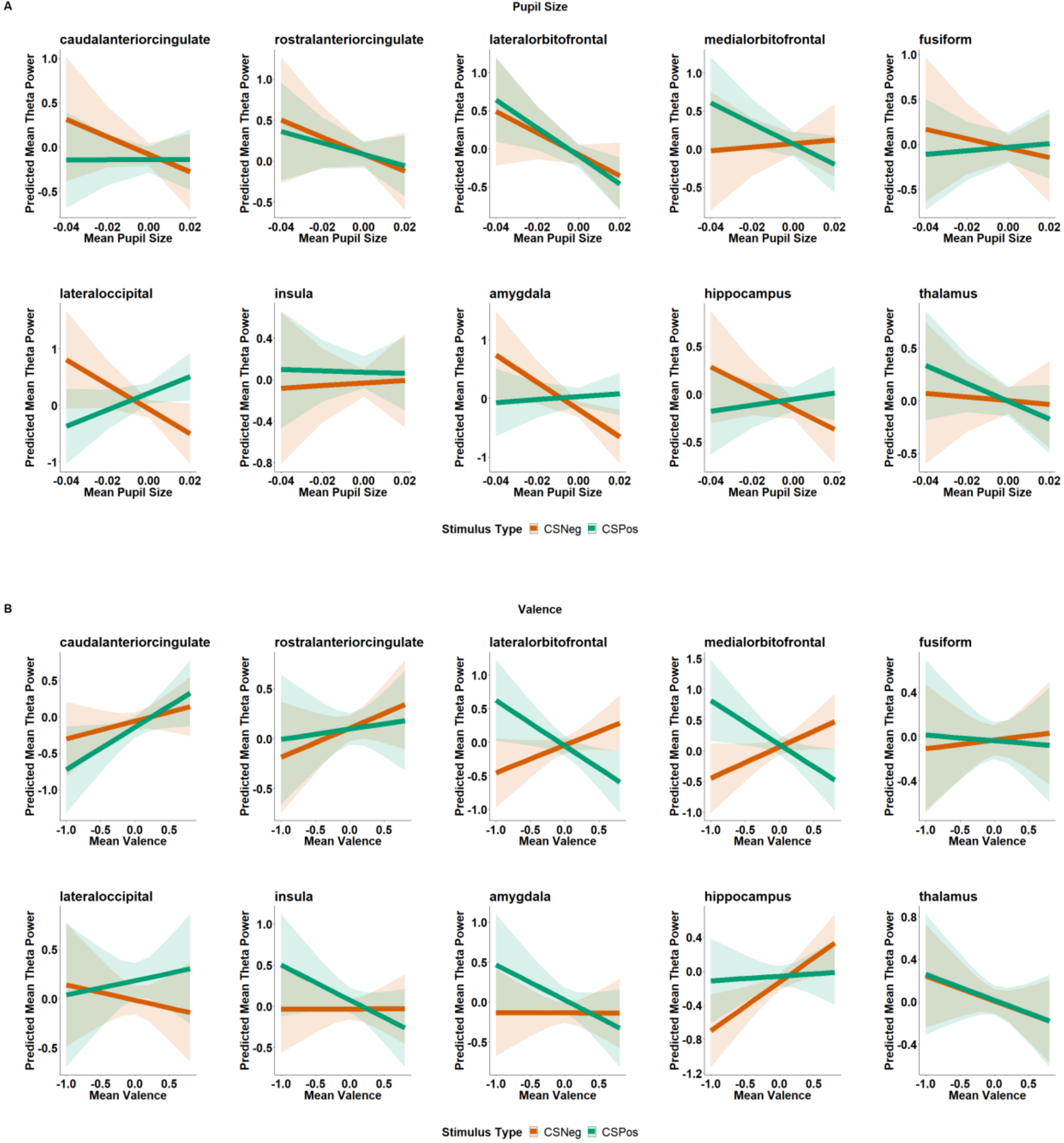
Relationship between conditioning effects and psychological self-report measures during Extinction

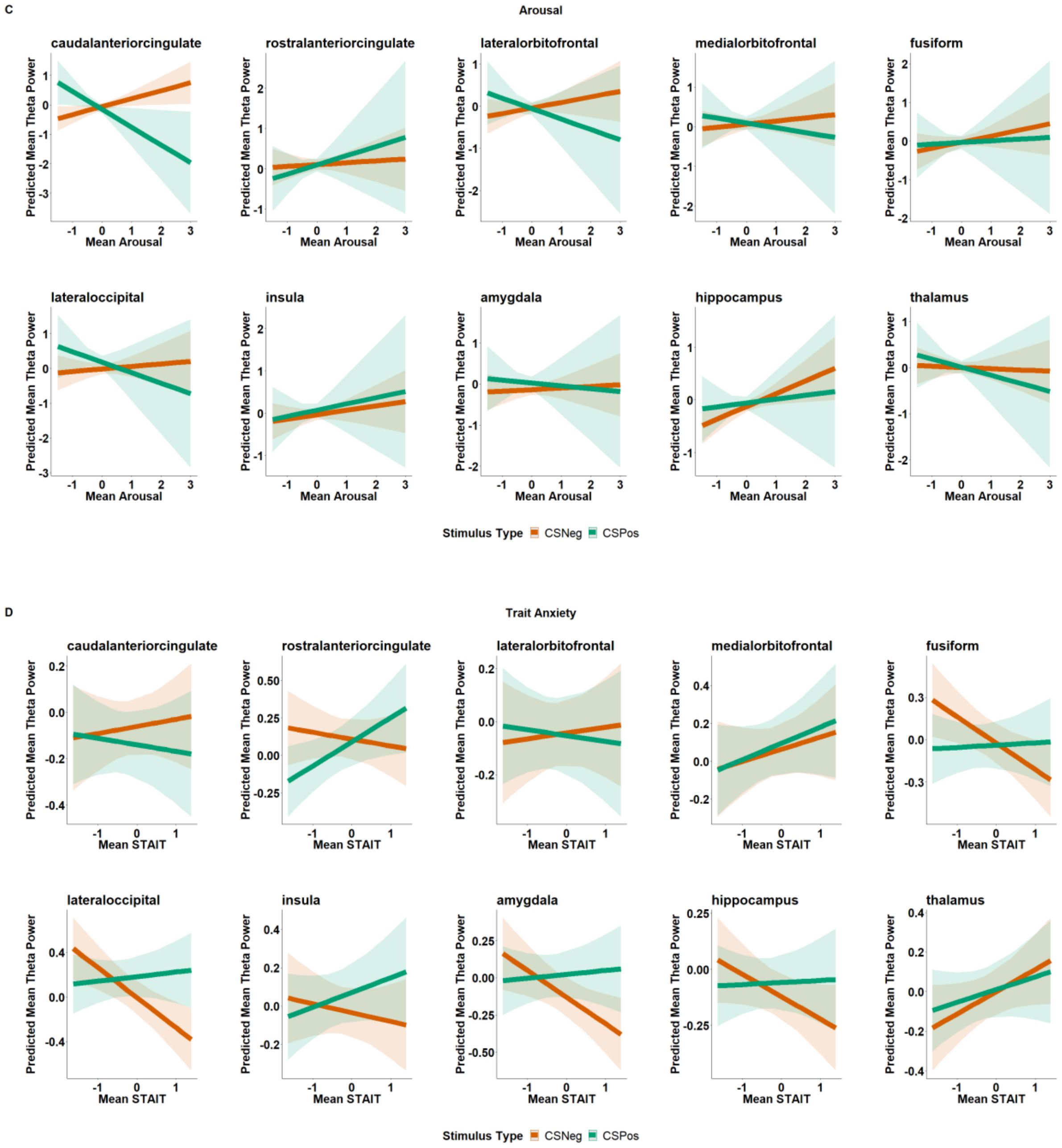


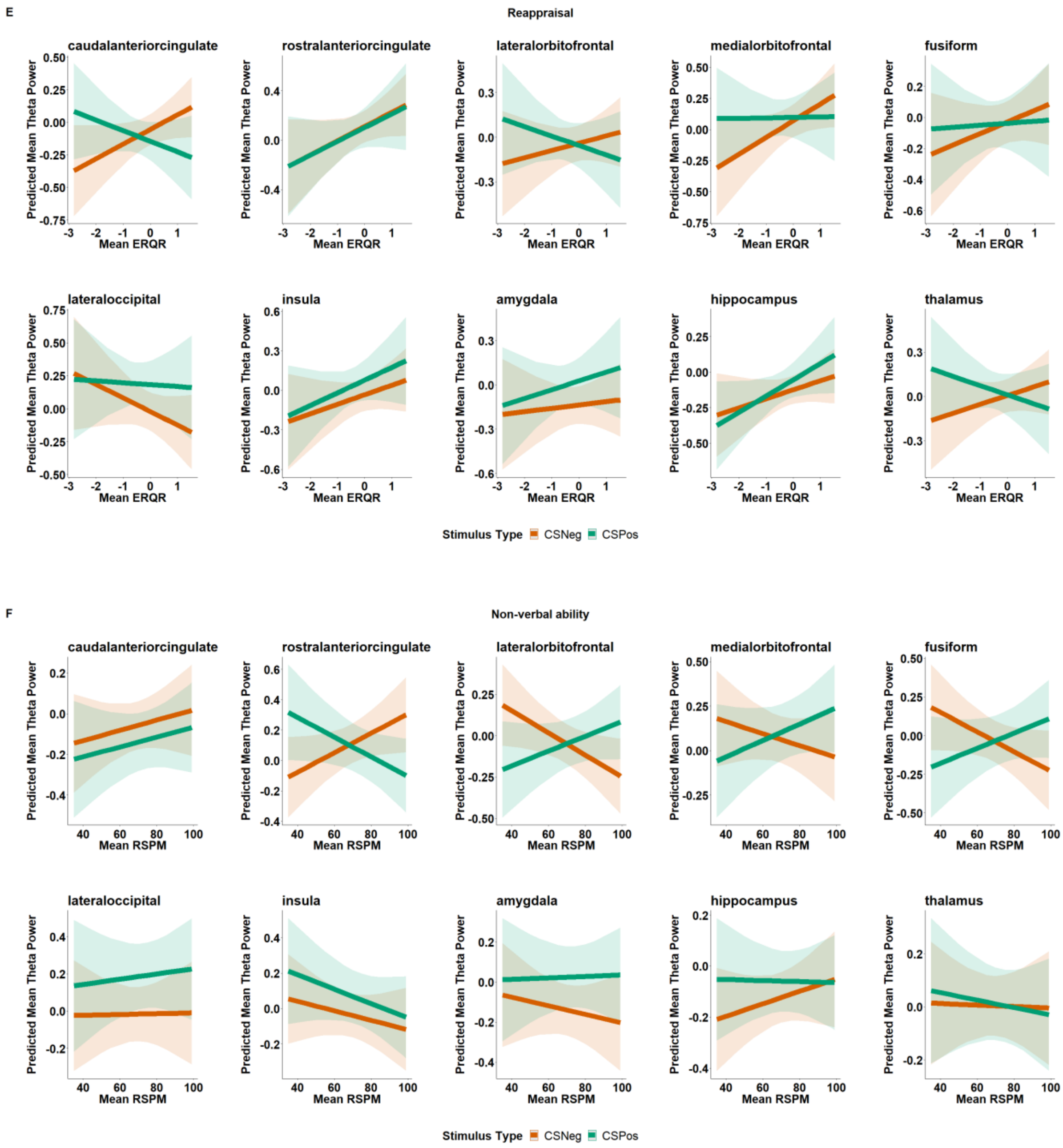
To examine potential moderators of the interaction between *Experimental Phase* and *Stimulus Type* in predicting mean theta power in ROIs, we conducted separate multiple linear regressions for each ROI. Since no trial information was retained during the MEG data cleaning it was not possible to conduct an LME model accounting for item and subject variability. The regression model included mean theta power as the outcome variable and the fixed effects and interaction between *Experimental Phase* and *Stimulus Type* as well as the fixed effects of valence, arousal ratings, trait anxiety, cognitive reappraisal, expressive suppression, and non-verbal ability and their interaction with *Experimental Phase* and *Stimulus Type*. We only focused on interactions including *Stimulus Type* since only these are theoretically relevant. None of the three -way interactions were significant in any of the ROIs (see Supplementary Table 6). A *Stimulus Type X Non-verbal ability* interaction was significant in the caudal and rostral ACC, lateral OFC and FFA models. The *Stimulus Type X Arousal* and the *Stimulus Type X Reappraisal* interactions were significant in the caudal ACC. *The Stimulus Type X Trait Anxiety* interaction was significant for the rostral ACC, lateral occipital area and the amygdala. The *Stimulus Type X Valence* interaction was significant in the lateral and middle OFC. *The Stimulus Type X Pupil size* interaction was significant in the lateral occipital area and the amygdala (see Supplementary Table 6). Post-hoc simple contrasts for those interactions were performed using package *emmeans*, however none of the contrasts were statistically significant (see Supplementary Table 7). These findings suggest that psychological self-report measures and behavioural and pupil measures of conditioning may be moderating some of the conditioning effects in mean theta power. However, considering that the effects within these interactions were very small and the confidence intervals very large (see Supplementary Figure 7), it is likely that the study is underpowered for disentangling any such effects.

Supplementary Figure 7

Predicted values of mean theta power by stimulus type.




C
Arousal
caudalanteriorcingulate
rostralanteriorcingulate
lateralorbitofrontal
medialorbitofrontal
fusiform
lateraloccipital
insula
amygdala
hippocampus
thalamus
Predicted Mean Theta Power
Mean Arousal
-1 0 1 2 3
Stimulus Type 🟧 CSNeg 🟩 CSPos
D
Trait Anxiety
caudalanteriorcingulate
rostralanteriorcingulate
lateralorbitofrontal
medialorbitofrontal
fusiform
lateraloccipital
insula
amygdala
hippocampus
thalamus
Predicted Mean Theta Power
Mean STAIT
-1 0 1
Stimulus Type 🟧 CSNeg 🟩 CSPos



Note. Predicted mean theta power for each ROI and stimulus type in the interaction between Stimulus Type and A) Pupil Size, B) Valence, C) Arousal, D) Trait Anxiety, E) Reappraisal and F) Non-verbal ability. Visualisaiton was performed for all ROIs if an interaction with Stimulus Type was significant in any one ROI.

Supplementary Table 6

Type III Wald Chi-square tests for each of the fixed effects derived from the multiple regression models of mean theta power within each ROI.

Caudal anterior cingulate				
	Sum Sq	Df	F value	P-value
(Intercept)	0.257	1.000	2.849	0.099
Experimental Phase	0.144	1.000	1.600	0.213
Stimulus Type	0.003	1.000	0.032	0.859
Pupil	0.066	1.000	0.737	0.395
Valence	0.452	1.000	5.022	0.030*
Arousal	0.114	1.000	1.267	0.266
STAIT	0.000	1.000	0.001	0.975
ERQS	0.381	1.000	4.226	0.046*
ERQR	0.010	1.000	0.107	0.745
RSPM	0.110	1.000	1.218	0.276
Experimental Phase X Stimulus Type	0.003	1.000	0.030	0.864
Experimental Phase X Pupil	0.025	1.000	0.281	0.599
Experimental Phase X Valence	0.556	1.000	6.168	0.017*
Experimental Phase X Arousal	0.096	1.000	1.070	0.307
Experimental Phase X STAIT	0.102	1.000	1.135	0.292
Experimental Phase X ERQS	0.119	1.000	1.322	0.256
Experimental Phase X ERQR	0.011	1.000	0.125	0.726
Experimental Phase X RSPM	0.101	1.000	1.122	0.295
Stimulus Type X Pupil	0.069	1.000	0.762	0.388
Stimulus Type X Valence	0.076	1.000	0.847	0.362
Stimulus Type X Arousal	0.785	1.000	8.716	0.005*
Stimulus Type X STAIT	0.037	1.000	0.407	0.527
Stimulus Type X ERQS	0.323	1.000	3.589	0.065
Stimulus Type X ERQR	0.379	1.000	4.205	0.046*
Stimulus Type X RSPM	0.000	1.000	0.000	0.984
Experimental Phase X Stimulus Type X Pupil	0.501	1.000	5.557	0.023*
Experimental Phase X Stimulus Type X Valence	0.077	1.000	0.858	0.359
Experimental Phase X Stimulus Type X Arousal	0.020	1.000	0.221	0.641
Experimental Phase X Stimulus Type X STAIT	0.015	1.000	0.162	0.689
Experimental Phase X Stimulus Type X ERQS	0.033	1.000	0.364	0.549
Experimental Phase X Stimulus Type X ERQR	0.018	1.000	0.203	0.655
Experimental Phase X Stimulus Type X RSPM	0.033	1.000	0.371	0.546
Residuals	3.963	44.000		

rostralanteriorcingulate				
	Sum Sq	Df	F value	P-value
(Intercept)	0.028	1.000	0.262	0.611
Experimental Phase	0.099	1.000	0.916	0.344
Stimulus Type	0.633	1.000	5.838	0.020*
Pupil	0.209	1.000	1.926	0.172
Valence	0.105	1.000	0.965	0.331
Arousal	0.076	1.000	0.701	0.407
STAIT	0.142	1.000	1.313	0.258
ERQS	0.443	1.000	4.085	0.049*
ERQR	0.515	1.000	4.752	0.035*
RSPM	0.000	1.000	0.000	0.990
Experimental Phase X Stimulus Type	0.002	1.000	0.016	0.899
Experimental Phase X Pupil	0.005	1.000	0.048	0.827
Experimental Phase X Valence	0.002	1.000	0.014	0.907
Experimental Phase X Arousal	0.001	1.000	0.006	0.939
Experimental Phase X STAIT	0.002	1.000	0.018	0.893
Experimental Phase X ERQS	0.024	1.000	0.217	0.644
Experimental Phase X ERQR	0.045	1.000	0.412	0.524
Experimental Phase X RSPM	0.098	1.000	0.907	0.346
Stimulus Type X Pupil	0.008	1.000	0.075	0.786
Stimulus Type X Valence	0.023	1.000	0.217	0.644
Stimulus Type X Arousal	0.034	1.000	0.314	0.578
Stimulus Type X STAIT	0.456	1.000	4.205	0.046*
Stimulus Type X ERQS	0.025	1.000	0.233	0.632
Stimulus Type X ERQR	0.000	1.000	0.001	0.972
Stimulus Type X RSPM	0.740	1.000	6.826	0.012*
Experimental Phase X Stimulus Type X Pupil	0.074	1.000	0.685	0.412
Experimental Phase X Stimulus Type X Valence	0.005	1.000	0.049	0.826
Experimental Phase X Stimulus Type X Arousal	0.106	1.000	0.974	0.329
Experimental Phase X Stimulus Type X STAIT	0.104	1.000	0.958	0.333
Experimental Phase X Stimulus Type X ERQS	0.034	1.000	0.314	0.578
Experimental Phase X Stimulus Type X ERQR	0.002	1.000	0.014	0.906
Experimental Phase X Stimulus Type X RSPM	0.001	1.000	0.006	0.939
Residuals	4.767	44.000		

lateralorbitofrontal				
	Sum Sq	Df	F value	P-value
(Intercept)	0.000	1.000	0.002	0.961
Experimental Phase	0.019	1.000	0.205	0.653
Stimulus Type	0.527	1.000	5.683	0.022*
Pupil	0.730	1.000	7.870	0.007**
Valence	0.046	1.000	0.498	0.484
Arousal	0.013	1.000	0.146	0.705
STAIT	0.000	1.000	0.000	0.999
ERQS	0.010	1.000	0.111	0.740
ERQR	0.002	1.000	0.023	0.880
RSPM	0.022	1.000	0.241	0.626
Experimental Phase X Stimulus Type	0.049	1.000	0.531	0.470
Experimental Phase X Pupil	0.137	1.000	1.479	0.230
Experimental Phase X Valence	0.260	1.000	2.804	0.101
Experimental Phase X Arousal	0.024	1.000	0.260	0.613
Experimental Phase X STAIT	0.023	1.000	0.246	0.623
Experimental Phase X ERQS	0.006	1.000	0.069	0.795
Experimental Phase X ERQR	0.004	1.000	0.040	0.843
Experimental Phase X RSPM	0.024	1.000	0.254	0.617
Stimulus Type X Pupil	0.013	1.000	0.136	0.714
Stimulus Type X Valence	0.781	1.000	8.428	0.006**
Stimulus Type X Arousal	0.147	1.000	1.584	0.215
Stimulus Type X STAIT	0.020	1.000	0.214	0.646
Stimulus Type X ERQS	0.117	1.000	1.265	0.267
Stimulus Type X ERQR	0.127	1.000	1.366	0.249
Stimulus Type X RSPM	0.555	1.000	5.987	0.018*
Experimental Phase X Stimulus Type X Pupil	0.003	1.000	0.037	0.849
Experimental Phase X Stimulus Type X Valence	0.038	1.000	0.411	0.525
Experimental Phase X Stimulus Type X Arousal	0.149	1.000	1.610	0.211
Experimental Phase X Stimulus Type X STAIT	0.012	1.000	0.129	0.721
Experimental Phase X Stimulus Type X ERQS	0.001	1.000	0.010	0.922
Experimental Phase X Stimulus Type X ERQR	0.000	1.000	0.003	0.957
Experimental Phase X Stimulus Type X RSPM	0.054	1.000	0.583	0.449
Residuals	4.080	44.000		

medialorbitofrontal

	Sum Sq	Df	F value	P-value
(Intercept)	0.002	1.000	0.018	0.892
Experimental Phase	0.002	1.000	0.014	0.907
Stimulus Type	0.263	1.000	2.359	0.132
Pupil	0.084	1.000	0.754	0.390
Valence	0.028	1.000	0.251	0.619
Arousal	0.002	1.000	0.017	0.897
STAIT	0.238	1.000	2.137	0.151
ERQS	0.034	1.000	0.308	0.581
ERQR	0.192	1.000	1.727	0.196
RSPM	0.007	1.000	0.060	0.808
Experimental Phase X Stimulus Type	0.017	1.000	0.154	0.697
Experimental Phase X Pupil	0.000	1.000	0.000	0.996
Experimental Phase X Valence	0.082	1.000	0.739	0.395
Experimental Phase X Arousal	0.004	1.000	0.035	0.852
Experimental Phase X STAIT	0.016	1.000	0.143	0.707
Experimental Phase X ERQS	0.068	1.000	0.610	0.439
Experimental Phase X ERQR	0.078	1.000	0.701	0.407
Experimental Phase X RSPM	0.011	1.000	0.095	0.760
Stimulus Type X Pupil	0.173	1.000	1.550	0.220
Stimulus Type X Valence	1.000	1.000	8.976	0.004**
Stimulus Type X Arousal	0.041	1.000	0.366	0.548
Stimulus Type X STAIT	0.005	1.000	0.045	0.833
Stimulus Type X ERQS	0.246	1.000	2.208	0.144
Stimulus Type X ERQR	0.175	1.000	1.567	0.217
Stimulus Type X RSPM	0.281	1.000	2.526	0.119
Experimental Phase X Stimulus Type X Pupil	0.003	1.000	0.026	0.874
Experimental Phase X Stimulus Type X Valence	0.042	1.000	0.374	0.544
Experimental Phase X Stimulus Type X Arousal	0.033	1.000	0.298	0.588
Experimental Phase X Stimulus Type X STAIT	0.090	1.000	0.807	0.374
Experimental Phase X Stimulus Type X ERQS	0.025	1.000	0.226	0.637
Experimental Phase X Stimulus Type X ERQR	0.001	1.000	0.005	0.944
Experimental Phase X Stimulus Type X RSPM	0.011	1.000	0.103	0.750
Residuals	4.901	44.000		

fusiform

	Sum Sq	Df	F value	P-value
(Intercept)	0.002	1.000	0.013	0.909
Experimental Phase	0.012	1.000	0.102	0.751
Stimulus Type	0.488	1.000	4.141	0.048
Pupil	0.007	1.000	0.063	0.803
Valence	0.000	1.000	0.004	0.949
Arousal	0.042	1.000	0.357	0.553
STAIT	0.306	1.000	2.601	0.114
ERQS	0.080	1.000	0.680	0.414
ERQR	0.078	1.000	0.660	0.421
RSPM	0.010	1.000	0.086	0.770
Experimental Phase X Stimulus Type	0.068	1.000	0.581	0.450
Experimental Phase X Pupil	0.019	1.000	0.164	0.687
Experimental Phase X Valence	0.043	1.000	0.365	0.549
Experimental Phase X Arousal	0.001	1.000	0.010	0.920
Experimental Phase X STAIT	0.034	1.000	0.286	0.595
Experimental Phase X ERQS	0.041	1.000	0.348	0.558
Experimental Phase X ERQR	0.064	1.000	0.543	0.465
Experimental Phase X RSPM	0.014	1.000	0.122	0.729
Stimulus Type X Pupil	0.036	1.000	0.309	0.581
Stimulus Type X Valence	0.011	1.000	0.094	0.760
Stimulus Type X Arousal	0.013	1.000	0.114	0.737
Stimulus Type X STAIT	0.432	1.000	3.667	0.062
Stimulus Type X ERQS	0.066	1.000	0.563	0.457
Stimulus Type X ERQR	0.038	1.000	0.320	0.574
Stimulus Type X RSPM	0.561	1.000	4.763	0.034*
Experimental Phase X Stimulus Type X Pupil	0.093	1.000	0.787	0.380
Experimental Phase X Stimulus Type X Valence	0.054	1.000	0.463	0.500
Experimental Phase X Stimulus Type X Arousal	0.063	1.000	0.538	0.467
Experimental Phase X Stimulus Type X STAIT	0.030	1.000	0.257	0.615
Experimental Phase X Stimulus Type X ERQS	0.016	1.000	0.135	0.715
Experimental Phase X Stimulus Type X ERQR	0.059	1.000	0.505	0.481
Experimental Phase X Stimulus Type X RSPM	0.037	1.000	0.314	0.578
Residuals	5.181	44.000		

lateraloccipital				
	Sum Sq	Df	F value	P-value
(Intercept)	0.002	1.000	0.014	0.906
Experimental Phase	0.028	1.000	0.209	0.650
Stimulus Type	0.030	1.000	0.221	0.641
Pupil	0.035	1.000	0.257	0.615
Valence	0.000	1.000	0.000	0.984
Arousal	0.054	1.000	0.394	0.534
STAIT	0.544	1.000	3.981	0.052
ERQS	0.211	1.000	1.543	0.221
ERQR	0.139	1.000	1.018	0.319
RSPM	0.012	1.000	0.087	0.770
Experimental Phase X Stimulus Type	0.012	1.000	0.087	0.770
Experimental Phase X Pupil	0.238	1.000	1.743	0.194
Experimental Phase X Valence	0.146	1.000	1.068	0.307
Experimental Phase X Arousal	0.004	1.000	0.029	0.866
Experimental Phase X STAIT	0.046	1.000	0.337	0.565
Experimental Phase X ERQS	0.249	1.000	1.827	0.183
Experimental Phase X ERQR	0.015	1.000	0.111	0.741
Experimental Phase X RSPM	0.030	1.000	0.217	0.644
Stimulus Type X Pupil	0.919	1.000	6.733	0.013*
Stimulus Type X Valence	0.061	1.000	0.445	0.508
Stimulus Type X Arousal	0.143	1.000	1.048	0.311
Stimulus Type X STAIT	1.009	1.000	7.390	0.009*
Stimulus Type X ERQS	0.074	1.000	0.540	0.466
Stimulus Type X ERQR	0.079	1.000	0.580	0.450
Stimulus Type X RSPM	0.006	1.000	0.045	0.832
Experimental Phase X Stimulus Type X Pupil	0.237	1.000	1.733	0.195
Experimental Phase X Stimulus Type X Valence	0.000	1.000	0.000	0.985
Experimental Phase X Stimulus Type X Arousal	0.007	1.000	0.048	0.828
Experimental Phase X Stimulus Type X STAIT	0.007	1.000	0.053	0.819
Experimental Phase X Stimulus Type X ERQS	0.000	1.000	0.000	0.989
Experimental Phase X Stimulus Type X ERQR	0.002	1.000	0.013	0.908
Experimental Phase X Stimulus Type X RSPM	0.015	1.000	0.113	0.738
Residuals	6.007	44.000		

insula

	Sum Sq	Df	F value	P-value
(Intercept)	0.211	1.000	2.151	0.150
Experimental Phase	0.024	1.000	0.243	0.624
Stimulus Type	0.031	1.000	0.311	0.580
Pupil	0.000	1.000	0.002	0.960
Valence	0.119	1.000	1.212	0.277
Arousal	0.064	1.000	0.652	0.424
STAIT	0.010	1.000	0.101	0.752
ERQS	0.135	1.000	1.374	0.248
ERQR	0.290	1.000	2.952	0.093
RSPM	0.203	1.000	2.069	0.157
Experimental Phase X Stimulus Type	0.077	1.000	0.788	0.380
Experimental Phase X Pupil	0.086	1.000	0.875	0.355
Experimental Phase X Valence	0.023	1.000	0.231	0.633
Experimental Phase X Arousal	0.029	1.000	0.291	0.592
Experimental Phase X STAIT	0.017	1.000	0.173	0.680
Experimental Phase X ERQS	0.005	1.000	0.052	0.820
Experimental Phase X ERQR	0.115	1.000	1.171	0.285
Experimental Phase X RSPM	0.011	1.000	0.111	0.740
Stimulus Type X Pupil	0.002	1.000	0.025	0.876
Stimulus Type X Valence	0.120	1.000	1.218	0.276
Stimulus Type X Arousal	0.002	1.000	0.020	0.890
Stimulus Type X STAIT	0.161	1.000	1.638	0.207
Stimulus Type X ERQS	0.088	1.000	0.900	0.348
Stimulus Type X ERQR	0.006	1.000	0.058	0.811
Stimulus Type X RSPM	0.008	1.000	0.086	0.770
Experimental Phase X Stimulus Type X Pupil	0.027	1.000	0.272	0.604
Experimental Phase X Stimulus Type X Valence	0.006	1.000	0.060	0.808
Experimental Phase X Stimulus Type X Arousal	0.028	1.000	0.290	0.593
Experimental Phase X Stimulus Type X STAIT	0.016	1.000	0.163	0.689
Experimental Phase X Stimulus Type X ERQS	0.011	1.000	0.109	0.743
Experimental Phase X Stimulus Type X ERQR	0.001	1.000	0.013	0.911
Experimental Phase X Stimulus Type X RSPM	0.106	1.000	1.079	0.305
Residuals	4.319	44.000		

amygdala				
	Sum Sq	Df	F value	P-value
(Intercept)	0.001	1.000	0.007	0.935
Experimental Phase	0.135	1.000	1.301	0.260
Stimulus Type	0.001	1.000	0.011	0.916
Pupil	0.301	1.000	2.895	0.096
Valence	0.131	1.000	1.260	0.268
Arousal	0.001	1.000	0.008	0.928
STAIT	0.246	1.000	2.367	0.131
ERQS	0.039	1.000	0.376	0.543
ERQR	0.067	1.000	0.650	0.424
RSPM	0.014	1.000	0.140	0.710
Experimental Phase X Stimulus Type	0.034	1.000	0.326	0.571
Experimental Phase X Pupil	0.054	1.000	0.524	0.473
Experimental Phase X Valence	0.039	1.000	0.376	0.543
Experimental Phase X Arousal	0.012	1.000	0.114	0.737
Experimental Phase X STAIT	0.017	1.000	0.167	0.685
Experimental Phase X ERQS	0.010	1.000	0.101	0.752
Experimental Phase X ERQR	0.214	1.000	2.061	0.158
Experimental Phase X RSPM	0.111	1.000	1.067	0.307
Stimulus Type X Pupil	0.460	1.000	4.435	0.041*
Stimulus Type X Valence	0.126	1.000	1.213	0.277
Stimulus Type X Arousal	0.012	1.000	0.112	0.740
Stimulus Type X STAIT	0.445	1.000	4.284	0.044*
Stimulus Type X ERQS	0.006	1.000	0.053	0.818
Stimulus Type X ERQR	0.014	1.000	0.132	0.718
Stimulus Type X RSPM	0.028	1.000	0.272	0.605
Experimental Phase X Stimulus Type X Pupil	0.209	1.000	2.016	0.163
Experimental Phase X Stimulus Type X Valence	0.020	1.000	0.195	0.661
Experimental Phase X Stimulus Type X Arousal	0.007	1.000	0.064	0.802
Experimental Phase X Stimulus Type X STAIT	0.020	1.000	0.195	0.661
Experimental Phase X Stimulus Type X ERQS	0.085	1.000	0.819	0.370
Experimental Phase X Stimulus Type X ERQR	0.000	1.000	0.000	0.989
Experimental Phase X Stimulus Type X RSPM	0.014	1.000	0.134	0.716
Residuals	4.568	44.000		

hippocampus				
	Sum Sq	Df	F value	P-value
(Intercept)	0.088	1.000	1.397	0.244
Experimental Phase	0.186	1.000	2.938	0.094
Stimulus Type	0.063	1.000	0.990	0.325
Pupil	0.041	1.000	0.653	0.423
Valence	0.262	1.000	4.142	0.048*
Arousal	0.102	1.000	1.609	0.211
STAIT	0.088	1.000	1.384	0.246
ERQS	0.913	1.000	14.426	0.000***
ERQR	0.317	1.000	5.009	0.030*
RSPM	0.022	1.000	0.354	0.555
Experimental Phase X Stimulus Type	0.085	1.000	1.346	0.252
Experimental Phase X Pupil	0.003	1.000	0.040	0.843
Experimental Phase X Valence	0.132	1.000	2.090	0.155
Experimental Phase X Arousal	0.129	1.000	2.047	0.160
Experimental Phase X STAIT	0.011	1.000	0.171	0.681
Experimental Phase X ERQS	0.012	1.000	0.185	0.669
Experimental Phase X ERQR	0.000	1.000	0.008	0.930
Experimental Phase X RSPM	0.166	1.000	2.619	0.113
Stimulus Type X Pupil	0.135	1.000	2.140	0.151
Stimulus Type X Valence	0.178	1.000	2.818	0.100
Stimulus Type X Arousal	0.029	1.000	0.461	0.501
Stimulus Type X STAIT	0.126	1.000	1.988	0.166
Stimulus Type X ERQS	0.006	1.000	0.099	0.754
Stimulus Type X ERQR	0.026	1.000	0.415	0.523
Stimulus Type X RSPM	0.031	1.000	0.491	0.487
Experimental Phase X Stimulus Type X Pupil	0.145	1.000	2.293	0.137
Experimental Phase X Stimulus Type X Valence	0.001	1.000	0.017	0.895
Experimental Phase X Stimulus Type X Arousal	0.053	1.000	0.830	0.367
Experimental Phase X Stimulus Type X STAIT	0.023	1.000	0.361	0.551
Experimental Phase X Stimulus Type X ERQS	0.010	1.000	0.163	0.688
Experimental Phase X Stimulus Type X ERQR	0.002	1.000	0.029	0.865
Experimental Phase X Stimulus Type X RSPM	0.073	1.000	1.156	0.288
Residuals	2.783	44.000		

thalamus				
	Sum Sq	Df	F value	P-value
(Intercept)	0.008	1.000	0.093	0.762
Experimental Phase	0.017	1.000	0.201	0.656
Stimulus Type	0.003	1.000	0.042	0.839
Pupil	0.072	1.000	0.874	0.355
Valence	0.153	1.000	1.851	0.181
Arousal	0.042	1.000	0.508	0.480
STAIT	0.328	1.000	3.978	0.052
ERQS	0.168	1.000	2.037	0.161
ERQR	0.000	1.000	0.001	0.976
RSPM	0.013	1.000	0.158	0.693
Experimental Phase X Stimulus Type	0.025	1.000	0.308	0.582
Experimental Phase X Pupil	0.032	1.000	0.385	0.538
Experimental Phase X Valence	0.006	1.000	0.070	0.792
Experimental Phase X Arousal	0.011	1.000	0.136	0.714
Experimental Phase X STAIT	0.152	1.000	1.848	0.181
Experimental Phase X ERQS	0.018	1.000	0.220	0.641
Experimental Phase X ERQR	0.013	1.000	0.161	0.690
Experimental Phase X RSPM	0.009	1.000	0.109	0.743
Stimulus Type X Pupil	0.031	1.000	0.379	0.542
Stimulus Type X Valence	0.000	1.000	0.001	0.982
Stimulus Type X Arousal	0.022	1.000	0.272	0.605
Stimulus Type X STAIT	0.024	1.000	0.294	0.591
Stimulus Type X ERQS	0.013	1.000	0.160	0.691
Stimulus Type X ERQR	0.151	1.000	1.831	0.183
Stimulus Type X RSPM	0.006	1.000	0.068	0.796
Experimental Phase X Stimulus Type X Pupil	0.086	1.000	1.045	0.312
Experimental Phase X Stimulus Type X Valence	0.007	1.000	0.088	0.768
Experimental Phase X Stimulus Type X Arousal	0.048	1.000	0.577	0.452
Experimental Phase X Stimulus Type X STAIT	0.003	1.000	0.033	0.856
Experimental Phase X Stimulus Type X ERQS	0.000	1.000	0.002	0.962
Experimental Phase X Stimulus Type X ERQR	0.013	1.000	0.163	0.688
Experimental Phase X Stimulus Type X RSPM	0.059	1.000	0.721	0.400
Residuals	3.630	44.000		

Supplementary Table 7

Simple contrasts for the significant two-way interactions in ROIs.

Pupil size

contrast	Pupil	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
lateraloccipital								
CS+>CS-	-0.002	-0.200	0.113	44.000	-0.429	0.028	-1.766	0.084
amygdala								
CS+>CS-	-0.002	-0.162	0.099	44.000	-0.361	0.037	-1.636	0.109

Valence

contrast	Valence	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
lateralorbitofrontal								
CS+>CS-	0.001	0.011	0.093	44.000	-0.177	0.200	0.120	0.905
medialorbitofrontal								
CS+>CS-	0.001	-0.032	0.102	44.000	-0.238	0.175	-0.311	0.757

Arousal

contrast	Arousal	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
caudalanteriorcingulate								
CS+>CS-	-0.011	0.082	0.092	44.000	-0.103	0.268	0.892	0.377

Trait Anxiety

contrast	STAIT	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
rostralanteriorcingulate								
CS+>CS-	0.040	0.010	0.101	44.000	-0.194	0.213	0.095	0.925
lateraloccipital								
CS+>CS-	0.040	-0.200	0.113	44.000	-0.429	0.028	-1.766	0.084
amygdala								
CS+>CS-	0.040	-0.162	0.099	44.000	-0.361	0.037	-1.636	0.109

Reappraisal

contrast	ERQR	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
caudalanteriorcingulate								
CS+>CS-	-0.037	0.082	0.092	44.000	-0.103	0.268	0.892	0.377

Non-verbal ability

contrast	RSPM	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
caudalanteriorcingulate								
CS+>CS-	68.789	-0.009	0.088	44.000	-0.187	0.169	-0.101	0.920
rostralanteriorcingulate								
CS+>CS-	68.789	0.010	0.101	44.000	-0.194	0.213	0.095	0.925
lateralorbitofrontal								
CS+>CS-	68.789	0.011	0.093	44.000	-0.177	0.200	0.120	0.905
fusiform								
CS+>CS-	68.789	0.006	0.105	44.000	-0.206	0.218	0.059	0.953

3.5.2 Supplementary Materials 5: Auditory Control Experiment 1

Historically, the most commonly used Unconditioned Stimulus (UCS) was an electric shock but due to the ethical issues of its administration to vulnerable populations (e.g. children), other forms of UCSs are now frequently employed (Lonsdorf et al., 2017). Auditory UCSs such as white noise and screams, have become a popular choice, potentially since other unpleasant stimuli such as odours and air puffs often present a procedural difficulty as they require specialised equipment for administration. While there have been some, comparative studies on the use of air puffs, odours, verbal and auditory stimuli as an alternative to electric shock (Busch & Evans, 1977; McEchron et al., 1992; Neumann & Waters, 2006a; Sperl, Panitz, Hermann, et al., 2016), a comprehensive investigation or normative data of sounds typically used as UCSs is lacking. Therefore, the goal of this study was to derive an unpleasant UCS based on participants' subjective evaluative judgements.

In addition, a crucial difference between an electric shock and auditory stimuli is that while an electric shock elicits only an expectancy response to an aversive stimulus, an auditory stimulus can also induce responses that purely driven by sound processing. For example, the acoustic properties of sounds such as intensity, pitch and frequency have been shown to drive physiological reactions such as pupil size and SCR changes (Gomez & Danuser, 2007; Liao et al., 2016), and to influence ratings of valence arousal (Bradley & Lang, 2000; Gomez & Danuser, 2007; Ma & Thompson, 2015; Västfjäll, 2013). In addition, sound intensity has been shown to modulate the relationship between pupil size and subjective sound perception, such as annoyance (Liao et al., 2016). Consequently, the present study also examined whether stimulus valence can predict pupil size, valence and arousal changes for sounds of short duration and of controlled, high intensity, under conditions similar to classical conditioning experiments. In addition, we examined whether the fundamental frequency of sounds can potentially mediate any such effects.

Methods

Participants

Fourteen subjects aged between 22 and 33 (M = 25.6, SD = 3.2, 3 males) took part in the study. One subject was excluded from the data analyses due to a corrupted eye tracking file. All participants had normal or corrected-to-normal vision and hearing.

Stimuli

A total of seventeen auditory stimuli were used in the control experiment. Seventeen of these were environmental or human sounds (12 negative, 5 positive), such as female scream and metal scrapes as well as bird chirping, bubbles, previously shown to elicit negative and positive valence respectively (Kumar et al., 2008). These stimuli were obtained from three online databases (Freesfx, Freesound and the CNBC Stimuli Repository). The remaining two stimuli were an unfiltered and low-pass filtered (1-3 kHz) white noise, created in Matlab R2016a. All sounds were trimmed to a length of 200 ms. To equalise the intensity of sounds, each sound was mean centred and then normalised to the same, maximum root mean square (RMS) amplitude without clipping using a RMS equaliser (The Phonetics Lab, University of Washington, https://depts.washington.edu/phonlab/resources/rmsLeveler.m). The first 20 ms of the signal of all stimuli was gradually faded in. The resulting normalised sounds were presented at a maximum intensity of approximately 85 -90 dB as measured by TENMA 72-6635 sound meter. The average fundamental frequency of each sound was computed in Matlab 2020b using the Audio Toolbox. Task scripts are available at https://osf.io/u6qza/.

Procedure

The task contained a total of 180 trials. Each sound was presented 8 times in a random order, with an additional 8 trials during which silence was presented. The auditory stimuli were administered through 4m plastic tubes and earpieces with a band pass frequency of 4 kHz. On each trial, a black fixation cross was presented on a gray background for 650 ms followed by the sound, with an inter-trial interval (ITI) of 1300 ms ±300 ms, comprising of a black fixation cross on a gray background. The sound was delivered through 4m tubes. To maintain subjects' attention, 3-4 trials were randomly selected and presented twice in succession. Participants were instructed to press a button when a sound was repeated. At the end of the task, all sounds were presented 3 times in random order and participants completed an auditory rating task where they were asked to rate each stimulus on valence and negative arousal using an 8-point response pad (1 not at all pleasant/arousing to 8 extremely pleasant/arousing).

Pupil response acquisition and pre-processing

Pupil response was recorded using EyeLink 1000 long-range eye tracker and pupil size was recorded continuously during each trial presentation with initial sampling rate of 1000 Hz. Pupil pre-processing was performed in Matlab 2017a using the Fieldtrip Toolbox (Oostenveld et al., 2011) and functions provided by Urai et al. (2017) using the same procedure as in chapter 3.

Results

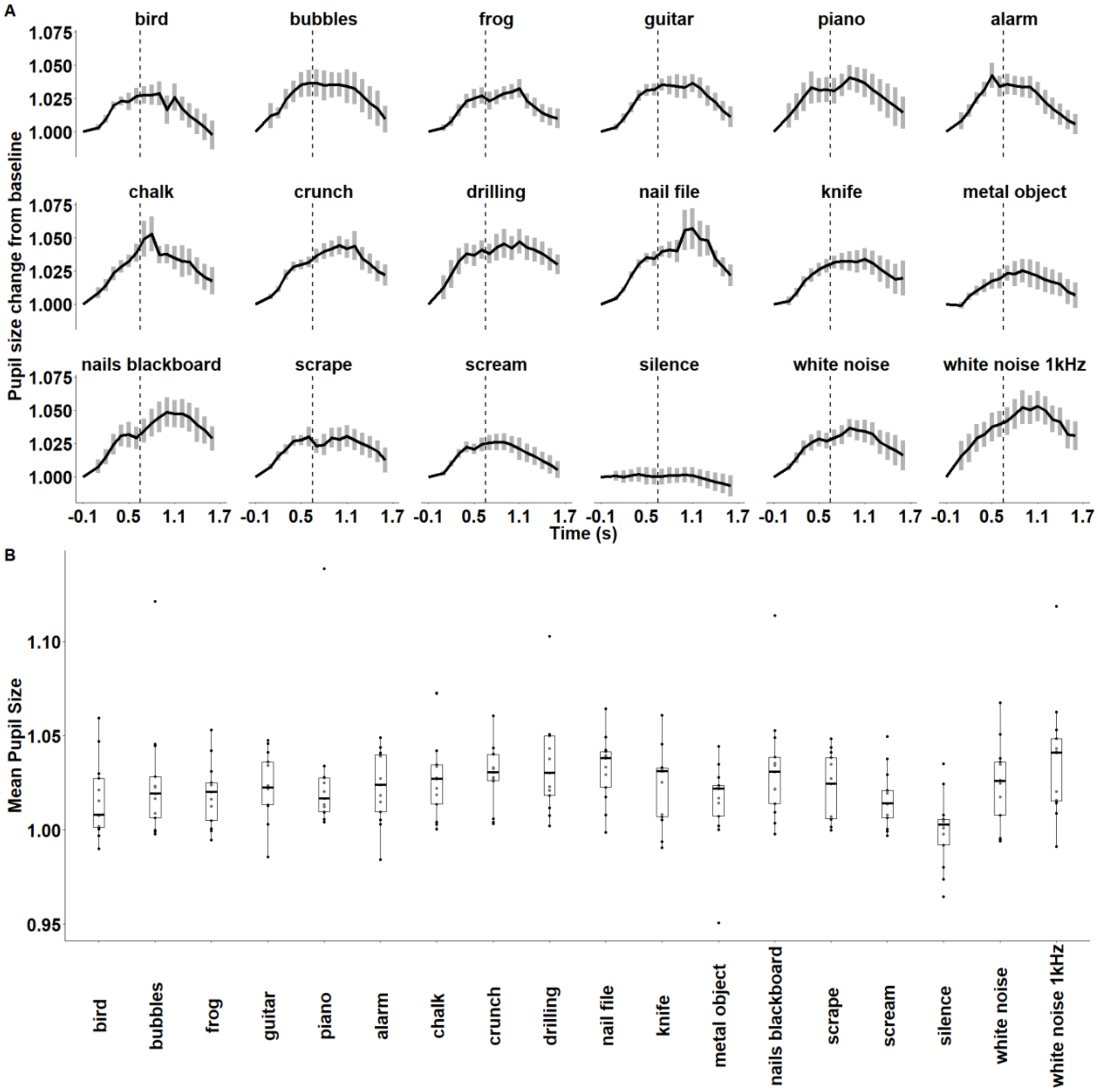
All analyses were performed in R. Analysis scripts are available at https://osf.io/u6qza.

Selection of unconditioned stimuli

As seen in Supplementary Figure 8, the average pupil size gradually increases post stimulus onset with a peak around 1.1 seconds. This trajectory is comparable across different sounds and regardless of the sound valence. Similarly, the mean pupil size across time for the different sounds does not appear to differ, at least a descriptive level. When looking at the behavioural ratings (see Supplementary Figure 9 and 10), a similar pattern emerges where all sounds regardless of their valence were rated as unpleasant in at least 70% of responses with the exception of the alarm sound which was rated as unpleasant in 97% of responses. In terms of arousal, the responses have a greater spread with approximately 40-50% of responses to sounds being not arousing with the remaining percentage belonging to arousing. Only one sound (drilling) was rated as arousing by 70% of responses. Since none of the measures elicited major differences between sounds, the most consistently rated sound on the valence scale (the alarm) was selected as the unconditioned stimulus.

Supplementary Figure 8

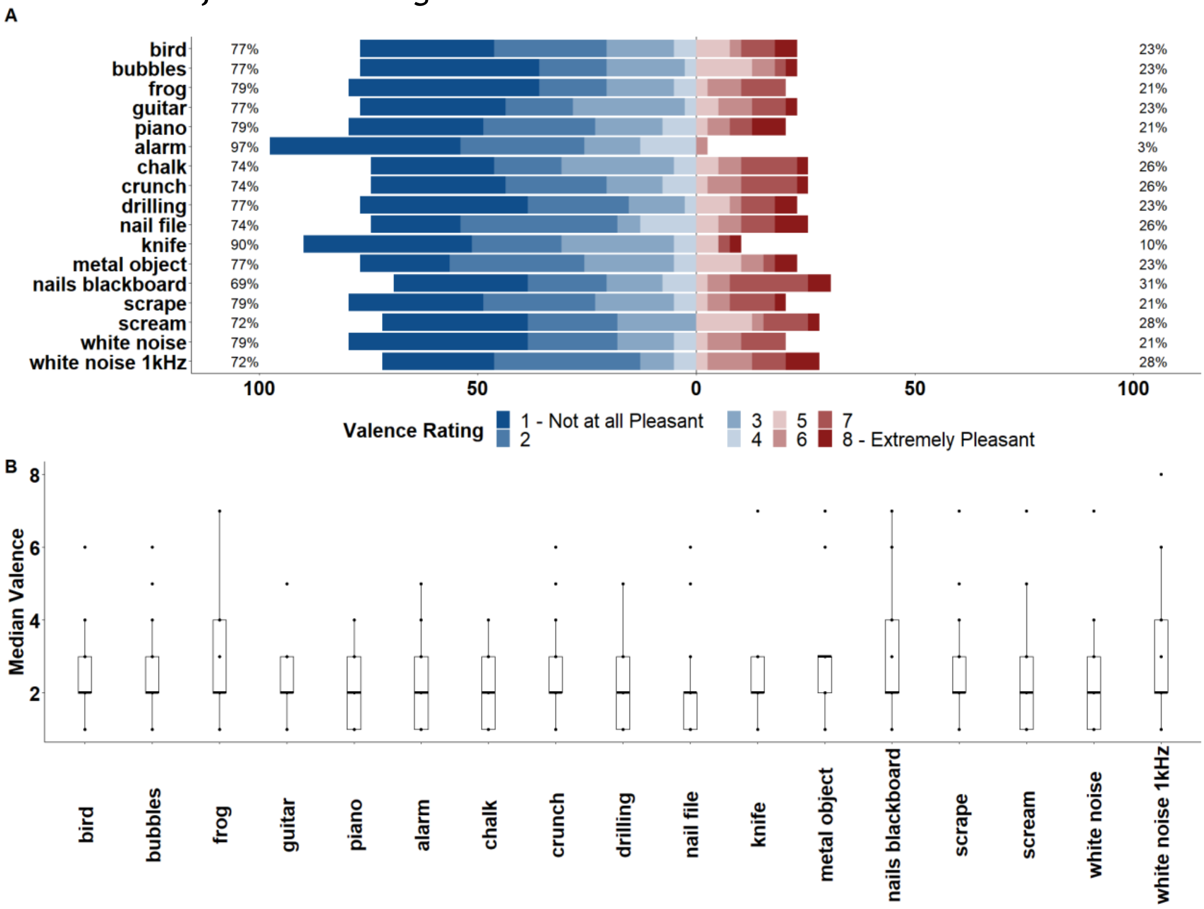
Pupil size responses to different sounds



Note: A) Mean pupil size over time. Vertical gray bars indicate the standard error of the mean B) Median of the mean pupil size across time.

Supplementary Figure 9

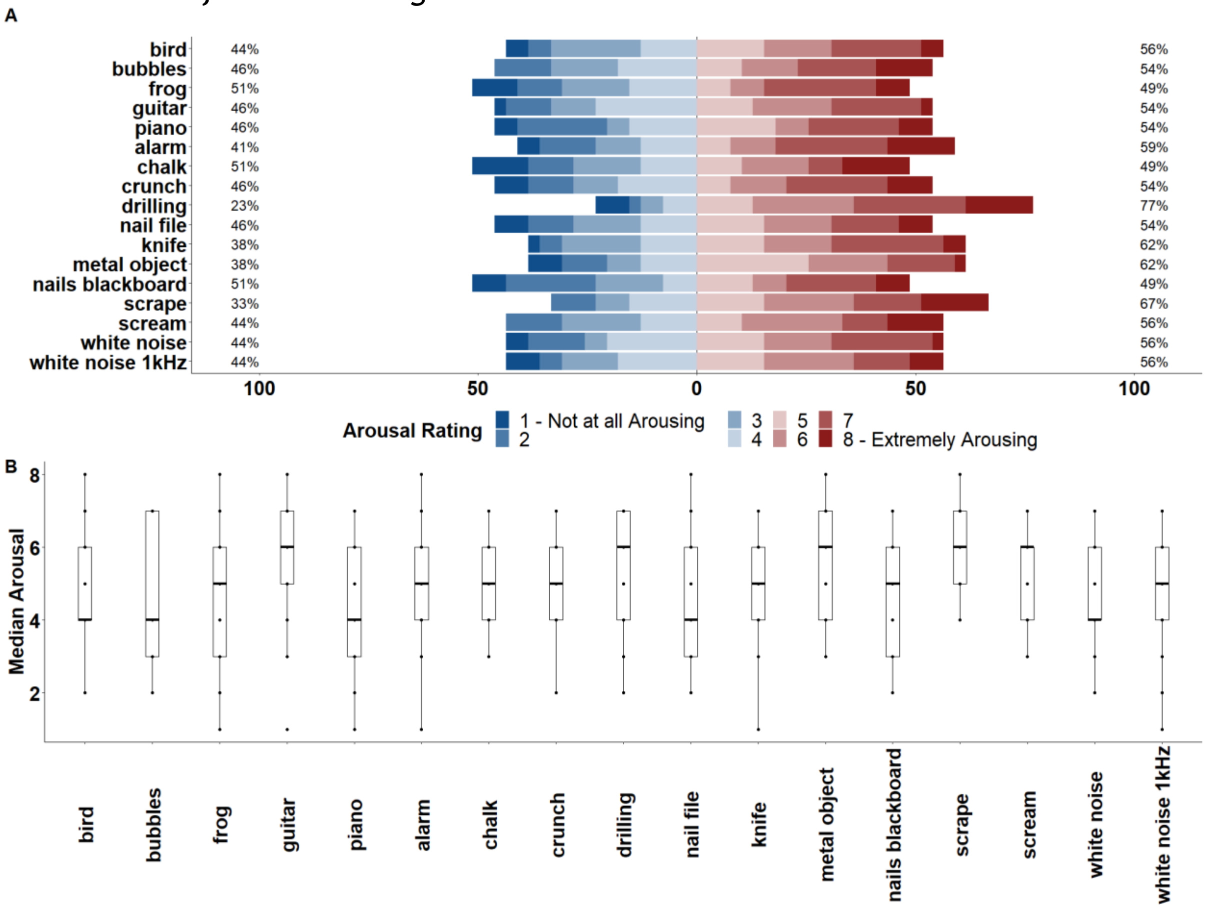
Distribution of valence ratings



Note: A) Percentage of ratings belonging to each Likert point (1-8) and B) median valence rating for each sound.

Supplementary Figure 10

Distribution of arousal ratings



Note: A) Percentage of ratings belonging to each Likert point (1-8) and B) median arousal rating for each sound.

Relationship between stimulus valence and fundamental frequency in predicting pupil size valence and arousal

As seen in Supplementary Figure 11A and 11B, pupil size is slightly larger for positive than negative stimuli, however, this difference is small due to the substantial overlap in the distributions between conditions. The fundamental frequency of sounds also does not appear to correlate with pupil size (see Supplementary Figure 11C). We performed linear mixed effects (LME) models (package *lme4*), to determine if the mean pupil size across the trial duration can be predicted from stimulus valence and the fundamental frequency of stimuli. The model included *Stimulus Valence* (Positive vs Negative) and *Fundamental Frequency* as fixed effects. Subjects were added as random intercepts with a random slope for *Stimulus Valence*. A random intercept was included for *Items*, accompanied by a random slope for the *Fundamental Frequency*. We observed no significant main effects or interactions (see Supplementary Table 8 and Supplementary Figure 11D), confirming that the valence and fundamental frequency of stimuli did not influence pupillary responses.

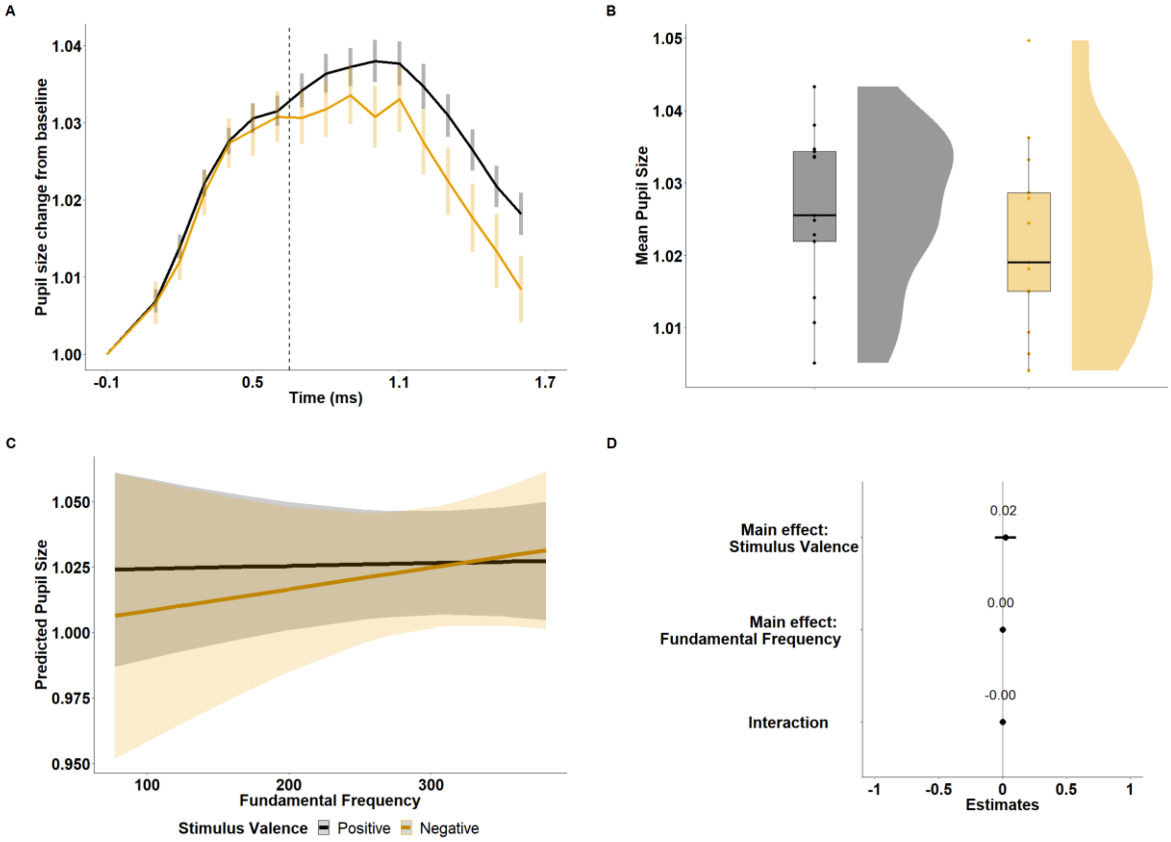
Supplementary Table 8

Type III Wald chi-square tests and R-squared values for the pupil model and each of the fixed effects.

	Chisq	Df	P-value	R^2 Fixed (CI)
Pupil				
Full model				0.009 (0.002-0.06)
Stimulus Valence	0.445	1.000	0.505	0.006 (0.00-0.04)
Fundamental Frequency	0.007	1.000	0.933	0.002 (0.00-0.03)
Stimulus Valence X Fundamental Frequency	0.254	1.000	0.614	0.001 (0.00-0.02)

Supplementary Figure 11

Summary of pupil size effects by valence and fundamental frequency of stimuli



Note. A) Proportional mean pupil size change from baseline over time. The vertical dashed line indicates sound onset. B) Mean pupil size averaged over time. C) Pupil size predicted from the valence and fundamental frequency of stimuli. D) Fixed effect estimates (labelled dots) derived from the linear mixed effects model of pupil size; bars represent 95% CIs for the estimates.

For valence and arousal ratings, instead LME models, we conducted cumulative-link mixed (CLM) models (package *ordinal*) to account for the ordinal nature of the data. The models included the same random effects structure as the pupil model. Again, we observed no significant main effects or interactions (see Supplementary Table 9 and Figures 12 and 13 C-D). As seen in Supplementary Figures 12A and 13A, the predicted probabilities for each rating category do not differ between positive and negative sounds for either valence or arousal ratings. Similarly, the fundamental frequency of sounds does not mediate the relationship between ratings and the stimulus valence (see Supplementary Figures 12B and 13B).

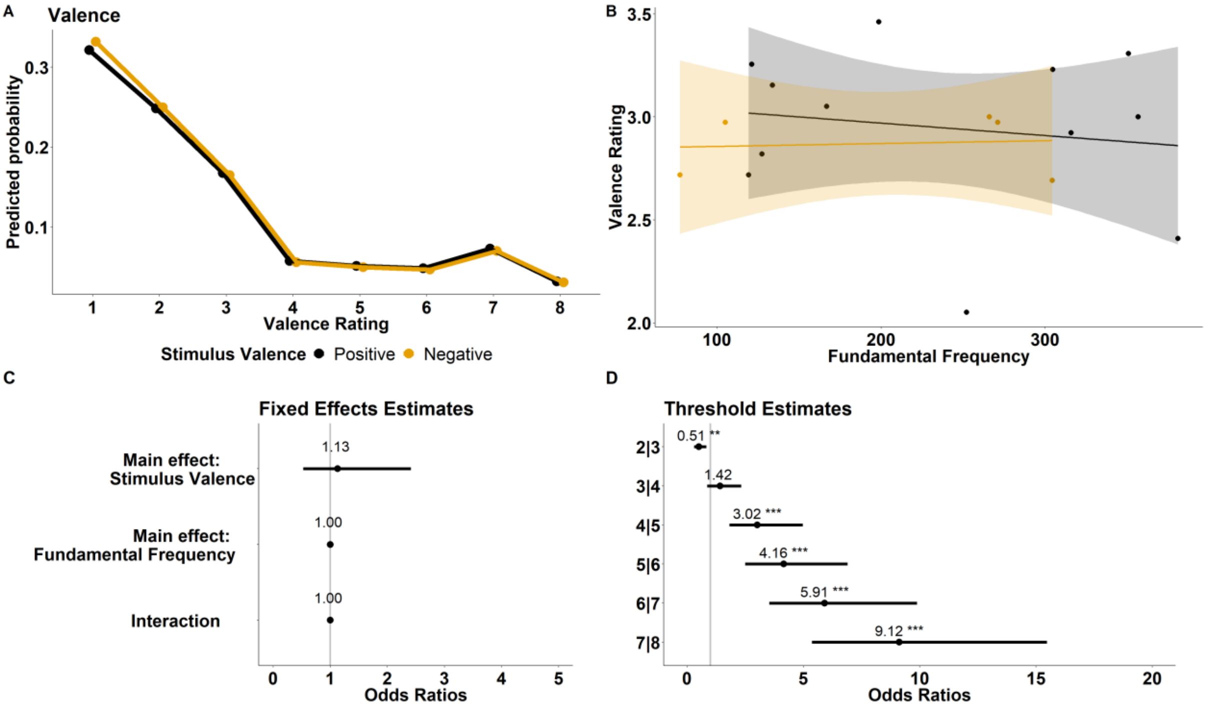
Supplementary Table 9

Type II Likelihood-ratio tests and R-squared values for the valence and arousal models and each of the fixed effects.

	LR Chisq	Df	P-value	McFadenn Pseudo R^2	Negelkerke Pseudo R^2
Valence					
Full model				0.0001	0.0003
Stimulus Valence	0.112	1.000	0.738		
Fundamental Frequency	0.035	1.000	0.852		
Stimulus Valence X Fundamental Frequency	0.039	1.000	0.843		
Arousal					
Full model					
Stimulus Valence	0.349	1.000	0.555	0.0003	0.001
Fundamental Frequency	0.225	1.000	0.635		
Stimulus Valence X Fundamental Frequency	0.113	1.000	0.736		

Supplementary Figure 12

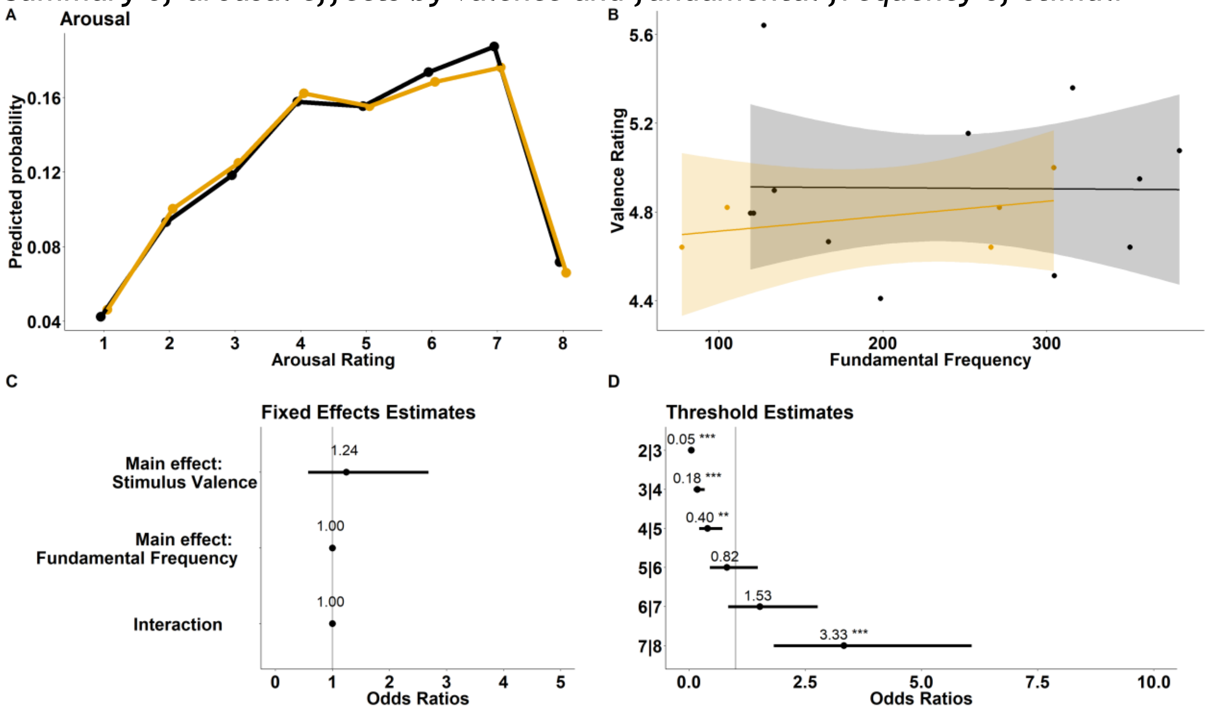
Summary of valence effects by valence and fundamental frequency of stimuli



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of valence ratings. B) Scatterplot of the relationship between valence ratings and fundamental frequency C) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates. D) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates.

Supplementary Figure 13

Summary of arousal effects by valence and fundamental frequency of stimuli



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of arousal ratings. B) Scatterplot of the relationship between arousal ratings and fundamental frequency C) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates. D) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates.

Overall, the findings from these analyses showed that sounds of short duration and high intensity cannot be differentiated by their valence from subjective ratings of valence and arousal or pupil size. The fundamental frequency of these sounds also does not appear to correlate with any of the outcome measures. These results suggest that the sounds and consequently their valence, may have been undistinguishable due to their short duration or that the high-volume intensity of the stimuli led them to be all perceived as unpleasant and arousing. This may also potentially limit the impact of the fundamental frequency on the sound perception. It is also possible that the combination of short stimulus duration and high intensity amplified the perceived low valence of the stimuli.

3.5.3 Supplementary Materials 6: Selection of an appropriate filter based on an examination of filtering artifacts

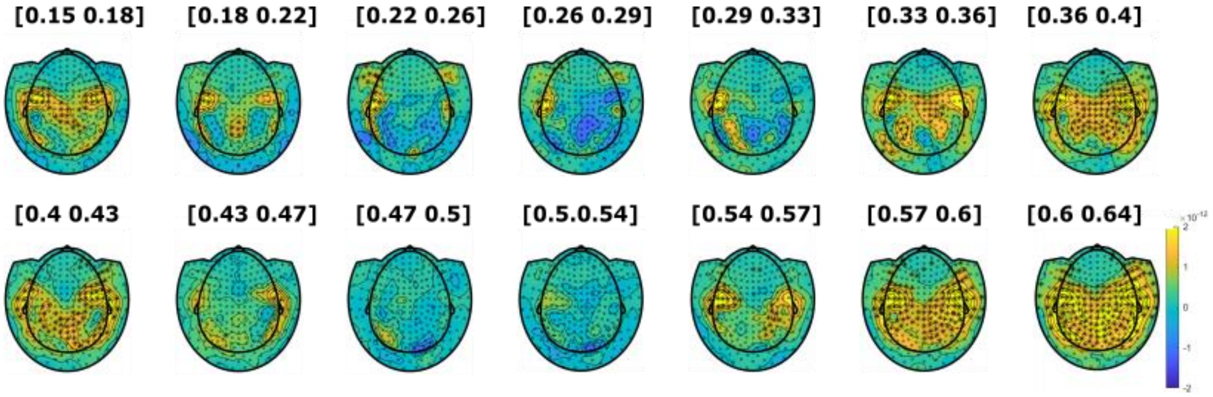
Applying a filter to E/MEG data is a widespread practice during pre-processing and analysis, aimed at reducing noise in the data. For example, high-pass filtering is used to remove slow drifts from the signal, while band-pass filters are often used to examine signals originating from specific frequencies (de Cheveigné & Nelken, 2019). However, filtering is shown to cause significant distortions of the shape and the timing of the signal. Specifically, non-causal filters, are typically applied forward and backward which can minimise phase delays but can introduce considerable time shifts and therefore, cause onsets to appear earlier (Subramaniyam, 2018). Such artifacts have been observed in both low-pass filtering (VanRullen, 2011b) and high-pass filtering (Acunzo et al., 2012; Rousselet, 2012). For high-pass filtering, the distortions are also shown to increase with an increase in the cut-off frequency (Acunzo et al., 2012). Consequently, some have argued against the use of filtering in general if examining onset latencies (VanRullen, 2011b), while others have suggested that the issue can be resolved through the application of filters based on an adequate consideration of the effects of filtering, specific to the data in question (Widmann et al., 2015). According to Rousselet (2012), causal filters which are often applied only forward may minimise timing artifacts and should the preferred option when attempting to make inferences regarding the onset of effects.

In the case of the present study, filtering may be even more problematic since for CS+ trials during Acquisition, the visual CS is followed by an auditory UCS. Therefore, two separate processes associated with two onsets are observed in a single trial, introducing the significant risk of detecting effects within the time window of the CR (before UCS onset) that may be driven by the UCR (after UCS onset). In an attempt to prevent this and to select the most appropriate filter that offers the least amount of signal distortion, we followed advice by Widmann et al. (2015) and filtered the data using different filtering parameters and filter types.

First, we applied a band-pass filter between 1 and 8 Hz using the default filter in Fieldtrip (a Butterworth IIR twopass filter, filter order=4). We chose this filter and high cut-off, high-pass frequency since based on the previous research discussed above, it would be a likely candidate for introducing significant timing distortions. The filter was applied on the pre-processed sensor level data for each subject and on each trial between -0.65 to 0.9 seconds. Trial-level data were baseline corrected using a baseline period of -0.65 to 0 seconds. A planar gradient transformation was then applied to facilitate interpretation of the MEG fields. To account for potential baseline differences during the Habituation phase, for each subject and condition, we subtracted the signal during Habituation from that during Acquisition. Next, we performed a two-tailed Monte-Carlo permutation paired t-test comparing CS+ and CS- trials during Acquisition. The test was performed in the time window between 0 and 0.65 s post-CS onset and thus, did not include the UCS presentation (i.e., only examined the CR effects). The results revealed two significant clusters of activation somewhere in the time windows between 0.15 and 0.45 s and between 0.52 and 0.65s (see Supplementary Figure 14). When comparing the time courses of CS+ and CS- trials in the channels identified as significant in the cluster-based permutation (see Supplementary Figure 15) these differences and may suggest that activity that this activity reflects an anticipation of the upcoming aversive stimulus. However, when looking at the topography of the unfiltered signal during the time when the UCS was presented (see Supplementary Figure 16), it becomes apparent that the topography is very similar to that observed in Supplementary Figure 14, around the significant time windows identified by the cluster-based permutation. This suggests that the filtering may have shifted the onset of the UCS, producing and artificial effect.

Supplementary Figure 14

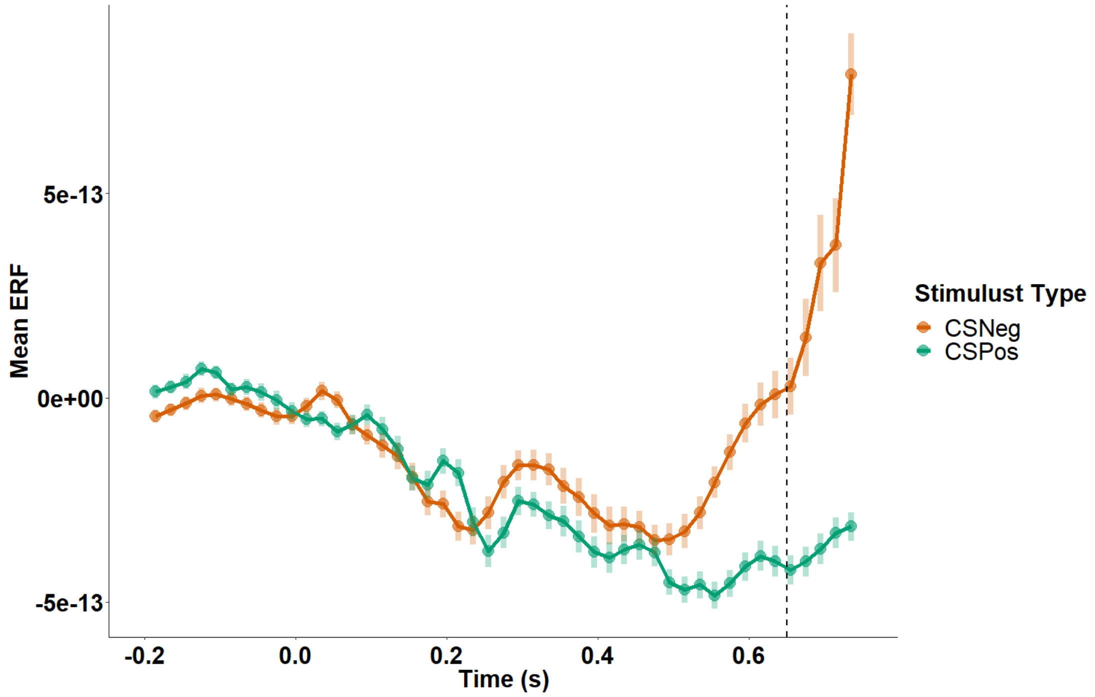
Topographical representation of the average ERFs between 1 and 8 Hz from around 0.15 to 0.65 seconds reflecting the CR.



Note. The white stars indicate clusters significantly different at the level of 0.05 between CS+ and CS- trials during Acquisition.

Supplementary Figure 15

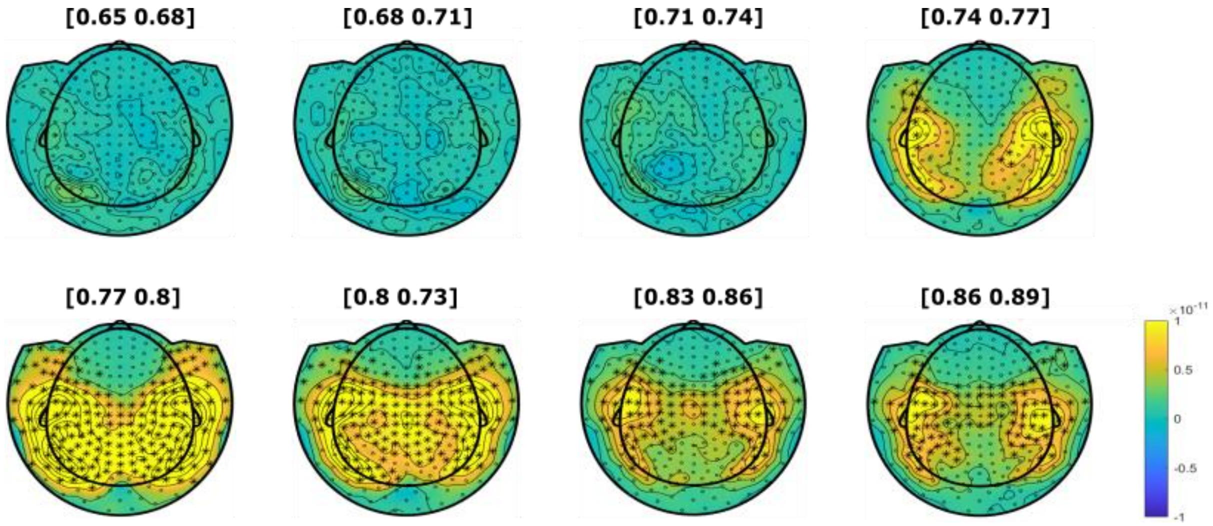
Average ERF time course between 1 and 8 Hz for CS+ and CS- trials during Acquisition.



Note. ERFs were averaged across the significant channels identified across the two significant clusters.

Supplementary Figure 16

Topographical representation of the ERFs between 0 and 8 Hz from 0.65 to 0.85 seconds reflecting the UCR.



To investigate this further and to find the least distortive filter parameters, we applied several infinite impulse response (IIR) and finite impulse response (FIR) causal, and non-causal filters within the theta frequency band. For the purposes of the present analysis, we define causal filters as those that are applied only forward (one-pass) and non-causal filters as filters that are applied forward and backward (two-pass). Specifically, we applied the following in addition to the initial Butterworth two-pass filter:

1) A low-pass Butterworth IIR two-pass filter (0 – 8 Hz)
2) A band-pass Butterworth IIR one-pass filter (1 – 8 Hz)
3) A low-pass FIR two-pass filter (0– 8 Hz)
4) A band-pass FIR one-pass filter (0.01 – 8 Hz)
5) A band-pass FIR one-pass filter (0.1 – 8 Hz)
6) A band-pass FIR one-pass filter (1 – 8 Hz)
7) A band-pass FIR two-pass filter (0.01 – 8 Hz)
8) A band-pass FIR two-pass filter (0.1 – 8 Hz)
9) A band-pass FIR two-pass filter (1 – 8 Hz)

These filters were applied using the same procedure as described for the initial filter (1-8 Hz Butterworth IIR two-pass filter). Here we present, the filtering effects for the CS+ and CS- conditions during Acquisition. To descriptively compare the impact of the different filtering parameters, for each filtered dataset, we down sampled the data to 50Hz and extracted sensor-level data from the channels that were identified as significant by the cluster-based permutation for the ERFs filtered using the 1-8 Hz Butterworth IIR two-pass filter. Each filter was then plotted against the unfiltered data (see Supplementary Figure 17 and 18).

To ease visualisation of the different onsets, Figure 17A includes the average ERFs during the CS+ trials and part of the UCS presentation (-0.2 to 0.75 s), thus reflecting both the CR and UCR. Figure 17B focuses only on the CR (-0.2 to 0.65 s). When examining the figure, several patterns emerge. First, the causal FIR filtering appears to cause large distortions, predominantly on the onset of effects. As seen in Figure 17A, compared to the unfiltered signal and the non-causal FIR filters, the onsets of peaks within the ERFs are shifted later in time by about 200 ms. The filters with high cut-off frequency of 1 Hz show high level of signal distortion that is also dependent on the filter type. The initially computed Butterworth non-causal filer leads to a shift of peak onsets earlier in time by around 100-200 ms but it also increases the amplitude of these peaks. The causal FIR and IIR (Butterworth) filters follow a similar trajectory to that of the causal FIR filters with lower cut-off frequencies. Finally, the non-causal FIR 1-8 Hz filter distortions mostly affect the amplitude of the peaks.

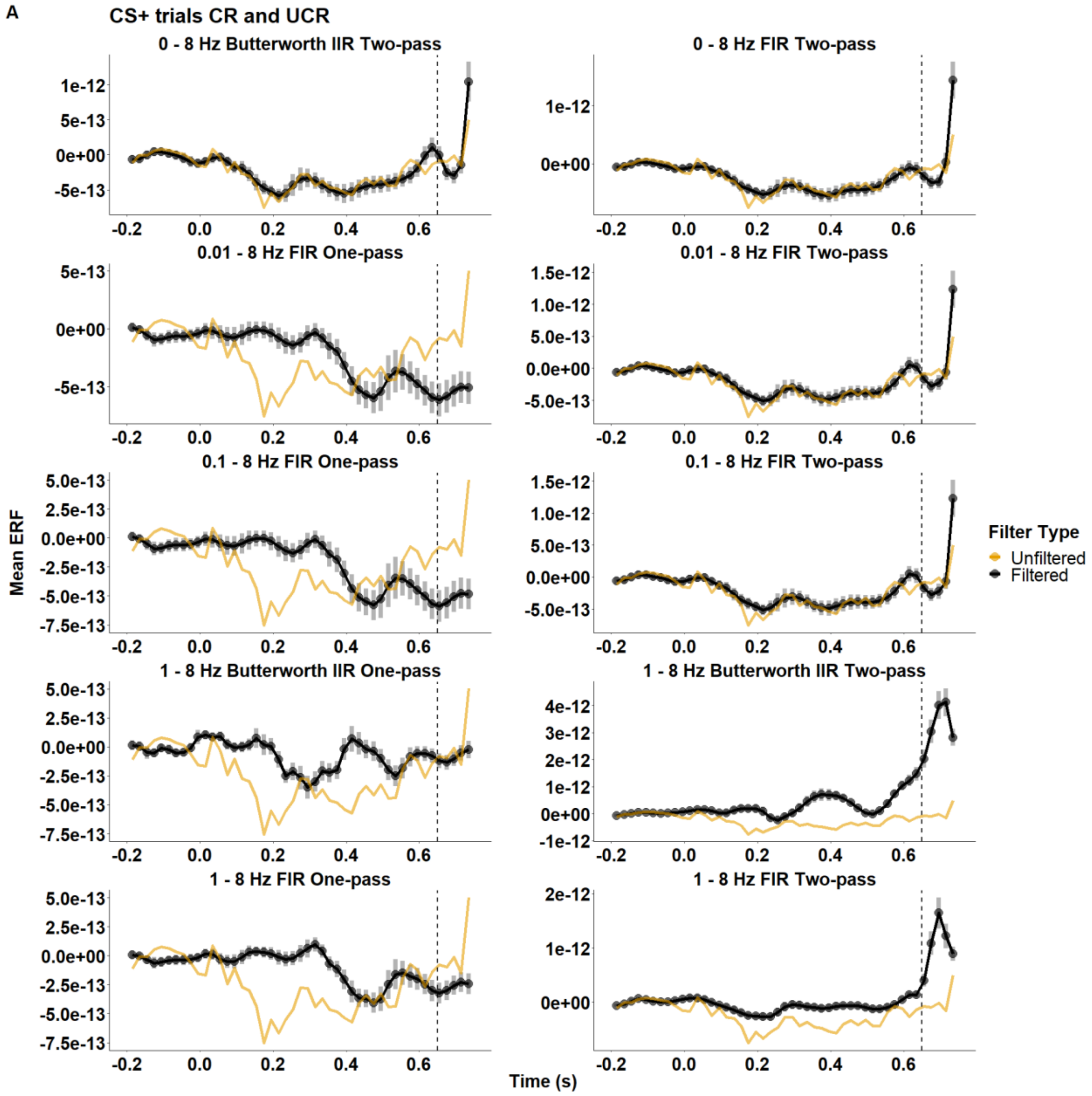
In contrast, the non-causal FIR filters with lower high-pass cut off frequency up to 0.1 seconds show the least amount of distortion and similar to the low-pass filters. When focusing on the CR only (-0.2-0.65 seconds, Figure 17B), however, a small amplitude difference can be seen between the filtered and unfiltered signal, immediately prior to UCS onset. When examining the average ERFs for CS- trials where no UCS was presented, no distortions are observed. This may suggest, that even with those filters, a signal leakage from the presentation of the UCS may be present in the CS+ condition which could contribute to observing an artificial difference between the two conditions.

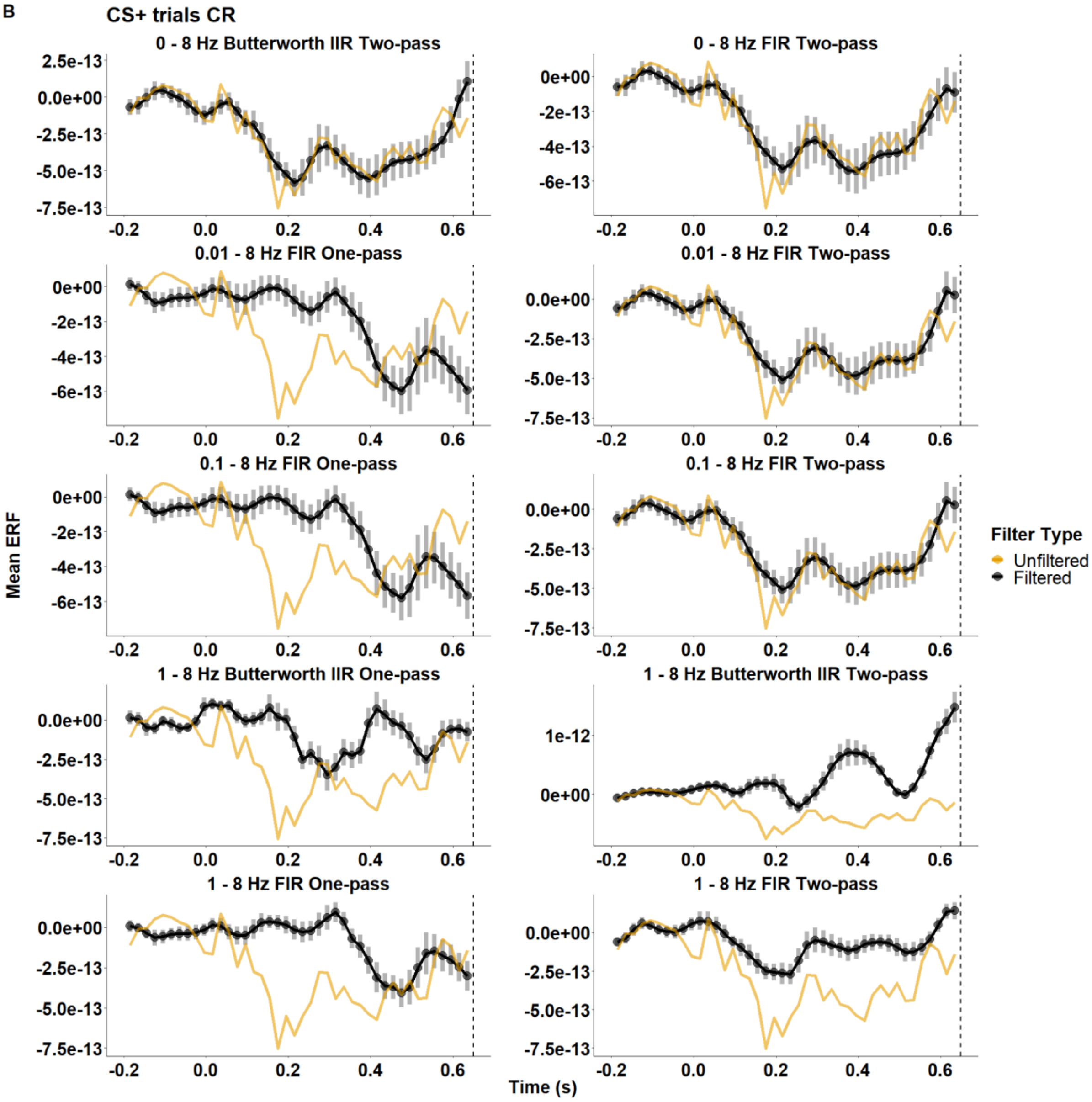
To formally test this descriptively, we performed the same filtering procedures but prior to filtering, we trimmed the data to -0.65 to 0.65 ms, thus excluding the onset and duration of the UCS. As seen in Figure 19, the small distortions that were observed for the low-pass and low cut-off band pass FIR filters earlier are no longer present.

Therefore, based on the present data, for trials that contain only one onset of interest low-pass filters do not cause significant distortion to the shape or timing of effects. Similar effects are also observed for band-pass filters with cut-off frequency of 0.01 and 0.1 Hz, with the exclusion of causal FIR filters which cause significant timing distortions. This is contrary to Rousselet's (2012) observation of high levels of distortion for causal high-pass filters and suggests that filter causality may have different impact on high compared to band pass filters. Finally, band-pass filters with a high cut-off frequency of 1 regardless of whether causal or non-causal also produce large artifacts. The distortions we observed using band-pass filtering are consistent with previous research on filtering artifacts suggesting that high-pass filtering creates distortions to the data that increases with the increased in the high-pass cut-off frequency (Acunzo et al., 2012; Rousselet, 2012; Widmann et al., 2015).

Supplementary Figure 17

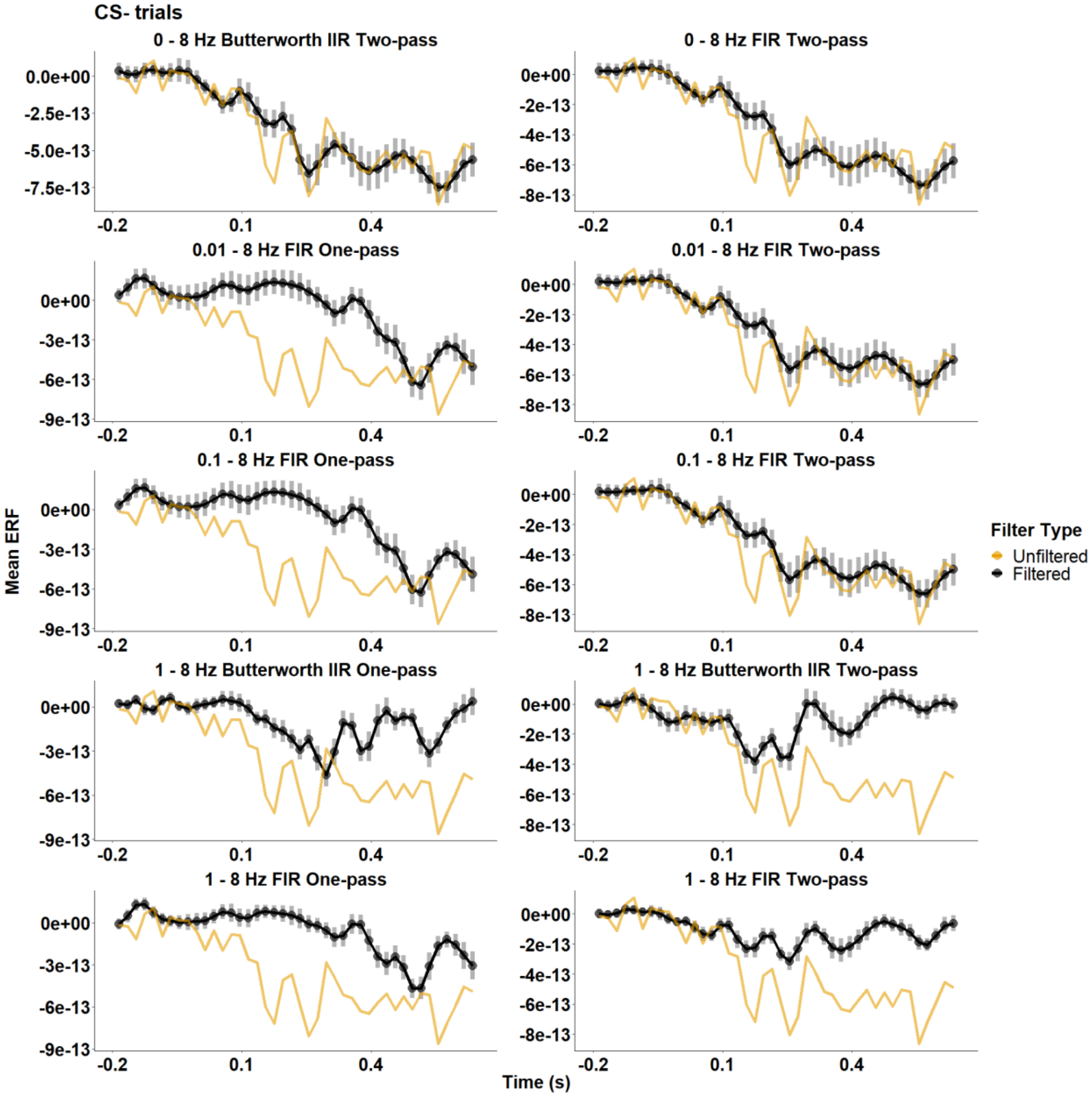
Comparison of filtered and unfiltered event-related fields for CS+ trials including A) both the CR and UCR (-0.2 to 0.75 s) *and B) only the CR* (-0.2 to 0.65 s)




B
CS+ trials CR
0 - 8 Hz Butterworth IIR Two-pass
0 - 8 Hz FIR Two-pass
0.01 - 8 Hz FIR One-pass
0.01 - 8 Hz FIR Two-pass
0.1 - 8 Hz FIR One-pass
0.1 - 8 Hz FIR Two-pass
1 - 8 Hz Butterworth IIR One-pass
1 - 8 Hz Butterworth IIR Two-pass
1 - 8 Hz FIR One-pass
1 - 8 Hz FIR Two-pass
2.5e-13
0.0e+00
-2.5e-13
-5.0e-13
-7.5e-13
0e+00
-2e-13
-4e-13
-6e-13
1e-12
-0.2
0.0
0.2
0.4
0.6
Mean ERF
Time (s)
Filter Type
Unfiltered
Filtered

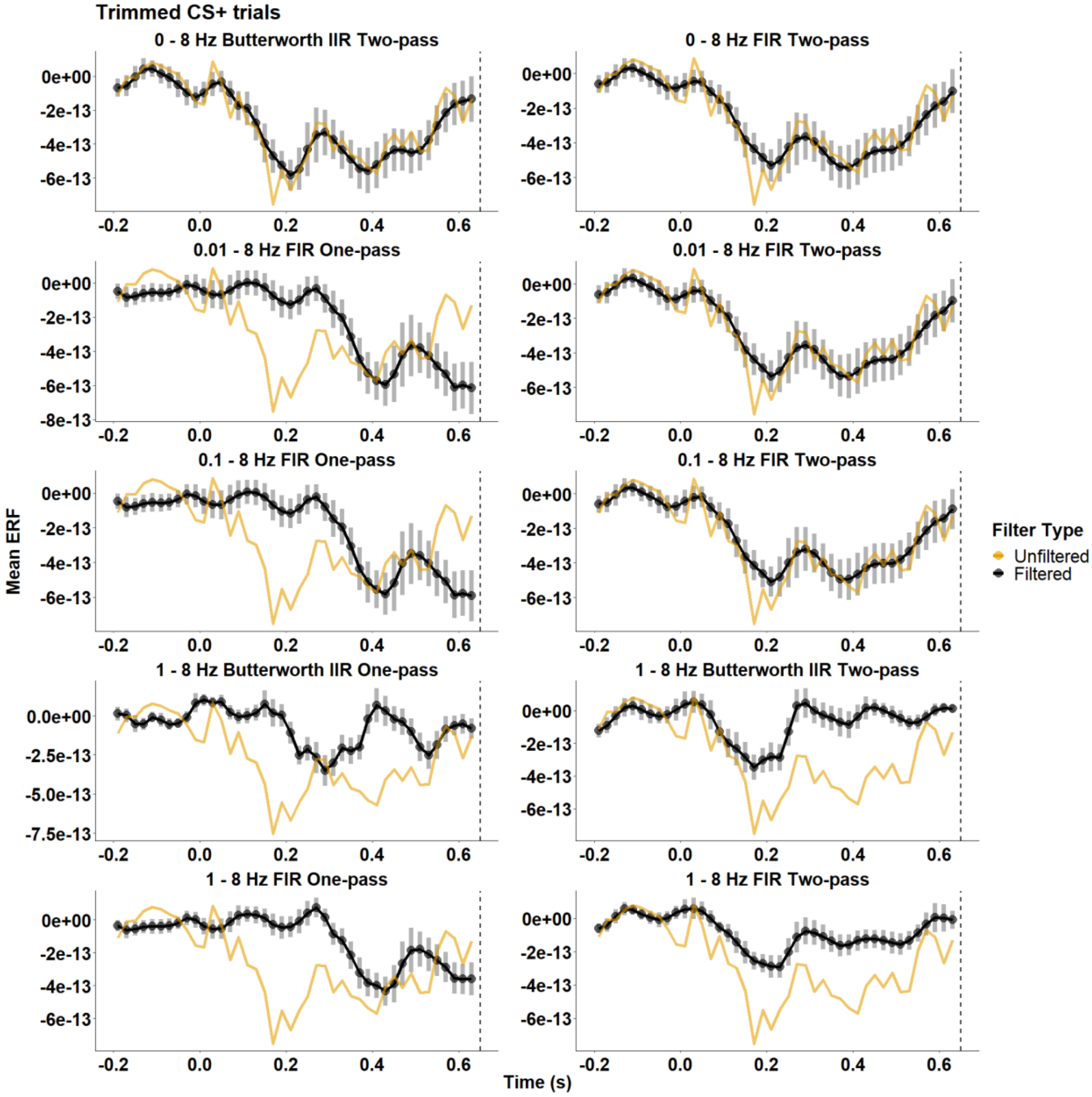
Supplementary Figure 18

Comparison of filtered and unfiltered event-related fields for CS- trials



Supplementary Figure 19

Comparison of filtered and unfiltered event-related fields for CS+ trials trimmed to 0.65 s prior to filtering.



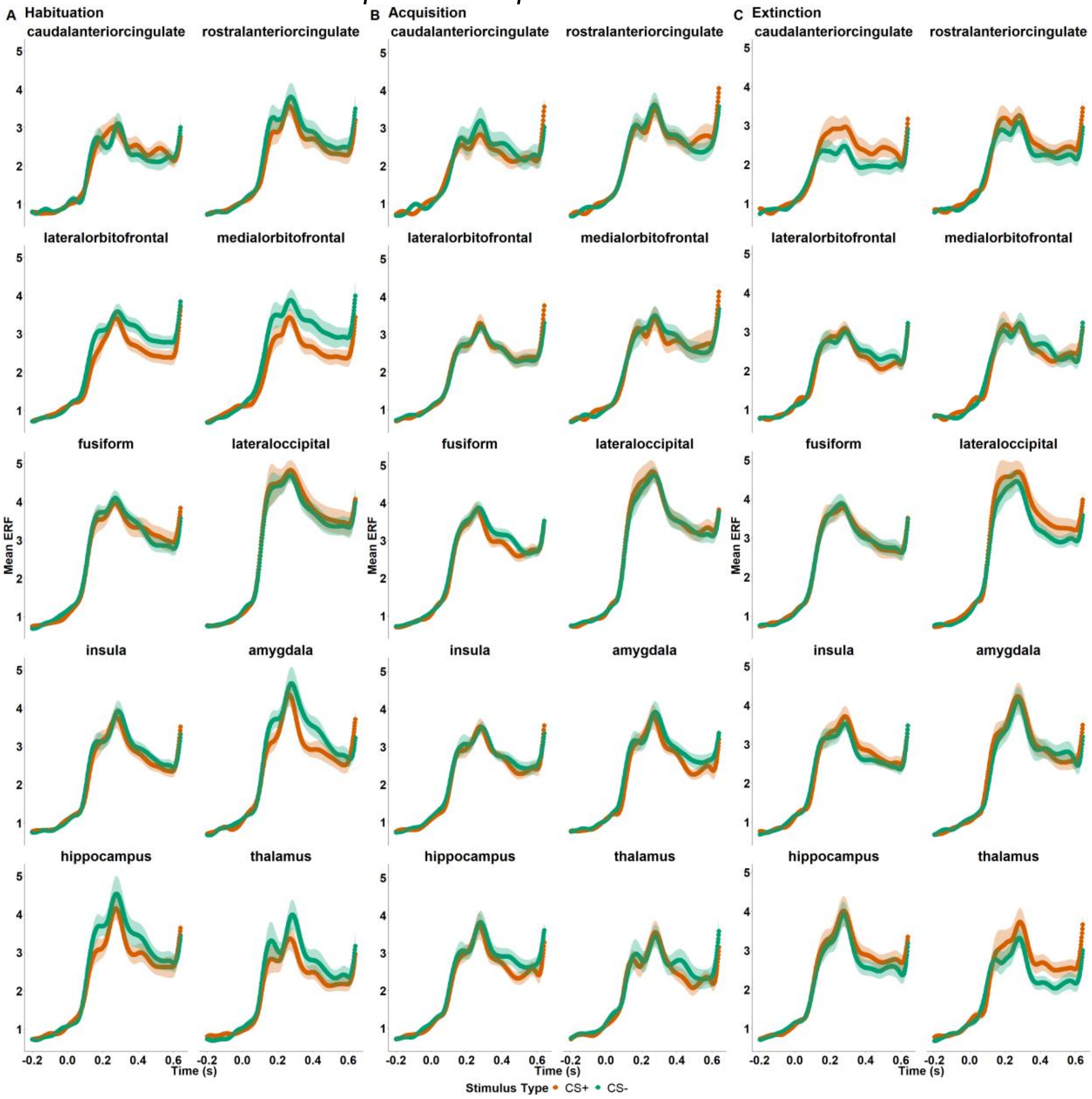
Overall, our findings suggest that a combination of factors can influence the shape and onset of events, including the cut-off frequency, filter causality as well as whether finite or infinite impulse filters. Furthermore, trials that contain the presentation of more than one stimulus require even more caution, as signal smearing can be observed between the onsets of the two events. These results demonstrate that when it comes to filtering, there is no one-size-fits-all and highlight the necessity to cautiously examine and select the most appropriate filter specific to data under investigation.

Based on the present results and recommendations by Widmann et al. (2015) for using low high-pass cut-off frequencies of 0.01 to 0.05, we selected the 0.01-8 Hz FIR causal (two-pass) filter for all analyses of ERFs. In addition, before filtering was performed, data were trimmed to -0.65 to 0.65 s to ensure that there would not be any signal smearing originating from the UCS.

3.5.4 Supplementary Materials 7: Source level visualization for each condition and experimental phase

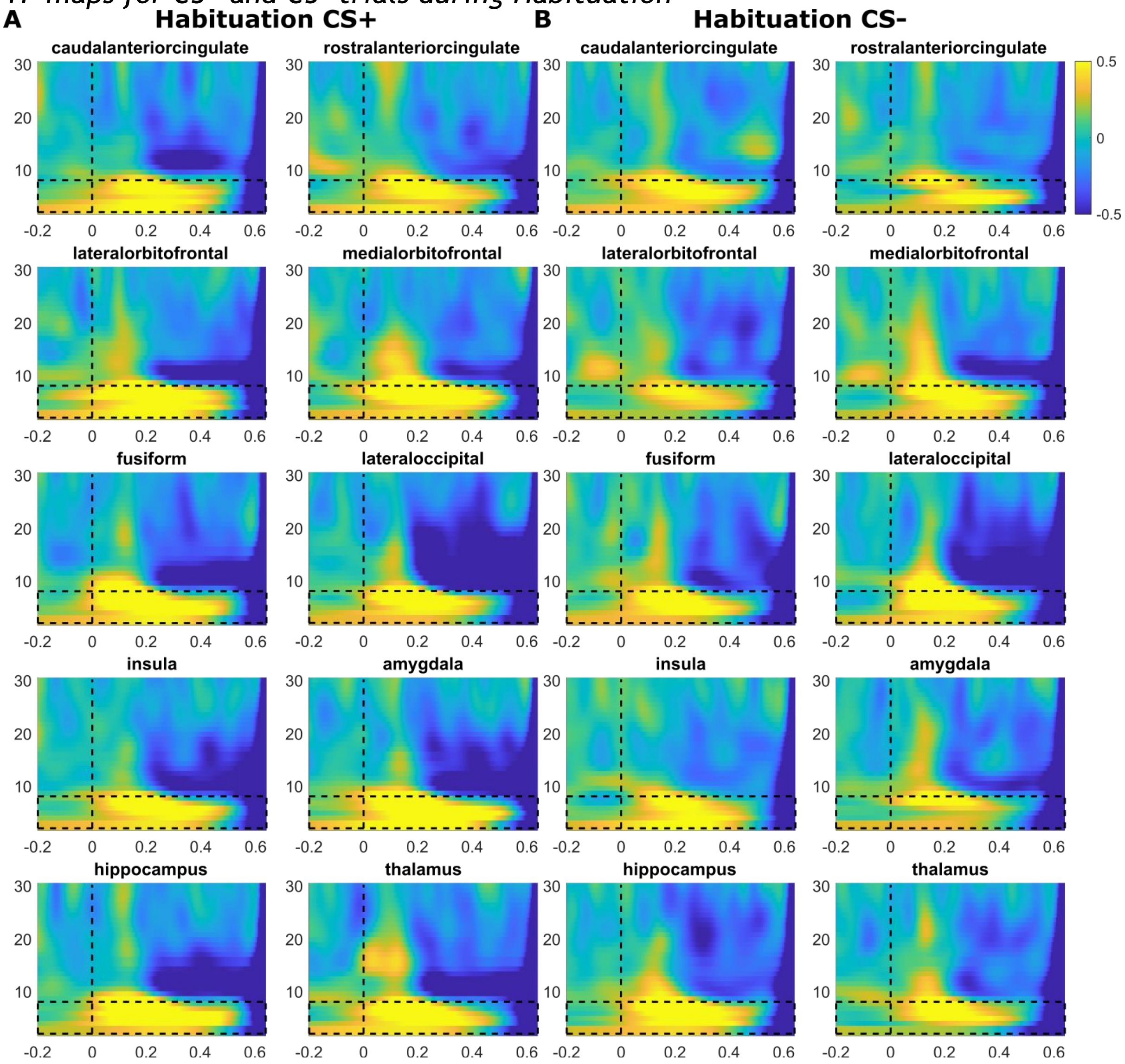
Supplementary Figure 20

Mean ERFs in ROIs across experimental phases



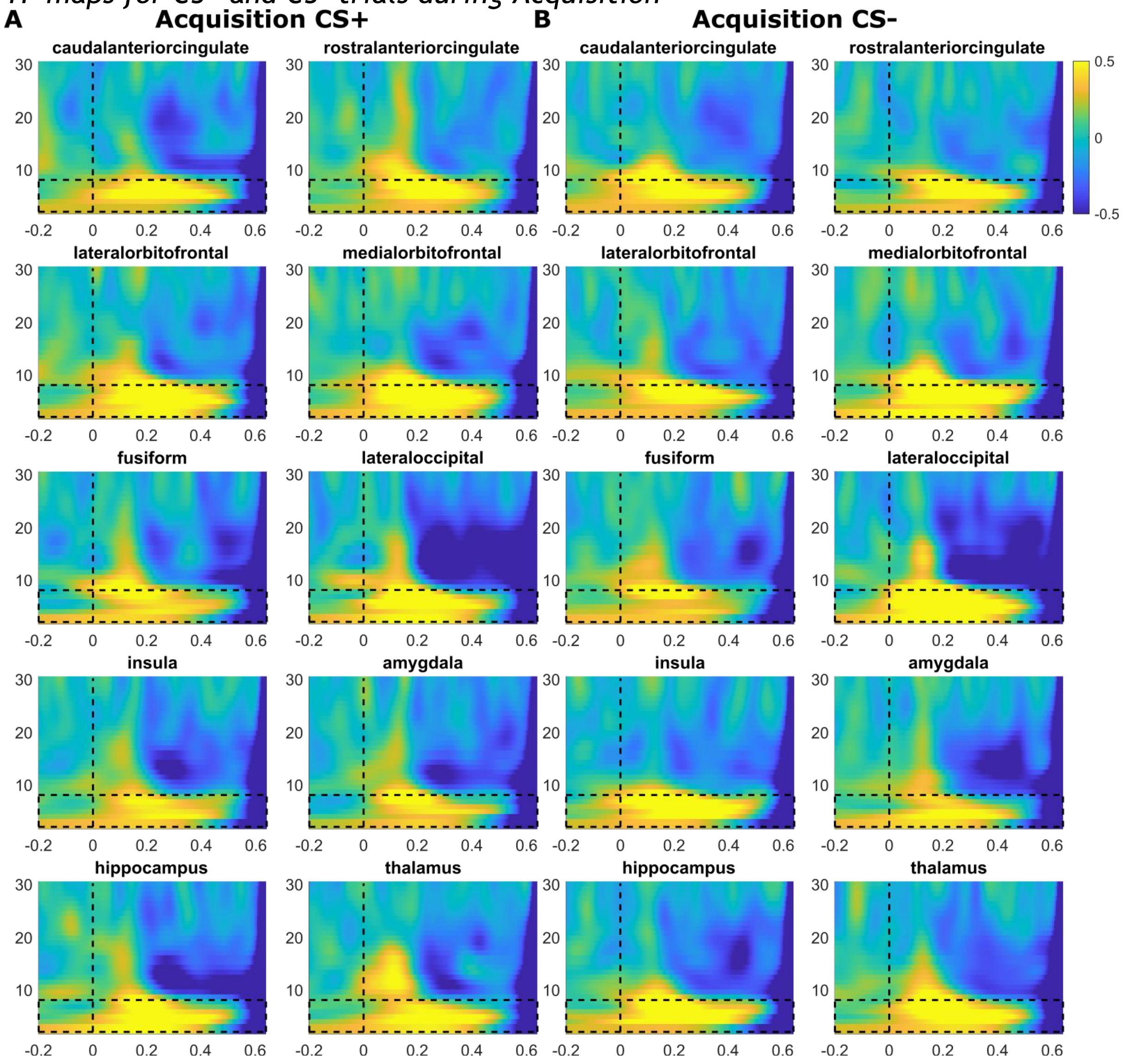
Supplementary Figure 21

TF maps for CS+ and CS- trials during Habituation



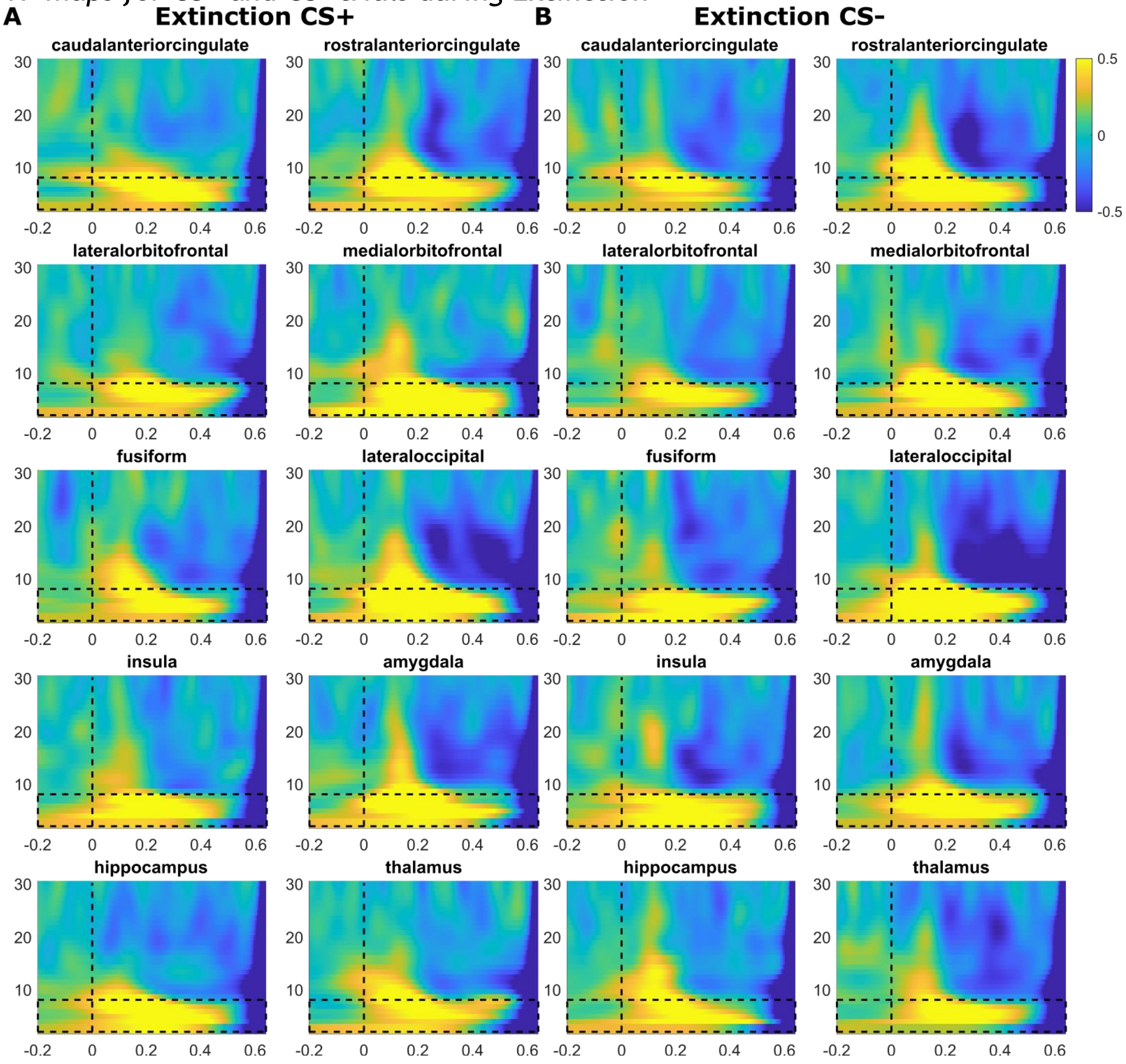
Supplementary Figure 22

TF maps for CS+ and CS- trials during Acquisition



Supplementary Figure 23

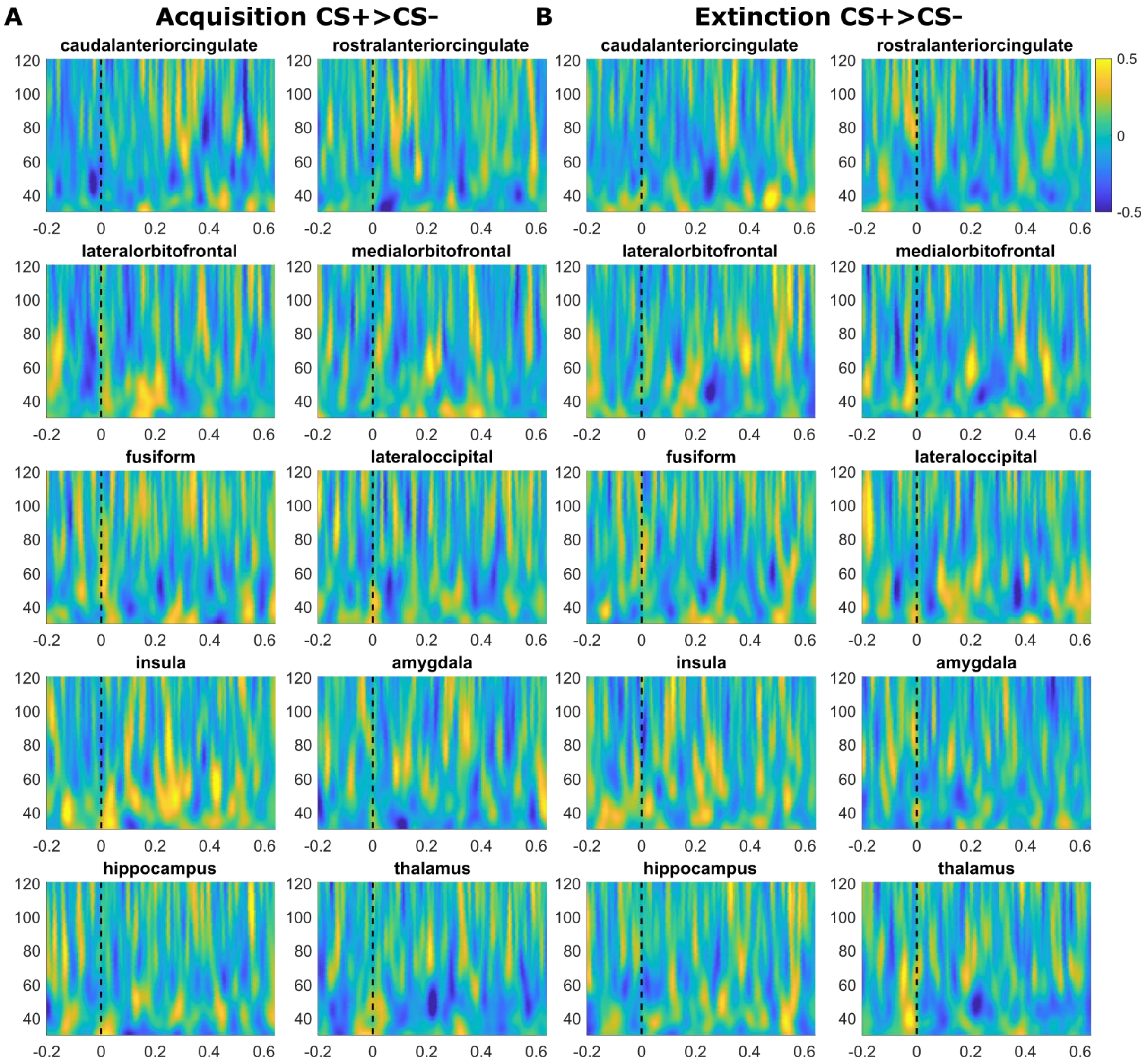
TF maps for CS+ and CS- trials during Extinction



3.5.5 Supplementary Materials 8: TF maps in higher frequencies (30-120 Hz)

Supplementary Figure 24

Habituation-baselined grand average time frequency maps, contrasting the difference between CS+ and CS- conditions during Acquisition and Extinction.



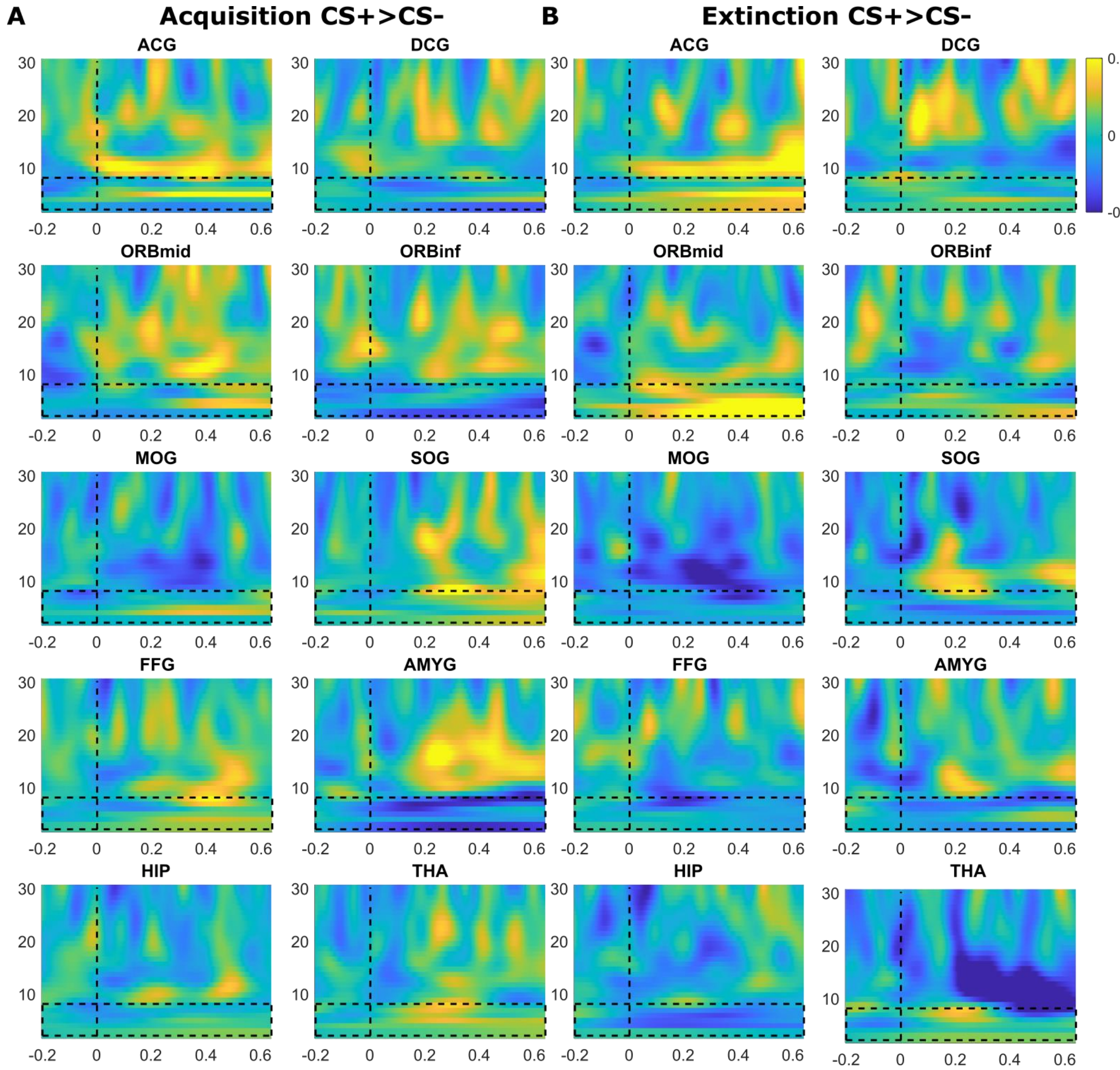
3.5.6 Supplementary Materials 9: Source reconstruction using LCMV beamforming on virtual channels

To validate the results obtained using MNE in Brainstorm, we performed a secondary source estimation using virtual channel LCMV beamforming in Fieldtrip. The two methods mainly differ in the assumptions they make in characterising sources, with MNE often being the preferred approach when attempting to localise deep structure activity such as the amygdala (Balderston et al., 2013; Dumas et al., 2011) as sources are reconstructed based on precisely defined subcortical regions and false detection of deep activity is less common than with other methods (Attal & Schwartz, 2013). In contrast, the main benefit of beamforming is that it estimates activity in a location of interest while blocking signals from other sources (Bourgeois & Minker, 2009).

MEG-MRI co-registration was performed using subject's digitised head shape and landmark information (nasion and peri-auricular points) using Fieldtrip. We used a single shell method for computing the head model. The inverse model was calculated for each trial using LCMV beamforming by reconstructing the sensor level data from MNI coordinates to regions from the AAL atlas (Tzourio-Mazoyer et al., 2002). Time frequency maps were computed at a trial-level in ROIs derived from the AAL atlas. Specifically we included the the anterior and median cingulate gyri (ACG and DCG), inferior and middle frontal gyri (ORBinf, ORBmid), middle and superior occipital gyri (MOG, SOG), fusiform gyrus (FFG), the amygdala, thalamus and hippocampus.(Amyg, Tha, Hip). We used an identical procedure for the computation of time frequency maps as that in the main analysis. As seen in Supplementary Figure 25 and 26, the pattern of results remains relatively similar to that observed from the MNE source estimation, specifically in the amygdala and the ACG. Similar to our main analyses, the cluster-based permutation on the habituation-baselined data also revealed no significant differences between conditions during either acquisition or extinction, not only at the theta band but across frequencies as well. These findings suggest that our results were not driven by the source reconstruction method.

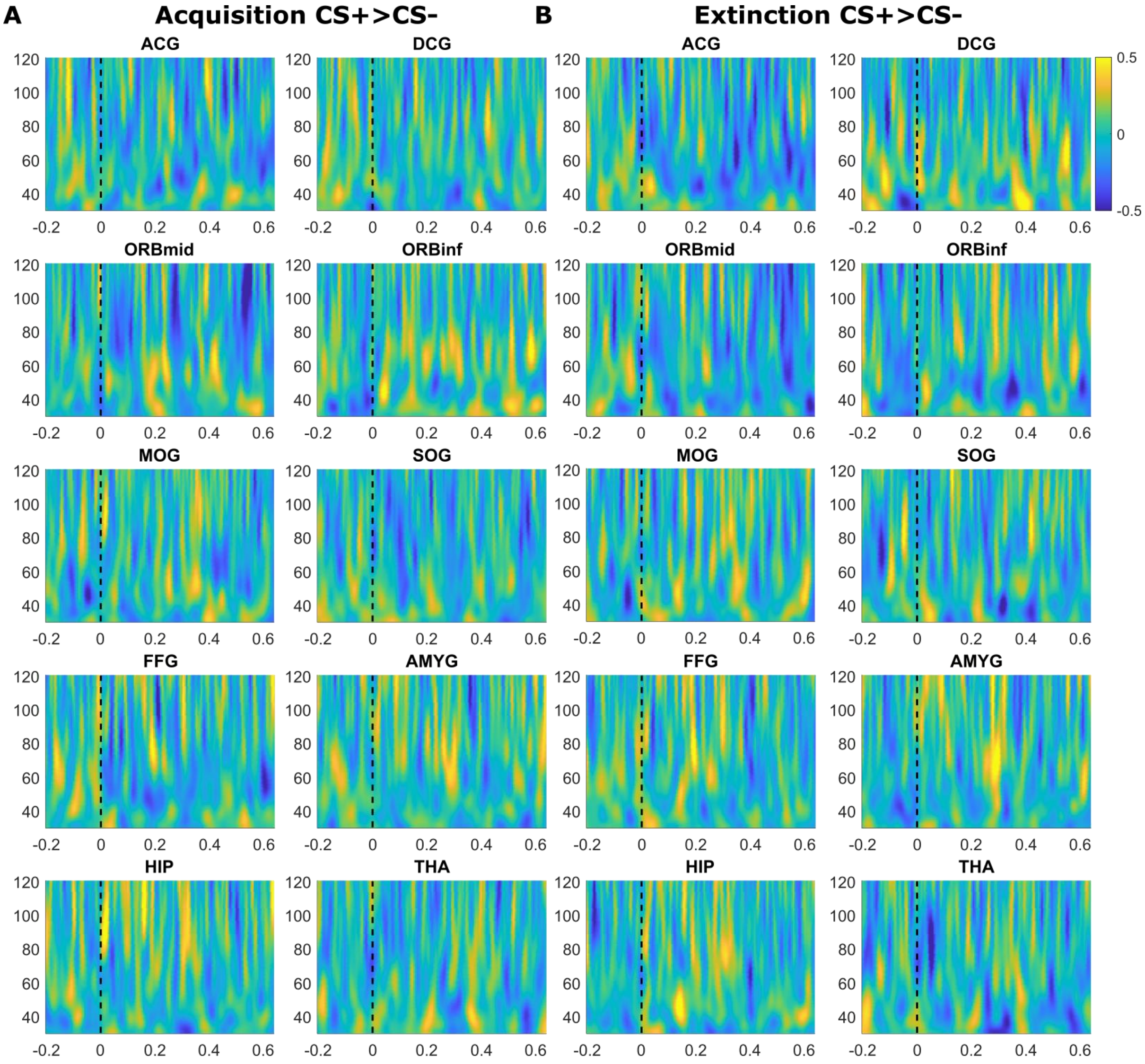
Supplementary Figure 25

Habituation-baselined grand average time frequency maps in frequencies below 30 Hz, contrasting the difference between CS+ and CS- conditions during Acquisition and Extinction.



Supplementary Figure 26

Habituation-baselined grand average time frequency maps in frequencies above 30 Hz, contrasting the difference between CS+ and CS- conditions during Acquisition and Extinction.



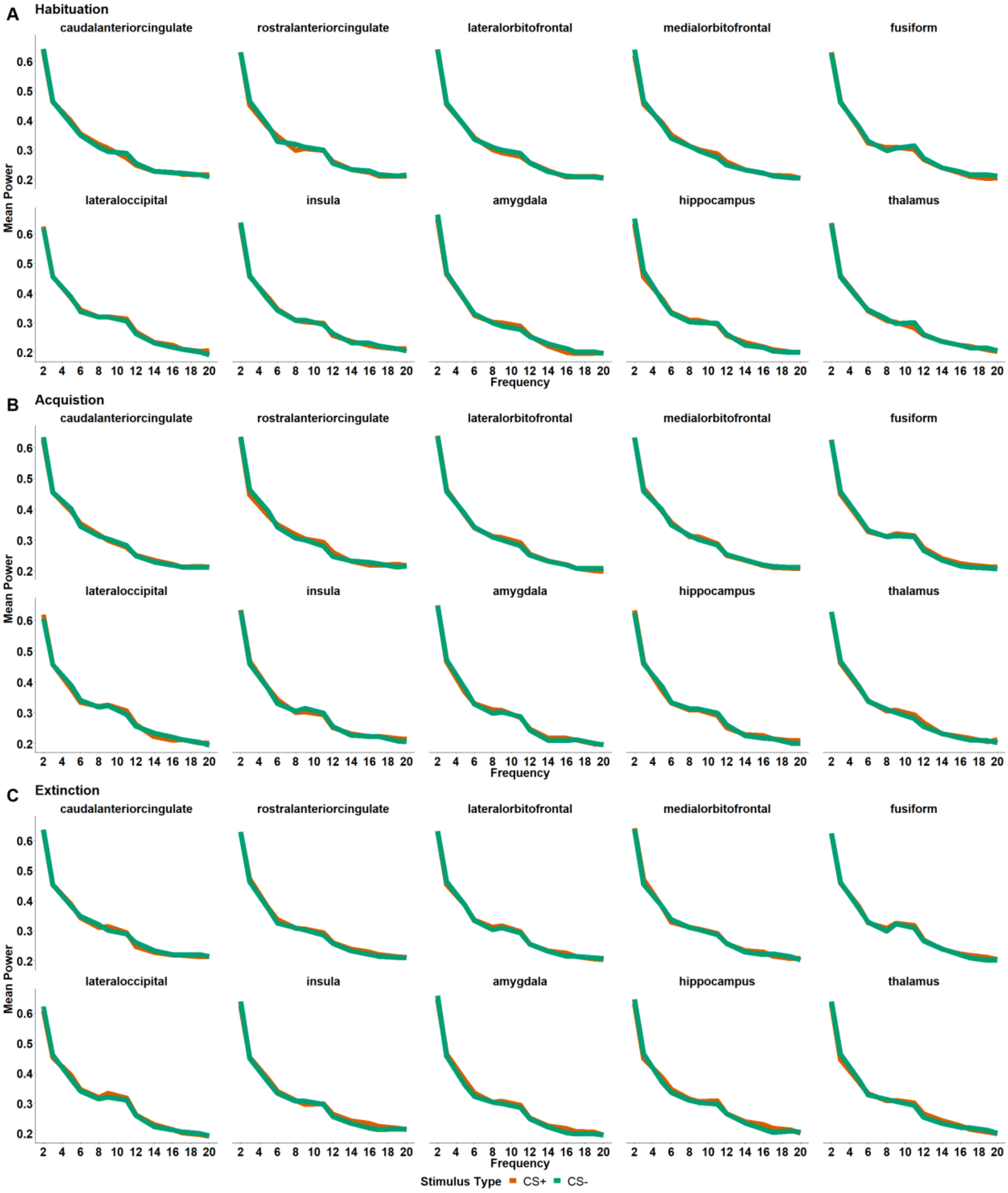
3.5.7 Supplementary Materials 10: Average oscillatory power in lower frequencies

To examine the specific frequencies that most strongly contribute to task-related changes, we performed a Fast Fourier Transformation (FFT) at trial level. The analysis used a hanning taper and was performed in the time window between 0 and 0.64 s post CS onset for frequencies between 1 and 20 Hz (frequency resolution of 1 Hz and frequency smoothing of 1 Hz). Supplemenrary Figure 27 shows mean oscillatory power for each condition and in each experimental phase. We employed a normalisation procedure similar to that adopted by Tzovara et al. (2019), in which the trial-level mean oscillatory power per frequency, was normalised by the maximum trial-level power across frequencies. This was performed for each ROI, condition, and participant. As seen in the Figure, the power is highest in the lowest frequencies, ranging between 2 and 4 Hz, with a small peak occurring between 8 and 12 Hz. This pattern is observed for all three experimental phases and for both CS+ and CS- conditions.

Similar findings of a reduction in oscillatory power with increasing frequencies within the theta band (2-8 Hz) was also reported by Tzovara et al. (2019) in the amygdala and the hippocampus. In addition, they showed that mean theta power was greater for CS- compared to CS+ trials. However, when quantifying differences in our data within Acquisition and Extinction, the two-sided paired Monte-Carlo permutation paired t-tests (2000 permutations, FDR corrected, alpha = 0.025), revealed no significant differences within any of the frequency bands. These analyses were performed on Habituation-baselined data as well as without baselining the data in respect to habituation, with similar findings across the two approaches.

Supplementary Figure 27

Mean oscillatory power between 1 and 20 Hz for each condition and experimental phase



3.5.8 Supplementary Materials 11: Random effects summaries derived from mixed models

Supplementary Table 10

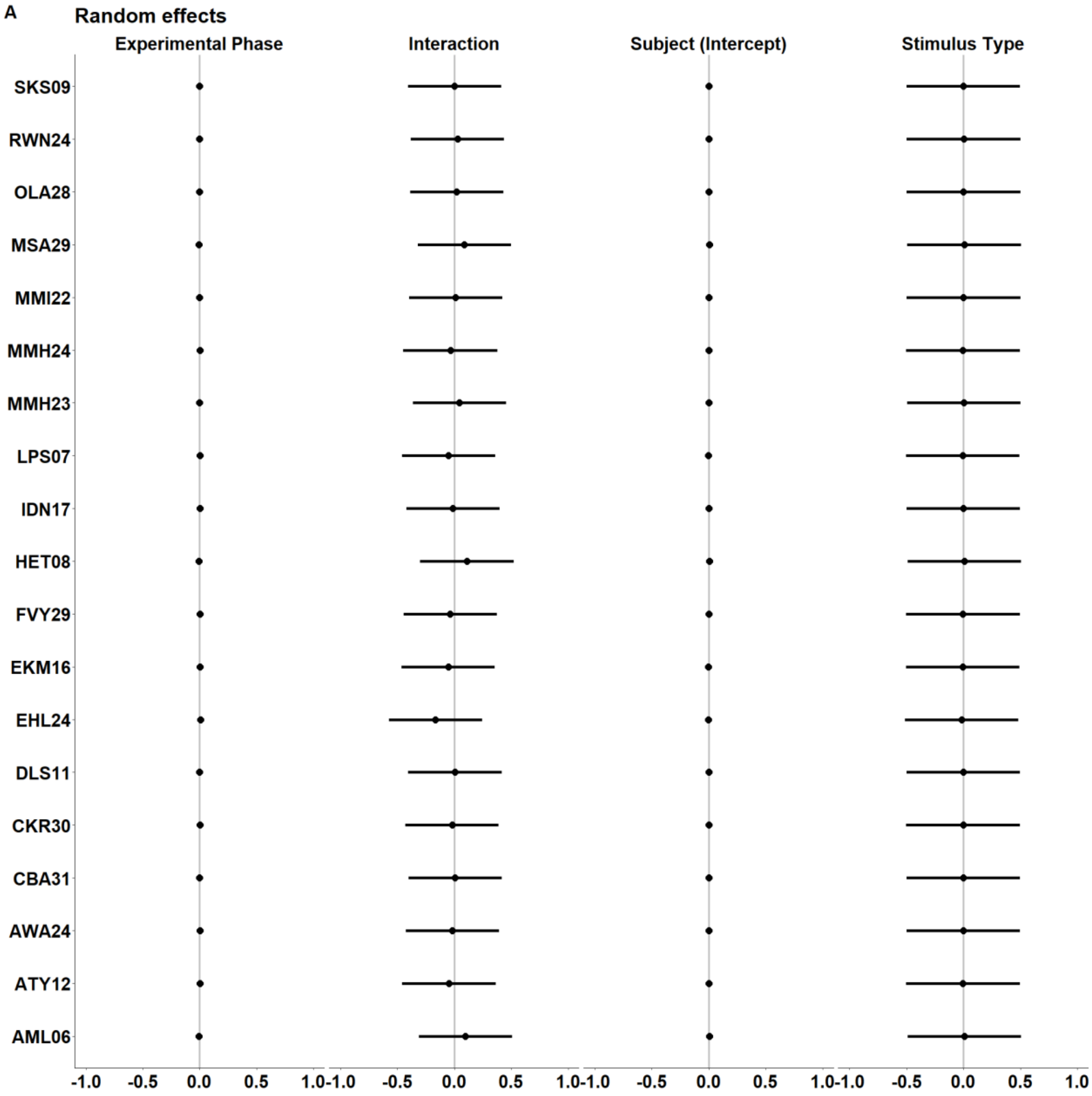
Summary of fixed estimates and random effect variance for the pupil model.

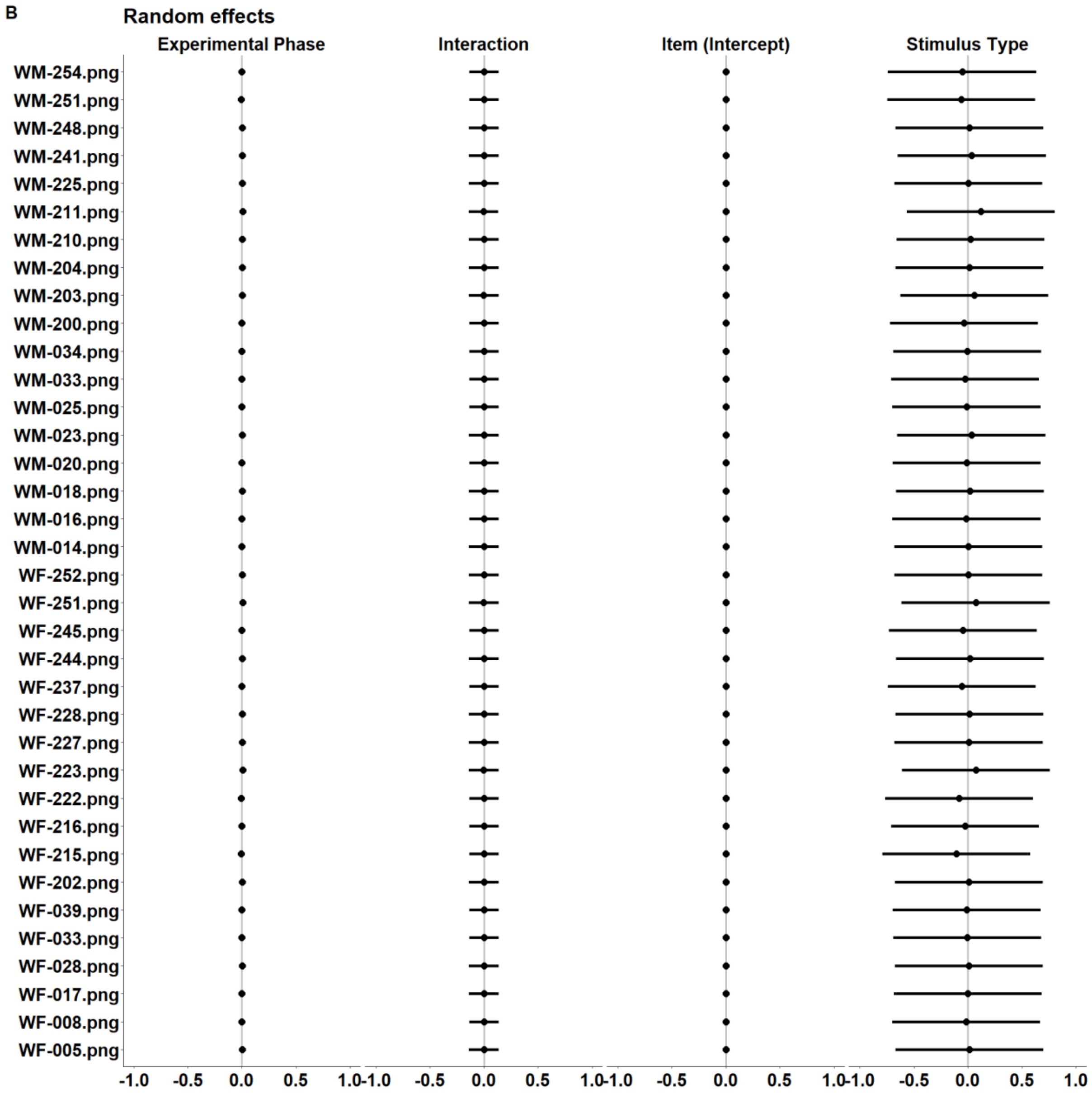
	Mean Pupil		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.00	-0.00 – 0.00	0.761
Experimental Phase	-0.01	-0.01 – -0.00	**0.004**
Stimulus Type	-0.00	-0.01 – 0.00	0.771
Interaction	0.00	-0.00 – 0.01	0.522
Random Effects			
σ^2	0.0082		
τ_{00} Subject	0.0009		
τ_{00} Item	0.0000		
τ_{11} Subject: Phase	0.0000		
τ_{11} Subject: Stimulus Type	0.0000		
τ_{11} Subject: Interaction	0.0064		
τ_{11} Item: Phase	0.0387		
τ_{11} Item: Stimulus Type	0.0000		
τ_{11} Item: Interaction	0.0049		
N Subject	0.1238		
N Item	0.0000		
Observations	0.0649		
Marginal R^2	0.0477		
τ_{00} Subject	19		
τ_{00} Item	36		
τ_{11} Subject: Phase	11808		
τ_{11} Subject: Stimulus Type	0.001 / NA		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 28

By-subject and by-item random coefficients and intercept for the pupil model.




B
Random effects
Experimental Phase
Interaction
Item (Intercept)
Stimulus Type
WM-254.png
WM-251.png
WM-248.png
WM-241.png
WM-225.png
WM-211.png
WM-210.png
WM-204.png
WM-203.png
WM-200.png
WM-034.png
WM-033.png
WM-025.png
WM-023.png
WM-020.png
WM-018.png
WM-016.png
WM-014.png
WF-252.png
WF-251.png
WF-245.png
WF-244.png
WF-237.png
WF-228.png
WF-227.png
WF-223.png
WF-222.png
WF-216.png
WF-215.png
WF-202.png
WF-039.png
WF-033.png
WF-028.png
WF-017.png
WF-008.png
WF-005.png
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0
-1.0
-0.5
0.0
0.5
1.0

Supplementary Table 11

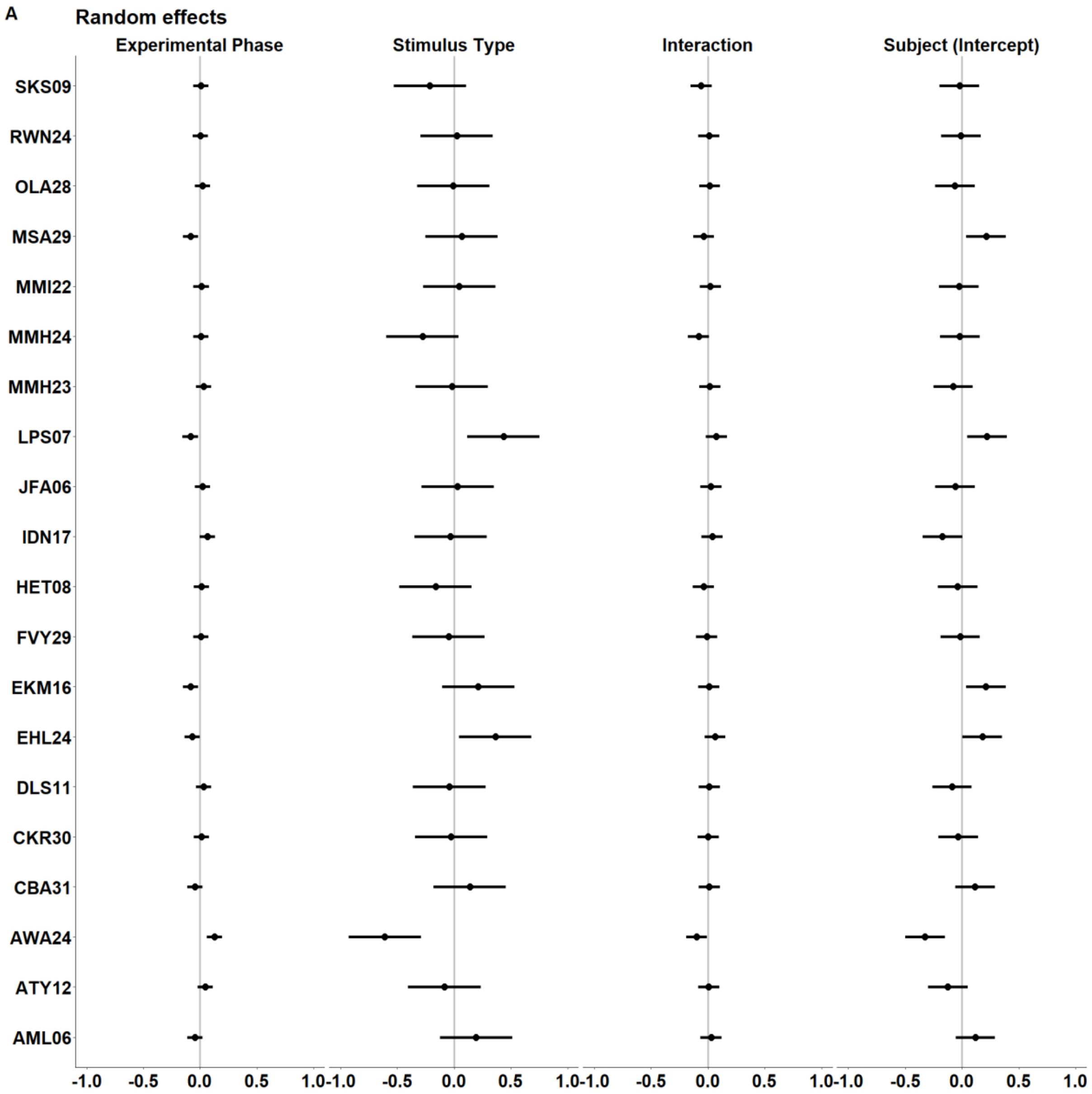
Summary of fixed estimates and random effect variance for the valence model.

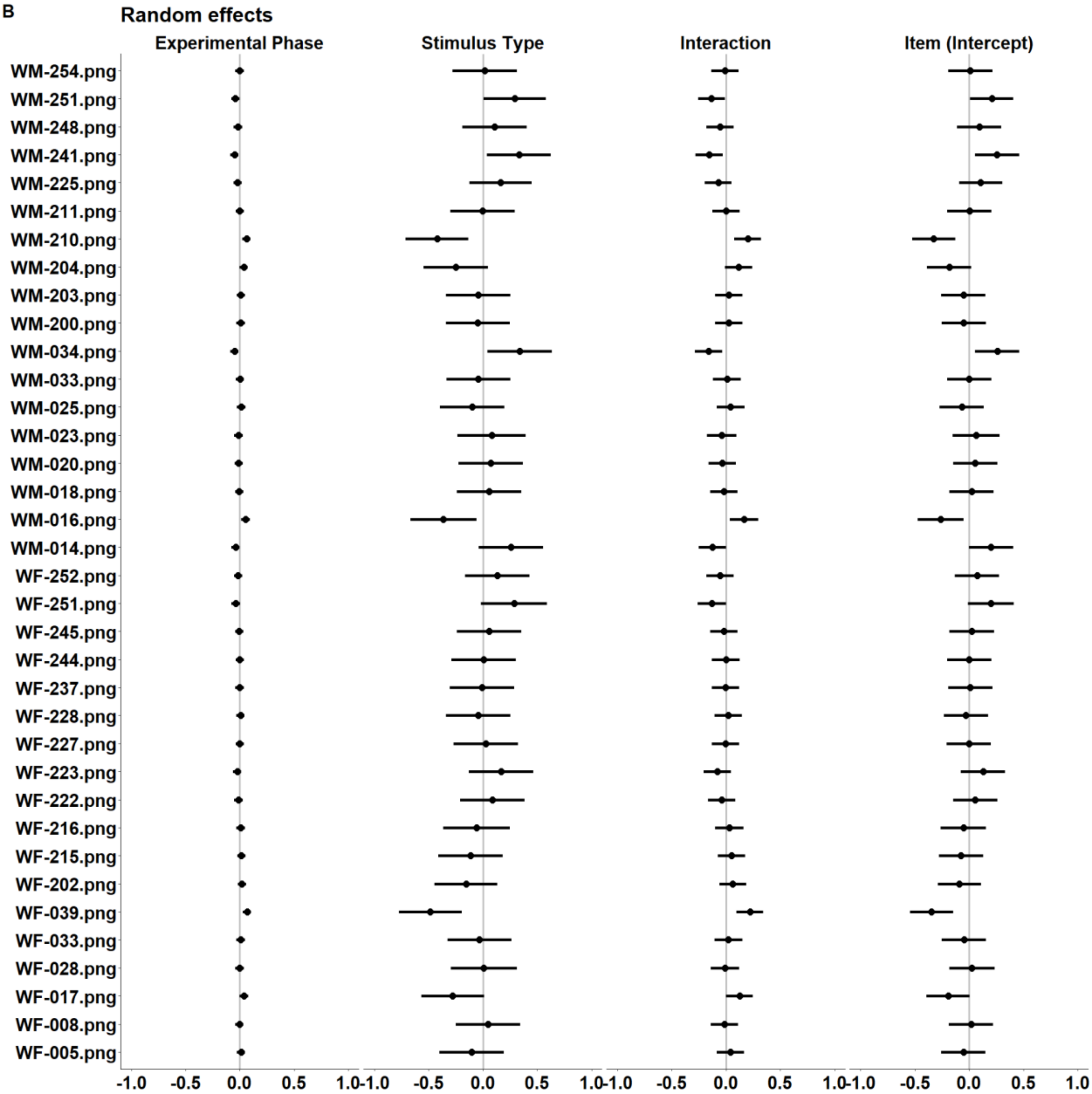
	Mean Valence		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.00	-0.11 – 0.10	0.950
Experimental Phase	-0.03	-0.13 – 0.07	0.534
Stimulus Type	-0.06	-0.23 – 0.12	0.528
Interaction	-0.08	-0.28 – 0.12	0.459
Random Effects			
σ^2	0.8954		
τ_{00} Subject	0.0309		
τ_{00} Item	0.0273		
τ_{11} Subject: Phase	0.0012		
τ_{11} Subject: Stimulus Type	0.0603		
τ_{11} Subject: Interaction	0.0120		
τ_{11} Item: Phase	0.0041		
τ_{11} Item: Stimulus Type	0.0767		
τ_{11} Item: Interaction	0.0042		
N Subject	20		
N Item	36		
Observations	1440		
Marginal R^2	0.002		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 29

By-subject and by-item random coefficients and intercept for the valence model.




B
Random effects
Experimental Phase
Stimulus Type
Interaction
Item (Intercept)
WM-254.png
WM-251.png
WM-248.png
WM-241.png
WM-225.png
WM-211.png
WM-210.png
WM-204.png
WM-203.png
WM-200.png
WM-034.png
WM-033.png
WM-025.png
WM-023.png
WM-020.png
WM-018.png
WM-016.png
WM-014.png
WF-252.png
WF-251.png
WF-245.png
WF-244.png
WF-237.png
WF-228.png
WF-227.png
WF-223.png
WF-222.png
WF-216.png
WF-215.png
WF-202.png
WF-039.png
WF-033.png
WF-028.png
WF-017.png
WF-008.png
WF-005.png
-1.0
-0.5
0.0
0.5
1.0

Supplementary Table 12

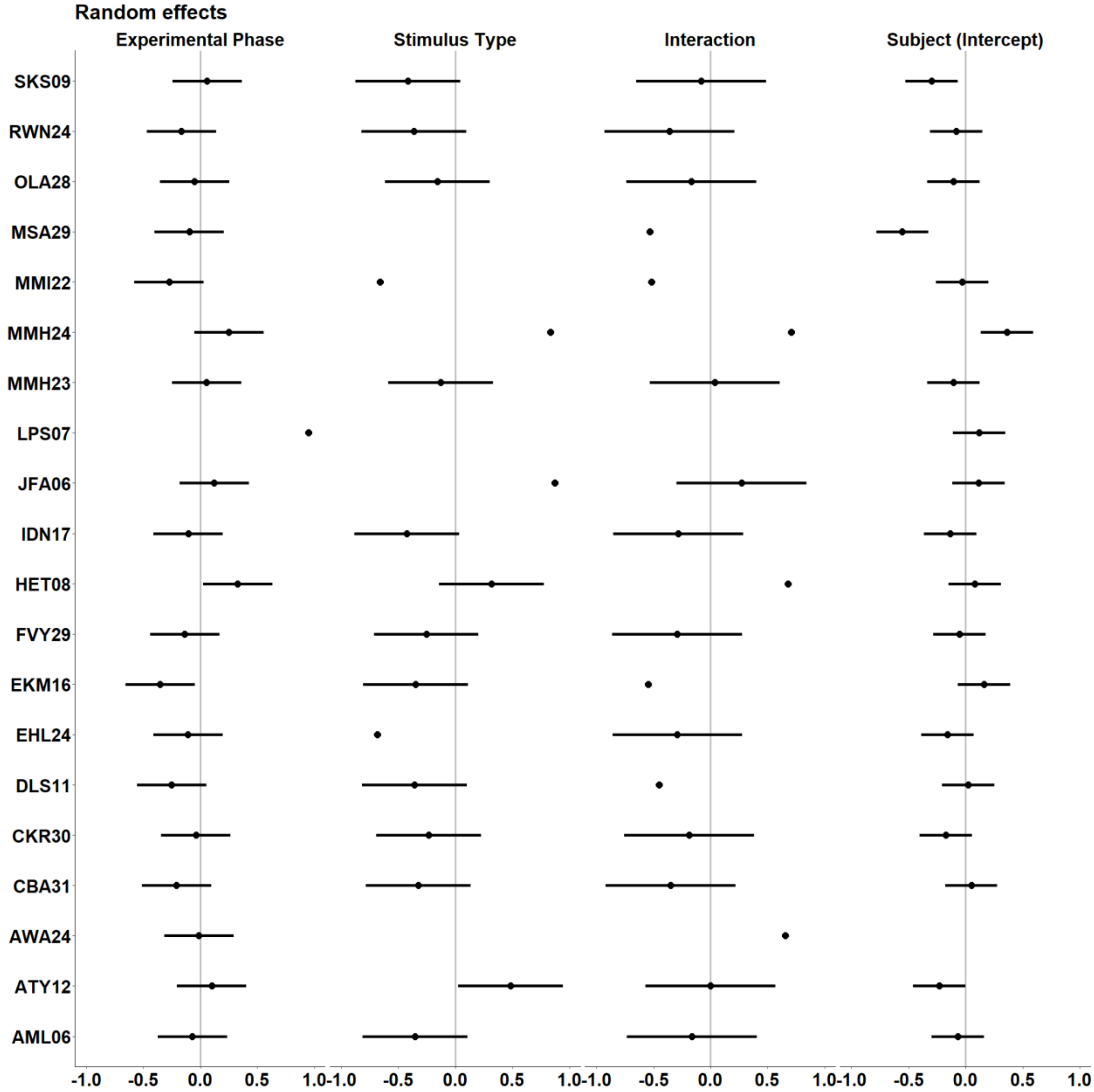
Summary of fixed estimates and random effect variance for the arousal model.

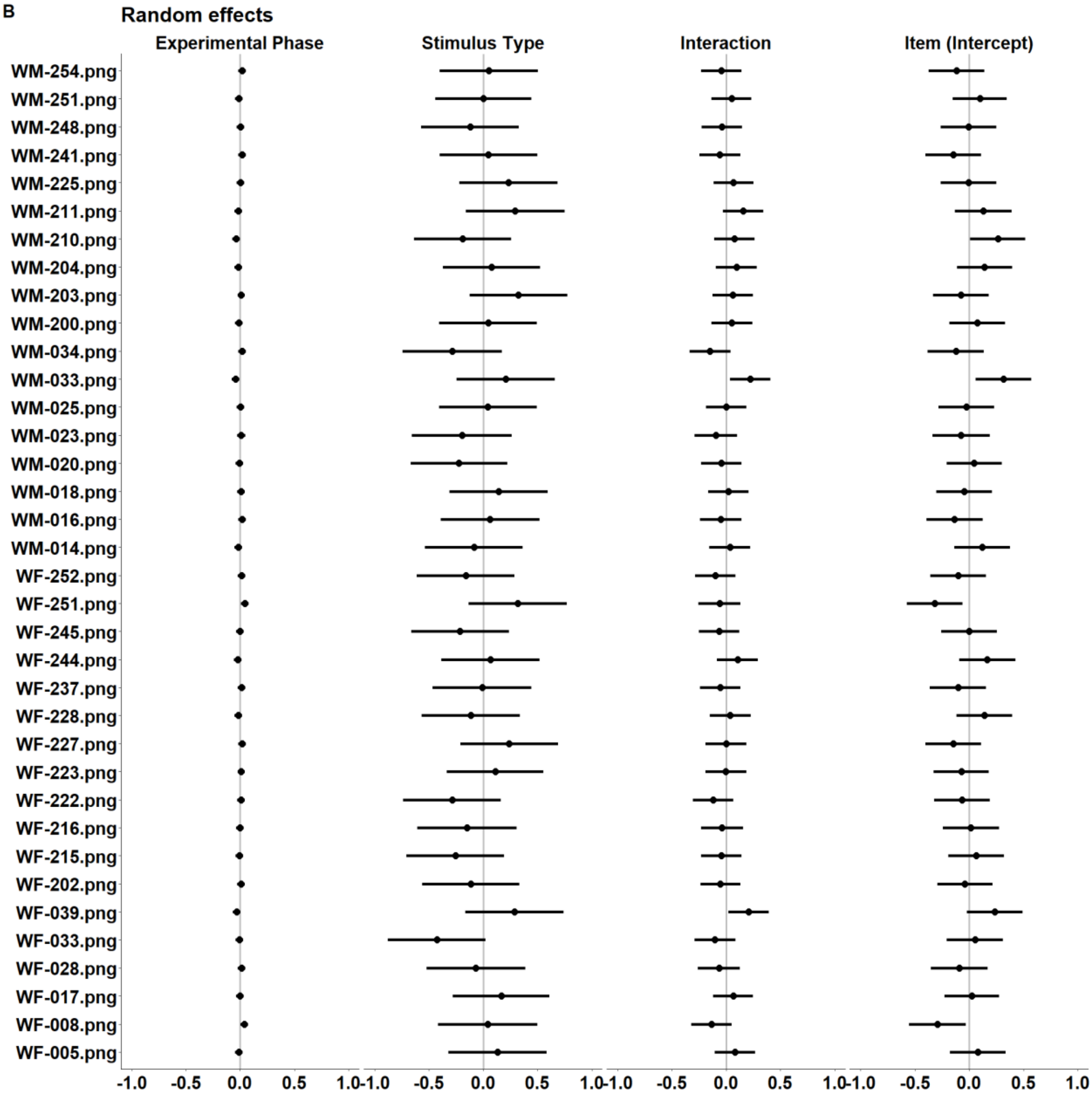
	Mean Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.01	-0.18 – 0.16	0.871
Experimental Phase	0.28	0.09 – 0.46	**0.003**
Stimulus Type	0.49	0.11 – 0.86	**0.011**
Interaction	0.30	-0.07 – 0.67	0.111
Random Effects			
σ^2	1.2516		
τ_{00} Subject	0.0366		
τ_{00} Item	0.1141		
τ_{11} Subject: Phase	0.0007		
τ_{11} Subject: Stimulus Type	0.0898		
τ_{11} Subject: Interaction	0.0172		
τ_{11} Item: Phase	0.1027		
τ_{11} Item: Stimulus Type	0.6107		
τ_{11} Item: Interaction	0.4302		
N Subject	20		
N Item	36		
Observations	1440		
Marginal R^2	0.063 / NA		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 30

By-subject and by-item random coefficients and intercept for the arousal model.




B
Random effects
Experimental Phase
Stimulus Type
Interaction
Item (Intercept)
WM-254.png
WM-251.png
WM-248.png
WM-241.png
WM-225.png
WM-211.png
WM-210.png
WM-204.png
WM-203.png
WM-200.png
WM-034.png
WM-033.png
WM-025.png
WM-023.png
WM-020.png
WM-018.png
WM-016.png
WM-014.png
WF-252.png
WF-251.png
WF-245.png
WF-244.png
WF-237.png
WF-228.png
WF-227.png
WF-223.png
WF-222.png
WF-216.png
WF-215.png
WF-202.png
WF-039.png
WF-033.png
WF-028.png
WF-017.png
WF-008.png
WF-005.png
-1.0
-0.5
0.0
0.5
1.0

3.5.9 Supplementary Materials 12: CLM modelling of the three-phased ordinal data

To account for the ordinal nature of the ratings we re-analysed the data using cumulative link mixed (CLM) models (package *ordinal*) on the three-phased datasets. Each model included a 3 *Experimental Phase* (Habituation, Acquisition vs Extinction) by 2 *Stimulus Type* (CS+ vs CS-) fixed effects design with mean-centred contrasts for the two categorical fixed effects. For the *Experimental Phase* fixed effect, the Habituation phase was used as the baseline level. *Subjects* and *Items* were added as random intercepts with a by-subject and by-item random slopes for the main effects and the interaction. Main effects and interactions were assessed using type II Likelihood-ratio test (package *RVAideMemoire*).

For valence, the only fixed effect significant at the level of 0.05 was the main effect of Stimulus Type (see Supplementary Table 13 and Figure 31). Post-hoc contrasts computed using asymptotic degrees of freedom approximation (package *emmeans*) showed that CS+ faces were rated as less pleasant than CS- faces ($z(\text{inf}) = -5.1$, $p < 0.001$), (see Supplementary Table 14). These findings are consistent with the descriptive analyses and confirm the likelihood of baseline differences between CS+ and CS- faces during Habituation. These findings also support the use of LME modelling on habituation-baselined data as means for partly accounting for these baseline differences. The results from the CLM arousal model were consistent with those from the LME model (see Supplementary Table 14 and Figure 32) and showed that significant main effects of *Experimental Phase* and *Stimulus Type* at the level of 0.05. Post-hoc contrasts revealed that CS+ faces were rated as more arousing than CS- faces and that faces were rated as more arousing during Acquisition compared to Habituation and Extinction and less arousing during Extinction compared to Habituation (see Supplementary Table 15 and Table 16). As seen in Supplementary Figure 32 showing the predicted probabilities for each rating category, these effects are driven by the higher probabilities for lower arousal ratings for CS- faces.

Supplementary Table 13

Type II Likelihood-ratio tests and R-squared values for the valence and arousal models and each of the fixed effects.

	LR Chisq	Df	P-value	McFadenn Pseudo R^2	Negelkerke Pseudo R^2
Valence					
Full Model				0.004	0.01
Experimental Phase	0.333	2.000	0.847		
Stimulus Type	26.677	1.000	0.000		
Experimental Phase X Stimulus Type	1.231	2.000	0.540		
Arousal					
Full Model					
Experimental Phase	12.383	2.000	0.002	0.009	0.03
Stimulus Type	46.667	1.000	0.000		
Experimental Phase X Stimulus Type	5.497	2.000	0.064		

Supplementary Table 14

Estimated marginal means and related contrasts derived for the main effect of Stimulus Type in the valence model.

Estimates

Stimulus Type	Cut	Estimate	SE	df	Lower CI	Upper CI
CS Unpaired	1|2	-4.469	0.265	Inf	-5.061	-3.877
CS Paired	1|2	-4.068	0.263	Inf	-4.656	-3.480
CS Unpaired	2|3	-2.304	0.230	Inf	-2.818	-1.790
CS Paired	2|3	-1.903	0.228	Inf	-2.413	-1.392
CS Unpaired	3|4	-0.839	0.225	Inf	-1.342	-0.336
CS Paired	3|4	-0.438	0.224	Inf	-0.939	0.063
CS Unpaired	4|5	0.341	0.224	Inf	-0.161	0.842
CS Paired	4|5	0.742	0.225	Inf	0.239	1.244
CS Unpaired	5|6	1.821	0.228	Inf	1.311	2.331
CS Paired	5|6	2.222	0.229	Inf	1.709	2.735
CS Unpaired	6|7	3.253	0.242	Inf	2.712	3.793
CS Paired	6|7	3.653	0.244	Inf	3.108	4.199
CS Unpaired	7|8	4.521	0.280	Inf	3.894	5.147
CS Paired	7|8	4.921	0.282	Inf	4.290	5.553

Contrasts

Contrast	Estimate	SE	df	Z ratio	P value

Estimates

Stimulus Type	Cut	Estimate	SE	df	Lower CI	Upper CI
CS Paired – CS Unpaired		-0.401	0.078	Inf	-5.167	<0.001

Supplementary Table 15

Estimated marginal means and related contrasts derived for main effect of Experimental Phase in the arousal model.

Estimated Marginal Means

Experimental Phase	Cut	Estimate	SE	df	Lower CI	Upper CI
Habituation	1|2	-2.951	0.427	Inf	-3.971	-1.930
Acquisition	1|2	-3.128	0.428	Inf	-4.149	-2.107
Extinction	1|2	-2.790	0.427	Inf	-3.809	-1.771
Habituation	2|3	-0.858	0.420	Inf	-1.862	0.146
Acquisition	2|3	-1.035	0.420	Inf	-2.039	-0.031
Extinction	2|3	-0.697	0.420	Inf	-1.700	0.306
Habituation	3|4	0.250	0.420	Inf	-0.752	1.253
Acquisition	3|4	0.073	0.420	Inf	-0.929	1.075
Extinction	3|4	0.411	0.420	Inf	-0.591	1.413
Habituation	4|5	1.263	0.420	Inf	0.259	2.266
Acquisition	4|5	1.085	0.420	Inf	0.083	2.088
Extinction	4|5	1.423	0.420	Inf	0.419	2.427
Habituation	5|6	2.112	0.422	Inf	1.105	3.119
Acquisition	5|6	1.934	0.421	Inf	0.929	2.940
Extinction	5|6	2.272	0.422	Inf	1.265	3.280
Habituation	6|7	3.191	0.425	Inf	2.175	4.206
Acquisition	6|7	3.013	0.425	Inf	1.999	4.027
Extinction	6|7	3.351	0.426	Inf	2.334	4.368
Habituation	7|8	4.851	0.444	Inf	3.792	5.911
Acquisition	7|8	4.674	0.443	Inf	3.616	5.732
Extinction	7|8	5.012	0.444	Inf	3.951	6.073

Contrasts

Contrast	Estimate	SE	df	Z ratio	P value
Acquisition - Habituation	0.178	0.097	Inf	1.839	0.157
Extinction - Habituation	-0.160	0.097	Inf	-1.661	0.221
Extinction - Acquisition	-0.338	0.096	Inf	-3.508	0.001

Supplementary Table 16

Estimated marginal means and related contrasts derived for main effect of Stimulus Type in the arousal model.

Estimated Marginal Means

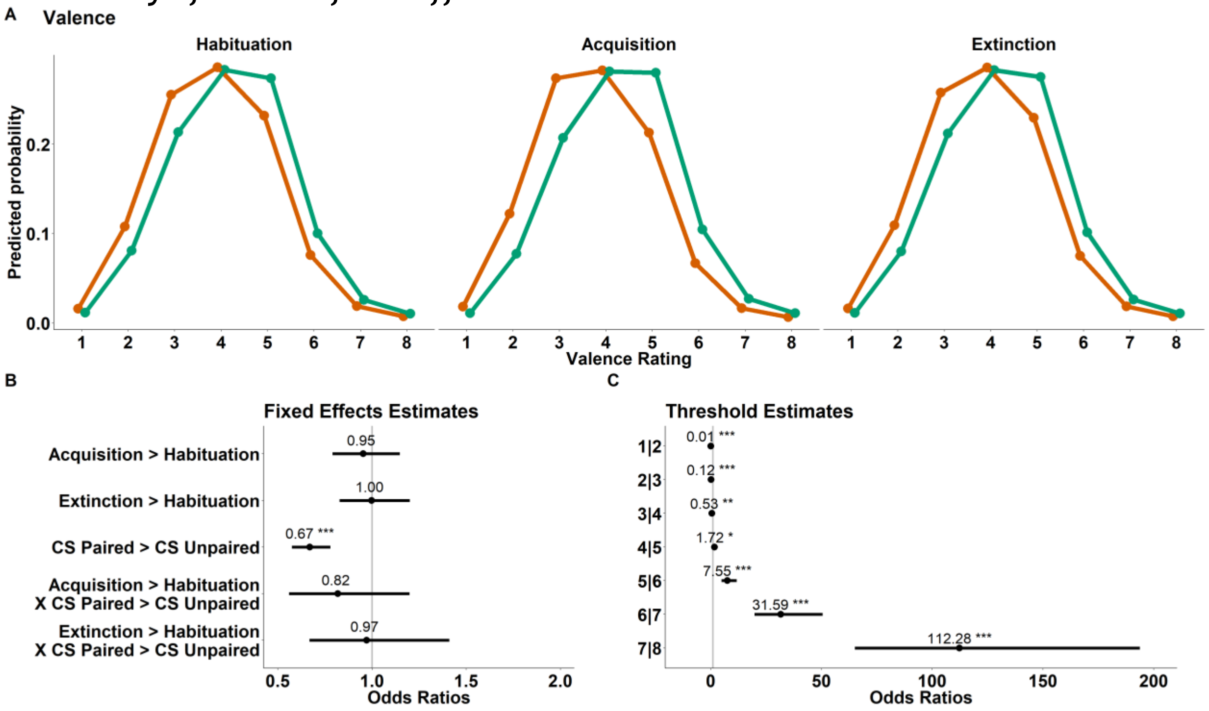
Stimulus Type	Cut	Estimate	SE	df	Lower CI	Upper CI
CS Unpaired	1|2	-2.686	0.425	Inf	-3.636	-1.736
CS Paired	1|2	-3.227	0.426	Inf	-4.180	-2.274
CS Unpaired	2|3	-0.593	0.418	Inf	-1.528	0.343
CS Paired	2|3	-1.134	0.419	Inf	-2.071	-0.197
CS Unpaired	3|4	0.515	0.418	Inf	-0.420	1.450
CS Paired	3|4	-0.026	0.418	Inf	-0.960	0.908
CS Unpaired	4|5	1.528	0.419	Inf	0.591	2.464
CS Paired	4|5	0.986	0.418	Inf	0.051	1.922
CS Unpaired	5|6	2.377	0.420	Inf	1.437	3.317
CS Paired	5|6	1.836	0.419	Inf	0.898	2.773
CS Unpaired	6|7	3.456	0.424	Inf	2.507	4.404
CS Paired	6|7	2.914	0.423	Inf	1.969	3.860
CS Unpaired	7|8	5.116	0.443	Inf	4.126	6.107
CS Paired	7|8	4.575	0.441	Inf	3.589	5.561

Contrasts

Contrast	Estimate	SE	df	Z ratio	P value
CS Paired – CS Unpaired	0.541	0.080	Inf	6.801	<0.0001

Supplementary Figure 31

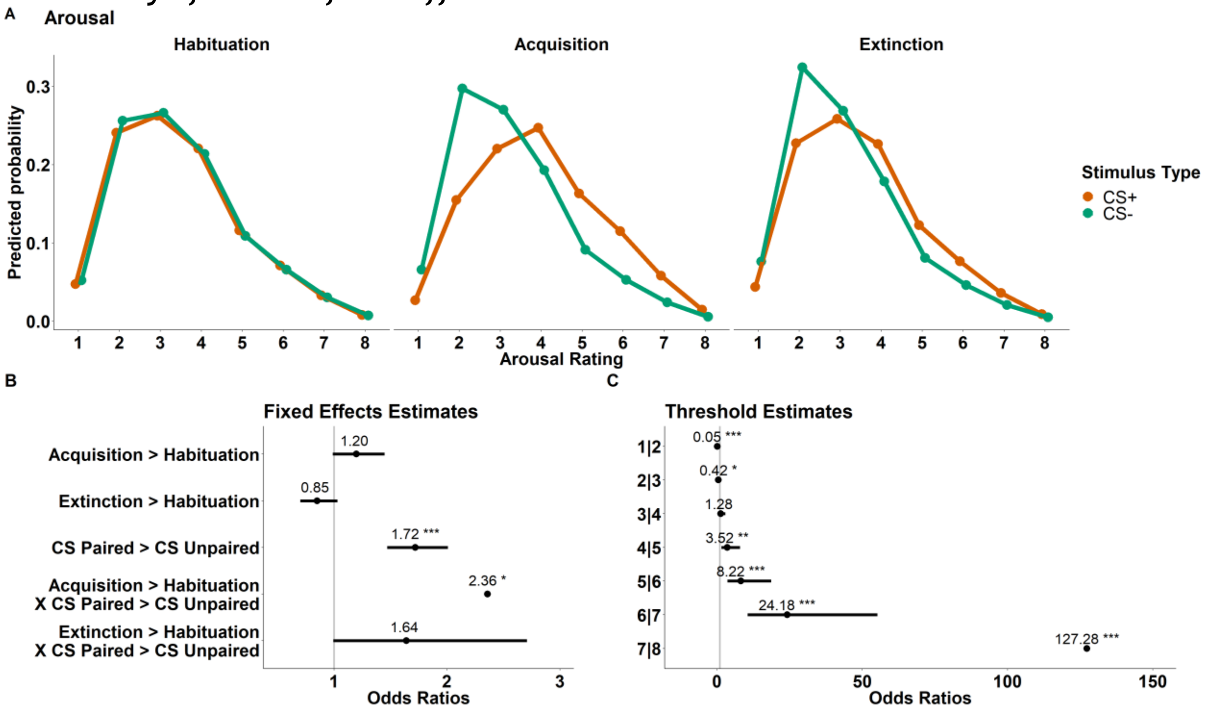
A summary of valence fixed effects.



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of valence ratings B) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates. C) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of valence ratings; bars represent 95% CIs for the estimates.

Supplementary Figure 32

A summary of arousal fixed effects.



Note. A) Predicted probability of each rating point per condition derived from the cumulative-link mixed effects model of arousal ratings B) Fixed effect estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates. C) Threshold estimates (labelled dots) derived from the cumulative-link mixed effects model of arousal ratings; bars represent 95% CIs for the estimates.

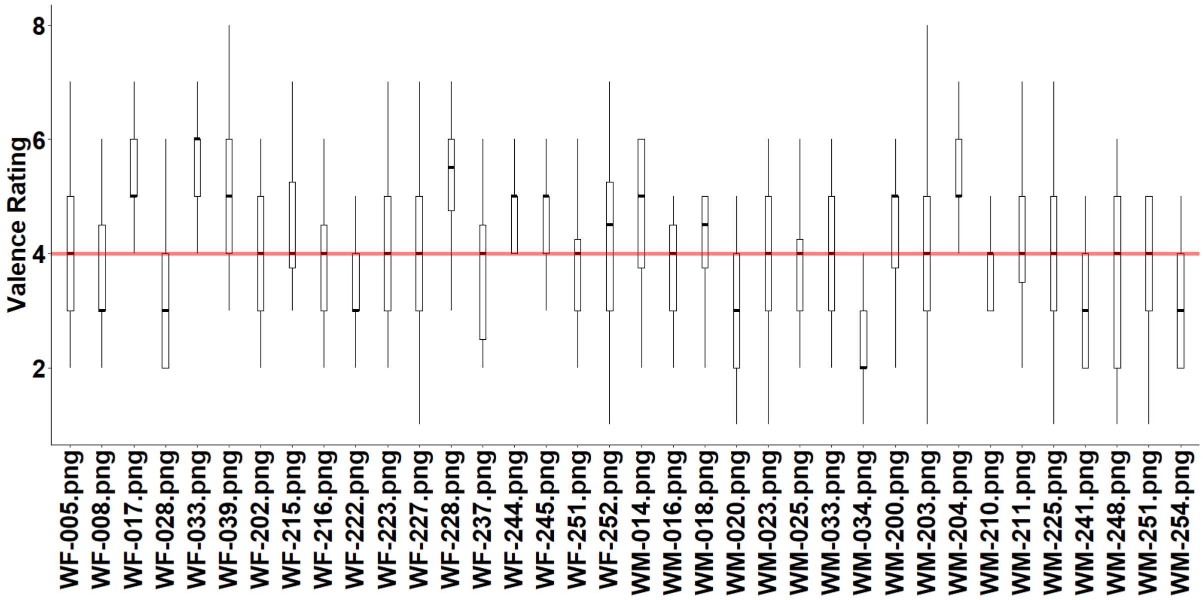
3.5.10 Supplementary Materials 13: Exploring baseline differences in valence ratings

The following exploratory analyses examine the potential factors driving the baseline valence differences between CS+ and CS- trials. We considered several potential factors, including variability of valence ratings per item, the counterbalancing procedure, and the normative ratings provided by the face database the stimuli were selected from.

Supplementary Figure 33 shows the median valence ratings per item. As seen in the figure, there are several items that had a median valence rating above the grand median (horizontal red line), with a few items below the grand median. This suggests that these stimuli were not perceived as neutral.

Supplementary Figure 33

Valence ratings per item.



Note. The horizontal red line indicates the grand median valence rating.

To investigate this further, we examined the mediating role of a number of ratings obtained from the normative database, including attractiveness, dominance, stimulus sex, and emotionality of the stimuli (anger, sadness, happiness, fear). These were included as interacting factors with the fixed effect of *Stimulus Type* in a linear mixed effects model. The outcome variable in this model was valence ratings during Habituation. The model also included 1) a by-subject random intercept, together with a random slope for the main effect of Stimulus Type, and 2) a by-item random intercept, with a random slope for the main effect of stimulus type and the main effects of the normative ratings. The model revealed no significant interactions. However, there were significant main effects of attractiveness, anger, and stimulus sex (see Supplementary Table 17). As seen in Supplementary figures 34-36, there was a positive correlation between valence and attractiveness, a negative correlation between valence and anger, and female faces were rated as slightly more pleasant than male faces. While these factors were not found to mediate the condition baseline differences, they can explain why some stimuli were rated as more or less pleasant as indicated in Supplementary Figure 33. These findings confirm that the faces were not perceived as completely neutral, even though they were selected so that they do not vary substantially.

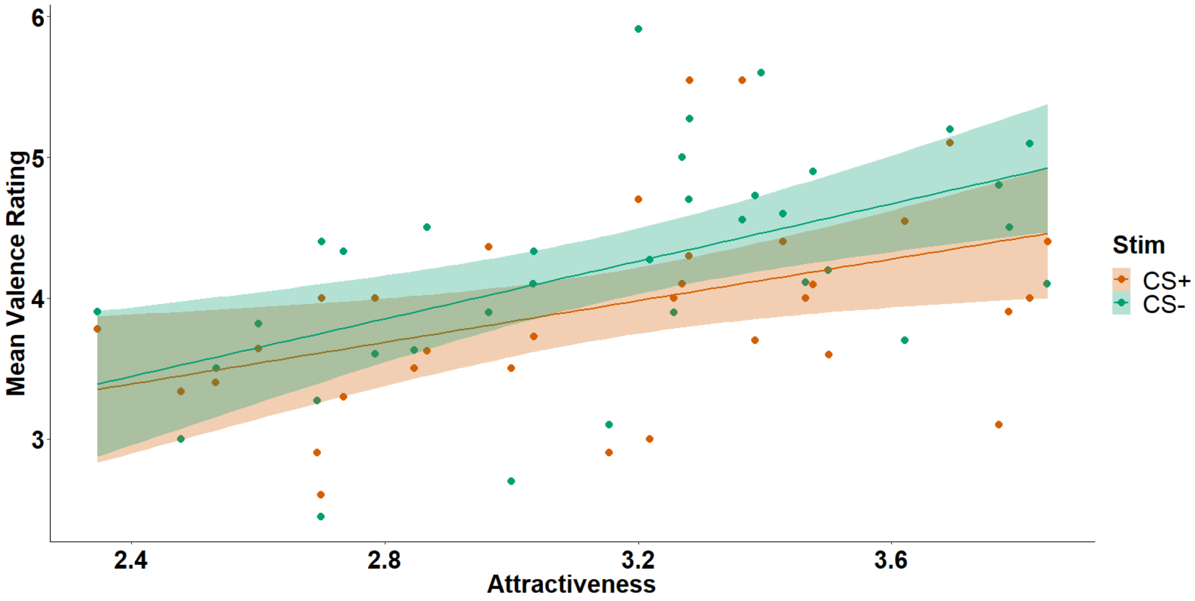
Supplementary Table 17

Type III Wald Chi-square tests for the main effects and interactions in the model examining the mediating role of normative ratings.

	Chisq	Df	P-value
Stimulus Type	0.168	1.000	0.682
Attractive	5.970	1.000	0.015
Dominant	3.680	1.000	0.055
Trustworthy	1.502	1.000	0.220
Angry	6.052	1.000	0.014
Sad	0.026	1.000	0.873
Happy	0.252	1.000	0.615
Afraid	0.811	1.000	0.368
Stimulus Sex	5.062	1.000	0.024
Stimulus Type X Attractive	0.031	1.000	0.860
Stimulus Type X Dominant	1.742	1.000	0.187
Stimulus Type X Trustworthy	0.041	1.000	0.839
Stimulus Type X Angry	2.723	1.000	0.099
Stimulus Type X Sad	0.486	1.000	0.486
Stimulus Type X Happy	0.254	1.000	0.615
Stimulus Type X Afraid	0.595	1.000	0.440
Stimulus Type X Stimulus Sex	1.022	1.000	0.312

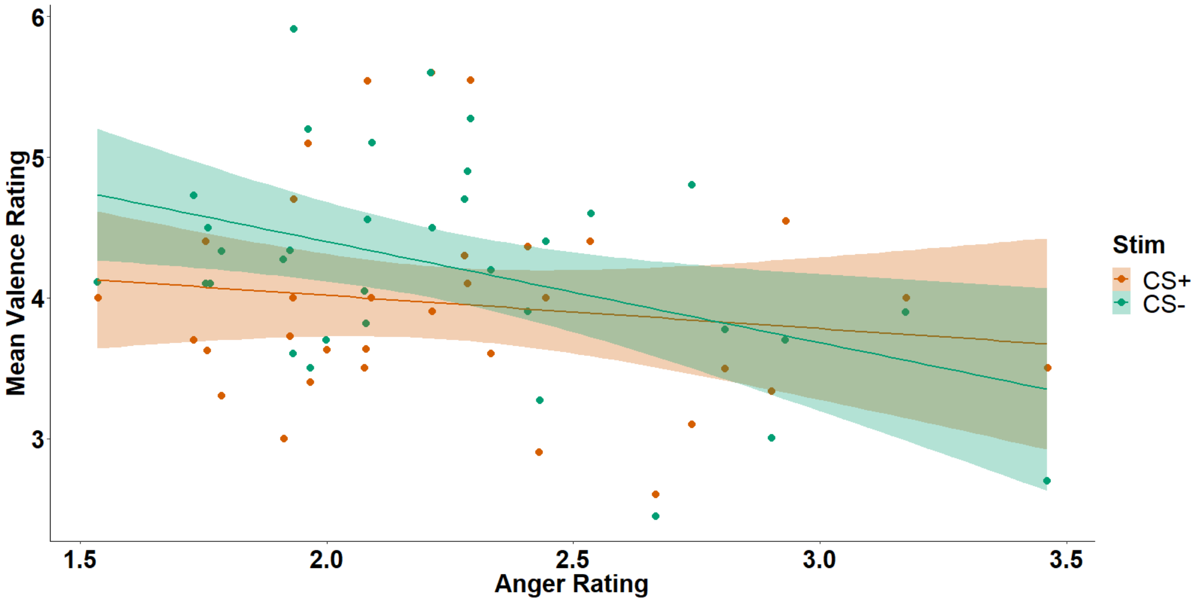
Supplementary Figure 34

Scatterplot of the relationship between valence and attractiveness ratings per stimulus type.



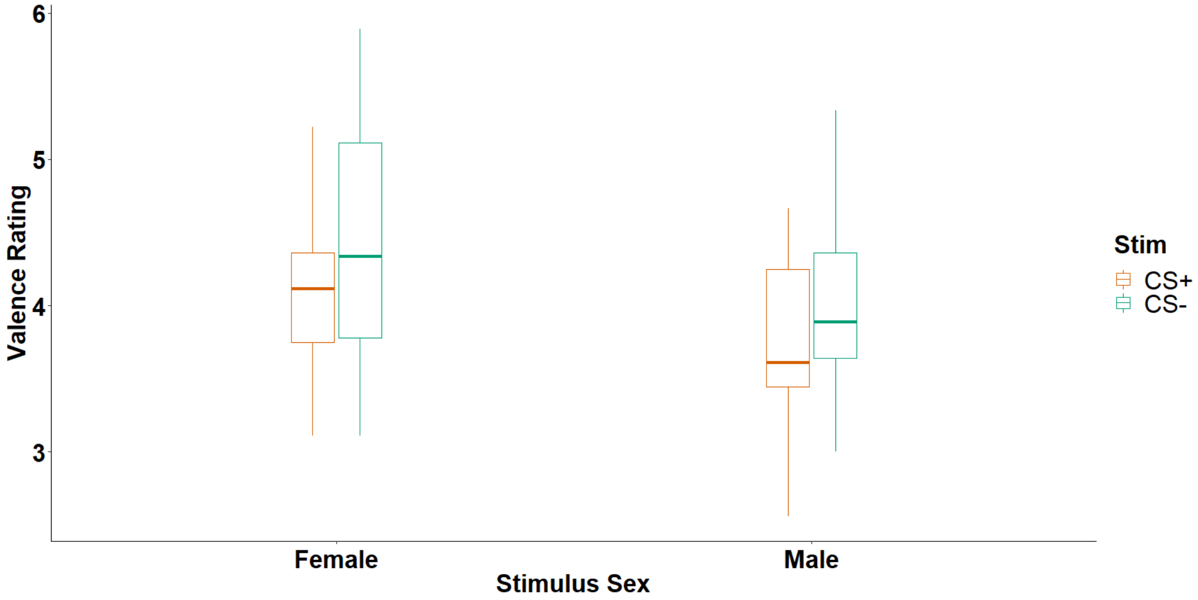
Supplementary Figure 35

Scatterplot of the relationship between valence and anger ratings per stimulus type.



Supplementary Figure 36

Valence ratings per stimulus sex and stimulus type.



Finally, we examined the potential impact of the counterbalancing procedure. The counterbalancing involved creating 10 stimulus sets within which stimuli for each block were randomly selected and randomly assigned to either the CS+ or the CS- condition. The assignment of stimuli to each condition was then counterbalanced across participants. Therefore, two participants completed each of the 10 stimulus sets. However, it is worth noting, that while the assignment of stimuli per block and condition was random, all participants were exposed to all stimuli at some point in the experiment. Yet, since the assignment of stimuli to each set was completely random, we examined whether the stimulus set that participants were assigned mediated the baseline valence differences. This was done using a linear mixed effects model with *Stimulus Type* and *Counterbalancing Group* as the fixed effects. We also included a by-item and by-subject random intercept together with a random slope for the main effect of *Stimulus Type*. As seen in Supplementary Table 18 and Figure 37, there were no significant differences, suggesting that the randomisation procedure did not mediate the baseline differences.

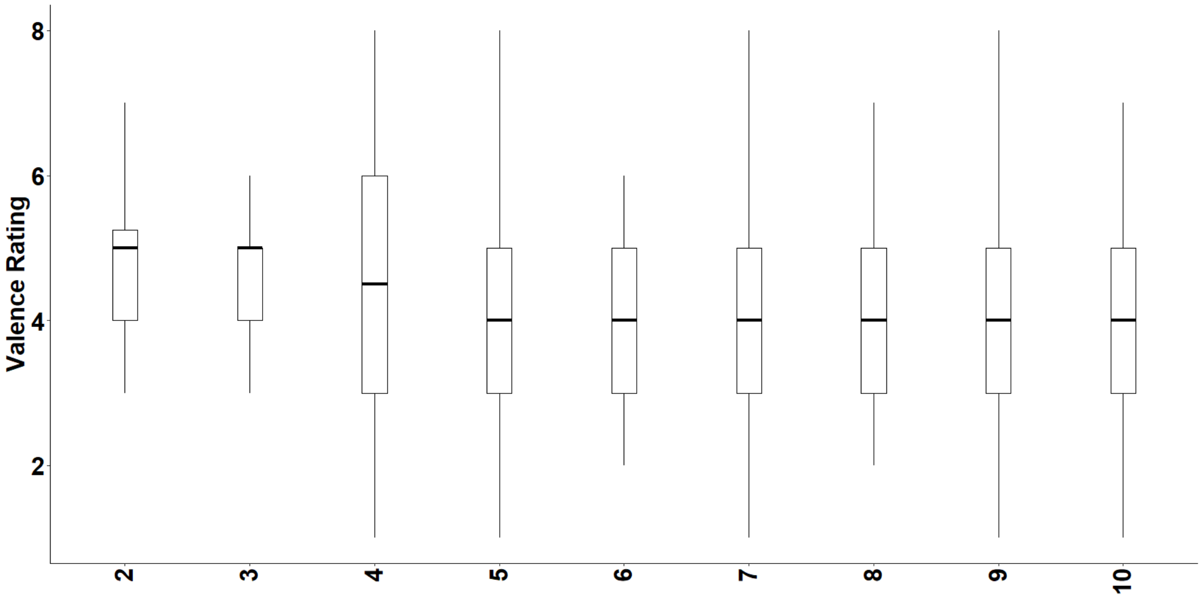
Supplementary Table 18

Type III Wald Chi-square tests for the main effects and interactions for the model examining the mediating role of the randomisation procedure.

	Chisq	Df	P-value
Stimulus Type	0.200	1.000	0.655
Counterbalancing Group	13.284	8.000	0.102
Stimulus Type X Counterbalancing Group	7.029	8.000	0.534

Supplementary Figure 37

Valence ratings per counterbalancing group.



3.5.11 Supplementary Materials 14: Development and habituation of pupillary and behavioural conditioning and extinction effects over blocks

The blocked conditioning paradigm is a 9-time multiplication of a standard fear conditioning set up, where each block represents a stand-alone experiment with a number of stimuli mirroring that of most available studies. As such, despite using different CSs in each block, it is possible that learning effects over time may occur. Supplementary Figure 38 explores the development and extinction of conditioned pupillary responses over the nine experimental blocks.

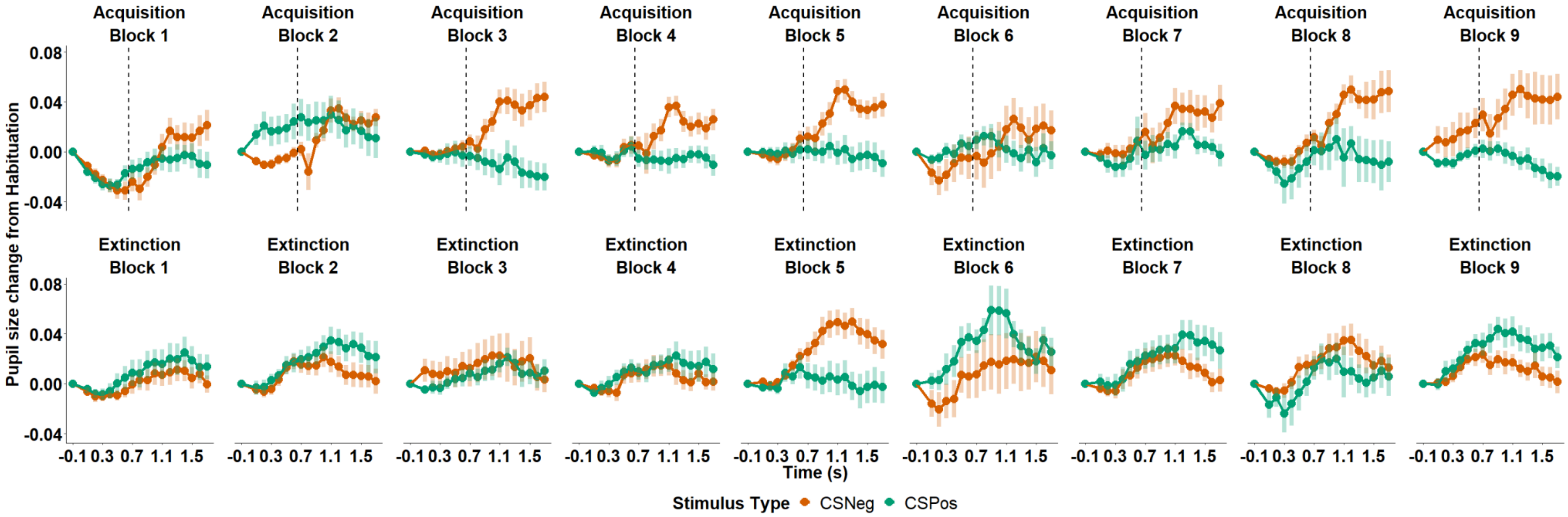
When examining the CR over blocks (pupil size before UCS onset), small changes in pupil size between conditions is observed, whereas for certain blocks pupil size is slightly larger for CS+ trials, while for other the opposite pattern (e.g., Block 2 and 6) is observed. However, overall, condition differences remain minimal and the magnitude of these differences remains relatively stable across blocks. A similar pattern of results is seen during extinction. When visualising the average pupil size across the CR, no differences are seen between blocks with a significant overlap in the distributions between conditions (see Supplementary Figure 39).

When examining the UCR over blocks (pupil size after UCS onset), the magnitude of the difference between CS+ and CS- during acquisition is also relatively stable suggesting low level of habituation to the UCS. During extinction, both increases and decreases in response to the CS+ can be seen, however, these differences are minimal in most blocks apart from blocks 2,5,7, and 9. Furthermore, these differences are negligible in magnitude when examining the mean pupil size over the entire time window of the UCR (see Supplementary Figure 40). If any learning or habituation effects were observed in either the conditioned or unconditioned responding, we would have expected to see a gradual decrease or increase in the magnitude of differences between conditions over time (blocks).

Finally, no evidence for block-related learning effects can be seen when looking at the subjective valence and arousal ratings across conditions and blocks (see Supplementary Figure 41 and 42).

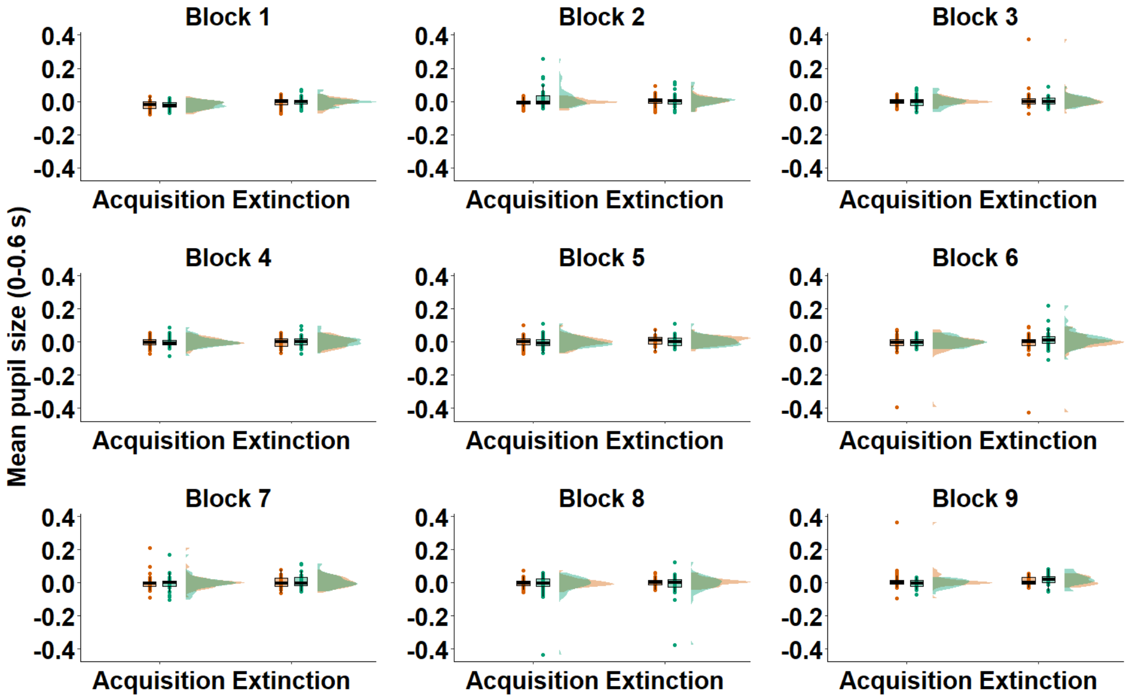
Supplementary Figure 38

Proportional mean pupil size over time and across blocks.



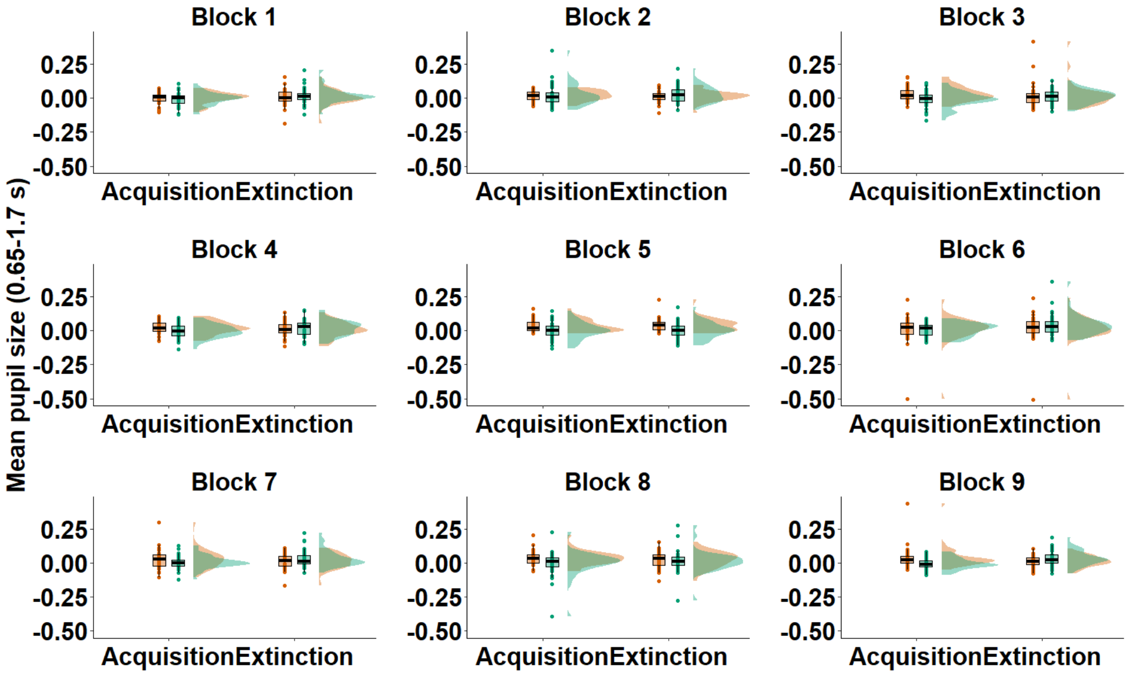
Supplementary Figure 39

Mean pupil size over the CR (0 – 0.6 s) over blocks.



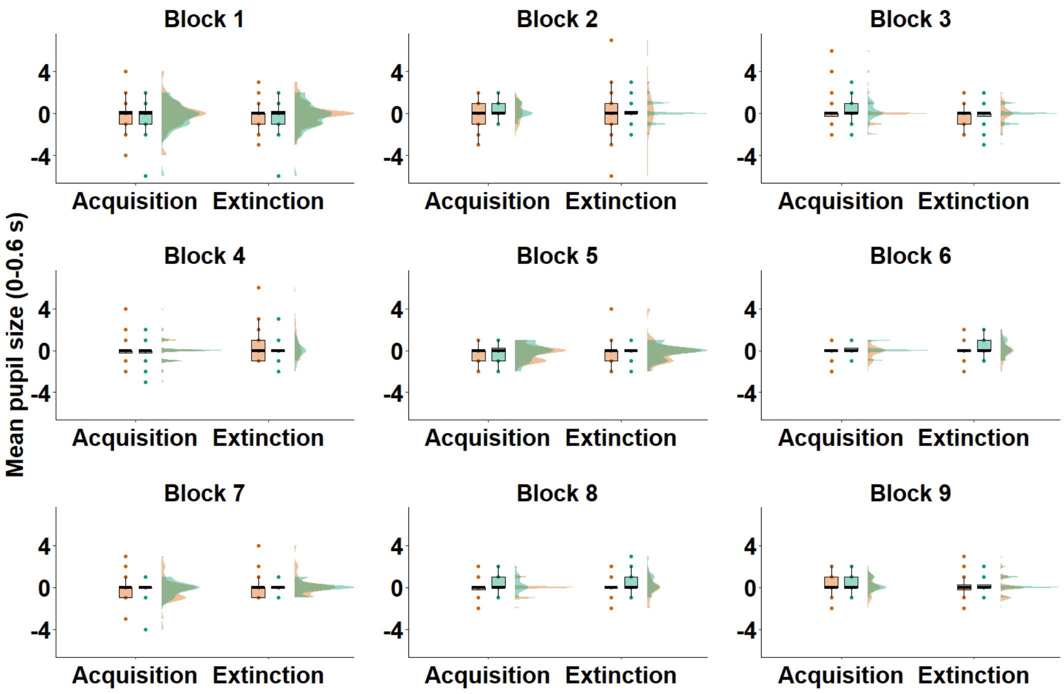
Supplementary Figure 40

Mean pupil size over the UCR (0.65 – 1.7 s) over blocks.



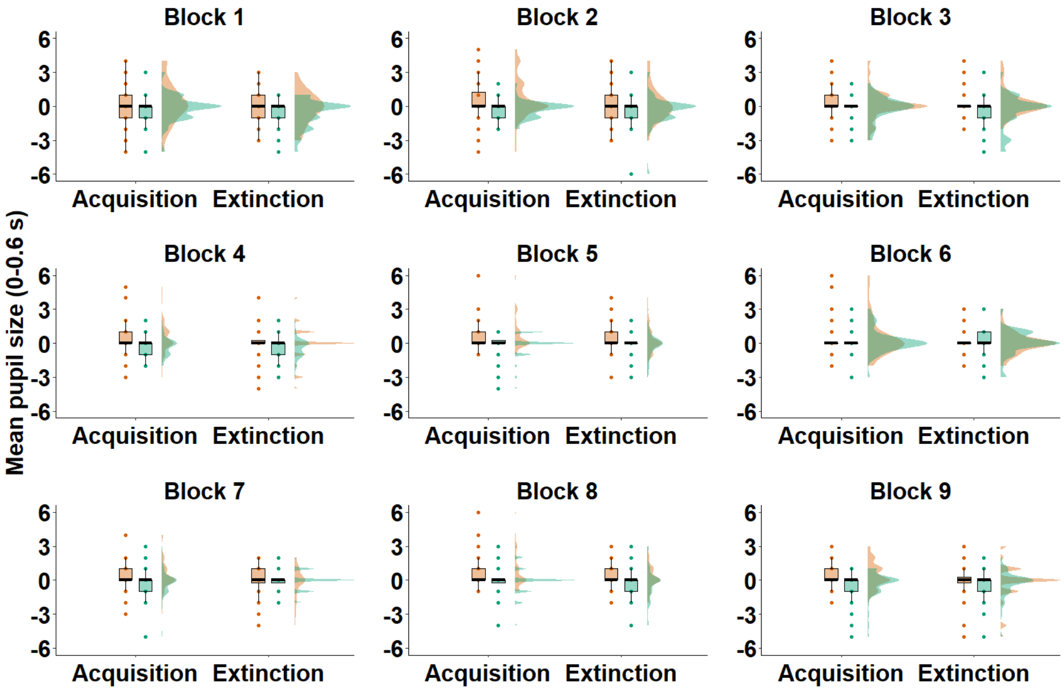
Supplementary Figure 41

Valence ratings over blocks.



Supplementary Figure 42

Arousal ratings over blocks.

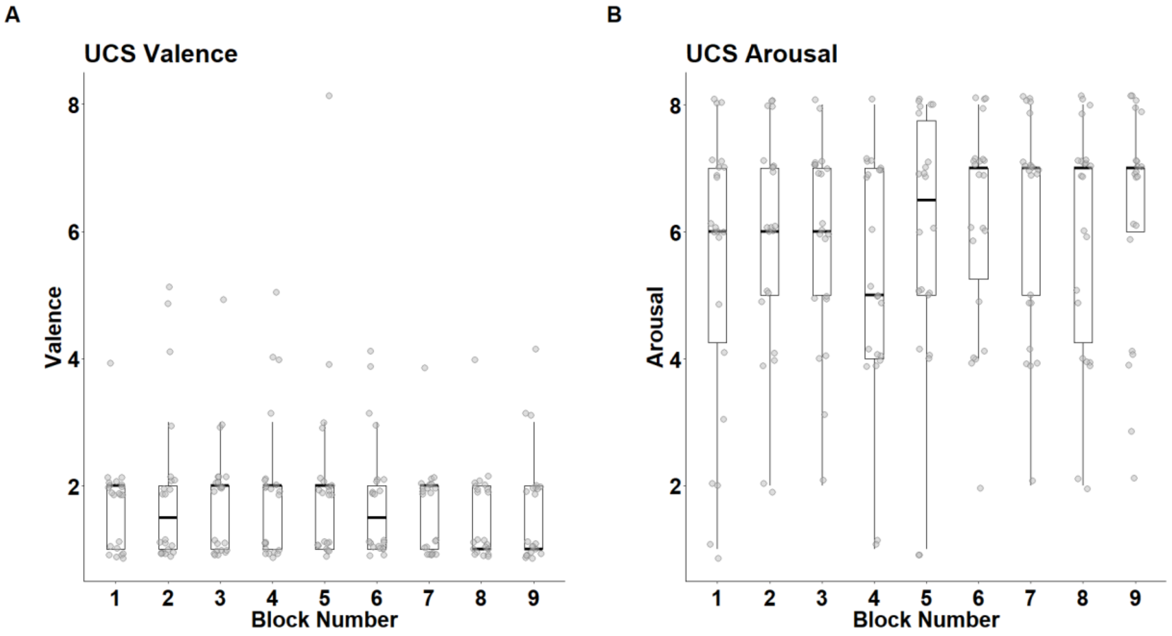


3.5.12 Supplementary Materials 15: Ratings of valence and arousal for the UCS over blocks

We examined participants' subjective ratings of valence and arousal for the UCS, to determine whether the UCS remained unpleasant over the duration of the experiment. The sound was rated on the scale of 1 to 8 ranging from not pleasant/not arousing to extremely pleasant/extremely arousing. As seen in Supplementary Figure 43, the median valence ratings remained at around 2 with relatively low variance that persisted over blocks. Arousal ratings also remained relatively high although there was a greater variance in responses, with a median arousal rating raging between 5 and 7. This suggests that the UCS remained arousing and unpleasant over the duration of the experiment.

Supplementary Figure 43

Valence and arousal UCS ratings for each block



4 Chapter 4 The more, the better: Auditory threat conditioning using multiple conditioned and unconditioned stimuli over many trials

4.1 Introduction

A wide range of physiological outcome measures have been employed to evidence the acquisition and extinction of threat related conditioned responses (CRs). The most commonly used CR readout (Lonsdorf et al., 2017) is the skin conductance response (SCR), which is an index of sympathetic nervous system activity and can be elicited through arousal (Jentsch et al., 2020; Tzovara et al., 2018). Studies have consistently demonstrated a larger SCR to CS+ than CS- during acquisition and a diminishing difference from early to late trials of extinction (Hopkins et al., 2015; Jentsch et al., 2020; Leuchs et al., 2019; Morriss et al., 2015; Reinhardt et al., 2010; Tzovara et al., 2018). Pupil size is a related, but in the context of conditioning, less frequently employed autonomic measure of arousal, reflecting both sympathetic (dilation) and parasympathetic (constriction) activation (Lonsdorf et al., 2017; Ojala & Bach, 2020). Previous pupillometry studies have reported patterns of conditioned responding similar to those in SCR (García-Palacios et al., 2018; Hopkins et al., 2015; Jentsch et al., 2020; Kluge et al., 2011; Korn et al., 2017; Leuchs et al., 2017, 2019; Reinhard & Lachnit, 2002; Tzovara et al., 2018; Visser et al., 2015, 2016, 2013). Specifically, these have shown more dilated pupils in response to CS+ than CS- trials, and a reduction in this difference during extinction. However, there is a large variety in the methods used for quantifying the CR in both SCR and pupil measurements, relying, for instance, on calculating peak responses, mean responses or areas under the curve within pre-defined time windows (Jentsch et al., 2020; Korn et al., 2017). For trials with longer durations, it is also common to subdivide trials into first (FIR) and second (SIR) interval responses[2], using pre-defined time

[2] Trial subdivisions into FIR and SIR have been motivated by early work suggesting that multiple CRs reflecting different underlying processes may be observable within a trial (Prokasy & Ebel, 1967). However, the empirical basis for such a distinction has been debated (Pineles et al., 2009).

windows that, again, vary greatly across studies (Jentsch et al., 2020; Pineles et al., 2009). Inconsistencies in pre-defining time windows of interest can make comparisons between outcomes from different studies extremely difficult. A potential solution to this problem, which we will apply in our own analyses, is to select time windows of interest using a purely data-driven approach.

Along with online (neuro)physiological measures of conditioning, behavioural CR indices are also commonly employed. These are typically obtained offline following conditioning and after extinction and include, but are not limited to, subjective ratings of valence and arousal (Bröckelmann et al., 2011; Gawronski & Mitchell, 2014; Glotzbach et al., 2012; Junghöfer et al., 2015a; Reinhardt et al., 2010; Sehlmeyer et al., 2011; Steinberg et al., 2013; Wendt et al., 2020) as well as self-reported fear/anxiety (Abend et al., 2020; Glotzbach et al., 2012; Morriss et al., 2015). Following conditioning, such studies typically (but see, e.g., Bröckelmann et al., 2011) report that CS+ trials are perceived as more unpleasant, arousing or fear-inducing than CS- trials (Gawronski & Mitchell, 2014; Glotzbach et al., 2012; Sehlmeyer et al., 2011; Morriss et al., 2015). This difference can diminish following extinction (Abend et al., 2020; Morriss et al., 2015). However, under certain conditions, cases of resistance to extinction have also been reported in both valence and arousal ratings (Gawronski & Mitchell, 2014; Sehlmeyer et al., 2011; Winn et al., 2018).

This extensive body of conditioning research is accompanied by common methodological practices that may affect both the replicability of previous affective conditioning research and the inferences that can be drawn from it. In recent years, there has been a growing interest in improving the methodological consistency, replicability, and validity of inferences drawn from conditioning studies. Efforts have focused on increasing data and reporting transparency, reducing excessive 'arbitrary' data reduction (e.g., through statistical modelling that incorporates all available data), and using evidenced-based data exclusion criteria (Lonsdorf et al., 2019; Ney et al., 2018). In this paper, we examine an additional set of inter-related experimental design and analytical decisions that may challenge the generalisability of findings both within and across studies.

While these will be discussed in the broad context of psychophysiological conditioning research, the study focuses specifically on pupillary and behavioural (valence and arousal) indices of learning.

First, many of the previous affective conditioning studies suffer from limitations in the generalisability of results across different items. This is because learning and extinction are typically investigated using only a single stimulus for CS+ and CS-, both of which are repeated over several trials (see Lonsdorf et al., 2017). Indeed, the use of several different CS-items within the same experiment is rather uncommon, because this may disrupt the acquisition of contingency awareness and thus delay, prevent, or reduce the CR of interest (Lonsdorf et al., 2017). However, the latter argument raises the question of whether results from studies with only one item per stimulus category are truly indicative of affective conditioning (which is thought of as an implicit process) or indeed of the potentially more strategic process of establishing contingency awareness.

More critically, experiments with such a limited number of items per condition cannot assess whether and to what extent results generalise to the population of items one could use for the same purposes (see Yarkoni, 2020 for a general discussion). This issue purely concerns the process of learning the initial association and is qualitatively different from the concept of generalisation of learning to related, unseen stimuli that have not previously been paired with an UCS (Dymond et al., 2015). Indeed, the point we are addressing here is that any conditioning effects observed between one CS+ (e.g., a circle) and one CS- (e.g., a triangle) are specific to those two items and any observed learning effects may not necessarily generalise to all shapes or even all triangles.

The problem with generalisability also applies to the UCS which in the context of threat conditioning is usually one unpleasant stimulus such as a mild electric shock or white noise [3] (Sperl, Panitz, & Hermann, 2016). Indeed, there is limited research examining how different UCSs would affect the CR. The available evidence so far suggests that white noise leads to a more sustained CR compared

[3] The differential effects of positive and negative UCSs on neutral CSs have also been studied in a form of classical conditioning often referred to as evaluative conditioning (De Houwer, 2007), however, there is little agreement on whether the two are qualitatively different.

to electric shocks, at least for studies that employ a very large number of trials (Sperl, Panitz, & Hermann, 2016). In standard paradigms, it has been shown that a fearful scream may be less effective than an electric shock (Glenn et al., 2012), whereas an unpleasant metal scrape sound can be as effective as white noise and/or electric shock in eliciting CRs (Neumann & Waters, 2006b). Nevertheless, these studies only used one item per condition, which cannot guarantee observing the same effect with another, similar item of the same category (e.g., female vs male scream).

Using a very limited set of items in the experimental design also extends to an analytical issue known as the *items-as-fixed-effect-fallacy* (Clark, 1973), i.e., failure to appropriately account for stimulus variability in the statistical analysis when participants are presented with multiple items per condition. Specifically, analytical approaches that rely on aggregating data up to the participant level (e.g., within-subjects t-tests and ANOVA) prevent the assessment of by-item generalisability even in experiments where several items per condition are being used. This is because analyses on by-subject means conflate by-item variability with residual noise. As a consequence, such analyses can lead to anticonservative inferences that do not generalise to new sets of stimuli of the same type. Such tests are also likely to produce 'too narrow' confidence intervals or 'too small' p-values, respectively (Judd et al., 2012; Yarkoni, 2020). For example, inflation of test statistics in standard 'by-subject only' analyses has been estimated to reach 50% or more in a number of publicly available fMRI datasets (Westfall et al., 2017), and up to 60% in simulated datasets (Judd et al., 2012). Simultaneously modelling participants and items as random factors, as well as all dependencies with the experimental conditions is a necessary requirement for generalising findings to populations of stimuli and participants as well as for reducing the risk of a Type I error (Barr et al., 2013; Judd et al., 2017). This is particularly important when the aim is to gain insights into general learning mechanisms that may underlie certain anxiety disorders, intended to inform generally applicable interventions in this domain.

A final point concerns the number of trials used within experiments. Since psychophysiological measures (e.g., SCR) can exhibit amplitude reductions with repeated stimulus presentations (Leuchs et al., 2019; Lonsdorf et al., 2017;

Ojala & Bach, 2020), studies usually employ a small number of trials (5-20 trials per condition, Lonsdorf et al., 2017) to prevent such habituation of the CR. At the same time, psychophysiological approaches are often noisy and therefore, require a large number of trials to achieve an acceptable signal-to-noise ratio (Ney et al., 2018; Steinberg et al., 2013; Tzovara et al., 2019). However, guidance for how to adequately consider the trade-off between signal-to-noise ratio and CR habituation is limited. For SCR, piloting is recommended as a method for estimating the number of observations required for the detection of a CR (Lonsdorf et al., 2017), while for pupillometry, there are no systematic investigations examining the precise effect of number of trials required to detect an effect, not only in the context of emotion and fear processing but in psychophysiological research in general.

The necessity of using many trials to compensate for intrinsically noisy signals also extends to M/EEG measures, where detecting activity in subcortical structures (e.g., amygdala, thalamus, hippocampus) suggested to underlie fear and emotion processing (Duvarci & Pare, 2014; Fossati, 2012; Fullana et al., 2016) can require hundreds or even thousands of trials (Attal et al., 2007). For example, a simulation study by Quraan et al. (2011) demonstrated that while evoked hippocampal activity can be detected with 10 trials, the localisation accuracy is very poor (i.e., 18 mm away from the actual source for 10 trials and 10 mm for 50 trials). Therefore, it is unlikely that employing a traditional conditioning paradigm with a small number of trials would allow for an accurate detection of neural learning and extinction-related processes. In addition, as discussed in Chapter 3, the currently available MEG paradigms offering large enough number of trials may risk hampering CR detection due to issues related to CR habituation and poor contingency awareness. A potential solution to these issues was recently reported in an EEG investigation by Sperl et al. (2021), who used a novel sequential conditioning task, in which learning was established three consecutive times using a different set of one CS+ and one CS- items. This was followed by a sequential extinction on the following day. This task was successful at eliciting differential CRs across multiple outcomes measures, including valence, arousal, SCR, heart rate changes and ERPs.

The goal of the present study was to test a qualitatively similar conditioning paradigm to that used by Sperl et al. (2021), that attempts to overcome many of the above methodological issues and to provide a means for measuring associative learning and extinction mechanisms robustly, using multiple outcome measures. We used continuous pupil size recordings, as well as ratings of valence, arousal, and contingency awareness to examine threat conditioning and extinction. The procedure was an auditory blocked conditioning in which learning and extinction were established several times using different sets of conditioned (pure tones) and unconditioned stimuli (environmental sounds). Both CSs and UCSs were auditory rather than visual since sounds were expected to interfere less with the measurement of pupil size. Since the present task used a different set of UCSs in each block, instead of exposing participants to successive acquisition blocks followed by successive extinction blocks (as in Sperl et al., 2021), in the present task each block contained the standard three experimental phases – habituation, acquisition, and extinction. In addition, to ensure that the CRs are driven by aversive anticipation rather than by valence-unspecific expectancy, we used both pleasant and unpleasant UCSs. To maximise the distinction between the UCSs, positive and negative stimuli were presented at low and high intensity levels, respectively. Establishing learning and extinction several times in this blocked design can offer a potential solution to the issue of small numbers of items and trials in conditioning research. And while the present study does not employ M/EEG, the design can theoretically be applied to neurophysiological measures as well, and potentially offer enough trials to make investigations of learning effects in deep structures viable.

To make inferences about pupil size changes in response to conditioning and extinction, we used a data driven approach to identify significant time windows of interest for averaging over time. Unlike traditional methods, this approach does not rely on prior knowledge for time-window selection, allows for using all available data, and maximises detection of unpredicted effects (Huang & Zhang, 2017). At the same time, the method also provides a powerful control of both Type I and Type II error (Sassenhagen & Draschkow, 2019). We then used design-appropriate linear mixed effects modelling to examine differences between conditions in the most generalisable manner for both the behavioural and pupillometry data. Finally, we demonstrate the potential risk of alpha inflation when using statistical tests that fail to account for both item and subject variation, by comparing our mixed model findings to the results from conventional repeated-measures ANOVA analyses.

We expected that during acquisition, the mean pupil size in the time window of interest would be greater for CS trials paired with a negative UCS (CS Negative) than for those paired with a positive UCS (CS Positive). During extinction, we expected this difference to be reduced in magnitude or completely eliminated. We also descriptively compared CRs during early and late trials during each experimental phase, to examine potential indications of a gradual increase (acquisition) and decrease (extinction) of the CR over time. For the behavioural ratings, we predicted lower valence and higher reported arousal for CS Negative than CS Positive trials during acquisition. In analogy to the pupil size predictions, we expected these differences to diminish in magnitude during extinction. Finally, we collected self-reported data on measures of state and trait anxiety and emotion regulation. These were used for descriptive and exploratory purposes that were not critical to the main hypotheses being tested.

4.2 Method

4.2.1 Participants

We collected data from 30 participants, whose general demographic characteristics are shown in Table 11. All participants had normal or corrected-to-normal vision and normal hearing. Participants were recruited through the University of Glasgow Psychology Subject pool and received £6 per hour for their time. Prior to participation, they provided written consent to take part. The study was ethically approved by the College of Science and Engineering ethics committee (300190006).

Two participants completed only 3 out of 4 blocks (either because of technical problems or because they cut the experiment short) but their data were retained for all analyses. Five participants had to be excluded because of excessive measurement error in the pupil size data (more details further below). Hence, only 25 participants were included in the pupil size analyses, whereas all

30 participants were included in the behavioural analyses (valence and arousal ratings).

Table 11

Demographic information.

Sex	N (N Pupil*)	N Native Speakers	Mean Age	Age Range	Mean STAI-Trait (SD)	Mean STAI-State (SD)	ERQ-Reappraisal (SD)	ERQ-Suppression (SD)
Female	15 (12)	8	22.93	19 - 31	45.53 (9.8)	35.07 (10.2)	4.5 (0.65)	3.15 (1.63)
Male	15 (14)	7	22.67	19 - 29	39.93 (11.4)	32.13 (8.9)	4.88 (1.28)	3.5 (1.48)

*N Pupil: Number of participants in the pupil size dataset after exclusion

4.2.2 Psychological Assessment

Participants were asked to complete a basic demographic information questionnaire as well as the State Trait Anxiety Inventory (STAI) (Spielberger et al., 1983) and the Emotion Regulation Questionnaire (ERQ), (J. J. Gross & John, 2003). Descriptive, exploratory analyses of the relationship between psychological and conditioning measures can be seen in Supplementary Materials 16. Overall, correlations between conditioning and psychological measures were low to moderate, but none were statistically significant.

4.2.3 Stimuli

Conditioned stimuli (CSs) were 16 sine-wave tones of 4 second duration. Eight tones had low (200 – 400 Hz, 28.57 Hz steps) and 8 had high (600 – 800 Hz, 28.57 Hz steps) constant frequency. Stimuli were created in Matlab 2017a using Psychtoolbox (*makeBeep*). Unconditioned Stimuli (UCSs) were 4 positive (bongos, guitar, harp, and bird chirping) and 4 negative (metal squeak, knife scrape, drilling, and a female scream) sounds with a duration of 1 second. The latter were selected based on their valence ratings (high and low respectively) in a separate norming study (N=30, different from the participants in the main study, see https://osf.io/dehxa/ and Supplementary Materials 17 for details about the stimuli). Positive sounds were selected based on the highest 25th percentile valence rated at 60 dB, whereas negative sounds were selected based on the lowest 75th percentile valence rated at 90 dB. To equalise the intensity of the CS and UCS stimuli, each audio file was mean-centred and then normalised to the same, maximum root mean square (RMS) amplitude, without inducing clipping, using a RMS equaliser (The Phonetics Lab, University of Washington, https://depts.washington.edu/phonlab/resources/rmsLeveler.m). The first and last 50 ms of the signal of all stimuli was gradually faded in and out in Audacity 2.1.2 (https://www.audacityteam.org/). The normalisation procedure was applied separately for CSs and UCSs. The resulting normalised CSs were presented at a maximum intensity of approximately 50 dBA. Positive and Negative UCSs were presented at approximately 60 dBA and 90 dBA, respectively. Sound intensity was measured by Cadrim sound level meter. All stimuli used in the present study can be found at https://osf.io/pnyrh/.

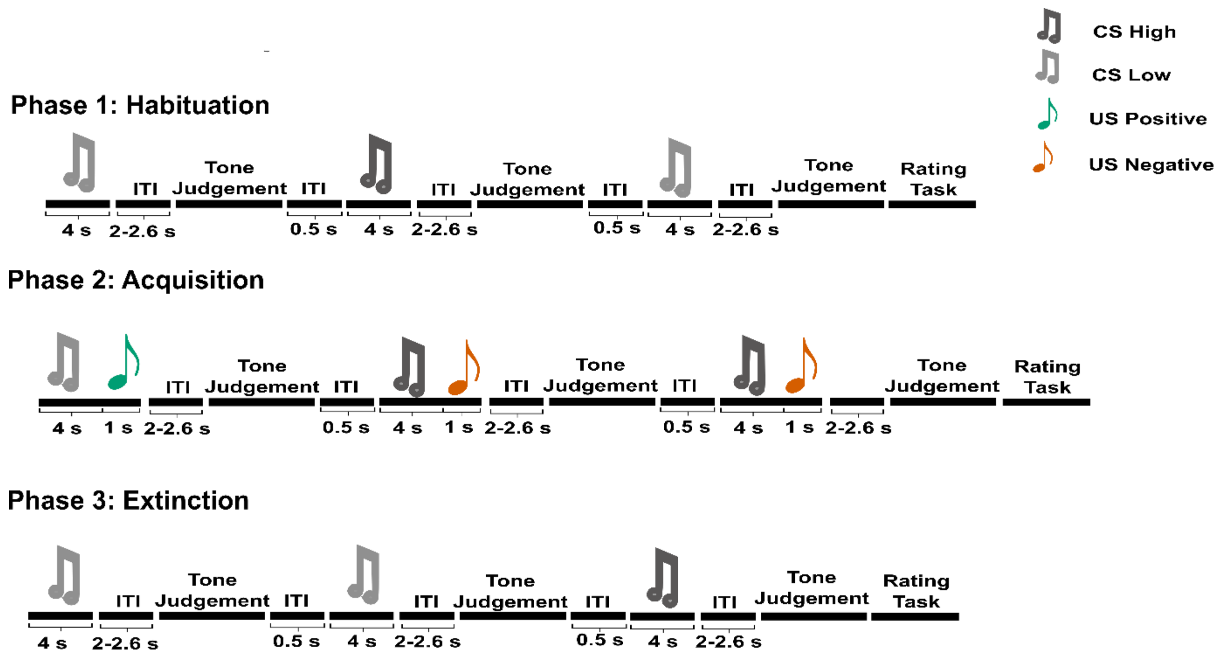
4.2.4 Procedure

Prior to completing the main task, participants were asked to fill in the self-report measures (see Psychological Assessment). The main task comprised 4 blocks, each containing three experimental phases – Habituation, Acquisition, and Extinction (see Figure 24). Each block contained a different subset of 2 low and 2 high frequency sine tones, selected from the total set so that the minimum difference *between* high and low frequency tones was 286 Hz, and the minimum difference *within* low or high frequency tones was 114 Hz. For example, one block contained low frequency tones of 200 and 314 Hz, and high frequency tones of 600 and 714 Hz. Assignment of low and high frequency tones to the CS Negative condition was Latin square counterbalanced, resulting in 4 stimulus sets. Four additional sets were constructed by swapping the assignment of low and high frequency tones to the CS Positive condition, leading to a total of 8 different stimulus sets (see Supplementary Materials 18). Block order and assignment of the 4 UCS positive and 4 UCS negative sounds to each block were also Latin square counterbalanced.

Each CS was presented together with a black fixation cross, positioned at the centre of the screen for 4 seconds. The inter-trial interval (ITI) had a duration of 2.3 sec ± 300 ms and was accompanied by a black fixation cross. To minimise ocular artifacts, participants were asked to always maintain fixation at the centre of the screen. In each block and phase, each stimulus was presented 10 times, resulting in a total of 20 trials per condition per block (a total of 80 trials per condition across blocks). Trial order was randomised across participants with the restrictions that the first trial was always a CS Positive and no more than two trials of the same stimulus type (e.g., a CS Negative) could occur consecutively. To maintain participants' attention during each block, they were asked to perform a tone judgement task. Participants were instructed that they will complete a 3-part task in which they will be presented with a series of tones that they have to listen to while maintaining fixation at the centre of the screen and indicate whether each tone is high or low in pitch (75% of trials) or hard or soft (25% of trials). They were informed that the only difference between the three parts would be that during part 2, the tones would be paired with positive and negative sounds. Participants were not informed about the contingency between CS and UCS. During Habituation and Extinction, the sine tones were presented on their own, and during Acquisition, CSs were paired with the UCSs, which occurred at CS offset for a duration of 1 second. At the end of each phase, participants completed a sound rating task where they were asked to rate each CS on valence ranging from unpleasant to pleasant and arousal ranging from boring to exciting using a slider scale ranging from -100 to 100. At the end of the Acquisition phase, participants also rated the UCS on valence and arousal (see Supplementary Materials 19). In order to determine participants' awareness of the relationship between CSs and UCSs, at the end of each block, participants completed a contingency awareness task in which they were asked to decide whether each pure tone was paired with a positive or a negative sound. Participants were also asked to provide a confidence rating on a slide scale ranging from "not at all confident" (-100) to "confident" (100).

Figure 21

Auditory blocked conditioning scheme for each of the three experimental phases (Habituation, Acquisition, and Extinction).



4.2.5 Pupil data acquisition and pre-processing

Pupil size data were obtained using an EyeLink 1000 eye tracker and recorded continuously during each trial presentation, with initial sampling rate of 250 Hz.

Raw .edf data were exported into .txt format using DataViewer and imported into R. Pre-processing was performed separately for each participant and block using the *PupilPre* package in R and following the recommended guidelines provided by the package (see https://bit.ly/3iBYCAR). Raw, continuous data were transformed into time series with a trial length of -0.5 to 5.5 s relative to CS onset. The first trial in each block was removed, leaving 119 trials per block. This was necessary since the continuous pupil size recording began at the onset of the first trial which meant that no baseline window could be acquired for this trial.

Blink removal and artifact correction were performed in several steps. First, blinks identified by the EyeLink software were padded with 150 ms on either side. Within the padded blink window, data were examined in two stages, using the *clean_blink* function. First, data points within the padded window were removed if the difference between subsequent pupil size values was larger than 5 (i.e., pupil size at time point 1 – pupil size at time point 2 > 5, package default). Next, any small runs of data points that were surrounded by missing values were identified and removed. Specifically, within the padded window, any 40 ms segment of data, surrounded by at least 2 missing values (default) on each side was removed.

Detected artifacts (extreme data points) were also removed using the *clean_artifact* function. Artifacts were detected using the defaults provided in the package vignette. First, to identify potential outliers, each trial was divided into bins of 100 ms within which the median absolute deviation (MAD) of the pupil size data was calculated. A bin was marked extreme if it had a sensitivity threshold (MAD constant) > 2 standard deviations of the pupil dilation. The larger the threshold, the more extreme data point is required to be considered as an outlier. Each extreme bin included 200 ms of padding within which a multidimensional distributional distance (Mahalanobis distance) was calculated using the horizontal and vertical velocity and acceleration of the pupil. A pupil size was marked as extreme if the Mahalanobis distance exceeded 2 SD. Runs of data points surrounded by missing values were removed in 2 stages again, using the same procedure as for blink removal.

Finally, remaining small blinks and artifacts that were undetected by the automatic cleaning procedure were manually removed (*user_cleanup_app*). Following artifact removal, sparse trials were identified and removed. A trial was discarded if it contained less than 20% of baseline (at least 0.1 s) and less than 70% of post-baseline data (at least 3.85 s). Five participant data sets were discarded from the analyses altogether because more than 50% of their trials showed excessive blinks or movement artifacts and were removed during cleaning. For the remaining 25 participants, the average data loss after cleaning was 12%. The missing values in the corresponding trials were replaced using spline interpolation (function *interpolate_nas* in *PupilPre*).

The cleaned and interpolated pupil size data were log10 transformed. Next, multiple linear regression was performed for each participant and block (log pupil size as outcome, and X and Y eye position as predictors) to remove small eye movement-related artifacts. From these regression analyses, the residual pupil sizes per trial and time point were extracted and used in subsequent analyses. Log10 pupil size change from baseline (mean pupil size between -0.5 and 0 s) was then calculated for each trial and time point using the *baseline* function (*change* = *pupil size* - *baseline*). The inverse of the baselined pupil was then calculated to obtain a measure of proportional change from baseline. Finally, data were down sampled to 10 Hz.

4.3 Results

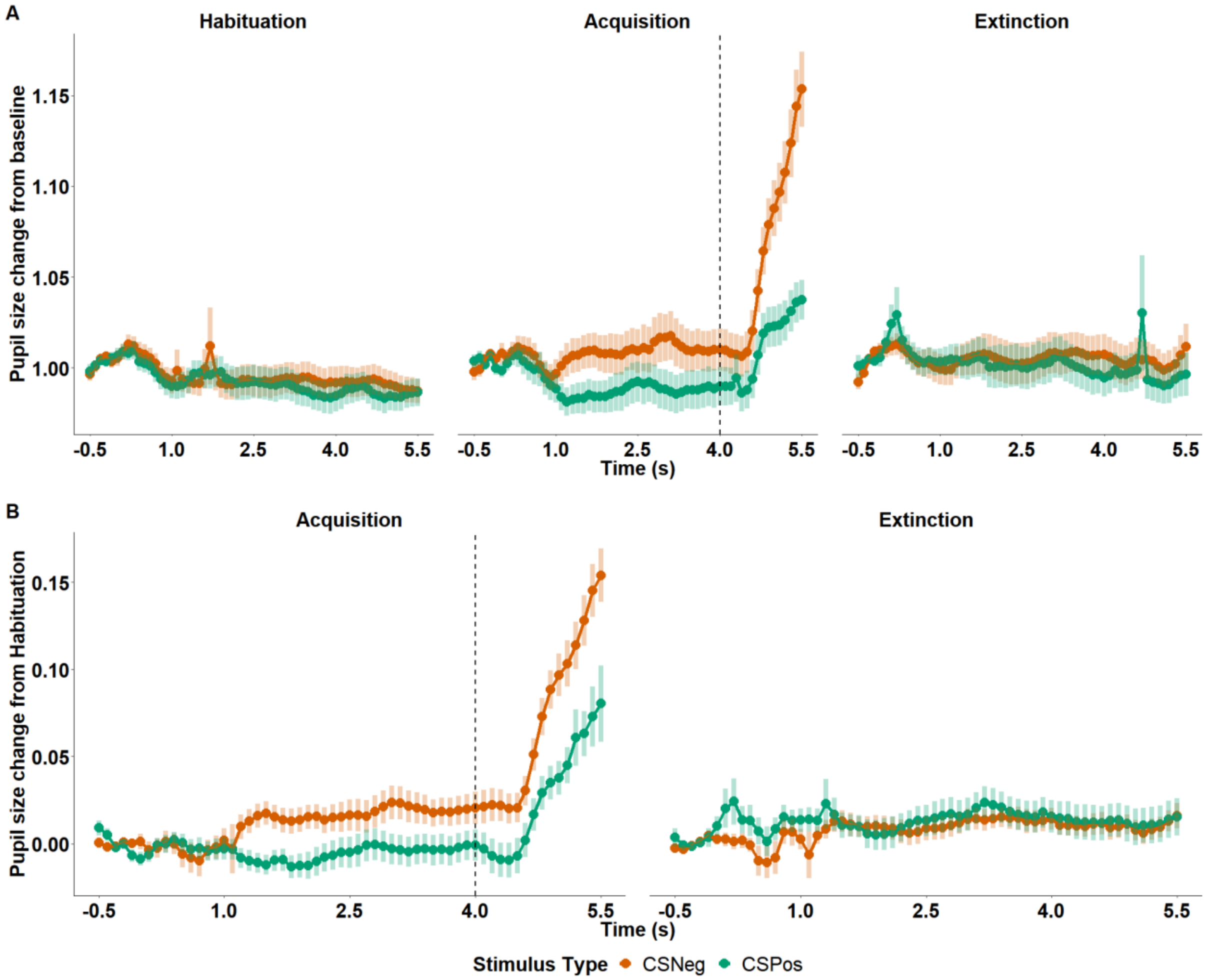
All analyses were performed in R version 4.0.2. The code and data associated with these analyses can be found at https://osf.io/pnyrh/.

4.3.1 Pupil size

Figure 25 shows, for each phase (Habituation, Acquisition, and Extinction) the average proportional pupil size change from baseline over a 5.5 s time window including both CS presentation (0- 4 seconds) and UCS presentation (4 – 5 seconds). As seen in the figure, differences in mean pupil size between CS Positive and CS Negative trials during Acquisition become apparent from around 1 s post CS onset, reflecting the conditioned response (CR) and increase in magnitude following UCS onset (4 s), reflecting the unconditioned response (UR). Since any anticipatory processes should occur prior to the onset of the UCS, the confirmatory analysis focused on the CR only (0- 4 s). To quantify the effects of the conditioned stimuli across the experimental phases, the analysis was performed in two stages. As explained in more detail below, we first employed a cluster-based permutation test to identify time windows of interest in a data-driven manner. Second, to account for subject and item-related random variation, a linear mixed effects (LME) model was built in which the mean pupil size across each time window of interest was used as the dependent variable. Since cluster-based permutation tests can presently only handle designs that are no more complex than 2 x 2, the current 3 x 2 design was reduced to 2 x 2 by calculating pupil size changes from Habituation. For each participant, block and item, a mean pupil size during Habituation was calculated across time and trials which was then subtracted from each time point during Acquisition and Extinction (see Figure 25B). The LME modelling was also performed on Habituation-baselined data.

Figure 22

Proportional mean pupil size over time.



Note. A) Changes in mean pupil size from baseline over time for each condition (green: CS Positive; orange: CS Negative) and in each of the three experimental phases (Habituation, Acquisition and Extinction). B) The same data for the Acquisition and Extinction phases after subtracting the Habituation phase baseline (reflecting the difference between Acquisition and Habituation and between Extinction and Habituation, respectively). Vertical, light-coloured bars indicate standard errors of the means. The vertical dashed lines indicate UCS onsets.

4.3.2 Identifying time window of interest using cluster-based permutation

The cluster-based permutation analysis was carried out using the R packages *clusterperm* and *exchangr* and involved the following steps. First, an *Experimental Phase* (Acquisition vs Extinction) * *Stimulus Type* (CS Positive vs CS Negative) within-subjects ANOVA (*aov_by_bin*) was performed for each time bin (0 – 4 s). Adjacent time bins with a $p < 0.05$ were combined into clusters, with the sum of the F values within each cluster serving as the cluster statistics for the two main effects and the interaction (*detect_clusters_by_effect*). Next, a Monte-Carlo test with 2000 permutations was performed (*cluster_nhds*) in which a cluster statistic was computed on trials that were randomly assigned to each condition within subjects, resulting in 2000 cluster statistics. For each of the main effects and the interaction, the permuted cluster statistics were compared against the observed statistic. Clusters were considered significant if they fell within the highest or lowest 2.5 % of the null distribution. The results from the cluster-permutation test can be seen in Table 12 and is visualised in Figure 26. The test revealed two significant time clusters for the *Experimental Phase* * *Stimulus Type* interaction: around 1.2 and 1.7 seconds and around 3 and 3.4 seconds, respectively. Since this test is only suggestive of where in time an effect may be observed (without providing clues about its generalisability across items and participants), the suggested interactions in these time clusters were further assessed inferentially using LME modelling (see below).

Table 12

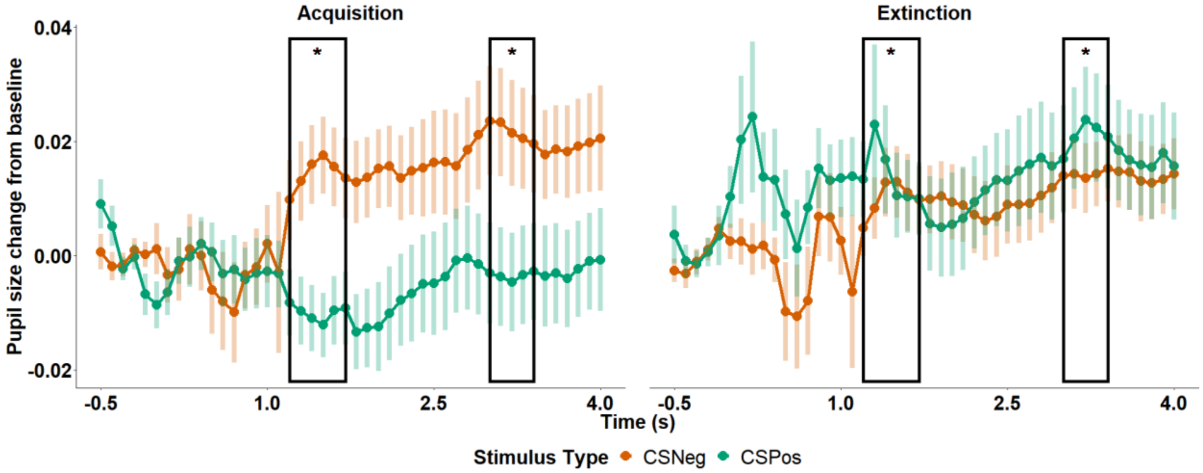
Cluster-based permutation results.

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

Effect	B0	B1	Sign	Cms	P-value
Experimental Phase	0	100	-1	11.29	1
Experimental Phase	800	900	-1	11.908	0.583
Stimulus Type	1500	2100	1	39.324	0.065
Experimental Phase X Stimulus Type	0	0	1	4.301	0.143
Experimental Phase X Stimulus Type	1200	1700	1	31.855	0.024*
Experimental Phase X Stimulus Type	2300	2400	1	8.819	0.09
Experimental Phase X Stimulus Type	3000	3400	1	27.68	0.028*
Experimental Phase X Stimulus Type	3900	3900	1	4.146	0.151

Figure 23

Significant Experimental Phase X Stimulus Type time cluster derived from the cluster-based permutation test.



Note. The vertical, light-coloured lines indicate the standard error of the mean.

4.3.3 Linear mixed effects modelling of data within the identified time clusters

To examine the development and extinction of anticipatory learning, an LME model (R package *lme4*) was built for each of the two significant time clusters identified in the cluster-based permutation analysis. From this point onward, the effect in the time window between 1.2 and 1.7 s post CS onset would be referred to as the first interval response (FIR) and that between 3 and 3.4 s post CS onset as the second interval response (SIR). For each subject, block and trial, a mean pupil size was calculated for the FIR (see Figure 27A) and SIR (see Figure 28 A). These served as the dependent variables in the two models. Each model consisted of a 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS Positive vs CS Negative) fixed effects design. The models included mean-centred contrasts (deviation coding) for the two categorical fixed-effect predictors. Following Barr et al. (2013), models with design-appropriate *maximal* random effects structure were fitted. Specifically, *Subjects* and *CS Items* were added as random intercepts. Since experimental manipulations were both within-subjects and within-items, the model also included by-subject and by-item random slopes for each main effect as well as for the interaction (see Supplementary Materials 20 for random effects summary). Since each item was presented multiple times, an interaction between the *Subjects* and *CS Items* random intercepts was also included, coupled with the main effects and the interaction of the by-subject and by-item random slopes. Random variability due to *UCS Item* variability was not modelled since UCSs were only present during Acquisition and not during Extinction. However, separate models, accounting for UCS variability were conducted on the Acquisition phase only (see Supplementary Materials 21). P-values for the fixed effects were determined via Type III Wald Chi-square tests. Table 13 shows the fixed-effects results and Figure 27C shows the model estimates, with associated 95% CIs. As shown, the interaction between *Experimental Phase* and *Stimulus Type* in the FIR model was significant. Simple effect analysis of the interaction was performed using estimated marginal mean simple contrasts (package *emmeans*) with Satterthwaite method for degrees of freedom approximation. Consistent with the descriptive data, these contrasts confirmed a reliably larger mean pupil size for CS Negative than CS Positive trials during Acquisition (t (56.6) = 3.2, p =

0.003), but no clear simple effect of *Stimulus Type* during Extinction (*t* (22.2) = -0.3, *p* = 0.79), (see Supplementary Materials 20 and Figure 27B). In addition, a larger mean pupil size was also observed for CS Positive trials during Acquisition compared to CS Positive trials during Extinction (*t* (22.6) = -2.4, *p* = 0.024). Similarly, the interaction between *Experimental Phase* and *Stimulus Type* in the SIR model was also significant. Again, the simple contrasts revealed a larger pupil size for CS Negative than CS Positive trials during Acquisition (*t* (23.3) = 2.42, *p* = 0.023), but no significant simple effects during Extinction (*t* (15.9) = -0.3, *p* = 0.75), (see Supplementary Materials 20 and Figure 28B).

Table 13

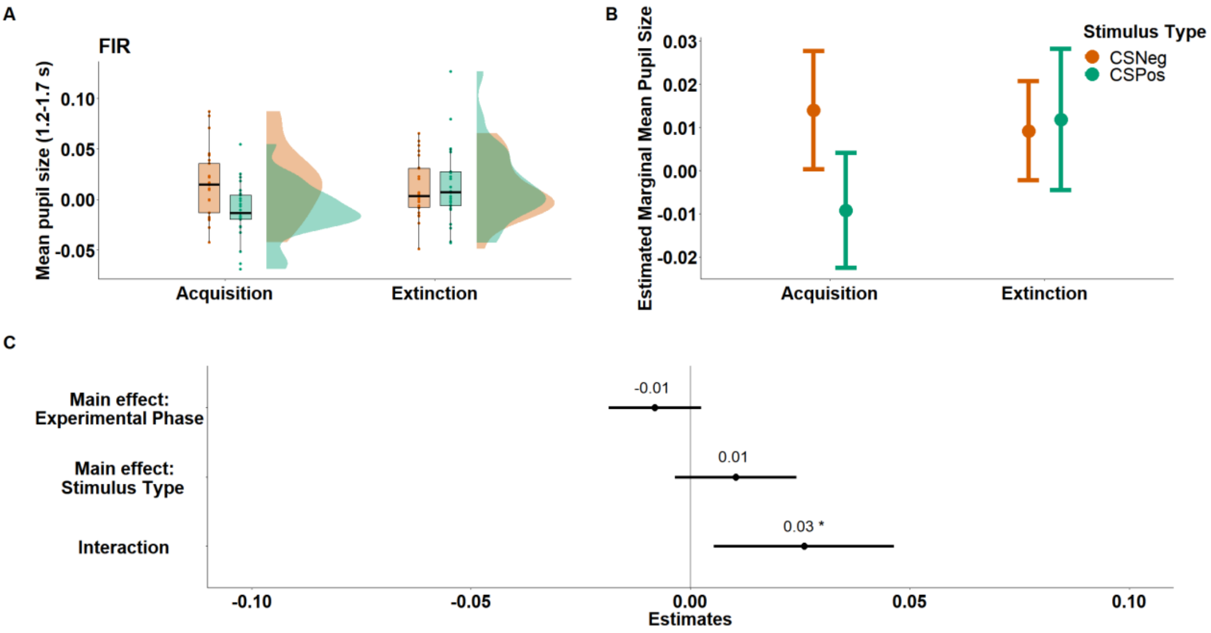
Type III Wald Chi-square tests and R-squared values for the complete pupil models and each of the fixed effects for the first and second interval responses.

	Chisq	Df	P-value	R^2 Fixed (CI)
FIR				
Full Model (Fixed)				0.003 (0.001 – 0.006)
Experimental Phase	2.29	1.000	0.13	0.001 (0-0.002)
Stimulus Type	2.10	1.000	0.15	0.001 (0-0.003)
Experimental Phase X Stimulus Type	6.08	1.000	0.01*	0.001 (0-0.004)
SIR				
Full Model (Fixed)				0.002 (0 – 0.005)
Experimental Phase	0.79	1.000	0.37	0.000 (0-0.002)
Stimulus Type	1.44	1.000	0.23	0.001 (0-0.002)
Experimental Phase X Stimulus Type	5.51	1.000	0.02*	0.001 (0-0.003)

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

Figure 24

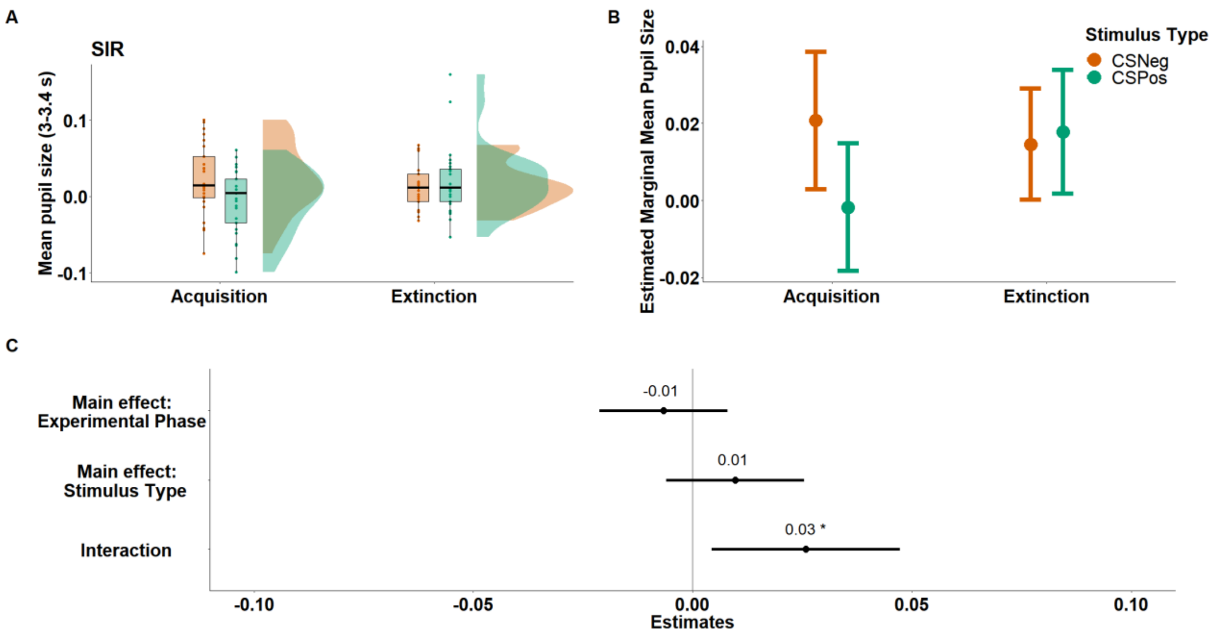
A summary of pupil size fixed effects for the first interval response (FIR)



Note. A) Distribution of mean pupil size between 1.2 and 1.7 s post CS onset of Habituation-baselined data. B) Estimated marginal means per condition derived from the mixed effects model of pupil size (error bars represent 95% CIs for the means conditioned on the random effects). C) Fixed effect estimates (labelled dots) derived from the mixed effects model of pupil size; bars represent 95% CIs for the estimates.

Figure 25

A summary of pupil size fixed effects for the second interval response (SIR)



Note. A) Distribution of mean pupil size between 3 and 4 s post CS onset of Habituation-baselined data. B) Estimated marginal means per condition derived from the mixed effects model of pupil size (error bars represent 95% CIs for the means conditioned on the random effects). C) Fixed effect estimates (labelled dots) derived from the mixed effects model of pupil size; bars represent 95% CIs for the estimates.

4.3.4 Valence and arousal ratings

For comparability with the pupil size analysis, the rating-data analysis was performed on Habituation-baselined valence and arousal data (see Figure 29B and 26D). Since each item was rated once per block, the baseline-adjustment was performed by subtracting a given item's rating during the Habituation phase from the same item's rating during the Acquisition and Extinction phases, respectively. As seen in Figure 29A and B, mean valence ratings were slightly lower for CS Negative than CS Positive trials during both Acquisition and Extinction, although the distributions overlapped to a substantial degree. The difference in the means between the two stimulus types was even more pronounced when the valence data were put in relation to the Habituation-phase baseline. Mean arousal ratings (Figure 29C and D) were slightly higher for CS Negative than CS Positive trials, but only after adjusting for the Habituation-phase baseline (Figure 29D).

In analogy to the pupil size analyses, we fitted two LME models - one predicting the valence ratings and the other one predicting the arousal ratings after subtracting the Habituation-phase baseline ratings. As with the pupil size model, each of the two rating models employed 2 *Experimental Phase* (Acquisition vs Extinction) by 2 *Stimulus Type* (CS Positive vs CS Negative) fixed effects design, using mean-centred contrasts for the two categorical fixed effect predictors. By-subject and by-item random intercepts were added, together with by-subject and by-item random slopes for both main effects and the interaction (see Supplementary Materials 20 for random effects summary).

For valence ratings, a Type III Wald Chi-square test (see Table 14 and Figure 30A for model estimates) revealed a significant *Stimulus Type* main effect. Since the interaction was non-significant, the estimated marginal means contrasts were computed by averaging over the levels of *Experimental Phase*. These revealed (see Supplementary Materials 20 for contrasts table) that overall, CS Negative trials had lower valence than CS Positive trials (*t* (18.03) -8.5, *p* = 0.0215). In terms of arousal ratings, no significant main effects or interactions were observed (see Table 14 and Figure 30C).

Table 14

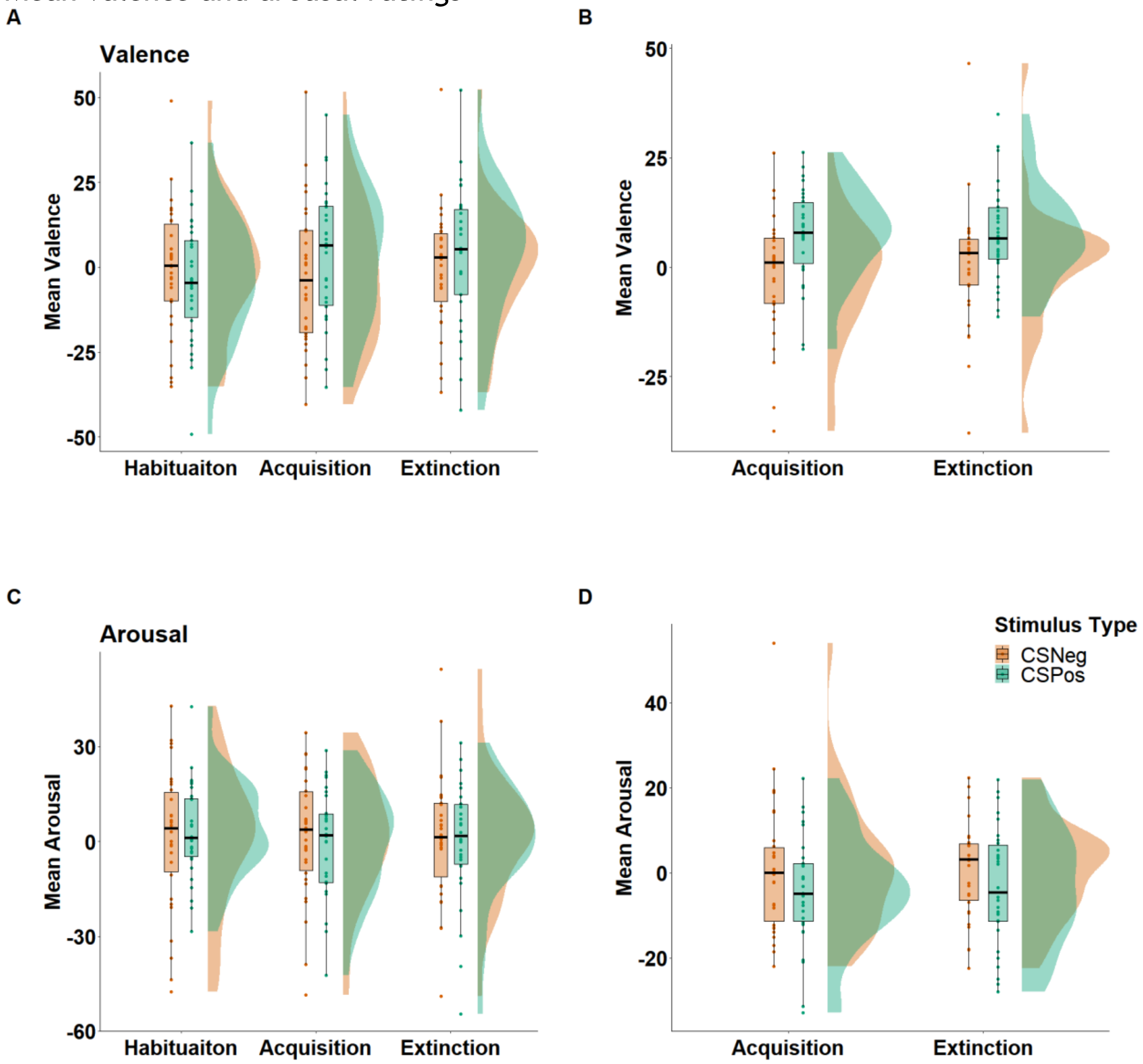
Type III Wald Chi-square tests and R-squared values for the complete valence and arousal models and each of the fixed effects.

	Chisq	Df	P-value	R^2 Fixed (CI)
Valence				
Full Model				0.014 (0.004 – 0.03)
Experimental Phase	0.27	1.000	0.6	0 (0 – 0.006)
Stimulus Type	6.46	1.000	0.01**	0 (0 – 0.007)
Experimental Phase X Stimulus Type	0.21	1.000	0.64	0.013 (0.003 - 0.032)
Arousal				
Full Model				0.005 (0.001-0.02)
Experimental Phase	0.029	1.000	0.866	0 (0 – 0.005)
Stimulus Type	1.519	1.000	0.218	0 (0 – 0.006)
Experimental Phase X Stimulus Type	0.095	1.000	0.757	0.005 (0 – 0.017)

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

Figure 26

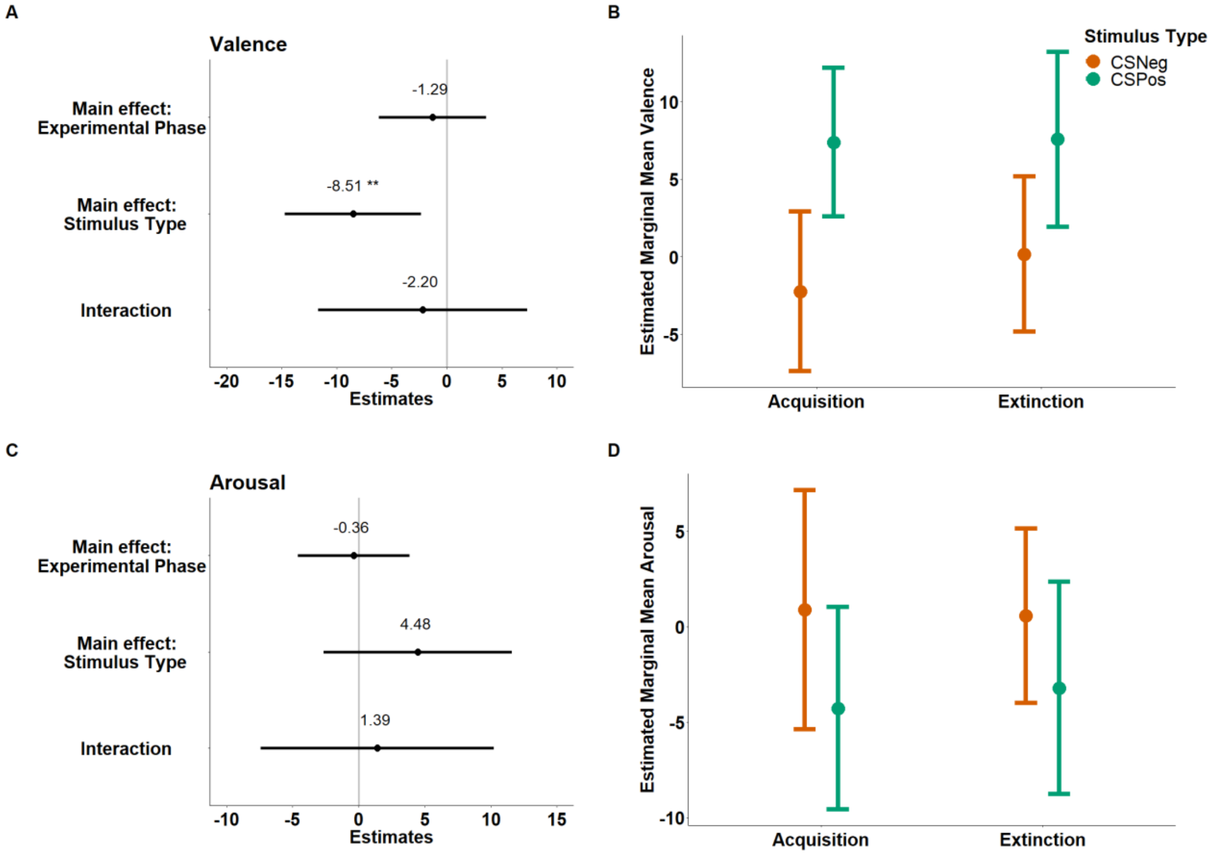
Mean valence and arousal ratings



Note. A) and C) Mean valence and arousal ratings across Habituation, Acquisition and Extinction phases. B) and D) Habituation-baselined mean valence and arousal ratings.

Figure 27

A summary of valence and arousal fixed effects.



Note. Fixed effect estimates (labelled dots) derived from the mixed effects model of valence (A) and arousal (C) ratings; bars represent 95% CIs for the estimates. Estimated marginal means per condition derived from the mixed effects model of valence (B) and arousal (D) ratings (error bars represent 95% CIs for the means conditioned on the random effects).

4.3.5 Contingency awareness

Detailed descriptions of the contingency awareness data, and related analyses can be seen in Supplementary Materials 22. Overall, we found that participants were clearly aware of the CS-UCS contingency. Importantly, however, there was no clear evidence for a relationship between contingency awareness and the conditioned response in any of our outcome measures.

4.3.6 Conventional Analysis approaches

For comparison, Table 15 shows the results from by-subject repeated-measures ANOVAs performed on pupil size, valence ratings, and arousal ratings as dependent variables. As seen in the table, the general pattern of results is consistent with that of the mixed effects modelling. However, in most instances the strength of association (standardized 'effect size') is larger, and associated p-values much smaller in the ANOVA outputs than in the mixed effects model analyses. These differences are particularly noticeable for the interaction effects, where the p-value for the pupil FIR and SIR interactions have decreased from 0.01 and 0.02 in the LME to 0.007 and 0.005 in the ANOVAs, respectively. The p-value for the main effect of valence rating has also decreased from 0.01 to 0.008. As seen in Figure 31, this is also reflected in the confidence intervals (CIs), whereas across outcome measures, the CIs derived from the mixed models tend to be wider. While the consistent patterns are reassuring, there is a suggestion of more anticonservativity in the ANOVA analyses, potentially because the latter only take by-subject but not by-item variability of effects into account (unlike our LME analyses which considered both simultaneously).

Table 15

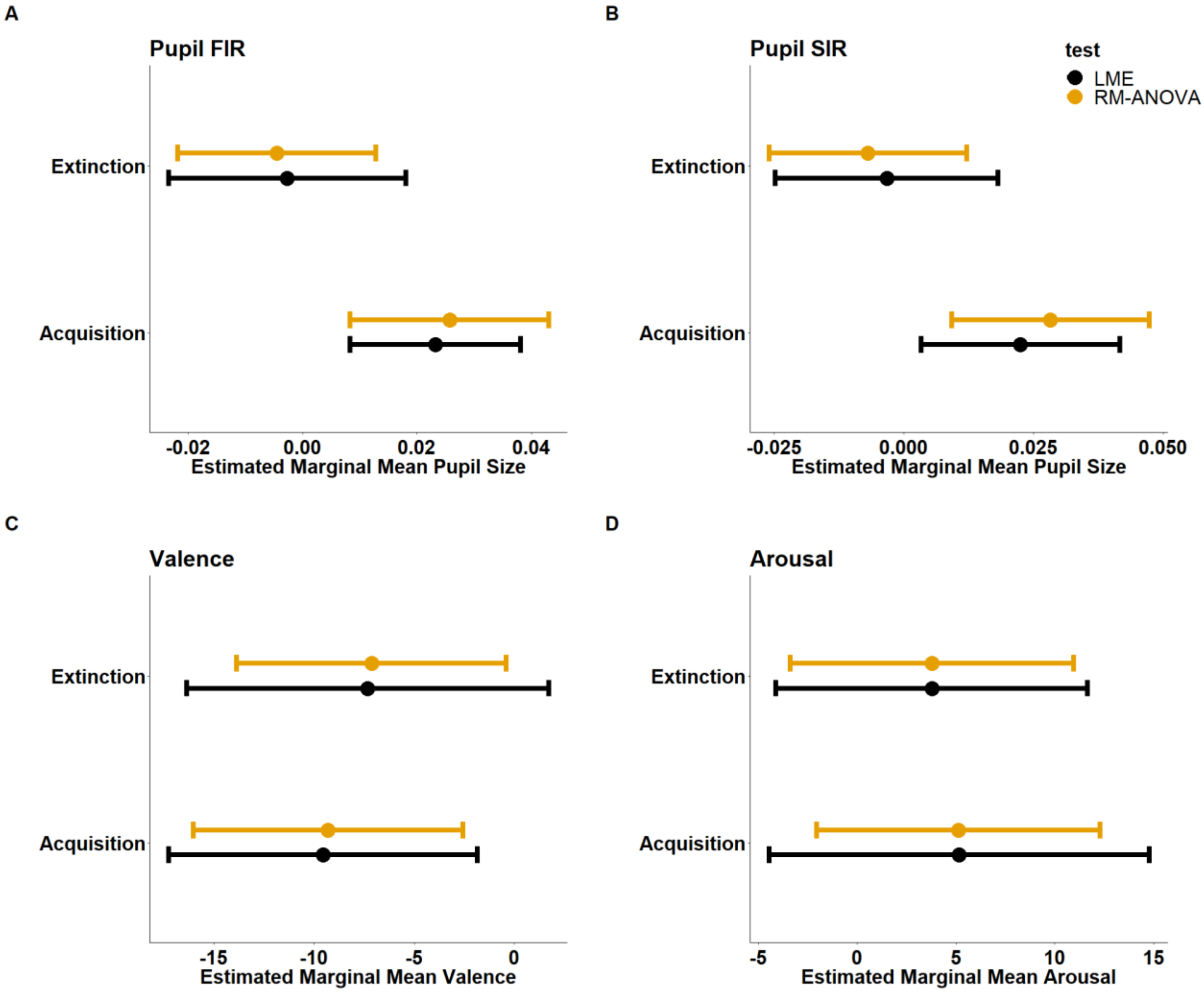
Repeated measures ANOVA and Generalised eta-squared for each effect.

Effect	DFn	DFd	MSE	F	p	ges
Pupil Size FIR						
Experimental Phase	1	24	0.001	3.3	0.08	0.03
Stimulus Type	1	24	0.001	2.4	0.14	0.03
Experimental Phase X Stimulus Type	1	24	0.002	8.5	0.008**	0.05
Pupil Size SIR						
Experimental Phase	1	25	0.002	1.52	0.3	0.01
Stimulus Type	1	25	0.001	1.98	0.2	0.02
Experimental Phase X Stimulus Type	1	25	0.001	9.8	0.005**	0.04
Valence						
Experimental Phase	1	29	79.4	0.769	0.388	0.003
Stimulus Type	1	29	244.9	8.251	0.008**	0.1
Experimental Phase X Stimulus Type	1	29	89.7	0.392	0.536	0.002
Arousal						
Experimental Phase	1	29	88.4	0.039	0.845	0
Stimulus Type	1	29	294.3	2.021	0.166	0.027
Experimental Phase X Stimulus Type	1	29	86.9	0.158	0.694	0.001

Note. The measure of effect size is generalised eta squared (ges).
* p < 0.05, ** p < 0.01, *** p < 0.001

Figure 28

Estimated marginal means and 95 % confidence intervals derived from the repeated-measures ANOVAs and LME models, computed on the pairwise stimulus type contrasts for each experimental phase



4.4 Discussion

The aim of this study was to assess the feasibility of a conditioning paradigm that uses multiple conditioned and unconditioned stimuli and many trials to examine learning and extinction mechanisms. A related objective was to determine the generalisability of results across both participants and items, thus going beyond previous studies in this area, which primarily only considered by-participant variation in their designs and analyses. Our paradigm demonstrated variable success in the detection of associative learning and extinction effects, depending on the outcome measure.

4.4.1 Pupil Data

During Acquisition, the task elicited a clearly measurable conditioned response, as manifested in significantly more dilated pupils for CS Negative than CS Positive stimuli before UCS onset. Our results corroborate previous findings (based on standard conditioning paradigms) whereby the CS+ elicits more pupil dilation than the CS- (Jentsch et al., 2020; Kluge et al., 2011; Korn et al., 2017; Leuchs et al., 2019). It has been argued that pupillary responses during threat conditioning reflect a valence-*unspecific* anticipation of the UCS, as pupil dilation-correlates with anterior cingulate activity have been independently observed during both threat and reward learning tasks (Leuchs et al, 2019). However, by showing larger pupil responses to CSs paired with unpleasant rather than pleasant UCSs, the present study provides direct, within-task evidence for valence-*specific* pupillary anticipation effects.

Importantly, instead of taking an average (or peak) pupil size over the entire trial period, or within pre-defined time bins (García-Palacios et al., 2018; Jentsch et al., 2020; Koenig et al., 2017; Leuchs et al., 2017), the differential pupil CRs we observed held true for time windows that were derived from the data themselves.

The cluster-based permutation results identified two temporal clusters of interest, a first interval (FIR) occurring between around 1.2 and 1.7 seconds, and a second interval (SIR) observed around 3 to 3.4 seconds. These findings were corroborated when mixed effects modelling was performed on the average pupil size within these windows, showing reliably increased pupil dilation for CS Negative than CS Positive trials during Acquisition in both the FIR and SIR. These results partly contrast with previous findings (using pre-defined time windows) which suggested only a *single* CR in pupillometry data, but a dual (FIR, SIR) pattern in skin conductance responses (Jentsch et al., 2020) - with the latter being more comparable to the present findings. It seems plausible that data-driven time window selection yields greater power for detecting multiple CRs in pupillometry data. At the same time, it is unclear to what extent *early* and *late* responses are qualitatively different, and how they compare across physiological measures. In the context of SCR, it has previously been suggested that early CRs (FIR) may reflect novelty responses as well as associative processes, whereas late CRs (SIR) could reflect the acquisition of the CS-UCS contingency and temporal prediction of UCS occurrence (Jentsch et al., 2020). Yet, despite SCR studies consistently reporting multiple CRs within pre-defined time windows, there is little or mixed empirical evidence to support a qualitative distinction between early and late responses. While Jentsch et al. (2020) provided some evidence for distinct underlying processes involved in FIR and SIR (i.e., only the early SCR was susceptible to extinction), they did not directly examine the relationship between the two CRs. In contrast, Pineles et al.(2009) showed that FIR and SIR derived from SCRs are correlated (Pineles et al., 2009). Similarly, in the present study we observed a high correlation (r=0.8) between the pupillary FIR and SIR (see Supplementary Materials 16), suggesting that at least in the context of pupillometry, these responses may not necessarily reflect independent processes.

Similar to Leuchs et al. (2019), we found no significant differences between CS Positive and CS Negative trials during the Extinction phase, indicating that the aversive association established during Acquisition was successfully extinguished. Studies attempting to trace the development of extinction commonly either compare the first and last extinction trials (Dunsmoor et al., 2019; Morriss et al., 2015; Sperl et al., 2018) or arbitrarily bin trials into small groups (Jentsch et al., 2020; Reinhard & Lachnit, 2002). For comparative purposes with previous research, the current dataset was examined descriptively by splitting trials into blocks of 5 (see Supplementary Materials 23), where we found no clear evidence of ‘extinction development’ in the form of gradually fading conditioned responses. However, there was a clear pattern in the development of the conditioned response during Acquisition, which appeared to be strongest from the 6th until the 15th repetition, following which the response appeared to diminish. We conjecture that the pupillary CR had already habituated by the time extinction was assessed, and that a slightly lower number of stimulus repetitions during acquisition may be preventative of such habituation.

It is also plausible that the extinction of the CR, as well as its acquisition, were influenced by type of instruction. In the present study, participants were told that the conditioned stimuli will only be paired with positive and negative sounds during the learning phase without receiving information about the CS-UCS contingency. As a result, participants expected positive and negative UCS delivery during learning but were aware that no UCS would be administered during extinction. Recent studies have demonstrated the differential impact of instruction type on cognitive and affective systems during extinction. In particular, lack of explicit instructions about the removal of the UCS results in a typical pattern of gradual reduction in the conditioned UCS expectancy, SCR and startle over time (Sevenster et al., 2012). However, the conditioned SCR and UCS expectancy diminish immediately if participants are instructed that the UCS will no longer be presented even when the device delivering the UCS (i.e., an electrode) is not removed. In contrast, the startle response diminishes at a slower rate compared to other measures (Sevenster et al., 2012), but physical removal of the UCS electrode facilitates its extinction (Wendt et al., 2020). According to Sevenster et al. (2012), these observations fit within the dual-process framework of fear learning and suggest the involvement of separate cognitive and affective systems. Within this context, the SCR may reflect anticipatory responses driven by the established CS-UCS contingency, while the startle response may be a more automatic, affective response linked to the valence of the CS, elicited through the UCS. (Sevenster et al., 2012). Assuming that the pupillary responses detected in the present study are anticipatory in nature (comparable to the SCR), such an “extinction by instruction” account may explain why we were unable to detect any sign of gradual attenuation of the pupillary conditioned response during the Extinction phase.

More recently, it has been shown that the CS+/CS- discrimination during learning is largest when explicit instructions about the CS-UCS contingency are provided, with the magnitude of the difference being smallest when no explicit contingency instruction is used (Mertens et al., 2020). Since our protocol combined “uninstructed acquisition” with “instructed extinction”, the relatively small effects we observed may not be overly surprising. Nevertheless, it is also worth noting that because we accounted for by-participant and by-item variability of effects simultaneously (see crossed random effects in our mixed effects models) our significance tests were arguably less anticonservative than the more traditional within-subjects ANOVA approach used in previous research. In particular, the latter approach had been shown to increase the risk of false positives when data are from populations in which effects vary not only across participants *but also across stimuli*, which is a very plausible general assumption (see, e.g., Barr et al., 2013; Clark, 1973; Yarkoni, 2020). We will return to this point further below.

Finally, to ensure that repetitively eliciting associative learning and extinction several times does not hamper the conditioning task, we examined the effect of block number on the elicitation of the conditioned response (see Supplementary Materials 23). We found no evidence of block-related learning effects during either acquisition or extinction which suggests that the blocked design of the study does not have a negative impact on the development of associative memories.

4.4.2 Rating Data

Indirect evidence for successful conditioning at the behavioural level was obtained from the contingency awareness task, suggesting that participants were likely aware of the relationship between conditioned and unconditioned stimuli. On the other hand, we found limited evidence in support of a relationship between the different measures of conditioned responding and contingency awareness (see Supplementary Materials 22). Even though a small proportion of participants failed to develop contingency awareness, all data were retained in the analyses to avoid introducing selection bias through exclusions based on arbitrary criteria (Lonsdorf & Merz, 2017).

In terms of evaluative signatures of conditioning and extinction, we found that arousal ratings did not significantly differ across CS Positive and CS Negative trials during either learning or extinction. This is not surprising, since the arousal measurement ('boring' to 'exciting') was valence-unspecific and therefore related to positive and negative stimuli in equal measures. This is consistent with the well-established U-shaped relationship between valence and arousal, in which arousal is high for extremely positive and negative stimuli but remains low for neutral-valence stimuli (Bradley et al., 1992; Bradley & Lang, 2000). In contrast, we observed a main effect in the valence ratings, with CS Negative trials being rated as more unpleasant than CS Positive trials regardless of experimental phase. This suggests that the valence differences elicited during learning did not (or at least not fully) extinguish during extinction training. This observation supports previous studies (Luck & Lipp, 2015a, 2015b; Wendt et al., 2020) which suggested the presence of independent processing systems where valence may reflect an evaluative process that is more difficult to extinguish, even when explicit extinction-supporting instructions are provided. Specifically, online conditioned valence ratings show resistance to extinction at the beginning of extinction training regardless of instruction type and even following removal of physical threat (Luck & Lipp, 2015a, 2015b). When ratings are obtained offline and regardless of instruction type, CS+ trials are still perceived as less pleasant, although there is some evidence to suggest that this resistance may be more prominent when no instructions are delivered (Wendt et al., 2020).

Across the literature, mixed findings have been reported when using subjective valence and arousal as measures of the conditioned response. Standard paradigms assessing only one CS+ and one CS-, typically demonstrated a reduction in valence and an increase in arousal for CS+ trials during learning (Gawronski & Mitchell, 2014; Reinhardt et al., 2010; Sehlmeyer et al., 2011). In contrast, Multi-CS conditioning paradigms in which many different stimuli are used, elicit much smaller and less consistent effects (Bröckelmann et al., 2011; Junghöfer et al., 2015a; Rehbein et al., 2014; Steinberg et al., 2013) which is more consistent with the findings of the present study. Indeed, inferences about associative memory mechanisms may not be very robust and generalisable when based on only two stimuli. On the other hand, when many different stimuli must be evaluated and/or contingency awareness is poor, subjective metacognitive judgements of affect may not be sufficiently powerful to detect the presence of learning and extinction.

Like the pupil size data, the valence and arousal effects observed in the present study were small, but comparable to previous findings based on larger trial numbers and varying stimuli. In particular, a reanalysis of the valence ratings from a study by Rehbein et al. (2014), who used a Multi-CS conditioning task to elicit learning, revealed an effect size for the *Experimental Phase X Stimulus Type* interaction that was comparable to that in the present study, even when a traditional repeated-measures ANOVA on aggregated data was performed (see Supplementary Materials 24).

4.4.3 Methodological considerations and conclusions

The present study offers a potentially more generalisable alternative to the hitherto available associative learning paradigms in the literature. In line with recent reports using a similar task (Sperl et al., 2021), we demonstrated that a blocked design in which learning and extinction are established several times, may allow for the use of a greater range of CSs and UCSs while still allowing for the development of a contingency awareness that seems crucial for the development of a CR (Mertens & Engelhard, 2020). Since there was no evidence that the blocked nature of the paradigm would negatively influence the development of a conditioned response, the task can theoretically also be applied to other noisy measurement modalities (such as M/EEG) and expanded to include a greater range of stimuli and number of trials.

In line with other investigations emphasising the necessity of modelling random variation across both subjects and items (Barr et al., 2013; Judd et al., 2012, 2017; Westfall et al., 2017; Yarkoni, 2020), the present study also indicated that conventional analysis methods, when applied to our design, may cause inflation of the test statistic. While it is reassuring that in our data, employing an analysis on aggregated data did not change the overall interpretation of the results, these analyses highlight the possible risk of making anti-conservative inferences at the expense of more generalisable conclusions. Furthermore, while the effects reported here appear weaker than those typically found in the threat conditioning literature, the present study offers a much greater degree of generalisability based on (a) data-driven identification of time windows of interest and (b) statistical modelling that takes both by-participant and by-item variation of effects into account.

We show that both early and late pupillary conditioned responses can be detected during acquisition, and that these vanish almost immediately during extinction. We show that early vs. late interval responses may not be reflecting independent processes as they are highly correlated. Yet, future work should aim at establishing the degree to which these findings can generalise across other physiological outcome measures. While our descriptive analysis suggests that habituation may have already developed during learning, it is possible that the type of instructions delivered to participants for the extinction phase also contributed to the quick onset of CR habituation. Future work should focus on establishing precisely how instructed versus uninstructed protocols affect the acquisition and extinction of conditioning using many conditioned and unconditioned stimuli. In addition, the paradigm may potentially benefit from a slight reduction in the number of stimulus repetitions, to avoid habituation of the CR prior to extinction. Finally, we show that our subjective measure of arousal was not sensitive enough to detect associative learning processes while subjective valence may potentially be less susceptible to extinction. Overall, our findings confirm previous suggestions (Lonsdorf et al., 2017; Sevenster et al., 2012) that different outcome measures may tap onto different processes related to associative learning and extinction.

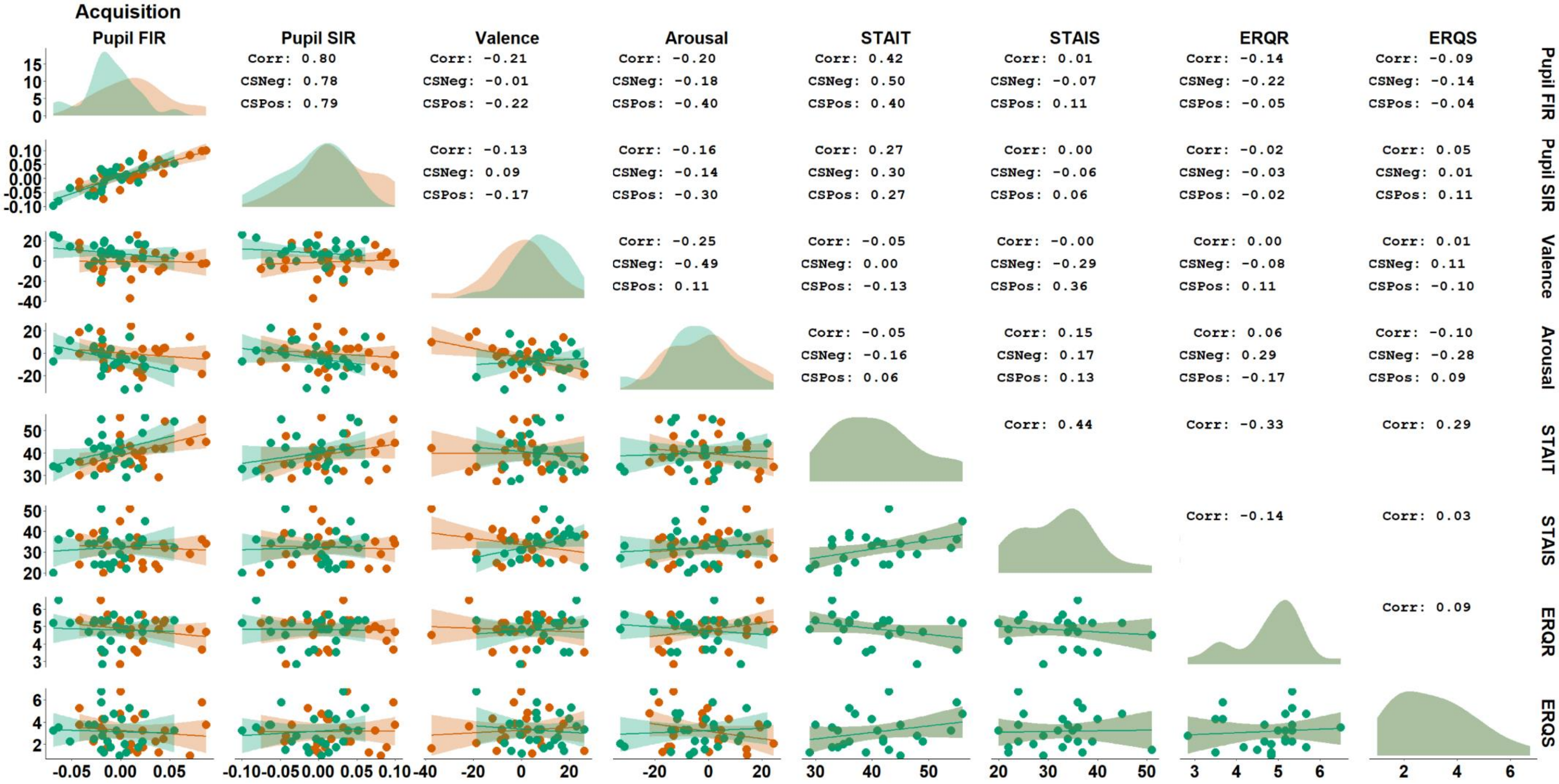
4.5 Chapter 4 Supplementary Materials

4.5.1 Supplementary Materials 16: Relationship between psychological and conditioning measures

The relationship between pupil size, valence and arousal during acquisition and extinction and self-reported anxiety and emotion regulation can be seen in Supplementary Figure 44 and 45 in the form of scatterplots and Pearson's r correlations. Overall, relationships between measures were low to moderate, but no correlations between CR and psychological measures were significant following Holm's multiple comparison corrections (see Supplementary Table19 and 20). There was, however, a positive correlation between FIR and SIR pupillary responses.

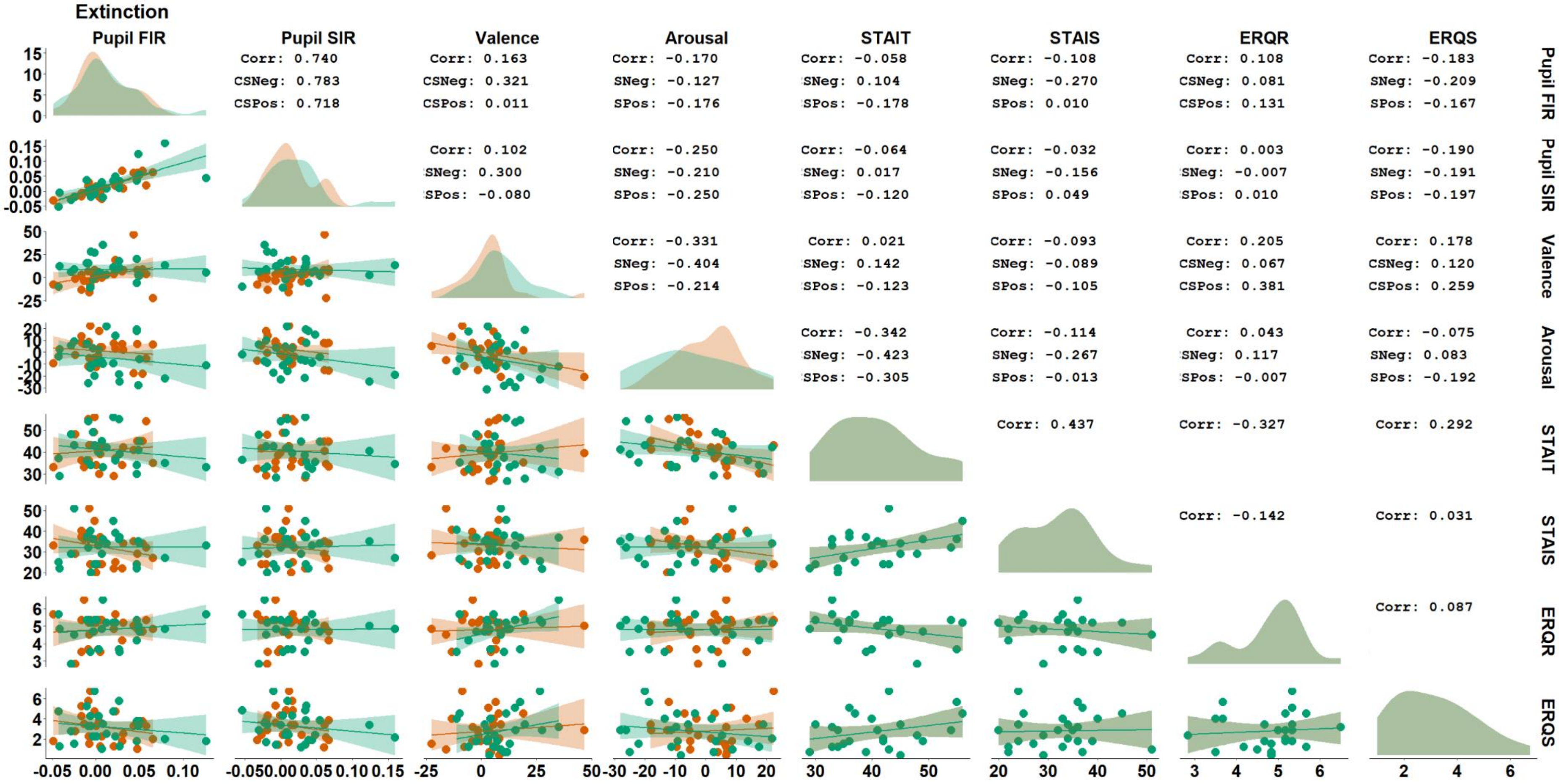
Supplementary Figure 44

Relationship between conditioning effects and measures of state and trait anxiety



Supplementary Figure 45

Relationship between extinction effects and measures of state and trait anxiety



Supplementary Table 19

Correlation coefficients and p-values within the Acquisition Phase.

CS Negative

Pearson's r

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	1.0	0.8	0.0	-0.2	-0.1	0.5	-0.2	-0.1
Pupil SIR	0.8	1.0	0.1	-0.1	-0.1	0.3	0.0	0.0
Valence	0.0	0.1	1.0	-0.5	-0.3	0.0	-0.1	0.1
Arousal	-0.2	-0.1	-0.5	1.0	0.2	-0.2	0.3	-0.3
STAIS	-0.1	-0.1	-0.3	0.2	1.0	0.4	-0.1	0.0
STAIT	0.5	0.3	0.0	-0.2	0.4	1.0	-0.3	0.3
ERQR	-0.2	0.0	-0.1	0.3	-0.1	-0.3	1.0	0.1
ERQS	-0.1	0.0	0.1	-0.3	0.0	0.3	0.1	1.0

P-value

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	0.00	0.00	1.00	1.00	1.00	0.33	1.00	1.00
Pupil SIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Valence	0.95	0.68	0.00	0.36	1.00	1.00	1.00	1.00
Arousal	0.39	0.49	0.01	0.00	1.00	1.00	1.00	1.00
STAIS	0.73	0.78	0.16	0.41	0.00	0.82	1.00	1.00
STAIT	0.01	0.16	0.99	0.44	0.03	0.00	1.00	1.00
ERQR	0.00	0.00	1.00	1.00	1.00	0.33	1.00	1.00
ERQS	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00

CS Positive

Pearson's r

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	1.0	0.8	-0.2	-0.4	0.1	0.4	-0.1	0.0
Pupil SIR	0.8	1.0	-0.2	-0.3	0.1	0.3	0.0	0.1
Valence	-0.2	-0.2	1.0	0.1	0.4	-0.1	0.1	-0.1
Arousal	-0.4	-0.3	0.1	1.0	0.1	0.1	-0.2	0.1
STAIS	0.1	0.1	0.4	0.1	1.0	0.4	-0.1	0.0
STAIT	0.4	0.3	-0.1	0.1	0.4	1.0	-0.3	0.3
ERQR	-0.1	0.0	0.1	-0.2	-0.1	-0.3	1.0	0.1
ERQS	0.0	0.1	-0.1	0.1	0.0	0.3	0.1	1.0

P-value

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Pupil SIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Valence	0.30	0.42	0.00	1.00	1.00	1.00	1.00	1.00
Arousal	0.05	0.15	0.61	0.00	1.00	1.00	1.00	1.00
STAIS	0.61	0.76	0.08	0.54	0.00	0.89	1.00	1.00
STAIT	0.06	0.20	0.53	0.76	0.03	0.00	1.00	1.00
ERQR	0.81	0.93	0.60	0.43	0.50	0.12	0.00	1.00
ERQS	0.83	0.60	0.62	0.68	0.88	0.17	0.68	0.00

Note. Multiple comparisons adjustment was performed for each condition separately.

Supplementary Table 20

Correlation coefficients and p-values within the Acquisition Phase.

CS Negative

Pearson's r

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	1.0	0.8	0.3	-0.1	-0.3	0.1	0.1	-0.2
Pupil SIR	0.8	1.0	0.3	-0.2	-0.2	0.0	0.0	-0.2
Valence	0.3	0.3	1.0	-0.4	-0.1	0.1	0.1	0.1
Arousal	-0.1	-0.2	-0.4	1.0	-0.3	-0.4	0.1	0.1
STAIS	-0.3	-0.2	-0.1	-0.3	1.0	0.4	-0.1	0.0
STAIT	0.1	0.0	0.1	-0.4	0.4	1.0	-0.3	0.3
ERQR	0.1	0.0	0.1	0.1	-0.1	-0.3	1.0	0.1
ERQS	-0.2	-0.2	0.1	0.1	0.0	0.3	0.1	1.0

P-value

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Pupil SIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Valence	0.12	0.14	0.00	1.00	1.00	1.00	1.00	1.00
Arousal	0.54	0.31	0.05	0.00	1.00	1.00	1.00	1.00
STAIS	0.19	0.46	0.67	0.20	0.00	0.89	1.00	1.00
STAIT	0.63	0.94	0.51	0.04	0.03	0.00	1.00	1.00
ERQR	0.70	0.97	0.75	0.58	0.50	0.12	0.00	1.00
ERQS	0.32	0.36	0.57	0.69	0.88	0.17	0.68	0.00

CS Positive

Pearson's r

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	1.0	0.7	0.0	-0.2	0.0	-0.2	0.1	-0.2
Pupil SIR	0.7	1.0	-0.1	-0.2	0.0	-0.1	0.0	-0.2
Valence	0.0	-0.1	1.0	-0.2	-0.1	-0.1	0.4	0.3
Arousal	-0.2	-0.2	-0.2	1.0	0.0	-0.3	0.0	-0.2
STAIS	0.0	0.0	-0.1	0.0	1.0	0.4	-0.1	0.0
STAIT	-0.2	-0.1	-0.1	-0.3	0.4	1.0	-0.3	0.3
ERQR	0.1	0.0	0.4	0.0	-0.1	-0.3	1.0	0.1
ERQS	-0.2	-0.2	0.3	-0.2	0.0	0.3	0.1	1.0

P-value

	Pupil FIR	Pupil SIR	Valence	Arousal	STAIS	STAIT	ERQR	ERQS
Pupil FIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Pupil SIR	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Valence	0.96	0.70	0.00	1.00	1.00	1.00	1.00	1.00
Arousal	0.40	0.23	0.30	0.00	1.00	1.00	1.00	1.00
STAIS	0.96	0.82	0.62	0.95	0.00	0.89	1.00	1.00
STAIT	0.40	0.58	0.57	0.15	0.03	0.00	1.00	1.00
ERQR	0.53	0.96	0.06	0.97	0.50	0.12	0.00	1.00
ERQS	0.43	0.35	0.21	0.36	0.88	0.17	0.68	0.00

4.5.2 Supplementary Materials 17: Auditory Control Experiment 2

The main goal of this study was to derive normative data for positive and negative sounds of low and high-volume intensity, to allow the selection of UCSs for the auditory blocked conditioning task. A secondary goal was to explore the relationship between stimulus valence and sound intensity in predicting ratings of valence and arousal. since exploratory analyses in the first auditory control experiment found that at short duration (0.1 s) and high intensity (~90 dB), positive and negative sounds do not differ in valence, arousal, or pupil size. To test whether intensity alone is sufficient to reduce valence of positive stimuli, in Experiment 2, sounds were played both at low (60 dB) and high (90 dB) intensity level, however at a longer duration (1 s). We hypothesised that if high sound intensity alone causes positive sounds to be perceived as unpleasant and arousing as negative sounds, the same pattern should be observed for sounds with longer duration (1 s), presented at high intensity. Consequently, positive, and negative sounds of long duration but low intensity should differ at least in valence ratings.

Methods

Participants

Thirty-one adults aged between 18 and 30 will take part in the study. One subject was discarded from the analyses as they continuously removed their headphones during the task.

Stimuli

A total of twenty auditory stimuli were used in the control experiment. Stimuli were comprised of environmental, human and animal sounds (10 negative, 10 positive), such as female scream and metal scrapes as well as bird chirping and bubbles, which have previously been shown to elicit negative and positive valence respectively (Kumar et al., 2008). These stimuli were obtained from different online databases (Freesfx, Freesound, Free sound effects, the CNBC Stimuli Repository and IADS-2/IADS-2E). All sounds had a duration of 1 sec. To equalise the intensity of sounds, each stimulus was mean centred and then normalised to the same, maximum root mean square (RMS) amplitude without inducing clipping using a RMS equaliser (The Phonetics Lab, University of

Washington, https://depts.washington.edu/phonlab/resources/rmsLeveler.m). The first and last 50 ms of the signal of all stimuli were gradually faded in and out. The resulting normalised sounds were presented at 2 different maximum intensity levels (~60 and 90 dBA) as measured by Cadrim sound level meter.

Procedure

The task contained 2 blocks (for each intensity level) with a total of 60 trials in each block. Block order was counterbalanced across participants. Each sound was presented 3 times in a random order, with the restriction that no sound was presented twice in succession. The auditory stimuli were administered through Sennheiser HD-202 headphones. On each trial, a black fixation cross was presented on a gray background for 500 ms followed by the sound (1 s duration), with an inter-trial interval (ITI) of 1300 ms ±300 ms, comprising of a black fixation cross on a gray background. Following each presentation, participants were required to rate each sound on valence and arousal using a slider scale ranging from boring/unpleasant to exciting/pleasant.

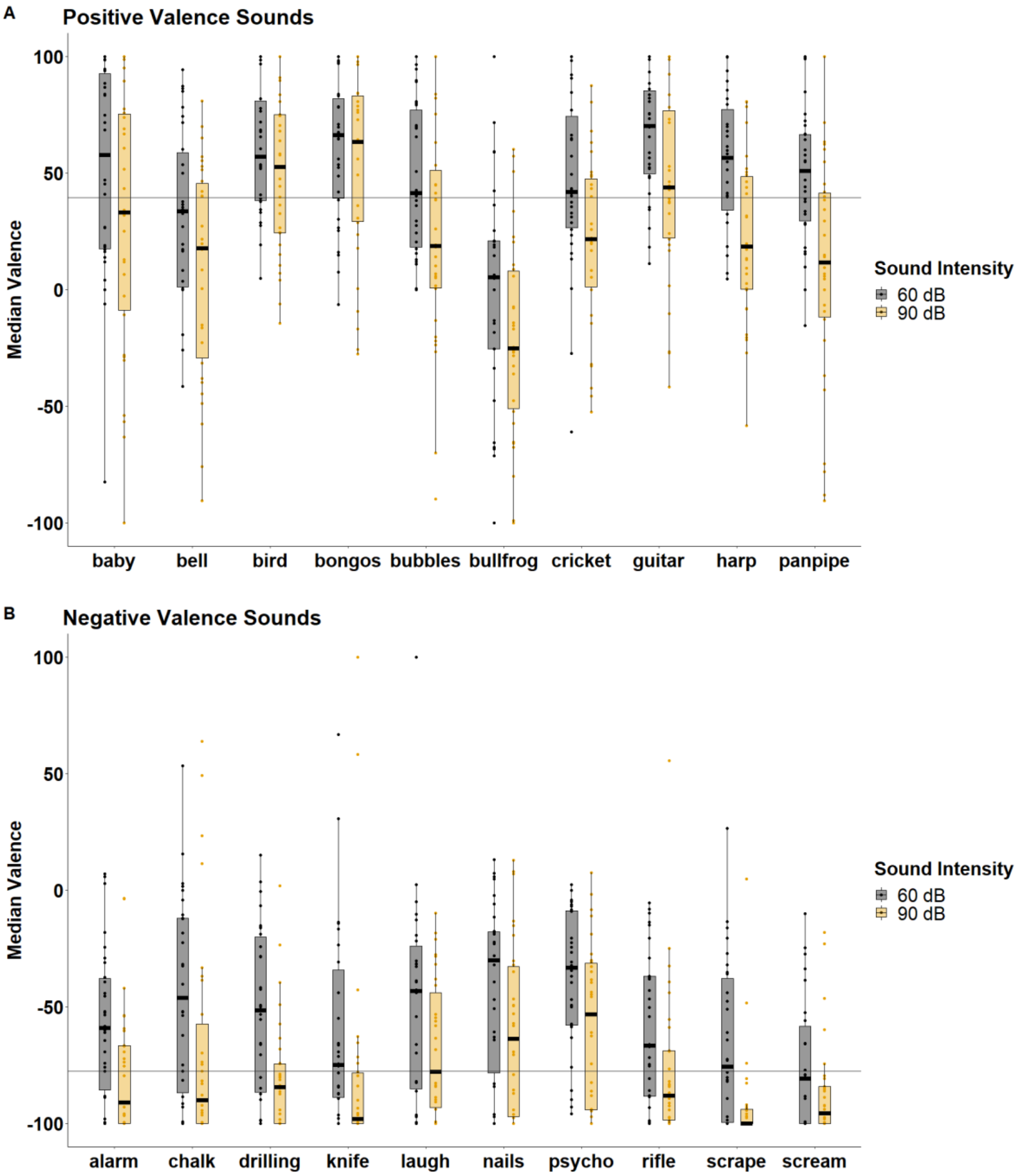
Results

Selection of unconditioned stimuli

For each participant, a median valence was calculated for each item and intensity level based on which the unconditioned stimuli were selected. Arousal ratings were not used in the selection process since pleasant stimuli could be perceived as both arousing as well as boring/calming (see Supplementary Figure 46 and 47). To ensure maximum difference between the pleasant and unpleasant stimuli, the selection of positive sounds was based on their valence ratings at 60 dB while of that of negative sounds was based on their valence at 90 dB. Specifically, for positive sounds, the four sounds in the highest 25th percentile were chosen. As seen in Figure 46A those were bongos, guitar, harp and a bird chirping. For, negative sounds the 4 sounds with the lowest 75th percentile valence were selected as negative unconditioned stimuli (knife, scrape, drilling, and a scream, see Supplementary Figure 46B). In terms of arousal ratings of the selected sounds, while negative unconditioned stimuli were also rated as highly arousing, positive stimuli were rated as arousing or neutral (see Supplementary Figure 47). Finally, the distribution of valence and arousal ratings for the selected stimuli appear to remain similar across item repetitions, suggesting lack of rapid habituation of perceived valence (see Supplementary Figure 48).

Supplementary Figure 46

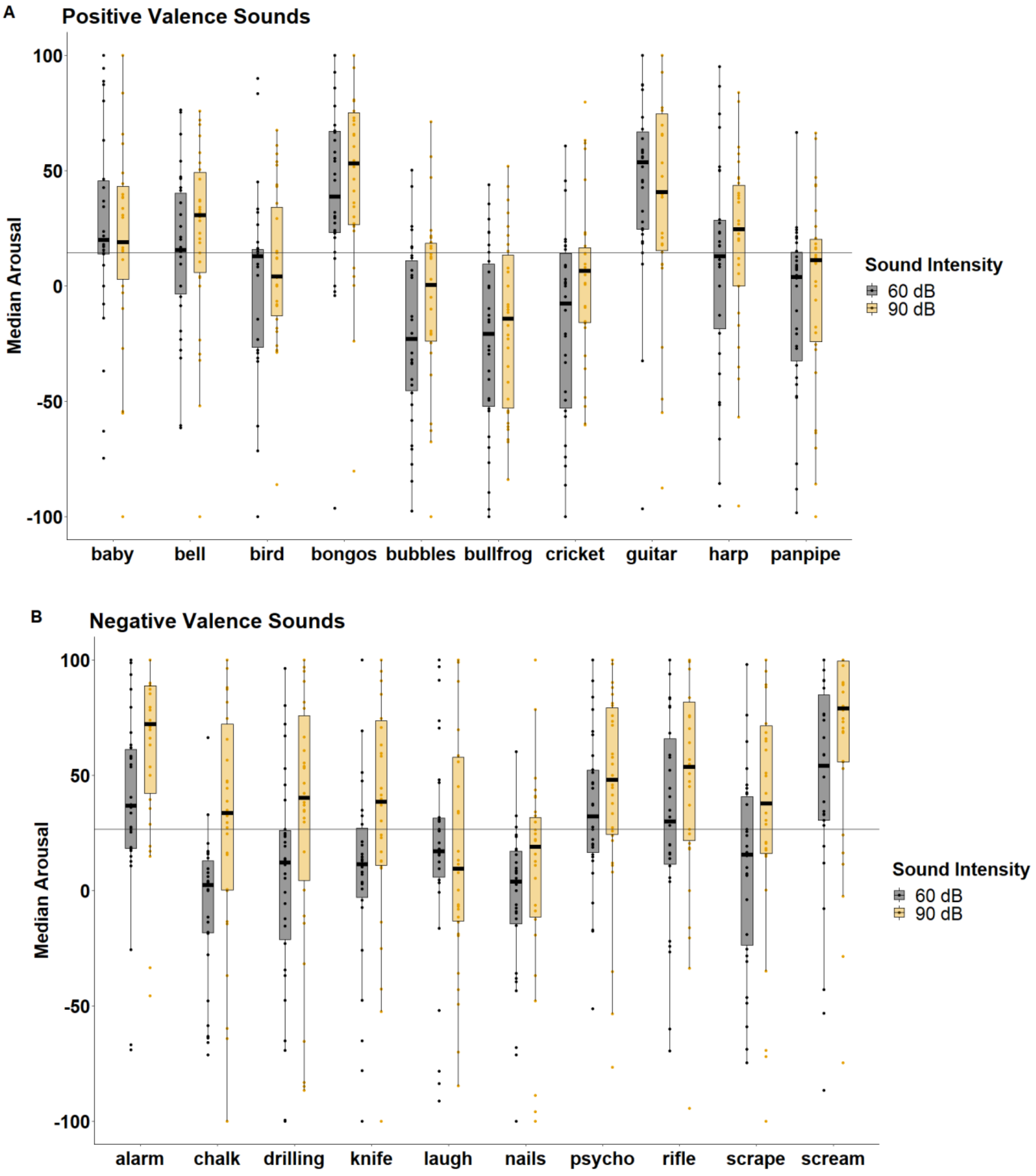
Distribution of median valence for individual positive and negative sounds rated at low and high intensity.



Note. The horizontal line indicates the overall median valence across items and intensity levels.

Supplementary Figure 47

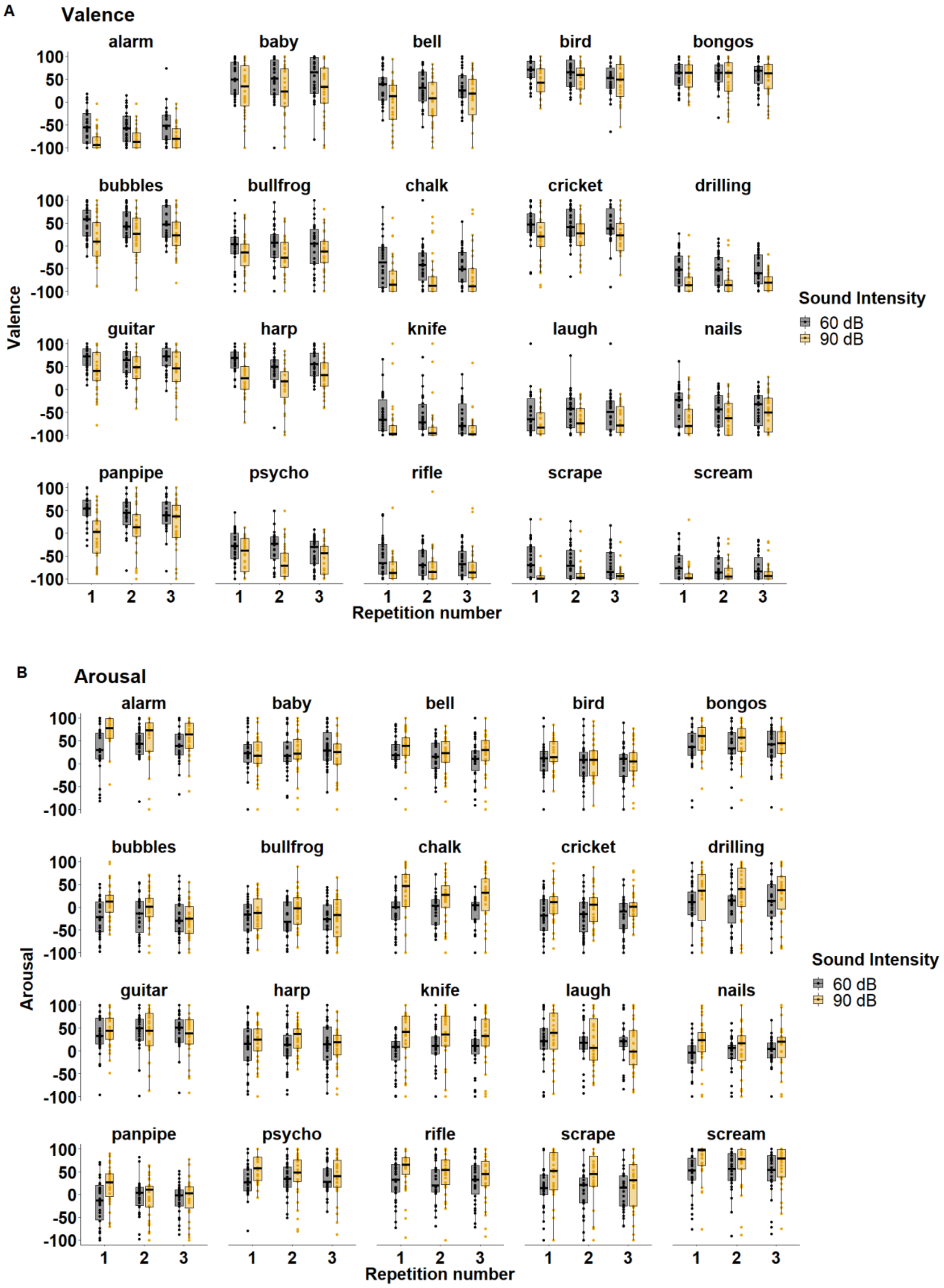
Distribution of median arousal for individual positive and negative sounds rated at low and high intensity.



Note. The horizontal line indicates the overall median arousal across items and intensity levels.

Supplementary Figure 48

Valence and arousal ratings across repetitions.



Relationship between stimulus valence and sound intensity

The results from the first auditory control experiment showed no differences in valence ratings between positive and negative sounds played at high intensity (~ 90 dB) for a brief period (0.1 s). These findings suggested that at a high volume and short duration, the valence of positive sounds may become as low as that of negative sounds. To test whether intensity alone is sufficient to reduce valence, in the second auditory control study, sounds were played both at low (60 dB) and high (90 dB) intensity level, however at a longer duration (1 s). Descriptively, as shown in Figure 49, the difference in median valence ratings appears to be driven to a greater extend by sound valence (positive/negative) than sound intensity (low/high). In contrast, sound valence and intensity do not elicit strong median differences in arousal ratings.

In order to quantify these differences, an LME model (package *lme4*) was built for valence and arousal separately. Each model comprised of a 2 *Stimulus Valence* (Positive vs Negative) by 2 *Sound Intensity* (Low vs High) fixed effects design. The model included mean-centred contrasts (deviation coding) for the two categorical fixed effects. A maximal model was fitted following guidelines by (Barr et al., 2013). Specifically, *Subjects* and *Items* were added as random intercepts. By-subject random slope was added for both main effects and a by-item random slope was added for sound intensity only (stimulus valence was not included as an item can only be positive or negative). Main effects and interactions were tested using type III Wald chi-square tests. Post-hoc analyses of main effects/interactions were performed using estimated marginal means contrasts (package *emmeans*) with Satterthwaite method for degrees of freedom approximation.

In terms of the valence model, a type III Wald chi-square test (see Supplementary Table 21 and Figure 50 for model estimates) revealed significant main effects at the level of $p < 0.05$ for stimulus valence and sound intensity but the interaction was non-significant. Post-hoc contrasts showed that sounds were rated as lower in valence if they were negative compared to positive (t (18) = 13.5, $p < 0.0001$). Sounds were also rated as more unpleasant when they were presented at high than low intensity (t (18) = 12.65, $p < 0.001$), (see Table 22 and Figure 52). As seen from the model estimates and R^2 values in Supplementary Table 22, most of the variance of the fixed effects in the model is explained by stimulus valence, and only ~6% by differences in sound intensity.

For arousal, the type III Wald chi-square test revealed a significant *Stimulus Valence* X *Sound Intensity* interaction at the level of 0. 05. (see Supplementary Table 21 and Figure 50 for model estimates). Post-hoc planned contrasts (see Supplementary Table 23) showed that sounds of high intensity were rated as more arousing than those of low intensity for sounds for both negative (t (27.1) = -6.58, p < 0.001) and positive sounds (t (27.1) = -2.9, p =0.03). At 90 dB, negative sounds were also rated as more arousing than positive sounds (t (19.3) =2.5, p < 0.02) but there was so no significant difference for sounds presented at 60 dB (t (18.9) = 1.06, p < 0.302. However, according to Wagenmakers et al (2012) this type of interaction which lacks a crossover effect, may be unstable and influenced by transformations of the measurement scale. Therefore, the observed effects may not reflect the underlying construct of arousal and should be interpreted with caution.

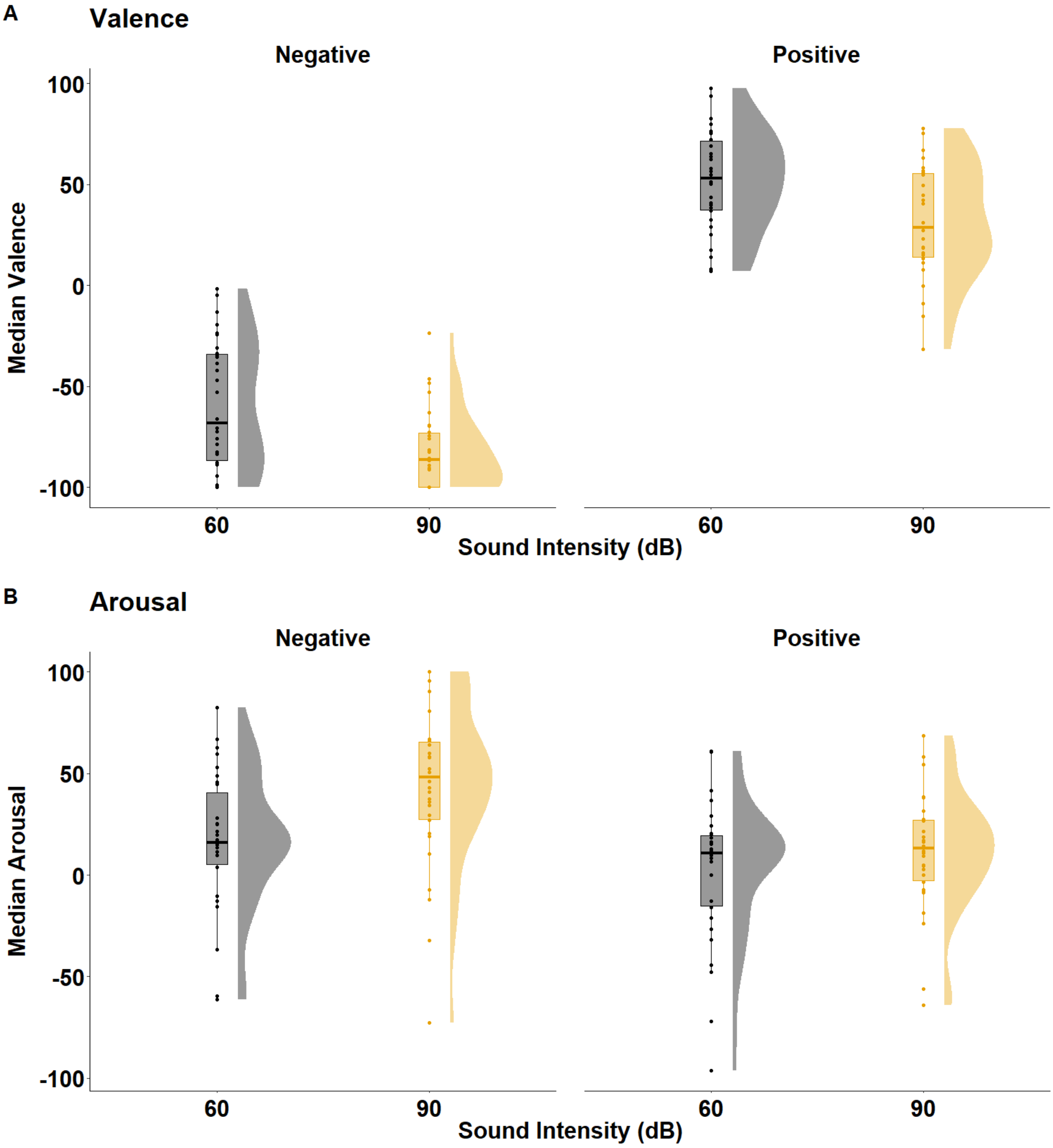
Summary of Results

In the first auditory control study, we found no valence and arousal differences for sounds of short duration and high intensity while the present study showed that when sounds are presented at a longer duration, valence and arousal differences between positive and negative stimuli can be observed. This suggests that short trial duration may prevent the identification of sounds and their valence.

In addition, we found that compared to low intensity sounds, high intensity sounds were also perceived as more unpleasant and arousing but sound intensity and stimulus valence did not interact in predicting subjective ratings. These findings suggest that the stimulus valence and its intensity independently influence evaluative judgements. Furthermore, when sounds are presented for long enough to be clearly distinguished, it appears that the effect of stimulus valence is more pronounced that of volume intensity. Therefore, sound duration may be a key factor to consider when designing auditory experiments that attempt to elicit response changes using positive and negative stimuli.

Supplementary Figure 49

Distribution of valence and arousal ratings across positive and negative sounds rated at low and high intensity level.



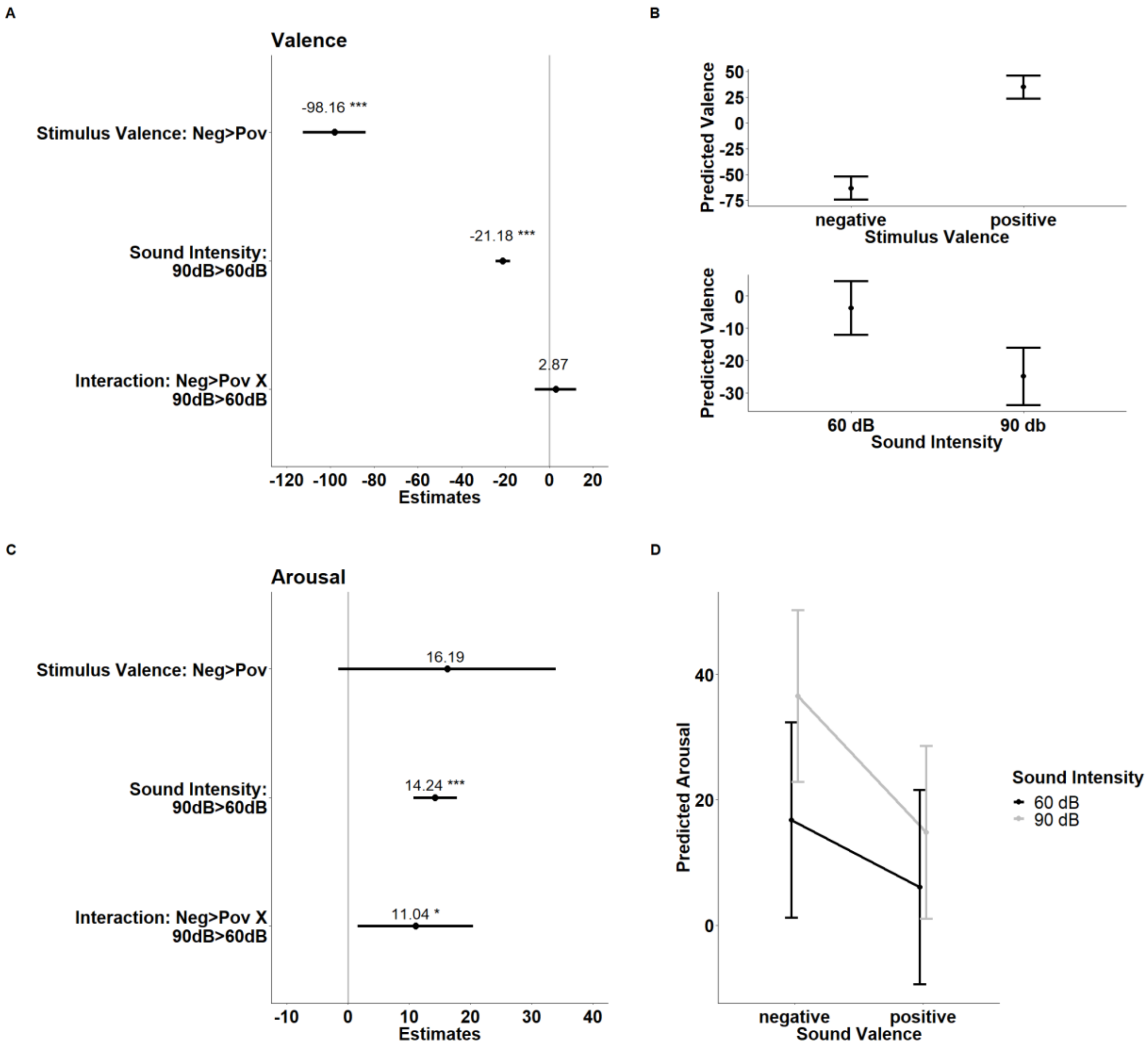
Supplementary Table 21

Type III ANOVA and R-squared values for the complete model and each of the fixed effects.

	Chisq	Df	P-value	R^2 Fixed (CI)
Valence				
Full Model (Fixed)				0.586 (0.57-0.6)
Stimulus Valence	181.102	1	0.000	0.575 (0.56 – 0.59)
Sound Intensity	160.085	1	0.000	0.059 (0.045-0.075)
Stimulus Valence X Sound Intensity	0.362	1	0.547	0.000 (0 – 0.002)
Arousal				
Full Model (Fixed)				0.047 (0.03-0.06)
Stimulus Valence	3.206	1	0.073	0.026 (0.02-0.04)
Sound Intensity	62.619	1	0.000	0.020 (0.012 – 0.03)
Stimulus Valence X Sound Intensity	5.265	1	0.022	0.003 (0.001 – 0.008)

Supplementary Figure 50

Fixed effects estimates and estimated marginal means.



Note. A) Fixed effect estimates derived from the pupil model. B) Estimated marginal mean pupil size derived from the pupil model.

Supplementary Table 22

Estimated marginal means and related contrasts derived for the valence model.

Main effect	Emmean	SE	df	Lower CI	Upper CI	
Estimated Marginal Means						
Stimulus Valence: Negative	-63.461	5.648	25.182	-75.090	-51.833	
Stimulus Valence: Positive	34.695	5.648	25.182	23.066	46.323	
Sound Intensity: Low	-3.792	4.234	32.255	-12.414	4.830	
Sound Intensity: High	-24.975	4.548	30.021	-34.262	-15.688	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P value
Negative – Positive	-98.156	7.294	18.000	-113.480	-82.832	0.000***
60 dB – 90 dB	21.183	1.674	18.000	17.666	24.701	0.000***

Notes. * p<0.05, ** p<0.01 *** p <0.001

Supplementary Table 23

Estimated marginal means and related contrasts derived for the arousal model.

Estimated Marginal Means

Stimulus Valence	Sound Intensity	Emmean	SE	df	Lower CI	Upper CI
Negative	60 dB	16.716	7.944	27.898	0.442	32.990
Positive	60 dB	6.040	7.886	27.274	-10.134	22.214
Negative	90 dB	36.479	6.968	30.471	22.258	50.699
Positive	90 dB	14.766	7.032	31.257	0.429	29.104

Contrasts

Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
negative,60 dB - positive,60 dB	10.676	10.061	18.920	-10.387	31.740	0.302
negative,90 dB - positive,90 dB	21.712	8.598	19.264	3.733	39.691	0.020*
negative,60 dB - negative,90 dB	-19.762	3.004	27.083	-25.926	-13.599	0.000***
positive,60 dB - positive,90 dB	-8.726	3.004	27.083	-14.890	-2.563	0.007**

Notes. * p<0.05, ** p<0.01 *** p <0.001

4.5.3 Supplementary Materials 18: Counterbalanced sets for pairing CS and UCS items.

Supplementary Table 24

Counterbalanced sets for pairing low and high frequency tones to positive and negative sounds.

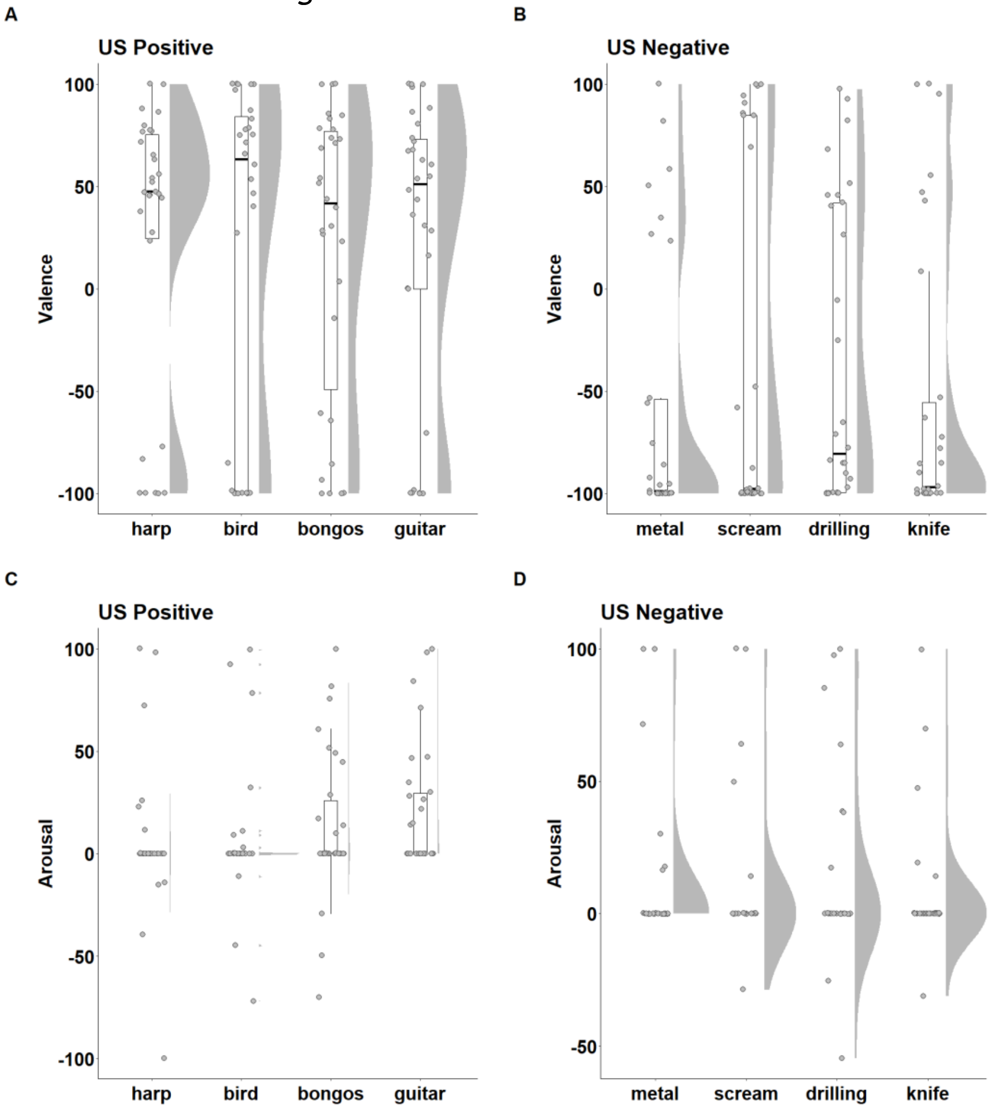
Participant/ Stimulus Set	Block 1	Block 2	Block 3	Block 4	Tone pair
1	low	low	high	high	Paired with UCS Negative
2	low	low	high	high	Paired with UCS Positive
3	low	high	high	low	Paired with UCS Negative
4	low	high	high	low	Paired with UCS Positive
5	high	high	low	low	Paired with UCS Negative
6	high	high	low	low	Paired with UCS Positive
7	high	low	low	high	Paired with UCS Negative
8	high	low	low	high	Paired with UCS Positive
… n = 30					

4.5.4 Supplementary Materials 19: Valence and arousal ratings of unconditioned stimuli

Supplementary Figure 51 show the valence and arousal ratings for each of the positive and negative UCSs. As the figure suggests, the positive sounds were consistently rated as pleasant whereas negative sounds were rated as unpleasant. In contrasts, arousal ratings show little variability, with the bulk of responses being centred around 0, suggesting that the sounds were not perceived as arousing.

Supplementary Figure 51

Valence and Arousal ratings to the unconditioned stimuli



Note. A) and B) Valence ratings for positive and negative sounds, C) and D) Arousal Ratings for positive and negative sounds

4.5.5 Supplementary Materials 20: Random effects and estimated marginal means summaries derived from mixed models

Supplementary Table 25

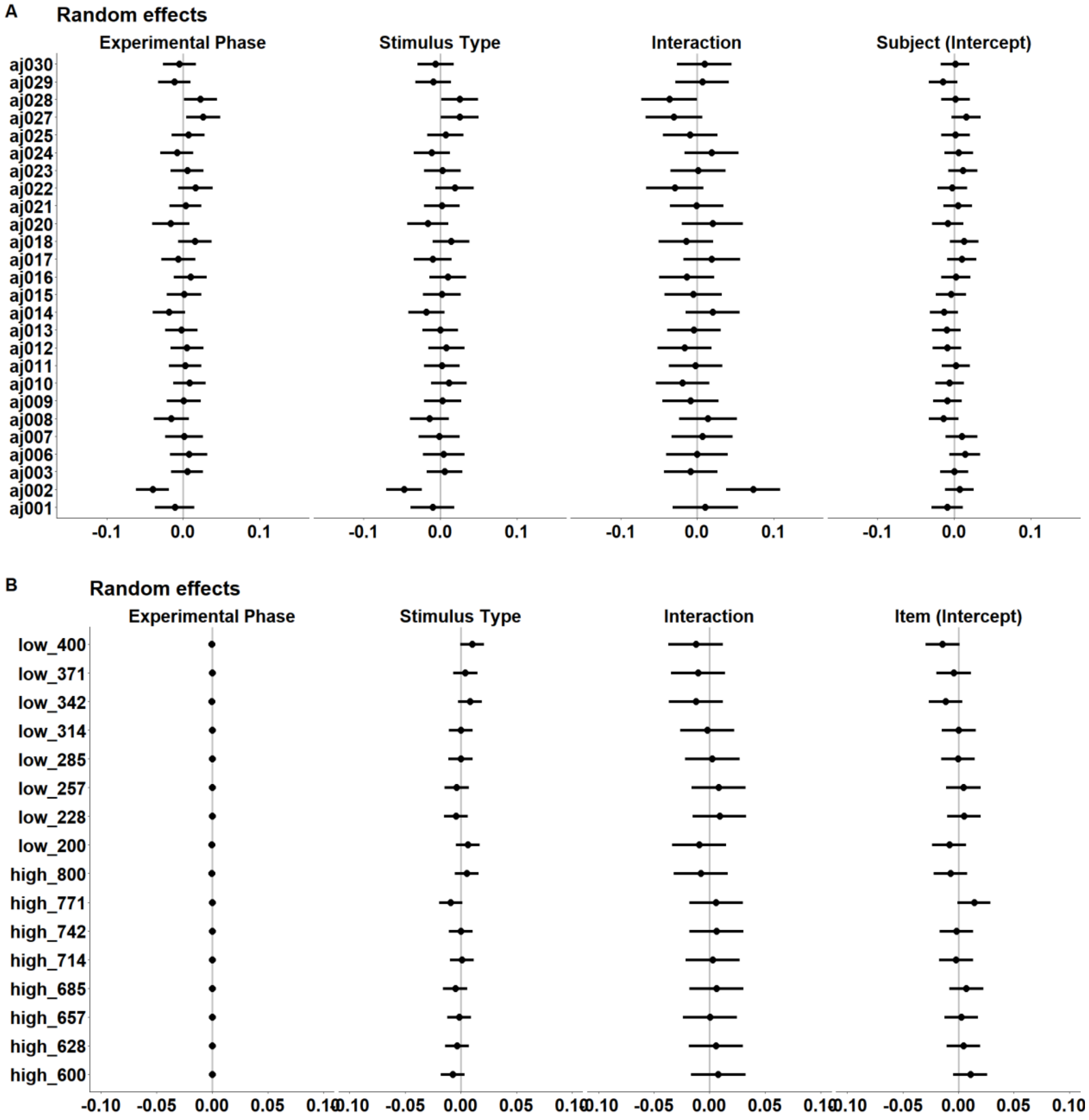
Summary of fixed estimates and random effect variance for the FIR pupil model

FIR	Mean Pupil Dilation		
Predictors	Estimates	CI	p
(Intercept)	0.01	-0.00 – 0.02	0.183
Experimental Phase	-0.01	-0.02 – 0.00	0.130
Stimulus Type	0.01	-0.00 – 0.02	0.147
Interaction	0.03	0.01 – 0.05	**0.014**
Random Effects			
σ^2	0.0269		
τ_{00} Subject	0.0000		
τ_{00} Item	0.0002		
τ_{00} Subject: Item	0.0001		
τ_{11} Subject: Item: Phase	0.0002		
τ_{11} Subject: Item: Stimulus Type	0.0064		
τ_{11} Subject: Item: Interaction	0.0008		
τ_{11} Subject: Phase	0.0003		
τ_{11} Subject: Stimulus Type	0.0003		
τ_{11} Subject: Interaction	0.0007		
τ_{11} Item: Phase	0.0000		
τ_{11} Item: Stimulus Type	0.0001		
τ_{11} Item: Interaction	0.0002		
N Subject	25		
N Item	16		
Observations	6749		
Marginal R^2	0.003		

Note. σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 52

By-subject and by-item random coefficients and intercept for the FIR pupil size model.



Note. A) By-subject random effects and B) By-item random effects.

Supplementary Table 26

Estimated marginal means and related contrasts derived from the mixed effects models of pupil size for A) first interval responses (FIR) and B) second interval response (SIR).

A) FIR

Estimated Marginal Means

Stimulus Type	Experimental Phase	Emmean	SE	df	Lower CI	Upper CI
CS Negative	Acquisition	0.014	0.007	21.76	-0.001	0.028
CS Positive	Acquisition	-0.009	0.007	21.00	-0.023	0.005
CS Negative	Extinction	0.009	0.006	22.80	-0.003	0.021
CS Positive	Extinction	0.012	0.008	20.92	-0.006	0.029

Contrasts

Contrast	Estimate	SE	df	Lower CI	Upper CI	P-values
CS Negative Acquisition - CS Positive Acquisition	0.023	0.007	56.595	0.008	0.038	0.003**
CS Negative Extinction - CS Positive Extinction	-0.003	0.010	22.199	-0.023	0.018	0.793
CS Negative Extinction - CS Positive Extinction	0.005	0.009	28.065	-0.008	0.027	0.438
CS Positive Acquisition - CS Positive Extinction	-0.021	0.009	22.696	-0.039	-0.003	0.024*

B) SIR

Estimated Marginal Means

Experimental Phase	Emmean	SE	df	Lower CI	Upper CI
Acquisition	0.021	0.009	23.91	0.002	0.040
Acquisition	-0.002	0.008	22.65	-0.019	0.016
Extinction	0.015	0.007	37.26	0.000	0.029
Extinction	0.018	0.008	16.32	0.000	0.035

Contrasts

Contrast	Estimate	SE	df	Lower CI	Upper CI	P-values
CS Negative Acquisition - CS Positive Acquisition	0.023	0.009	23.326	-0.003	0.042	0.023*
CS Negative Extinction - CS Positive Extinction	-0.003	0.010	15.923	-0.025	0.018	0.753
CS Negative Acquisition - CS Negative Extinction	0.006	0.009	28.086	-0.012	0.025	0.497
CS Positive Acquisition - CS Positive Extinction	-0.020	0.010	19.284	-0.039	0.000	0.20

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

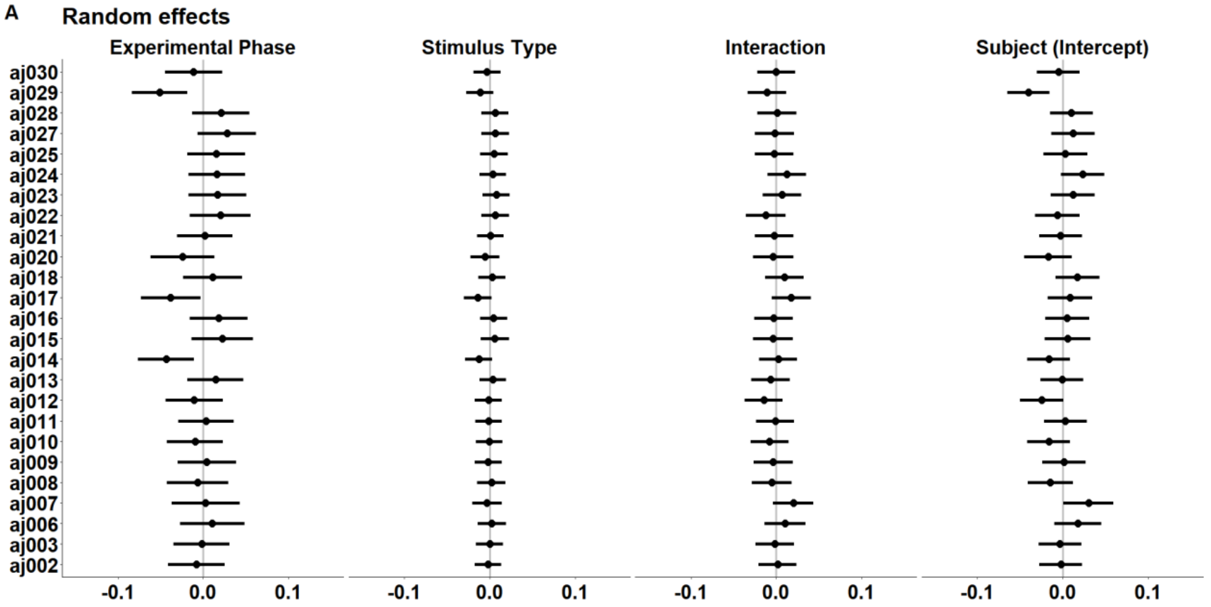
Supplementary Table 27

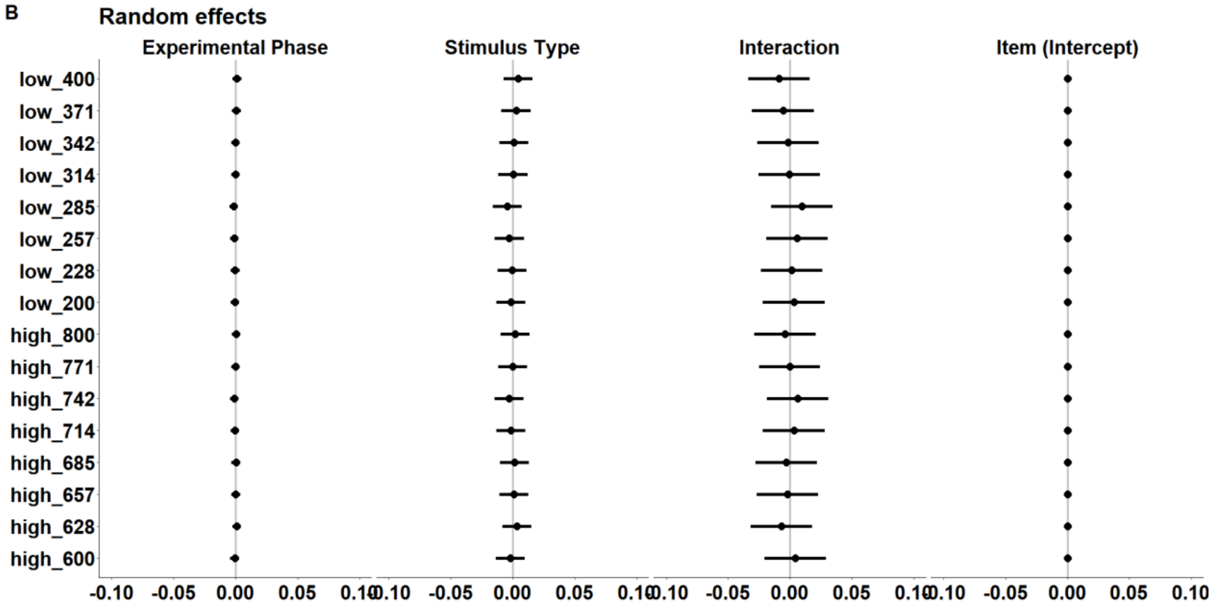
Summary of fixed estimates and random effect variance for the SIR pupil model

SIR	Mean Pupil Dilation		
Predictors	Estimates	CI	p
(Intercept)	0.01	0.00 – 0.02	**0.022**
Experimental Phase	-0.01	-0.02 – 0.01	0.372
Stimulus Type	0.01	-0.01 – 0.03	0.230
Interaction	0.03	0.00 – 0.05	**0.019**
Random Effects			
σ^2	0.0413		
τ_{00} Subject	0.0000		
τ_{00} Item	0.0004		
τ_{00} Subject: Item	0.0000		
τ_{11} Subject: Item: Phase	0.0001		
τ_{11} Subject: Item: Stimulus Type	0.0127		
τ_{11} Subject: Item: Interaction	0.0002		
τ_{11} Subject: Phase	0.0008		
τ_{11} Subject: Stimulus Type	0.0001		
τ_{11} Subject: Interaction	0.0002		
τ_{11} Item: Phase	0.0000		
τ_{11} Item: Stimulus Type	0.0000		
τ_{11} Item: Interaction	0.0002		
N Subject	25		
N Item	16		
Observations	6749		
Marginal R^2	0.002		

Supplementary Figure 53

By-subject and by-item random coefficients and intercept for the SIR pupil size model.





Notes. A) By-subject random effects and B) By-item random effects.

Supplementary Table 28

Summary of fixed estimates and random effect variance for the valence model

	Mean Valence		
Predictors	Estimates	CI	p
(Intercept)	3.21	0.90 – 5.52	**0.006**
Experimental Phase	-1.29	-6.17 – 3.58	0.603
Stimulus Type	-8.51	-14.69 – -2.33	**0.007**
Interaction	-2.20	-11.70 – 7.29	0.649
Random Effects			
σ^2	1310.9466		
τ_{00} Subject	0.0000		
τ_{00} Item	0.0000		
τ_{11} Subject: Phase	0.4668		
τ_{11} Subject: Stimulus Type	69.1598		
τ_{11} Subject: Interaction	1.1487		
τ_{11} Item: Phase	9.8813		
τ_{11} Item: Stimulus Type	33.3728		
τ_{11} Item: Interaction	19.4397		
N Subject	30		
N Item	16		
Observations	944		
Marginal R^2	0.014		

Note: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Table 29

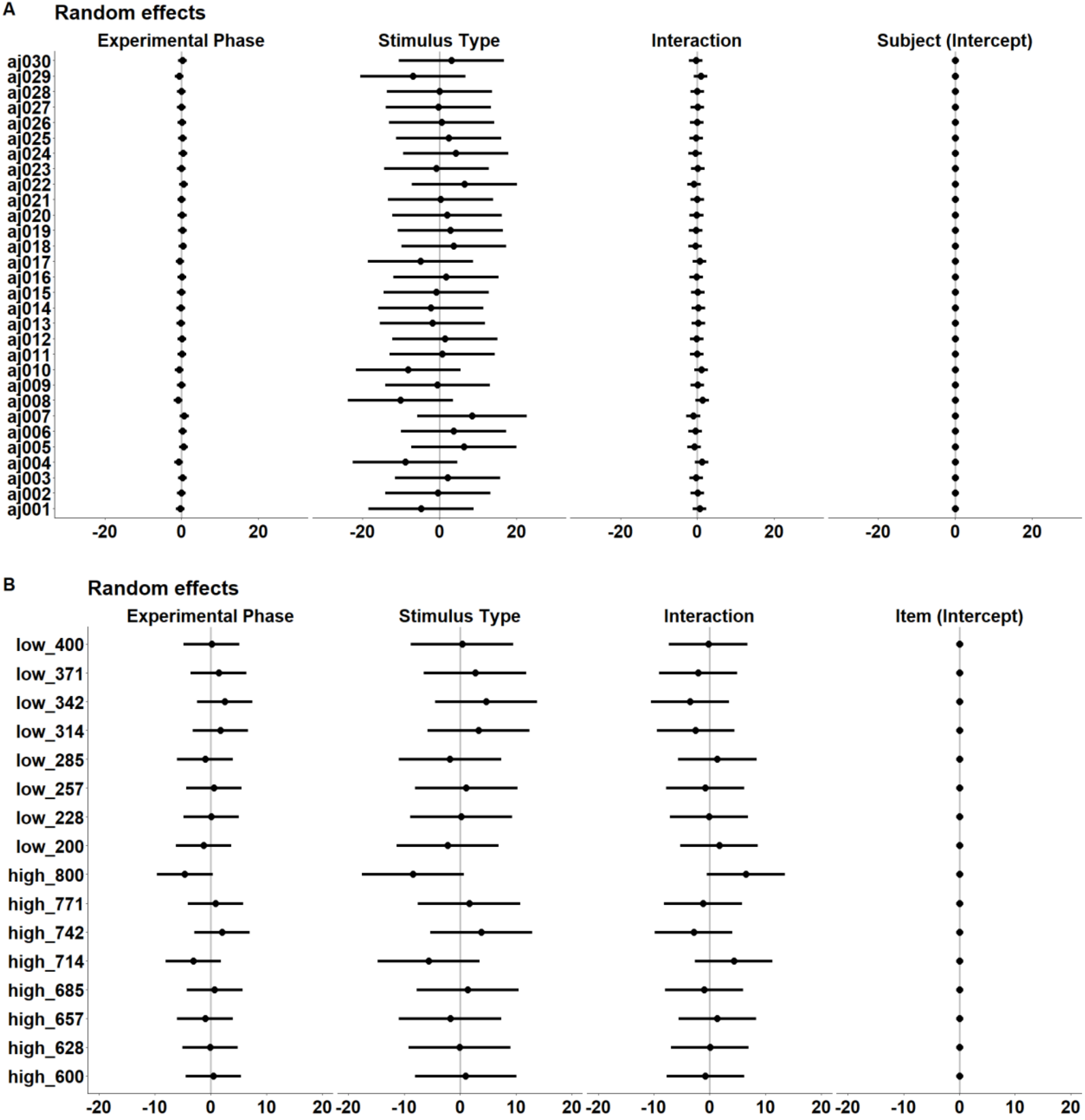
Estimated marginal means and related contrasts derived for the valence model.

Estimated Marginal Means						
Stimulus Type	Emmean	SE	df	Lower CI	Upper CI	
CS Negative	-1.043	1.97	10.07	-5.42	3.34	
CS Positive	7.467	1.97	16.50	340	11.63	
Contrasts						
Contrast	Estimate	SE	df	Lower CI	Upper CI	P-value
CS Negative – CS Positive	-8.51	3.15	18.03	-15.14	-1.88 0.015*	

Note. * p < 0.05, ** p < 0.01, *** p < 0.001

Supplementary Figure 54

By-subject and by-item random coefficients and intercept for the valence model



Supplementary Table 30

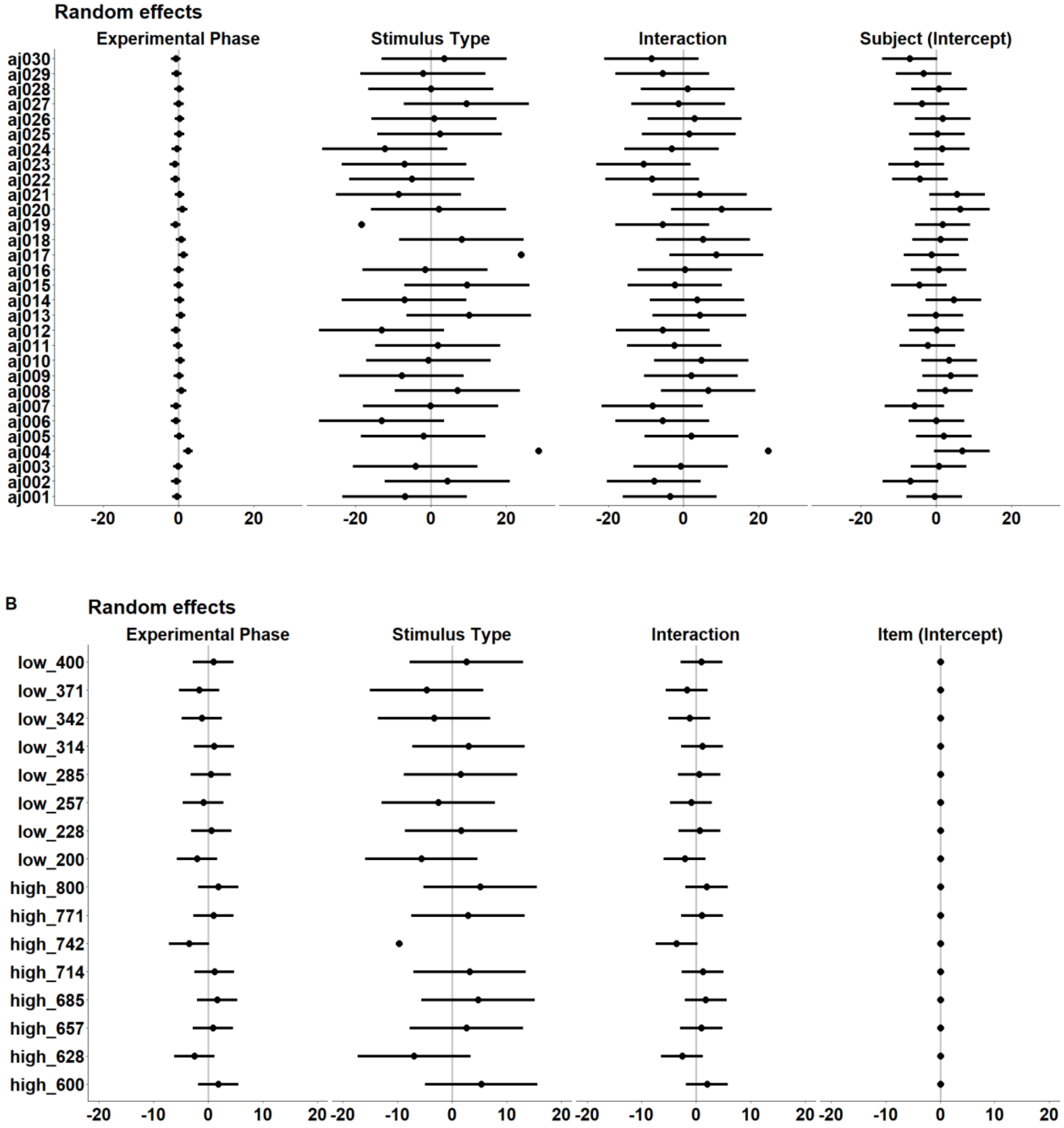
Summary of fixed estimates and random effect variance for the valence model

	Mean Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-1.52	-4.30 – 1.26	0.285
Experimental Phase	-0.36	-4.58 – 3.86	0.866
Stimulus Type	4.48	-2.64 – 11.59	0.218
Interaction	1.39	-7.43 – 10.20	0.757
Random Effects			
σ^2	993.1611		
τ_{00} Subject	28.6754		
τ_{00} Item	0.0000		
τ_{11} Subject: Phase	1.0020		
τ_{11} Subject: Stimulus Type	175.3835		
τ_{11} Subject: Interaction	88.8345		
τ_{11} Item: Phase	6.3467		
τ_{11} Item: Stimulus Type	49.9351		
τ_{11} Item: Interaction	6.8804		
N Subject	30		
N Item	16		
Observations	944		
Marginal R^2	0.005		

Notes: σ^2 Mean Random Effect Variance, τ_{00} Random Intercept Variance, τ_{11} Random Slope Variance

Supplementary Figure 55

By-subject and by-item random coefficients and intercept for the arousal model.



4.5.6 Supplementary Materials 21: Modelling UCS Item variability during Acquisition

In the main *Experimental Phase* by *Stimulus Type* mixed models, it was only possible to account for *Subjects* and *CS Items* but not and not *UCS Items* random variability, since the UCSs were delivered only during Acquisition and not during Extinction. To ensure generalisability across UCSs as well, we built a second set of models focusing on the main effect of *Stimulus Type* during Acquisition on pupil dilation first and second interval responses, valence and arousal ratings. In all models, *Stimulus Type* was included as the fixed effect. *Subjects* and *CS Items* were added as random intercepts, accompanied by random slopes for the main effects of stimulus type. In addition, we included US Items as a random intercept. For the pupil model only, we added a *Subjects* by *CS Item*s interaction intercept together with a random slope for *Stimulus Type*, as well as a *Subjects* by *UCS Item* interaction intercept. P-values for the fixed effects were determined via Type III Wald Chi-square tests (see Supplementary Table 31). Estimated marginal means and contrasts for the significant main effects were computed using *emmeans* and Satterthwaite method for degrees of freedom approximation (see Supplementary Table 32). As seen from the tables, the Acquisition-only models accounting for both CS and UCS item variability mirror the effects observed in the main models.

Supplementary Table 31

Type III Wald Chi-square tests *and R-squared values for the Acquisition-only mixed models.*

	Chisq	Df	P-value	R^2 Fixed (CI)
Pupil FIR				
Stimulus Type	10.63	1.000	0.001	0.006 (0.003 -0.01)
Pupil SIR				
Stimulus Type	6.56	1.000	0.010	0.004 (0.002-0.008)
Valence				
Stimulus Type	8.25	1.000	0.004	0.002 (0.005 -0.04)
Arousal				
Stimulus Type	1.13	1.000	0.3	0.006 (0.000 -0.02)

Supplementary Table 32

Estimated marginal means and related contrasts derived for each model.

Pupil FIR

Estimated Marginal Means

Stimulus Type	Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	
CS Negative	Acquisition	0.014	0.007	20.9	-0.001	0.029	
CS Positive	Acquisition	-0.010	0.006	19.4	-0.024	0.004	

Contrasts

Contrast	Experimental Phase	Estimate	SE	df	Lower CI	Upper CI	P-value
CS Negative – CS Positive	Acquisition	0.024	0.007	50.7	0.01	0.039 0.002	

Pupil SIR

Estimated Marginal Means

Stimulus Type	Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	
CS Negative	Acquisition	0.022	0.009	23.8	0.003	0.027	
CS Positive	Acquisition	-0.004	0.008	23.5	-0.021	0.66	

Contrasts

Contrast	Experimental Phase	Estimate	SE	df	Lower CI	Upper CI	P-value
CS Negative – CS Positive	Acquisition	0.025	0.01	23.5	0.005	0.046 0.02	

Valence

Estimated Marginal Means

Stimulus Type	Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	
CS Negative	Acquisition	-2.1	3.06	24.8	-8.41	4.19	
CS Positive	Acquisition	7.42	2.24	18.8	2.73	12.12	

Contrasts

Contrast	Experimental Phase	Estimate	SE	df	Lower CI	Upper CI	P-value
CS Negative – CS Positive	Acquisition	-9.54	3.32	26.75	-16.35	-2.72	0.01

Arousal

Estimated Marginal Means

Stimulus Type	Experimental Phase	Emmean	SE	df	Lower CI	Upper CI	

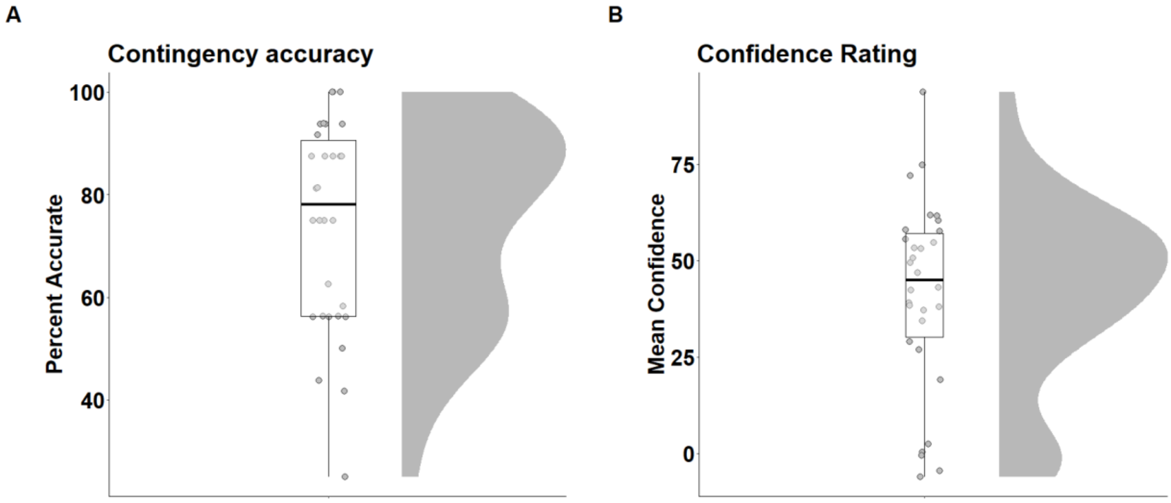
CS Negative	Acquisition	0.85	3.78	31.7	-6.84	8.55
CS Positive	Acquisition	-4.34	2.68	22.9	-9.88	1.2

4.5.7 Supplementary Materials 22: Contingency awareness analyses.

As seen in Supplementary Figure 56, only a small proportion of participants (~13%) were less than 50% accurate or showed negative average confidence ratings. This suggests that most participants were aware of the CS-UCS contingency. In order to corroborate these descriptive results, contingency awareness was estimated by calculating the detection of CS Negative trials using d-prime. First, trials were sorted into hits, misses, false alarms, and correct rejections in the following way. Correctly identified CS Negative trials were classed as hits, while correctly identified CS Positive trials as correct rejections. CS Positive trials falsely identified as CS Negative were classed as false alarms while CS Negative trials falsely identified as CS Positive were classed as misses. Next, two indices were calculated (package *sdt.rmcs*): a sensitivity index (d'), representing the difference of the hit-rate (zH) and miss-rate (zM) z-values and a bias index (c), representing the number of standard deviations from the midpoint between zH and zM (see Supplementary Figure 57). Finally, one sample t-tests were performed to determine whether d-prime and c indices were significantly greater than 0. These revealed that participants performed above chance at detecting CS Negative trials (t (df) = 6.25, p < 0.001) and that there was no significant bias towards CS Negative or CS Positive judgements (t (df) = 0.37, p < 0.71). Overall, these results suggest that participants were aware of the CS-UCS contingency.

Supplementary Figure 56

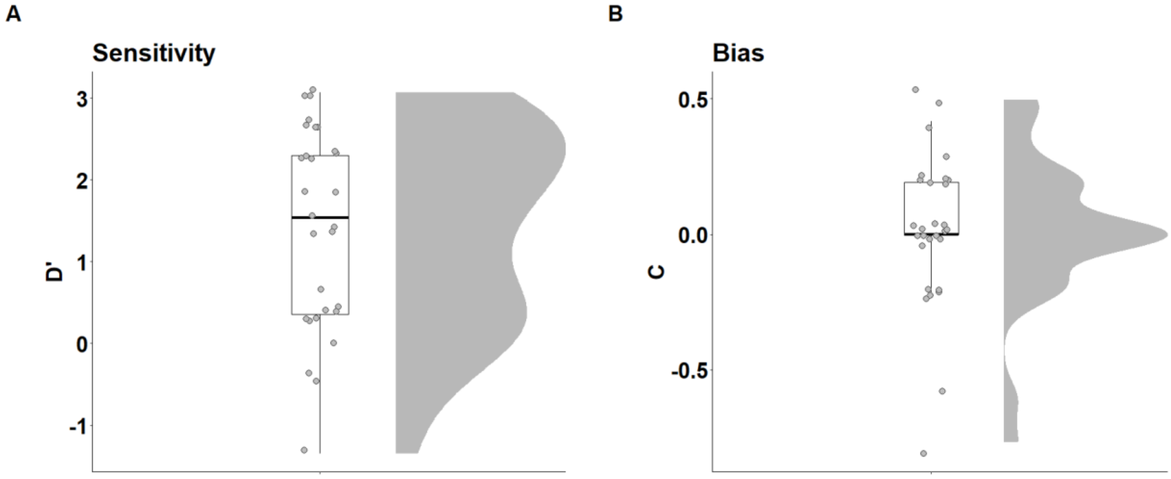
Distribution of contingency awareness and confidence ratings



Notes. Percentage accuracy was calculated by summing the number of correct responses and dividing by the overall number of trials * 100.

Supplementary Figure 57

Distribution of sensitivity and bias estimates

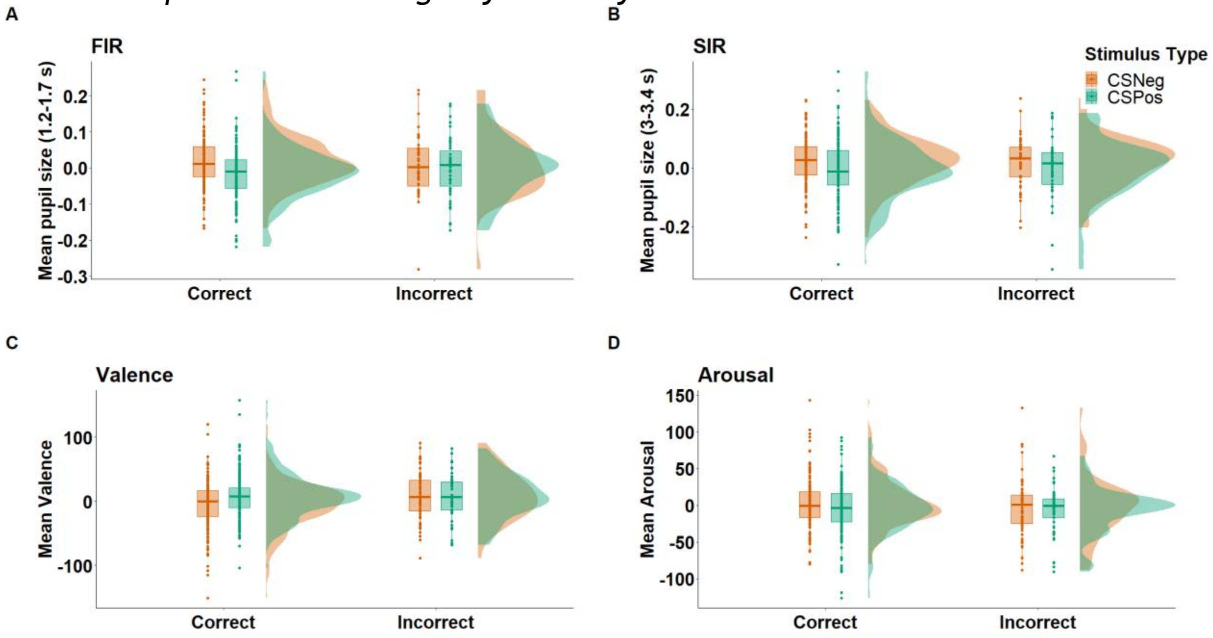


To examine whether conditioned responding is related to contingency awareness, we built another set of models focusing on the Acquisition phase only. The first set included the main effect of *Stimulus Type* (CS Positive/ CS Negative) and *Accuracy* (Correct/Incorrect) as well as their interaction. The second set included the main effect of *Stimulus Type* (CS Positive/ CS Negative) and participants' confidence ratings in relation to their contingency judgement as well as their interaction. The models were built for pupil dilation first and second interval responses, valence and arousal ratings. In all models *Subjects* and *CS Items* were added as random intercepts, accompanied by random slopes for the main effects and the interaction. P-values for the fixed effects were determined via Type III Wald Chi-square tests. In terms of contingency accuracy, there were no significant effects at the level of 0.05 (see Supplementary Table 33 and Figure 58). For, the models including the confidence ratings, the only significant effect was the main effect of valence (see Supplementary Table 34 and Figure 59).

If contingency awareness had an impact on the development of the CR on any of the outcome measures, we would have expected to detect a significant interaction between *Stimulus Type* and the contingency awareness measures. If contingency awareness did not influence CRs, we would have expected to only observe a main effect of *Stimulus Type*. The latter, was only the case for the model examining valence and confidence, suggesting that confidence in contingency responses had no relationship with the valence CR. The failure to observe any effects in the other models might be simply because there is no actual relationship between contingency awareness and conditioned responding using the current paradigm. It is also likely that our offline measures of awareness may not be sensitive enough to truly assess contingency awareness as participants completed those at the end of each block. Alternatively, since the main effects of *Stimulus Type* observed in our main models diminished when modelling the additional contribution of contingency information, it is possible that we did not have sufficient power to detect such a relationship.

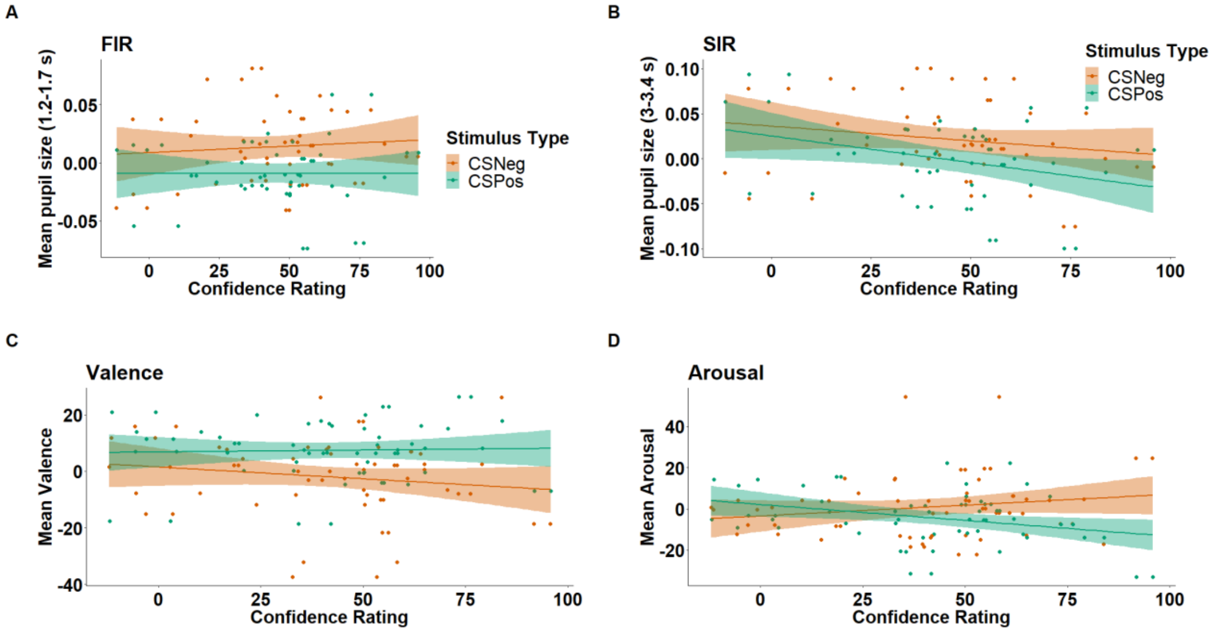
Supplementary Figure 58

Relationship between contingency accuracy and mean CR across measures.



Supplementary Figure 59

Relationship between mean confidence in contingency judgements and mean CR across measures.



Supplementary Table 33

Type III Wald Chi-square tests for each effect derived from the contingency accuracy response model.

	Chisq	Df	P-value
FIR			
Accuracy	1.9	1.000	0.15
Stimulus Type	3.5	1.000	0.06
Accuracy X Stimulus Type	0.9	1.000	0.33
SIR			
Accuracy	0.002	1.000	0.96
Stimulus Type	3.3	1.000	0.07
Accuracy X Stimulus Type	0.93	1.000	033
Valence			
Accuracy	3.4	1.000	0.06
Stimulus Type	3.1	1.000	0.07
Accuracy X Stimulus Type	1.9	1.000	0.16
Arousal			
Accuracy	0.3	1.000	0.56
Stimulus Type	1.3	1.000	0.24
Accuracy X Stimulus Type	0.2	1.000	0.61

Supplementary Table 34

Type III Wald Chi-square tests for each effect derived from the confidence ratings model.

	Chisq	Df	P-value
FIR			
Confidence	0.01	1.000	0.91
Stimulus Type	0.5	1.000	0.47
Confidence X Stimulus Type	0.6	1.000	0.44
SIR			
Confidence	0.2	1.000	0.6
Stimulus Type	1.4	1.000	0.24
Confidence X Stimulus Type	0.03	1.000	0.87
Valence			
Confidence	1.4	1.000	0.23
Stimulus Type	4.4	1.000	0.04*
Confidence X Stimulus Type	0.5	1.000	0.46
Arousal			
Confidence	0.4	1.000	0.53
Stimulus Type	0.9	1.000	0.35
Confidence X Stimulus Type	0.1	1.000	0.73

4.5.8 Supplementary Materials 23: Descriptive exploratory analysis of pupil and behavioural effects

A series of exploratory analyses were carried out to gain insight into the factors that may modulate conditioning and extinction effects. Any results reported here should be interpreted with caution due to their explorative nature and potential power limitations.

Development and habituation of pupillary conditioning effects over stimulus repetition

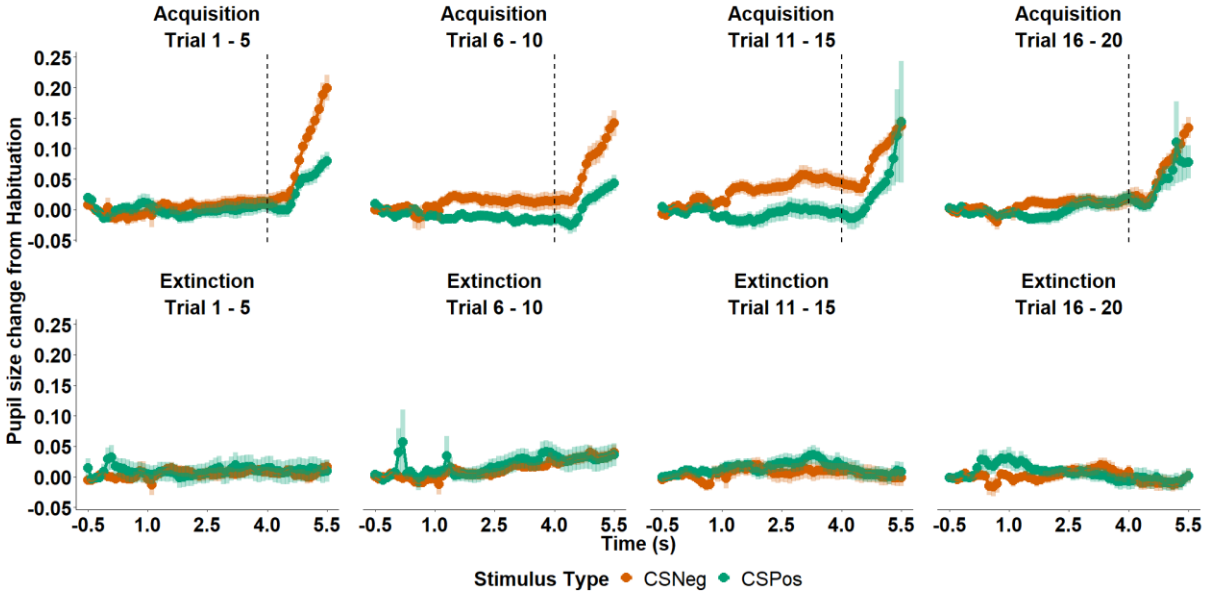
The conditioned response during fear acquisition elicited in experimental settings had previously been shown to habituate over repeated presentations, at least for some outcome measures (Leuchs et al., 2019). Likewise, when the UCS is removed, new associations are formed rapidly, resulting in fast extinction of the conditioned response (Lonsdorf et al., 2017). We therefore descriptively examined the trajectory of acquisition and extinction effects over stimulus repetitions in our study (Supplementary Figure 60).

Specifically, we calculated the mean pupil size over time across every 5 trials (i.e., trials 1-5, 6-10, 11-15 and 16-20). This was computed separately for CS Positive and CS Negative trials. As expected, during Acquisition the difference in average pupil size between CS Positive and CS Negative trials *before* UCS onset (the *conditioned* response) became apparent from trials 6-10. This Stimulus Type contrast increased during trial 11-15, but decreased again (to almost zero) during the final trials 16-20. Interestingly, the difference in average pupil size between CS Positive and CS Negative trials *after* UCS onset (the *unconditioned* response) appeared to decrease over trial repetitions in a more monotonic fashion.

During Extinction, a small difference between CS Positive and CS Negative trials may be expected during the first few stimulus repetitions. However, the average pupil size over time appeared to remain similar between CS Positive and CS Negative trials, at least at a descriptive level.

Supplementary Figure 60

Proportional mean pupil size over time and across repetitions of conditioned stimuli.



Note. The light-coloured areas indicate the standard error of the mean and the and the dashed vertical lines indicate UCS onset.

Development and habituation of pupillary conditioning and extinction effects over blocks

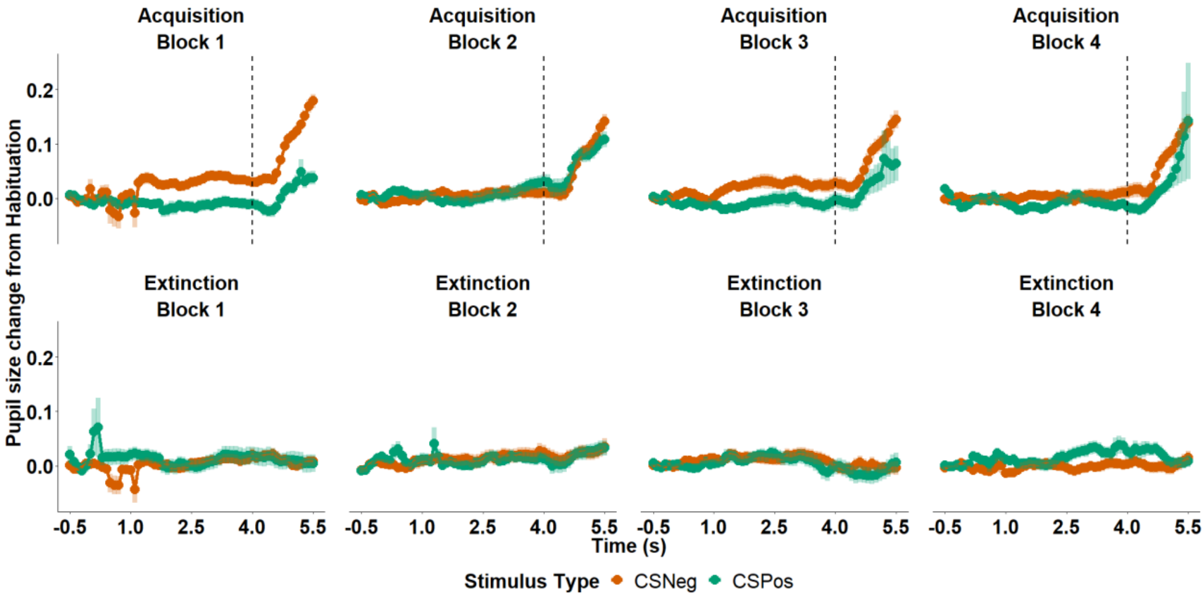
Since the current task represents a standard conditioning paradigm repeated 4 times over 4 blocks, it is possible that learning or repetition effects may have occurred despite the use of different conditioned and unconditioned stimuli in each block. Supplementary Figure 61 explores the trajectory of acquisition and extinction over the four experimental blocks.

In terms of the conditioned response (*before* stimulus onset), the difference in average pupil size between CS Positive and CS Negative trials was most apparent in block 1 and 3 and very small in block 2 and 4. If a repetition effect was to be present, any reduction in average difference would likely manifest in a more continuous manner. In contrast, the mean difference between CS Positive and CS Negative trials in the unconditioned response (*after* stimulus onset) was largest and occurred at the earliest timepoint during block 1 but reduced in magnitude during later blocks, suggesting a possible learning or habituation effect.

It is unlikely that this pattern was caused by differences in valence across US items as the order of US items across blocks was Latin-Square counterbalanced and randomised. During extinction, the average difference between CS Positive and CS Negative trials over time and blocks appears to remain comparable in magnitude.

Supplementary Figure 61

Proportional mean pupil size over time and blocks.



4.5.9 Supplementary Materials 8: Secondary analysis of valence and arousal ratings in Rehbein et al. (2014)

For comparative purposes, a re-analysis of the open dataset provided by Rehbein et al. (2014) was performed in order to obtain a measure of effect size. Identical to, the original analysis, a repeated measures ANOVA was performed for valence and arousal ratings separately, accompanied by generalized eta-squared estimates of effect size (see Supplementary Table 35).

Supplementary Table 35

Repeated measures ANOVA and Generalised eta-squared for each effect.

Effect	DFn	DFd	SSn	SSd	F	p	ges
Valence							
Experimental Phase	1	47	943.57	22924.90	1.93	0.171	0.005
Stimulus Type	1	47	102.73	10062.46	0.48	0.492	0.001
Experimental Phase X Stimulus Type	1	47	311.99	2857.21	5.13	0.028	0.002
Arousal							
Experimental Phase	1	47	1.46	21559.69	0.003	0.95	0
Stimulus Type	1	47	121.86	13121.26	0.437	0.51	0.0003
Experimental Phase X Stimulus Type	1	47	108	4603.60	1.10	0.29	0.0003

6 Chapter 5 - General Discussion

The goal of this thesis was to provide a contribution to recent research aiming at improving methodological and analytical practices in the study of threat learning. The present work focused on establishing the potential utility of several multi-trial classical conditioning tasks for the investigation of psychophysiological and behavioural indices of learning and extinction. In addition, we discussed these paradigms in the context of a number of widely debated theoretical and methodological topics in conditioning and psychological research in general, including the role of contingency awareness in learning, the role of neural oscillations as an underlying mechanism of learning, the potential for detecting deep structure learning indices using MEG, as well as improving generalisability of inferences through employing data-driven tools and analytical approaches which consider random variation between individuals and experimental items.

6.1 Summary of main findings

Chapter 1 provided a historical overview of the empirical study of threat learning, fear, and anxiety that has laid the foundations of our ongoing understanding of adaptive and maladaptive learning, and of the potential treatments for anxiety disorders. The chapter also reviewed more recent work in the area, aiming to improve replicability and inferences derived from classical conditioning, through raising awareness of theoretical and practical limitations in the field, and increasing methodological and analytical consistency, transparency, and open science practice.

Chapter 2 provided an indirect behavioural replication of the Multi-CS conditioning paradigm (Steinberg et al., 2013), and considered the extent to which contingency awareness is a necessary component for establishing a CR in a task with a large number of items and trials. Our study failed to replicate the results of previous investigations that have observed valence and arousal effects using this paradigm (e.g., Bröckelmann et al., 2011, 2013; Rehbein et al., 2014; Steinberg et al., 2012). The present findings appeared to offer support for theoretical and empirical contributions (Lipp & Purkis, 2005; Lovibond & Shanks, 2002; Mertens & Engelhard, 2020) suggesting that conditioned responding can only develop if there is a subjective awareness of the CS-UCS relationship. In line with recent findings indicating that methodological limitations can equally explain the variation in empirical support for unaware conditioning (Mertens & Engelhard, 2020), Chapter 2 examined several methodological and analytical factors that may contribute to the inconsistent evidence for eliciting subjective behavioural CRs using Multi-CS conditioning (Bröckelmann et al., 2011; Rehbein et al., 2014, 2015). First, we raised awareness of how differences in construct operationalisation can complicate cross-study comparisons. Next, we discussed the issue of conventional analyses on aggregated data in preventing generalisability of inferences to populations of items of the same type, and offered design-appropriate linear mixed effects modelling as a potential alternative, suggested to provide a better control of false positive rates (Westfall et al., 2017). In addition, we considered the role of individual variability, in line with recent work raising awareness of the high level of individual differences in learning about and responding to CS-UCS contingencies (Lonsdorf et al., 2019; Lonsdorf & Merz, 2017). Specifically, we re-examined a Multi-CS conditioning dataset provided by Rehbein et al. (2014) using robust and transparent graphical representations that go beyond depicting mean differences. We showed that the inconsistency in reported findings may be driven not only by the presence of individual differences in conditioned responding but also by insufficient consideration of the underlying data distributions. Specifically, we demonstrated the utility of detailed data visualisation in revealing important patterns in the data (e.g., baseline differences and outliers), and in guiding analytical decisions and interpretation. Finally, we discussed the possibility that the measurements that were used both in previous research and in our study may not be sensitive enough to detect subtle condition differences in the absence of CS-UCS contingency awareness.

Chapter 3 examined the potential utility of a novel visual blocked conditioning paradigm as a means for providing a large number of trials. The task was designed with the goal of aiding MEG investigations of the cortical and subcortical oscillatory dynamics of learning, which so far have been limited by significant technical challenges. In addition, we took into account previous suggestions of the benefits in employing multiple outcome measures in classical conditioning research (Lonsdorf et al., 2017). As such, we examined pupil size, subjective valence, and arousal ratings as additional potential indices of learning. The task employed a large number of trials and attempted to establish learning and extinction successively in multiple blocks, using a different set of CSs in each block. However, our findings provided limited support for conditioned responding in any of the measures.

The results obtained from the MEG data were based on a comprehensive set of analyses aimed at understanding the role of theta oscillations in learning and extinction across the fear network. While these analyses failed to observe any differential brain activation patterns, they are informative both from a theoretical and from a design perspective, as they revealed several aspects that may be important to consider in future investigations. For instance, our descriptive results in the time-frequency domain aligned with those provided by a recent high-precision MEG study (Tzovara et al., 2019) showing a reduced theta power to CS+ stimuli in deep structures such as the amygdala. Chapter 3 discussed the theoretical and methodological implications of this observation due to its inconsistency with the typical pattern of increased theta power found in the rodent conditioning literature (Karalis et al., 2016; Lesting et al., 2011; McCullough et al., 2016). We argued that this discrepancy may be indicative of potential cross-species functional differences in the theta range or can be representative of an instance in which cross-species procedural differences may hamper comparisons between human and animal data.

In terms of pupillary signatures of conditioning, we found little evidence of a CR in pupillary responses, even though the measure was clearly able to detect a UCR during the Acquisition phase. Failure to detect a CR was potentially due to the short trial duration. For subjective ratings of valence, we observed no significant differences in responding. However, we found increased arousal ratings for CS+ trials and during Acquisition. While the results in relation to arousal were not modulated by an interaction between experimental phase and stimulus type, they were suggestive of the paradigm's potential to elicit conditioning. Yet, as presently observed, these effects were too weak to allow for any firm conclusions.

Importantly, our exploratory evaluation of the task showed no evidence that the blocked nature of the design itself (potentially causing UCS-habituation via repeated exposure) was responsible for the lack of clear conditioning effects. Instead, we argued that the findings reported in Chapter 3 highlight several other design factors that may require consideration in future paradigm development. These include 1) the necessity for increasing trial duration, especially for enabling the detection of pupil size and time-frequency decomposition effects, 2) considering the implementation of simple CSs to avoid the potential confound of complex, higher-order processes involved in face discrimination, and 3) utilising a better strategy for maintaining participants' attention and ensuring the presence of contingency awareness.

These design factors were considered in **Chapter 4.** The study reported in this chapter aimed at refining the blocked conditioning task and examine its utility in the context of pupil size and subjective behavioural indices of learning vs. extinction (with an eye on potential MEG investigations in the future). At the same time, the study in Chapter 4 paid closer attention to the issue of generalisability, specifically in relation to the benefits of employing a greater variety of items (and consequently, a higher number of trials) and to the potential of this paradigm to allow for more reliable inferences through the use of robust analytical tools. The study adopted simple auditory stimuli as CSs and unlike the visual blocked conditioning, it included both positive and negative UCSs in attempt to elicit a more pronounced CR. In addition, the duration of both the CSs and the UCSs were extended significantly from 0.65 and 0.2 s to 3 and 1 s respectively. The paradigm also included a secondary tone differentiation task designed to maintain participants' attention.

In terms of analysis, we employed a robust data-driven approach to determine time windows of interest in the pupil size data. Moreover, inferential analyses for all measures (pupil size as well as rating data) were based on design-appropriate linear mixed-effects models taking both by-participant and by-item variation into account.

As a result of the above design and analysis features, we found clearly measurable CRs which were simultaneously generalisable across participants and items. In the context of pupil size, we reported both early and late pupillary CRs that were highly correlated, potentially indexing the same underlying mechanism. These CRs were found to have already disappeared during extinction training. Considering the latter, we discussed a slight reduction in the number of trials per condition as a potential solution for allowing the detection of more 'gradual' extinction learning processes. In the discussion, we also considered other factors that may influence extinction effects such as the type of instructions provided to participants. In terms of ratings, while subjective arousal judgements were insensitive to the critical experimental manipulations, the study revealed a valence CR that appeared to be resistant to extinction. These effects were discussed in the context of previous findings suggesting that evaluative judgements may be less susceptible to extinction (Luck & Lipp, 2015a, 2015b; Wendt et al., 2020), and that different outcome measures may be sensitive to different aspects of learning (Lonsdorf et al., 2017; Sevenster et al., 2012). The findings in this chapter also provided some further evidence for the importance of contingency awareness. Although awareness was not found to modulate any of the CR indices, it was found that the majority of participants were aware of the CS-UCS relationships.

Finally, considering the important role of individual differences in contextualising conditioned responding (Lonsdorf et al., 2019; Lonsdorf & Merz, 2017), we conducted a set of exploratory analyses (in both Chapters 3 and 4) to establish potential links between the strength of conditioned responding on the one hand and differences on various participant-specific dimensions on the other. These dimensions included *trait* as well as *state anxiety*, use of *emotion regulation strategies* (i.e., expressive suppression and cognitive reappraisal), and *non-verbal ability* (the latter was examined in Chapter 3 only). We found hardly any evidence in support of a relationship between any of these person-specific variables and CR outcome measures (including pupil size, theta activity in ROIs, and valence and arousal ratings). This, however, does not necessarily suggest that psychological variables are irrelevant for conditioned responding. Note, for example, that we observed relatively low variability across participants in any of the psychometric variables considered. A less homogeneous participant sample (I.e., with more inter-individual variation on these dimensions) would increase the prospect of uncovering potential relationships between person-specific variables and the strength of conditioning.

6.2 Limitations

One of the most limiting factors on reproducibility, as well as the quality and reliability of statistical inferences, is the *sample size* used in an empirical study (Clayson et al., 2019; Lakens, 2021; Larson, 2020; Szucs & Ioannidis, 2019). A common approach for justifying sample sizes is power analysis, whereby one can estimate the minimum sample size that is needed to detect an effect of a certain magnitude. Yet, conducting power analysis can be challenging, especially for repeated-measures designs and for certain types of measurement (e.g., neuroimaging). Moreover, the availability of power estimation tools for linear mixed-effects modelling is still rather limited, although recent efforts have been made towards the development of statistical packages for simulating data with crossed random effects (DeBruine & Barr, 2019), which can allow for power estimation in designs like the ones used in this thesis and related literature. To be truly valid, however, a large amount of prior information on various (usually unknown) population parameters is needed to give such simulations (or power analyses in general) sufficient credibility. Such parameters include, for instance, expected *effect sizes* and various *population variances* (and co-variances) at participant and item level. These parameters are difficult to obtain without prior research, a problem that has an analogue in determining *priors* for Bayesian analysis. In addition, the relevant tools are still under development and limited to relatively simple experimental designs and models. Indeed, currently available tools cannot effectively account for complex fixed and random effect structures (and their correlations) in experimental designs that go beyond 2 x 2 complexity. The issue of power estimation is even more prominent in the context of neuroimaging, where the degree of unknown parameters is even higher due to the multivariate nature of the data. Furthermore, while such tools are more widely available for fMRI (Mumford, 2012), this is not the case for MEG research, which only very recently has seen attempts to develop such tools (Chaumon et al., 2019).

Due to the complexity of our designs, and in the absence of strong prior expectations in terms of effect magnitude and other population parameters for the paradigms we have used, the sample sizes for the experiments in this thesis were not based on power analyses. Instead, our decisions were guided by common practice in the fields under investigation, as well as by resource and time constraints due to the extensive duration of each experiment (ranging between 2 to 5 hours per participant). Specifically, the sample size for the Multi-CS conditioning and visual blocked conditioning tasks were based on the sample sizes of previous MEG investigations as reviewed in Chapters 2 and 3 (mean ~ 19, median ~ 19, range 5 - 48). The sample size for the auditory blocked conditioning task was guided by other pupillometry studies in the field (mean ~37, median ~ 25, range 18-135). Also note that while the initially intended sample size for this study was 40 participants, we were only able to collect data from 30 participants, because government restrictions in relation to the Covid-19 pandemic came into effect before the study was completed. Due to measurement noise, only data from 25 participants were used in the pupil size analysis. Even though we did not acquire the desirable number of participants in this study, the sample is still representative of the median sample size used in pupillometry conditioning studies and is also of a size similar to that used by Tzovara et al. (2018), who also employed mixed effects modelling in their analyses. While adopting a ‘rule of thumb’ approach is not always optimal in the context of paradigm development, choices are often limited to standard practices when there is insufficient prior knowledge and limited time. With the successful development of the auditory blocked conditioning task, however, there are now estimates of both fixed and random effects parameters that can be used when planning sample size in future investigations using this design.

It is also worth considering the design and analytical limitations of the studies presented in this thesis. Experiment 2 (Chapter 3) suffered from a set of design limitations which we attempted to account for in the subsequent experiment.

These were related to the short trial duration which may have prevented the detection of a CR in the time-frequency domain and in the pupil size measurements. In the subsequent chapter, we demonstrated that increasing the trial duration may be beneficial for pupillary CR detection, which makes sense considering the dynamics of pupillary responses in general. Another substantial limitation in Experiment 2 was that it relied on the assumption that contingency awareness had been established. Specifically, since we employed only two unique items per condition, we expected that it will be easy for participants to remember the relationship between the CSs and the UCS. However, contingency awareness was not explicitly measured to avoid prolonging the already substantial experimental duration (~5 hours per participant). Again, this was rectified in Experiment 3 where the measurement demands were not as substantial as those in MRI and MEG data collection, and where a better balance between trial duration and trial number was achieved to allow for the inclusion of a contingency awareness task.

In addition, we found that the visual blocked conditioning task in Experiment 2 (Chapter 2) elicited baseline differences in valence ratings. It was not clear, however, what caused these differences given that the face stimuli were selected from a normative database of ‘neutral’ faces similar in valence and arousal norms. While all analyses were performed on habituation-baselined data that aimed to account for the potential of such baseline differences, it cannot be guaranteed that the initial differences in the perceived valence of the stimuli had no impact on the responses in the other outcome measures.

In terms of the MEG analysis and the detection of subcortical activity, our source estimation approach was not as sophisticated as some of the more recently developed approaches (i.e., Tzovara et al., 2019), but nonetheless, it is a commonly accepted technique that is considerably more cost- and time-effective. Indeed, the option of adopting high-precision methodology in this study was considered and briefly tested. However, we estimated that such an approach would not be feasible in relation to the resources we had, since developing a high-precision MEG pipeline and the 3D printing of individual head casts is incredibly time consuming and requires additional equipment that comes at a high cost.

The main limitation in the study reported in Chapter 4 is related to the use of positive and negative UCSs in the absence of a completely neutral control condition (i.e., CS trials that are not followed by a UCS). This can be problematic since both positive and negative UCSs elicit an UCR and therefore, any inferences about differential responses to CSPos and CSNeg conditions are relative to a valent stimulus (either of positive or negative valence). This may create a situation in which the absence of a differential CR, as evidenced by a non-significant difference between CSPos and CSNeg, does not necessarily imply absence of a CR. In particular, it is likely that a differential CR may be obtained if comparing responses to CSs paired with valent UCSs to responses that were unpaired or paired with a truly neutral stimulus. From a design perspective, however, including a third stimulus type (e.g., CS Neutral or CS-) may be impractical due to the challenges in designing neutral experimental stimuli that clearly vary and can be distinguished along three dimensions. Requiring participants to distinguish between multiple items within three conditions may also prove problematic for establishing contingency awareness. A more practical alternative would be to use a partial reinforcement protocol whereby the CSs are not always followed by a UCS. The utility of such a design modification needs to be studied in more detail in future work.

It is important to consider yet another set of design constraints that are present when studying conditioning in a blocked design manner. For instance, our design may make cross-study comparisons difficult as it is still unknown exactly how the CRs obtained from our task compare to more conventional tasks or other multi-trial conditioning tasks. This issue also applies to the simultaneous use of positive and negative UCSs, as this is not a common practice in the related literature. More importantly, the blocked design prevents the investigation of other commonly studied mechanisms in associative learning research that have important implications for the treatment of anxiety disorders, such as long-term extinction and the conditions under which fear can return (i.e., spontaneous recovery, reinstatement, and renewal protocols). Notwithstanding, our task was not designed for these purposes, as its primary aim was to allow for the study of conditioned responding and the development of extinction in a specific set of experimental contexts that require many trials. The next section will discuss two alternative multi-trial paradigms that may allow for the examination of non-immediate extinction effects. These are based on the design of the blocked conditioning we reported in Chapter 4 and the recently developed sequential conditioning task (Sperl et al., 2021).

6.3 Contributions and implications for future research

The present thesis contributes to methodological research in threat learning in several ways, as discussed below.

6.3.1 Replicability and transparency

In the context of Multi-CS conditioning and the inconsistent evidence for conditioned responding without subjective awareness, our findings largely agree with previous reports on unaware conditioning (Mertens & Engelhard, 2020), and suggest that different methodological practices may confound the perpetuating nature of the debate regarding the role of contingency awareness in threat learning. The findings from our replication study and secondary analysis of the Rehbein et al. (2014) dataset also add to the body of research highlighting the importance of using transparent data practices in overcoming the issue of replicability in threat learning. Based on the observations derived from these analyses, we proposed a set of recommendations that may guide future Multi-CS conditioning research. These were aimed at improving clarity when (operationally) defining constructs and their measurement, analytical and data visualisaiton practices as well as measurement methods. We believe that such guidelines can improve the accuracy and reliability of future studies utilising not only this task but any threat learning paradigm. Consequently, throughout this thesis, an important goal was to provide a highly transparent reporting through adequate visualisation tools (providing more than just means per condition) and offer clarity regarding the magnitude of the observed effects. Furthermore, we conducted our experiments in a reproducible manner by providing open access to the code necessary to reproduce our findings as well as experimental tasks.

6.3.2 Informed and generalisable analytical practice

Throughout this thesis, we reinforced the benefits of using analytical tools that can potentially improve the generalisability and reliability of findings in several ways. In Chapter 2, we focused on the utility of robust estimates of central tendency (Rousselet et al., 2017; Wilcox, 2017) as a means to deal with issues related to outliers when performing conventional analysis on aggregated data. Across chapters, we also discussed the importance of modelling both by-item and by-participant random variability, especially in the case of multi-trial paradigms using a large number of unique items. We emphasised the potential risks of making anti-conservative inferences that may not generalise beyond the study under investigation, when employing conventional analytical strategies relying on aggregated data (e.g., repeated-measures ANOVAs and t-tests). This focus on the utility of mixed effects modelling was based upon on an increasing body of research highlighting the ongoing issues of poor generalisability of findings across studies, driven by analytical tools that produce inflated test-statistics and narrow confidence intervals, ultimately giving poor control over false positives (Barr et al., 2013; Judd et al., 2012, 2017; Westfall et al., 2017; Yarkoni, 2020).

In line with recent research highlighting the consequences of variable and often arbitrary criteria for data exclusion in the analysis of psychophysiological data (Lonsdorf et al., 2019), Chapter 4 offered means to avoid arbitrary selection of time windows of interest in pupil size data analysis. Specifically, we adopted a data-driven cluster permutation approach to identify time windows that exhibit true cross-condition differences. While this approach is not novel and frequently adopted for the analysis of neuroimaging data, it is not commonly applied to other psychophysiological measures. As detailed in Chapter 4, a data-driven approach can increase cross-study comparability by reducing the excessive arbitrariness in time window selection that currently prevails in the literature. Moreover, it facilitates detection of unpredicted effects while appropriately controlling for Type I and Type II errors (Huang & Zhang, 2017; Sassenhagen & Draschkow, 2019). While the combination of data-driven time window selection and mixed effects modelling for inferential analysis in psychophysiological measures is an approach that may not necessarily be as sophisticated as other recently proposed techniques for estimating learning indices (i.e., PSPM, Bach et al., 2018), it offers considerable benefits over more common approaches. Compared to PSPM, our approach also provides a higher degree of flexibility as it does not constrain data pre-processing and analysis to a specific software.

6.3.3 Multi-trial paradigm development

As evidenced by the work detailed in this thesis, the design of a multi-trial paradigm to assess threat learning has proven a significant technical and resource-intensive challenge, which may explain the limited popularity of such tasks across the literature. A significant barrier to multi-trial paradigms is the issue of experimental length. Specifically, utilising a large number of trials where the minimum trial duration is restricted by the demands of psychophysiological processes, can often be impractical both in terms of time and costs to both the experimenter and participants. Consequently, it can be difficult to obtain a balance between trial number and trial duration, that allows for the detection of reliable learning indices in multiple outcome measures and results in an experiment with an acceptable duration. As demonstrated in this thesis, the optimisation of such a task may often require multiple revisions. Nonetheless, as discussed extensively throughout this thesis, there are significant benefits in establishing a reliable measure of conditioning in a multi-trial context, particularly in relation to improving generalisability via a greater range of unique items and the SNR in noisy psychophysiological measures.

To this end, we believe that the results obtained in the last experimental chapter of this thesis are encouraging, in that they demonstrate the possible utility of auditory blocked conditioning as a means for investigating associative learning in a multi-trial context. Unlike other commonly employed multi-trial paradigms that repeat the same CS+ and CS- stimulus many times, this task can potentially offer greater generalisability across stimuli by providing a higher degree of variability in unique CSs as well as UCSs, while also allowing for this variability to be modelled at the analytical stage.

While the results obtained in Experiment 3 were promising, the task may still benefit from additional design modifications, such as reducing the number of stimulus repetitions within a block to reduce CR habituation and to allow for the examination of extinction training effects. To maintain the number of trials provided currently, however, this repetition reduction will require the inclusion of an additional block including a new set of stimuli. Furthermore, there remains a considerable amount of future work to be done on validating the reliability of this task in different design contexts, and to replicate its findings across psychophysiological and behavioural outcome measures. As identified in Chapter 4, it is also important to explicitly consider the role of instructions in mediating learning and extinction effects. This can be achieved in a between-subjects design, similar to previous studies (e.g., Luck & Lipp, 2015b) in which the level of instruction is manipulated across groups.

In addition, future work should establish whether the effects observed in the auditory domain would generalise to CSs in other modalities such as visual stimuli, as these are the most commonly adopted across the conditioning literature. Importantly, it will be beneficial to examine how design variation in relation to the UCS will affect the blocked conditioning task. For example, it is important to consider whether the results can be replicated with other types of UCSs (e.g., olfactory, electric stimulation) and across different reinforcement protocols (e.g., partial reinforcement), but also in the absence of a positive UCS, resembling more closely the blocked design detailed in Chapter 3 and the sequential conditioning task in the study by Sperl et al. (2021).

In relation to the pupillary signatures of conditioning, further research should focus on systematically examining the conditions under which multiple interval CRs are detectable, specifically in relation to trial duration. In the context of SCR, it has been shown that early and late intervals are only observable in trials with longer duration (Jentsch et al., 2020). It is therefore possible that this may occur in pupil size measurement as well. If this is the case, then reducing the trial duration will diminish the presence of multiple interval CRs while increasing the trial duration should not affect their detectability. It is worth noting, however, that a significant reduction of the trial duration beyond what was implemented in Chapter 4 may pose a significant risk to the detectability of the pupillary CR, similar to the problems observed in Chapter 3. In line with recent efforts in understanding the commonalities and differences between pupillary and SCR learning indices (Jentsch et al., 2020; Leuchs et al., 2019), future work should also consider whether these early and late interval pupillary CRs would replicate when using SCR as an outcome measure.

Another future avenue of research should evaluate the potential utility of an inter-mixed CSs and UCSs design as an alternative to the blocked design presented in this thesis. Specifically, examining conditioning in a single block where multiple different CSs and UCSs are presented would resolve the caveat of the blocked design in preventing the investigation of long-term extinction as well as the return of fear. Alternatively, the task can also be modified to resemble more closely the design utilised by Sperl et al. (2021), where sequential conditioning across multiple blocks was followed by sequential extinction the following day. It is worth noting that this design may not allow for the use of multiple UCSs, as it is unknown whether participants can feasibly acquire and retain information about multiple different CS-UCS contingencies. In addition, both of the proposed task modifications would require a substantial amount of work to optimise the balance between number of unique items and the preservation of contingency awareness.

6.3.4 Threat learning using MEG

The work presented in this thesis has laid solid foundations for the study of the oscillatory signatures of learning and extinction in cortical and subcortical regions. While the technical limitations of the visual blocked conditioning task (Chapter 3) prevented the reliable quantification of the neural signatures of associative learning and extinction, there was limited evidence that the analytical strategy itself was ineffective. In particular, the use of realistic anatomical information of deep brain structures, depth weighted MNE source estimation and baseline subtraction to reduce the effects of leakage have already been successful in localising amygdala and hippocampal activity (Attal & Schwartz, 2013; Balderston et al., 2014a; Quraan et al., 2011a). However, considering that the magnitude of previously reported effects using high-precision MEG (Tzovara et al., 2019) were small even though consistent with some of the present descriptive findings, it is likely that the strategies used in our study to maximise deep structure detection may not be sufficiently sensitive to detect very small effects in deep structures. Future work attempting to localise deep source activity in MEG should also examine the extent to which depth weighted MNE and baseline subtraction alone are sufficient to deal with the issue of leakage, by comparing its effectiveness to the generative models used by Tzovara et al. (2019), as these were shown to exhibit high sensitivity to individual anatomies, evidenced by poor model fits when minimal displacement of deep structures was performed.

Last but not least, the results reported in Chapter 4 are encouraging in the context of future multi-trial MEG investigations and consistent with recent findings (Sperl et al., 2021), demonstrating that measurable physiological CRs can be elicited in a blocked design context. In addition, the current design may hold potential benefits compared to the sequential conditioning paradigm proposed by Sperl et al. (2021) since it allows for a greater range of unique CSs as well as for the use of more than one UCS. In its current form, however, the task will not provide a trial number large enough for studying the MEG correlates of learning and extinction, especially in deep structures. Theoretically, it would be possible to increase the number of blocks since the exploratory analyses of this study revealed no evidence that the blocked nature of the conditioning had any impact on the magnitude of the CRs. Such increase in blocks should provide the desirable number of trials, although this would also require conducting the MEG recordings over multiple testing sessions. Resource-intensive, multi-day recordings are not uncommon in classical conditioning research. Such recordings are also common in the context of MEG, precisely because obtaining high number of trials in any MEG task typically requires a considerable time commitment.

List of References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Abend, R., Gold, A. L., Britton, J. C., Michalska, K. J., Shechner, T., Sachs, J. F., Winkler, A. M., Leibenluft, E., Averbeck, B. B., & Pine, D. S. (2020). Anticipatory Threat Responding: Associations With Anxiety, Development, and Brain Structure. *Biological Psychiatry*, *87*(10), 916–925. https://doi.org/10.1016/j.biopsych.2019.11.006

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. W. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. https://doi.org/10.1016/j.jneumeth.2012.06.011

Akpan, B. (2020). Classical and Operant Conditioning—Ivan Pavlov; Burrhus Skinner. In B. Akpan & T. J. Kennedy (Eds.), *Science Education in Theory and Practice* (pp. 71–84). Springer, Cham. https://doi.org/10.1007/978-3-030-43620-9_6

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). *Raincloud plots: a multi-platform tool for robust data visualization*. https://doi.org/10.7287/peerj.preprints.27137v1

Ardiel, E. L., & Rankin, C. H. (2010). An elegant mind: Learning and memory in Caenorhabditis elegans. In *Learning and Memory* (Vol. 17, Issue 4, pp. 191–201). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/lm.960510

Attal, Y., Bhattacharjee, M., Yelnik, J., Cottereau, B., Lefèvre, J., Okada, Y., Bardinet, E., Chupin, M., & Baillet, S. (2007). Modeling and detecting deep brain activity with MEG & EEG. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 4937–4940. https://doi.org/10.1109/IEMBS.2007.4353448

Attal, Y., Maess, B., Friederici, A., & David, O. (2012). Head models and dynamic causal modeling of subcortical activity using magnetoencephalographic/electroencephalographic data. In *Reviews in the Neurosciences* (Vol. 23, Issue 1, pp. 85–95). Rev Neurosci. https://doi.org/10.1515/rns.2011.056

Attal, Y., & Schwartz, D. (2013). Assessment of Subcortical Source Localization

Using Deep Brain Activity Imaging Model with Minimum Norm Operators: A MEG Study. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059856

Bach, D. R., Castegnetti, G., Korn, C. W., Gerster, S., Melinscak, F., & Moser, T. (2018). Psychophysiological modeling: Current state and future directions. *Psychophysiology*, *55*(11). https://doi.org/10.1111/psyp.13209

Bach, D. R., & Friston, K. J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, *50*(1), 15–22. https://doi.org/10.1111/j.1469-8986.2012.01483.x

Bach, D. R., Gerster, S., Khemka, S., Korn, C., Moser, T., Paulus, Philipp, C., & Staib, M. (2015). *PsPM: Psychophysiological Modelling*. *Dcm*, 4–5.

Bach, D. R., & Melinscak, F. (2020). Psychophysiological modelling and the measurement of fear conditioning. *Behaviour Research and Therapy*, *127*, 103576. https://doi.org/10.1016/j.brat.2020.103576

Balderston, N. L., Schultz, D. H., Baillet, S., & Helmstetter, F. J. (2013). How to detect amygdala activity with magnetoencephalography using source imaging. *Journal of Visualized Experiments : JoVE*, *76*, 1–10. https://doi.org/10.3791/50212

Balderston, N. L., Schultz, D. H., Baillet, S., & Helmstetter, F. J. (2014a). Rapid amygdala responses during trace fear conditioning without awareness. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0096803

Balderston, N. L., Schultz, D. H., Baillet, S., & Helmstetter, F. J. (2014b). *Rapid Amygdala Responses during Trace Fear Conditioning without Awareness*. *9*(5). https://doi.org/10.1371/journal.pone.0096803

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*(5227), 1115–1118. https://doi.org/10.1126/science.7652558

Becket Ebitz, R., & Moore, T. (2017). Selective modulation of the pupil light reflex by microstimulation of prefrontal cortex. *Journal of Neuroscience*, *37*(19), 5008–5018. https://doi.org/10.1523/JNEUROSCI.2433-16.2017

Bennett, M., Vervoort, E., Boddez, Y., Hermans, D., & Baeyens, F. (2015). Perceptual and conceptual similarities facilitate the generalization of instructed fear. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 149–155. https://doi.org/10.1016/j.jbtep.2015.03.011

Betella, A., & Verschure, P. F. M. J. (2016). The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLOS ONE*, *11*(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

Boakes, R. A. (2003). The Impact of Pavlov on the Psychology of Learning in English-Speaking Countries. *Spanish Journal of Psychology*, *6*(2), 93–98. https://doi.org/10.1017/S1138741600005242

Bourgeois, J., & Minker, W. (Eds.). (2009). Linearly Constrained Minimum Variance Beamforming. In *Time-Domain Beamforming and Blind Source Separation* (pp. 27–38). https://doi.org/10.1007/978-0-387-68836-7_3

Boyle, S., Roche, B., Dymond, S., & Hermans, D. (2016). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion*, *30*(2), 340–352. https://doi.org/10.1080/02699931.2014.1000831

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 379–390. https://doi.org/10.1037/0278-7393.18.2.379

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, *37*(2), 204–215. http://www.ncbi.nlm.nih.gov/pubmed/10731770

Bröckelmann, A., Steinberg, C., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Jungh??fer, M. (2013). Affect-specific modulation of the N1m to shock-conditioned tones: Magnetoencephalographic correlates. *European Journal of Neuroscience*, *37*(2), 303–315. https://doi.org/10.1111/ejn.12043

Bröckelmann, A., Steinberg, C., Elling, L., Zwanzger, P., & Pantev, C. (2011). *Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing : Magnetoencephalographic Correlates*. *31*(21), 7801–7810. https://doi.org/10.1523/JNEUROSCI.6236-10.2011

Burkhouse, K. L., Owens, M., James, K., & Gibb, B. E. (2019). Age differences in electrocortical reactivity to fearful faces following aversive conditioning in youth. *Journal of Experimental Child Psychology*, *188*, 104676. https://doi.org/10.1016/j.jecp.2019.104676

Busch, C. J., & Evans, I. M. (1977). The effectiveness of electric shock and foul odor as unconditioned stimuli in classical aversive conditioning. *Behaviour Research and Therapy, 15*(2), 167–175. https://doi.org/10.1016/0005-7967(77)90101-2

Camfield, D. A., Mills, J., Kornfeld, E. J., & Croft, R. J. (2016). Modulation of the N170 with Classical Conditioning: The Use of Emotional Imagery and Acoustic Startle in Healthy and Depressed Participants. *Frontiers in Human Neuroscience, 10,* 337. https://doi.org/10.3389/fnhum.2016.00337

Carter, R. M. K., Hofstötter, C., Tsuchiya, N., & Koch, C. (2003). Working memory and fear conditioning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1399–1404. https://doi.org/10.1073/pnas.0334049100

Chaumon, M., Puce, A., & George, N. (2019). Statistical power: Implications for planning MEG studies. In *bioRxiv* (p. 852202). bioRxiv. https://doi.org/10.1101/852202

Ciocchi, S., Herry, C., Grenier, F., Wolff, S. B. E., Letzkus, J. J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M. B., Müller, C., & Lüthi, A. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature, 468*(7321), 277–282. https://doi.org/10.1038/nature09559

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, *56*(11). https://doi.org/10.1111/psyp.13437

Cohen, M. X. (2016). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2016.02.001

Corneille, O., & Mertens, G. (2020). Behavioral and Physiological Evidence Challenges the Automatic Acquisition of Evaluations. *Current Directions in Psychological Science*, 096372142096411. https://doi.org/10.1177/0963721420964111

Corneille, O., & Stahl, C. (2019). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dawson, M. E., Rissling, A. J., Schell, A. M., & Wilcox, R. (2007). Under What Conditions Can Human Affective Conditioning Occur Without Contingency Awareness?: Test of the Evaluative Conditioning Paradigm. *Emotion*, *7*(4), 755–766. https://doi.org/10.1037/1528-3542.7.4.755

de Cheveigné, A., & Nelken, I. (2019). Filters: When, Why, and How (Not) to Use Them. *Neuron*, *102*(2), 280–293. https://doi.org/10.1016/j.neuron.2019.02.039

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. In *Spanish Journal of Psychology* (Vol. 10, Issue 2, pp. 230–241). Cambridge University Press. https://doi.org/10.1017/S1138741600006491

de Jong, R., Lommen, M. J. J., de Jong, P. J., & Nauta, M. H. (2019). Using Multiple Contexts and Retrieval Cues in Exposure-Based Therapy to Prevent Relapse in Anxiety Disorders. *Cognitive and Behavioral Practice*, *26*(1), 154–165. https://doi.org/10.1016/j.cbpra.2018.05.002

DeBruine, L. M., & Barr, D. (2019). Understanding mixed effects models through data simulation. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.31234/osf.io/xp5cy

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural Circuitry Underlying the Regulation of Conditioned Fear and Its Relation to Extinction. *Neuron*, *59*(5), 829–838. https://doi.org/10.1016/j.neuron.2008.06.029

Diekhof, E. K., Geier, K., Falkai, P., & Gruber, O. (2011). Fear is only as deep as the mind allows. A coordinate-based meta-analysis of neuroimaging studies on the regulation of negative affect. *NeuroImage*, *58*(1), 275–285. https://doi.org/10.1016/j.neuroimage.2011.05.073

Dolan, R. J., Heinze, H. J., Hurlemann, R., & Hinrichs, H. (2006). Magnetoencephalography (MEG) determined temporal modulation of visual and auditory sensory processing in the context of classical conditioning to faces. *NeuroImage*, *32*(2), 778–789. https://doi.org/10.1016/j.neuroimage.2006.04.206

Dumas, T., Attal, Y., Dubal, S., Jouvent, R., & George, N. (2011). Detection of activity from the amygdala with magnetoencephalography. *IRBM*, *32*(1), 42–47. https://doi.org/10.1016/J.IRBM.2010.11.001

Dumas, T., Dubal, S., Attal, Y., Chupin, M., Jouvent, R., Morel, S., & George, N. (2013). MEG Evidence for Dynamic Amygdala Modulations by Gaze and Facial Emotions. *PLoS ONE*, *8*(9), 1–11. https://doi.org/10.1371/journal.pone.0074145

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., Mitroff, S. R., & LaBar, K. S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning and Memory*, *16*(7), 460–469. https://doi.org/10.1101/lm.1431609

Dunsmoor, J. E., & Paz, R. (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biological Psychiatry*, *78*, 336–343. https://doi.org/10.1016/j.biopsych.2015.04.010

Duvarci, S., & Pare, D. (2014). Amygdala microcircuits controlling learned fear. *Neuron*, *82*(5), 966–980. https://doi.org/10.1016/j.neuron.2014.04.042

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behavior Therapy*, *46*, 561–582. https://doi.org/10.1016/j.beth.2014.10.001

Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 278, Issue 1718, pp. 2553–2561). Proc Biol Sci. https://doi.org/10.1098/rspb.2011.0836

European Meeting of Human Fear Conditioning. (n.d.). Retrieved January 15, 2021, from https://emhfc.blogs.uni-hamburg.de/dfg-funded-research-network/

Ferreira de Sá, D. S., Michael, T., Wilhelm, F. H., & Peyk, P. (2019). Learning to see the threat: temporal dynamics of ERPs of motivated attention in fear conditioning. *Social Cognitive and Affective Neuroscience*, *14*(2), 189–203. https://doi.org/10.1093/scan/nsy103

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion*, *19*(2), 217–243. https://doi.org/10.1080/02699930441000292

Fossati, P. (2012). Neural correlates of emotion processing: From emotional to social brain. *European Neuropsychopharmacology*, *22*(SUPPL3), S487–S491. https://doi.org/10.1016/j.euroneuro.2012.07.008

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., Dunsmoor, J. E., Schruers, K. R. J., Savage, H. S., Bach, D. R., & Harrison, B. J. (2020). Human fear conditioning: From neuroscience to the clinic. *Behaviour Research and Therapy*, *124*(November 2019), 103528. https://doi.org/10.1016/j.brat.2019.103528

Fullana, M. A., Harrison, B. J., 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*(4), 500–508. https://doi.org/10.1038/mp.2015.88

Gaffey, A. E., & Wirth, M. M. (2014). Psychophysiological Measures. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 5181–5184). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_2315

García-Palacios, A., Costa, A., Castilla, D., Del Río, E., Casaponsa, A., & Duñabeitia, J. A. (2018). The effect of foreign language in fear acquisition. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-19352-8

Gawronski, B., & Mitchell, D. G. V. (2014). Simultaneous conditioning of valence and arousal. *Cognition and Emotion*, *28*(4), 577–595. https://doi.org/10.1080/02699931.2013.843506

Glenn, C. R., Lieberman, L., & Hajcak, G. (2012). Comparing electric shock and a fearful screaming face as unconditioned stimuli for fear learning. *International Journal of Psychophysiology*, *86*(3), 214–219. https://doi.org/10.1016/j.ijpsycho.2012.09.006

Glotzbach, E., Ewald, H., Andreatta, M., Pauli, P., & Mühlberger, A. (2012). Contextual fear conditioning predicts subsequent avoidance behaviour in a virtual reality environment. *Cognition & Emotion*, *26*(7), 1256–1272. https://doi.org/10.1080/02699931.2012.656581

Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, *7*(2), 377–387. https://doi.org/10.1037/1528-3542.7.2.377

Gonzalez, S. T., & Fanselow, M. S. (2020). The Role of the Ventromedial Prefrontal Cortex and Context in Regulating Fear Learning and Extinction. *Psychology and Neuroscience*, *13*(3), 459–472. https://doi.org/10.1037/pne0000207

Goodman, A. M., Harnett, N. G., & Knight, D. C. (2018). Pavlovian conditioned diminution of the neurobehavioral response to threat. *Neuroscience and Biobehavioral Reviews*, *84*(April 2017), 218–224. https://doi.org/10.1016/j.neubiorev.2017.11.021

Grillon, C., Baas, J. M. P., Cornwell, B., & Johnson, L. (2006). Context Conditioning and Behavioral Avoidance in a Virtual Reality Environment: Effect of Predictability. *Biological Psychiatry*, *60*(7), 752–759.

https://doi.org/10.1016/j.biopsych.2006.03.072

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J. R., van Wassenhove, V., Wibral, M., & Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363. https://doi.org/10.1016/j.neuroimage.2012.10.001

Gross, J. J., & John, O. P. (2003). Emotion Regulation Questionnaire. *NeuroImage*, *48*(10), 9–9. https://doi.org/10.1037/0022-3514.85.2.348

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: An overview and methodological challenges. In *Learning and Memory* (Vol. 21, Issue 9, pp. 424–440). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/lm.036053.114

Haaker, J., Maren, S., Andreatta, M., Merz, C. J., Richter, J., Richter, S. H., Meir Drexler, S., Lange, M. D., Jüngling, K., Nees, F., Seidenbecher, T., Fullana, M. A., Wotjak, C. T., & Lonsdorf, T. B. (2019). Making translation work: Harmonizing cross-species methodology in the behavioural neuroscience of Pavlovian fear conditioning. In *Neuroscience and Biobehavioral Reviews* (Vol. 107, pp. 329–345). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2019.09.020

Halladay, L. R., Zelikowsky, M., Blair, H. T., & Fanselow, M. S. (2012). Reinstatement of extinguished fear by an unextinguished conditional stimulus. *Frontiers in Behavioral Neuroscience*, *6*(MAY). https://doi.org/10.3389/fnbeh.2012.00018

Harrison, B. J., Fullana, M. A., Via, E., Soriano-Mas, C., Vervliet, B., Martínez-Zalacaín, I., Pujol, J., Davey, C. G., Kircher, T., Straube, B., & Cardoner, N. (2017). Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *NeuroImage*, *152*, 12–18. https://doi.org/10.1016/j.neuroimage.2017.02.080

Hawkins, R. D., & Byrne, J. H. (2015). Associative learning in invertebrates. *Cold Spring Harbor Perspectives in Biology*, *7*(5), 1–18. https://doi.org/10.1101/cshperspect.a021709

Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014). Time-frequency analysis of event-related potentials: A brief tutorial. In *Brain Topography* (Vol. 27, Issue 4, pp. 438–450). Springer New York LLC. https://doi.org/10.1007/s10548-013-0327-5

Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. In *Clinical Psychology Review* (Vol. 28, Issue 2, pp. 199–210). Elsevier Inc. https://doi.org/10.1016/j.cpr.2007.04.009

Hofmann, S. G., & Hay, A. C. (2018). Rethinking avoidance: Toward a balanced approach to avoidance in treating anxiety disorders. In *Journal of Anxiety Disorders* (Vol. 55, pp. 14–21). Elsevier Ltd. https://doi.org/10.1016/j.janxdis.2018.03.004

Hopkins, L. S., Schultz, D. H., Hannula, D. E., & Helmstetter, F. J. (2015). Eye movements index implicit memory expression in fear conditioning. *PLoS ONE*, *10*(11), 1–19. https://doi.org/10.1371/journal.pone.0141949

Huang, G., & Zhang, Z. (2017). Improving sensitivity of cluster-based permutation test for EEG/MEG data. *International IEEE/EMBS Conference on Neural Engineering, NER*, 9–12. https://doi.org/10.1109/NER.2017.8008279

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*(3), 539–557. https://doi.org/10.1037/a0026477

Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents: Implications for models of spatial navigation and memory. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 369, Issue 1635). Royal Society. https://doi.org/10.1098/rstb.2013.0304

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Jentsch, V. L., Wolf, O. T., & Merz, C. J. (2020). Temporal dynamics of conditioned skin conductance and pupillary responses during fear acquisition and extinction. *International Journal of Psychophysiology*, *147*(November 2019), 93–99. https://doi.org/10.1016/j.ijpsycho.2019.11.006

Jones, M. (1924). A laboratory study of fear: The case of peter. *Pedagogical Seminary and Journal of Genetic Psychology*, *31*(4), 308–315. https://doi.org/10.1080/08856559.1924.9944851

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, *68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Junghöfer, M., Bröckelmann, A. K., Küppers, K., Ohrmann, P., & Pedersen, A. (2015a). Abnormal, affect-specific modulatory effects on early auditory processing in schizophrenia: Magnetoencephalographic evidence.

Schizophrenia Research, *161*(2–3), 308–313. https://doi.org/10.1016/j.schres.2014.11.025

Junghöfer, M., Bröckelmann, A., Küppers, K., Ohrmann, P., & Pedersen, A. (2015b). Abnormal , affect-specific modulatory effects on early auditory processing in schizophrenia : Magnetoencephalographic evidence. *Schizophrenia Research*, *161*(2–3), 308–313. https://doi.org/10.1016/j.schres.2014.11.025

Junghöfer, M., Rehbein, M. A., Maitzen, J., Schindler, S., & Kissler, J. (2017). *An evil face ? Verbal evaluative multi-CS conditioning graphic responses. December 2016*, 695–705. https://doi.org/10.1093/scan/nsw179

Karalis, N., Dejean, C., Chaudun, F., Khoder, S., Rozeske, R. R., Wurtz, H., Bagur, S., Benchenane, K., Sirota, A., Courtin, J., & Herry, C. (2016). 4-Hz oscillations synchronize prefrontal–amygdala circuits during fear behavior. *Nature Neuroscience*, *19*(4), 605–612. https://doi.org/10.1038/nn.4251

Kattner, F. (2012). Revisiting the relation between contingency awareness and attention: Evaluative conditioning relies on a contingency focus. *Cognition & Emotion*, *26*(1), 166–175. https://doi.org/10.1080/02699931.2011.565036

Keil, J., & Senkowski, D. (2018). Neural Oscillations Orchestrate Multisensory Processing. In *Neuroscientist* (Vol. 24, Issue 6, pp. 609–626). SAGE Publications Inc. https://doi.org/10.1177/1073858418755352

Khemka, S., Barnes, G., Dolan, R. J., & Bach, D. R. (2017). Dissecting the Function of Hippocampal Oscillations in a Human Anxiety Model. *The Journal of Neuroscience*, *37*(29), 6869–6876. https://doi.org/10.1523/JNEUROSCI.1834-16.2017

Kindt, M. (2014). A behavioural neuroscience perspective on the aetiology and treatment of anxiety disorders. *Behaviour Research and Therapy*, *62*, 24–36. https://doi.org/10.1016/j.brat.2014.08.012

Klein, S. (2019). An Introduction to Learning. In *Learning: Principles and applications* (8th ed., pp. 1–15). Sage Publications, Inc.

Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., & Stark, R. (2009). Neural, electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture–picture conditioning paradigm. *Neuroscience*, *158*(2), 721–731. https://doi.org/10.1016/j.neuroscience.2008.09.049

Kluge, C., Bauer, M., Leff, A. P., Heinze, H.-J., Dolan, R. J., & Driver, J. (2011). Plasticity of human auditory-evoked fields induced by shock conditioning and contingency reversal. *Proceedings of the National Academy of Sciences*, *108*(30), 12545–12550. https://doi.org/10.1073/pnas.1016124108

Koenig, S., Nauroth, P., Lucke, S., Lachnit, H., Gollwitzer, M., & Uengoer, M. (2017). Fear acquisition and liking of out-group and in-group members: Learning bias or attention? *Biological Psychology*, *129*, 195–206. https://doi.org/10.1016/j.biopsycho.2017.08.060

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

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

Krypotos, A.-M. (2015). Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*(July), 189. https://doi.org/10.3389/fnbeh.2015.00189

Kumar, S., Forster, H. M., Bailey, P., & Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *The Journal of the Acoustical Society of America*, *124*(6), 3810–3817. https://doi.org/10.1121/1.3006380

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. *Neuron*, *20*(5), 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4

Lakens, D. (2021). *Sample Size Justification*. https://doi.org/10.31234/OSF.IO/9D3YF

Larson, M. J. (2020). Improving the Rigor and Replicability of Applied Psychophysiology Research: Sample Size, Standardization, Transparency, and Preregistration. *Biofeedback*, *48*(1), 2–6. https://doi.org/10.5298/1081-5937-48.1.2

Le Pelley, M. E. (2004). The Role of Associative History in Models of Associative Learning: A Selective Review and a Hybrid Model. *The Quarterly Journal of Experimental Psychology Section B*, *57*(3b), 193–243. https://doi.org/10.1080/02724990344000141

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience*, *23*(1), 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, *111*(8), 2871–2878. https://doi.org/10.1073/pnas.1400335111

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

Leopold, D. A., & Rhodes, G. (2010). A Comparative view of face perception. *Journal of Comparative Psychology*, *124*(3), 233–251. https://doi.org/10.1037/a0019460

Lesting, J., Daldrup, T., Narayanan, V., Himpe, C., Seidenbecher, T., & Pape, H. C. (2013). Directional Theta Coherence in Prefrontal Cortical to Amygdalo-Hippocampal Pathways Signals Fear Extinction. *PLoS ONE*, *8*(10), 17–19. https://doi.org/10.1371/journal.pone.0077707

Lesting, J., Narayanan, R. T., Kluge, C., Sangha, S., Seidenbecher, T., & Pape, H.-C. (2011). Patterns of Coupled Theta Activity in Amygdala-Hippocampal-Prefrontal Cortical Circuits during Fear Extinction. *PLoS ONE*, *6*(6), e21714. https://doi.org/10.1371/journal.pone.0021714

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017). Neural correlates of pupil dilation during human fear learning. *NeuroImage*, *147*, 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

Liao, H. I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin and Review*, *23*(2), 412–425. https://doi.org/10.3758/s13423-015-0898-0

Likhtik, E., Stujenske, J. M., Topiwala, M. A., Harris, A. Z., & Gordon, J. A. (2014). Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. *Nature Neuroscience*, *17*(1), 106–113. https://doi.org/10.1038/nn.3582

Lin, F.-H., Witzel, T., Raij, T., Ahveninen, J., Tsai, K. W.-K., Chu, Y.-H., Chang, W.-T., Nummenmaa, A., Polimeni, J. R., Kuo, W.-J., Hsieh, J.-C., Rosen, B. R., & Belliveau, J. W. (2013). fMRI hemodynamics accurately reflects neuronal timing in the human brain measured by MEG. *NeuroImage*, *78*, 372–384. https://doi.org/10.1016/j.neuroimage.2013.04.017

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition and Emotion*, *19*(2), 269–282. https://doi.org/10.1080/02699930441000319

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391–1424. https://doi.org/10.1016/j.brat.2004.10.007

Lithari, C., Moratti, S., & Weisz, N. (2015). *Thalamocortical Interactions Underlying Visual Fear Conditioning in Humans*. *4603*(February), 4592–4603. https://doi.org/10.1002/hbm.22940

Lithari, C., Moratti, S., & Weisz, N. (2016). Limbic areas are functionally decoupled and visual cortex takes a more central role during fear conditioning in humans. *Scientific Reports*, *6*(July), 29220. https://doi.org/10.1038/srep29220

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910–916. https://doi.org/10.1038/nn909

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V. L., Drexler, S. M., Mertens, G., Richter, J., Sjouwerman, R., Wendt, J., & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning research. *ELife*, *8*. https://doi.org/10.7554/eLife.52465

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. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience and Biobehavioral Reviews*, *80*(April), 703–728. https://doi.org/10.1016/j.neubiorev.2017.07.007

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. In *Journal of Experimental Psychology: Animal Behavior Processes* (Vol. 28, Issue 1, pp. 3–26). https://doi.org/10.1037/0097-7403.28.1.3

Luck, C. C., & Lipp, O. V. (2015a). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy*, *66*, 18–31. https://doi.org/10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2015b). To remove or not to remove? Removal of the unconditional stimulus electrode does not mediate instructed extinction effects. *Psychophysiology*, *52*(9), 1248–1256. https://doi.org/10.1111/psyp.12452

Lueschow, A., Weber, J. E., Carbon, C.-C., Deffke, I., Sander, T., Grüter, T., Grüter, M., Trahms, L., & Curio, G. (2015). The 170ms Response to Faces as Measured by MEG (M170) Is Consistently Altered in Congenital Prosopagnosia. *PLOS ONE*, *10*(9), e0137624. https://doi.org/10.1371/journal.pone.0137624

Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), 14563–14568. https://doi.org/10.1073/pnas.1515087112

Maren, S. (2014). Nature and causes of the immediate extinction deficit: A brief review. In *Neurobiology of Learning and Memory* (Vol. 113, pp. 19–24). Academic Press Inc. https://doi.org/10.1016/j.nlm.2013.10.012

Marschner, A., Kalisch, R., Vervliet, B., Vansteenwegen, D., & Buchel, C. (2008). Dissociable Roles for the Hippocampus and the Amygdala in Human Cued versus Context Fear Conditioning. *Journal of Neuroscience*, *28*(36), 9030–9036. https://doi.org/10.1523/JNEUROSCI.1651-08.2008

McCullough, K. M., Morrison, F. G., & Ressler, K. J. (2016). Bridging the Gap: Towards a cell-type specific understanding of neural circuits underlying fear behaviors. *Neurobiology of Learning and Memory*, *135*, 27–39. https://doi.org/10.1016/j.nlm.2016.07.025

McEchron, M. D., McCabe, P. M., Green, E. J., Llabre, M. M., & Schneiderman, N. (1992). Air puff versus shock unconditioned stimuli in rabbit heart rate conditioning. *Physiology & Behavior*, *51*(1), 195–199. https://doi.org/10.1016/0031-9384(92)90223-O

Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage*, *49*(2), 1760–1768. https://doi.org/10.1016/j.neuroimage.2009.09.040

Mertens, G., Boddez, Y., KrypotosIris, A.-M., & Engelhard, I. (2020). *PsyArXiv Preprints | Human fear conditioning depends on stimulus contingency instructions*. https://psyarxiv.com/by964/

Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. In *Neuroscience and Biobehavioral Reviews* (Vol. 108, pp. 254–268). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2019.11.012

Meyer, S. S., Bonaiuto, J., Lim, M., Rossiter, H., Waters, S., Bradbury, D., Bestmann, S., Brookes, M., Callaghan, M. F., Weiskopf, N., & Barnes, G. R. (2017). Flexible head-casts for high spatial precision MEG. *Journal of*

Neuroscience Methods, *276*, 38–45. https://doi.org/10.1016/j.jneumeth.2016.11.009

Meyer, S. S., Rossiter, H., Brookes, M. J., Woolrich, M. W., Bestmann, S., & Barnes, G. R. (2017). Using generative models to make probabilistic statements about hippocampal engagement in MEG. *NeuroImage*, *149*, 468–482. https://doi.org/10.1016/j.neuroimage.2017.01.029

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., Zeidan, M. A., Handwerger, K., Orr, S. P., & Rauch, S. L. (2009). Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder. *Biological Psychiatry*, *66*(12), 1075–1082. https://doi.org/10.1016/j.biopsych.2009.06.026

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annual Review of Psychology*, *63*, 129–151. https://doi.org/10.1146/annurev.psych.121208.131631

Mills, T., Lalancette, M., Moses, S. N., Taylor, M. J., & Quraan, M. A. (2012). Techniques for detection and localization of weak hippocampal and medial frontal sources using beamformers in MEG. *Brain Topography*, *25*(3), 248–263. https://doi.org/10.1007/s10548-012-0217-2

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: It's not what you thought it was. *American Psychologist*, *61*(1), 10–26. https://doi.org/10.1037/0003-066X.61.1.10

Miskovic, V., & Keil, A. (2012). Acquired fears reflected in cortical sensory processing: A review of electrophysiological studies of human classical conditioning. In *Psychophysiology* (Vol. 49, Issue 9, pp. 1230–1241). Blackwell Publishing Inc. https://doi.org/10.1111/j.1469-8986.2012.01398.x

Morand-Ferron, J. (2017). Why learn? The adaptive value of associative learning in wild populations. In *Current Opinion in Behavioral Sciences* (Vol. 16, pp. 73–79). Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2017.03.008

Moratti, S., Giménez-Fernández, T., Méndez-Bértolo, C., & de Vicente-Pérez, F. (2017). Conditioned inhibitory and excitatory gain modulations of visual cortex in fear conditioning: Effects of analysis strategies of magnetocortical responses. *Psychophysiology*, *54*(6), 882–893. https://doi.org/10.1111/psyp.12841

Moratti, S., & Keil, A. (2005). *Cortical activation during Pavlovian fear conditioning depends on heart rate response patterns : An MEG study*. *25*, 459–471. https://doi.org/10.1016/j.cogbrainres.2005.07.006

Moratti, S., Keil, A., & Miller, G. A. (2006). Fear but not awareness predicts enhanced sensory processing in fear conditioning. *Psychophysiology*, *43*(2), 216–226. https://doi.org/10.1111/j.1464-8986.2006.00386.x

Morriss, J., Christakou, A., & van Reekum, C. M. (2015). Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. *Biology of Mood and Anxiety Disorders*, *5*(1), 4. https://doi.org/10.1186/s13587-015-0019-8

Moses, S. N., Houck, J. M., Martin, T., Hanlon, F. M., Ryan, J. D., Thoma, R. J., Weisend, M. P., Jackson, E. M., Pekkonen, E., Tesche, C. D., Centre, B., & Ma, O. (2007). *Dynamic neural activity recorded from human amygdala during fear conditioning using magnetoencephalography*. *71*, 452–460. https://doi.org/10.1016/j.brainresbull.2006.08.016

Moses, S. N., Martin, T., Houck, J. M., Ilmoniemi, R. J., & Tesche, C. D. (2005). The C50m response: Conditioned magnetocerebral activity recorded from the human brain. *NeuroImage*, *27*(4), 778–788. https://doi.org/10.1016/J.NEUROIMAGE.2005.05.017

Mowrer, O. H. (1947). On the dual nature of learning—a re-interpretation of “conditioning” and “problem-solving.” - PsycNET. *Harvard Educational Review, 17,* 102–148.

Mumford, J. A. (2012). A power calculation guide for FMRI studies. *Social Cognitive and Affective Neuroscience*, *7*(6), 738–742. https://doi.org/10.1093/scan/nss059

Murphy, E. S., & Lupfer, G. J. (2014). Basic Principles of Operant Conditioning. In *The Wiley Blackwell Handbook of Operant and Classical Conditioning* (pp. 165–194). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468135.ch8

Neumann, D. L., & Waters, A. M. (2006a). The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. *Biological Psychology*, *73*(2), 175–185. https://doi.org/10.1016/j.biopsycho.2006.03.004

Neumann, D. L., & Waters, A. M. (2006b). The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. *Biological Psychology*, *73*(2), 175–185. https://doi.org/10.1016/j.biopsycho.2006.03.004

Ney, L. J., Wade, M., Reynolds, A., Zuj, D. V., Dymond, S., Matthews, A., & Felmingham, K. L. (2018). Critical evaluation of current data analysis strategies for psychophysiological measures of fear conditioning and extinction in humans. In *International Journal of Psychophysiology* (Vol. 134, pp. 95–107). Elsevier B.V. https://doi.org/10.1016/j.ijpsycho.2018.10.010

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning and*

Memory, *13*(6), 681–685. https://doi.org/10.1101/lm.393906

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. In *Neuroscience and Biobehavioral Reviews* (Vol. 114, pp. 96–112). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2020.04.019

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pastor, M. C., Rehbein, M. A., Junghõfer, M., Poy, R., López, R., & Moltó, J. (2015). Facing challenges in differential classical conditioning research: Benefits of a hybrid design for simultaneous electrodermal and electroencephalographic recording. *Frontiers in Human Neuroscience*, *9*(June), 336. https://doi.org/10.3389/fnhum.2015.00336

Pavlov, I. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. I. P. Pavlov , G. V. Anrep* (Issue 6). Oxford University Press. https://doi.org/10.1086/214896

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology*, *46*(5), 984-995. https://doi.org/10.1111/j.1469-8986.2009.00852.x

Pischek-Simpson, L. K., Boschen, M. J., Neumann, D. L., & Waters, A. M. (2009). The development of an attentional bias for angry faces following Pavlovian fear conditioning. *Behaviour Research and Therapy*, *47*(4), 322–330. https://doi.org/10.1016/j.brat.2009.01.007

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Prokasy, W. F., & Ebel, H. C. (1967). Three components of the classically conditioned GSR in human subjects. *Journal of Experimental Psychology*, *73*(2), 247–256. https://doi.org/10.1037/h0024108

Quraan, M. A., Moses, S. N., Hung, Y., Mills, T., & Taylor, M. J. (2011a). *Detection and Localization of Hippocampal Activity Using Beamformers with*

MEG : A Detailed Investigation Using Simulations and Empirical Data. *827*, 812–827. https://doi.org/10.1002/hbm.21068

Quraan, M. A., Moses, S. N., Hung, Y., Mills, T., & Taylor, M. J. (2011b). Detection and localization of hippocampal activity using beamformers with MEG: A detailed investigation using simulations and empirical data. *Human Brain Mapping*, *32*(5), 812–827. https://doi.org/10.1002/hbm.21068

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

Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). Human auditory cortex is activated by omission of auditory stimuli. *Brain Research*, *745*(1–2), 134–143. https://doi.org/10.1016/S0006-8993(96)01140-7

Raven, J. C. (1941). Standardization of the progressive matrices, 1938. *British Journal of Medical Psychology*, *19*(1), 137–150. https://doi.org/10.1111/j.2044-8341.1941.tb00316.x

Rehbein, M. A., Steinberg, C., Wessing, I., Pastor, M. C., Zwitserlood, 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

Rehbein, M. A., Wessing, I., Zwitserlood, P., Steinberg, C., Eden, A. S., Dobel, C., Junghöfer, M., Beck, K. D., & Flor, H. (2015). *Rapid prefrontal cortex activation towards aversively paired faces and enhanced contingency detection are observed in highly trait-anxious women under challenging conditions*. *9*(June), 1–19. https://doi.org/10.3389/fnbeh.2015.00155

Reinhard, G., & Lachnit, H. (2002). Differential conditioning of anticipatory pupillary dilation responses in humans. *Biological Psychology*, *60*(1), 51–68. https://doi.org/10.1016/S0301-0511(02)00011-X

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

Reips, U. D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in internet-based research: VAS generator. *Behavior Research Methods*, *40*(3), 699–704. https://doi.org/10.3758/BRM.40.3.699

Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y., & LeDoux, J. E. (2001). Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neuroscience*, *4*(7), 724–731.

https://doi.org/10.1038/89512

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. In *Schizophrenia Bulletin* (Vol. 34, Issue 5, pp. 907–926). Oxford Academic. https://doi.org/10.1093/schbul/sbn093

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

Rossion, B. (2014). "Understanding face perception by means of human electrophysiology" Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, *18*(6), 310–318. https://doi.org/10.1016/j.tics.2014.02.013

Rousselet, G. A. (2012). Does Filtering Preclude Us from Studying ERP Time-Courses? *Frontiers in Psychology*, *3*(May), 1–9. https://doi.org/10.3389/fpsyg.2012.00131

Rousselet, G. A., Pernet, C. R., & Wilcox, R. (2017). Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13610

Rowles, M. E., Lipp, O. V., & Mallan, K. M. (2012). On the resistance to extinction of fear conditioned to angry faces. *Psychophysiology*, *49*(3), 375–380. https://doi.org/10.1111/j.1469-8986.2011.01308.x

Ruan, X., & Wu, X. (2013). The skinner automaton: A psychological model formalizing the theory of operant conditioning. *Science China Technological Sciences*, *56*(11), 2745–2761. https://doi.org/10.1007/s11431-013-5369-0

Sánchez, J. A., & Halliday, D. M. (2013). Reducing the effect of correlated brain sources in MEG using a linearly constrained spatial filter based on Minimum Norm. *Conference Record - Asilomar Conference on Signals, Systems and Computers*, 1828–1832. https://doi.org/10.1109/ACSSC.2013.6810618

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schultz, D. H., & Helmstetter, F. J. (2010). Classical Conditioning of Autonomic Fear Responses Is Independent of Contingency Awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(4), 495–500. https://doi.org/10.1037/a0020263

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., Zwitserlood, P., Schiffbauer, H., Heindel, W., Arolt, V., & Konrad, C. (2011). Neural correlates of trait anxiety in fear extinction. *Psychological Medicine*, *41*(4), 789–798. https://doi.org/10.1017/S0033291710001248

Sehlmeyer, C., Schö Ning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human Fear Conditioning and Extinction in Neuroimaging: A Systematic Review. *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005865

Seidenbecher, T., Laxmi, T. R., Stork, O., & Pape, H. C. (2003). Amygdalar and Hippocampal Theta Rhythm Synchronization During Fear Memory Retrieval. *Science*, *301*(5634), 846–850. https://doi.org/10.1126/science.1085818

Sevenster, D., Beckers, T., & Kindt, M. (2012). Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. *Psychophysiology*, *49*(10), 1426–1435. https://doi.org/10.1111/j.1469-8986.2012.01450.x

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

Shechner, T., Hong, M., Britton, J. C., Pine, D. S., & Fox, N. A. (2014). Fear conditioning and extinction across development: Evidence from human studies and animal models. In *Biological Psychology* (Vol. 100, Issue 1, pp. 1–12). Elsevier. https://doi.org/10.1016/j.biopsycho.2014.04.001

Shin, L. M., McNally, R. J., Kosslyn, S. M., Thompson, W. L., Rauch, S. L., Alpert, N. M., Metzger, L. J., Lasko, N. B., Orr, S. P., & Pitman, R. K. (1999). Regional cerebral blood flow during script-driven imagery in childhood sexual abuse-related PTSD: A PET investigation. *American Journal of Psychiatry*, *156*(4), 575–584. https://doi.org/10.1176/ajp.156.4.575

Singh, K., Dawson, M. E., Schell, A. M., Courtney, C. G., & Payne, A. F. H. (2013). Can human autonomic classical conditioning occur without contingency awareness? The critical importance of the trial sequence. *Biological Psychology*, *93*(1), 197–205. https://doi.org/10.1016/j.biopsycho.2013.02.007

Sjouwerman, R., & Lonsdorf, T. B. (2019). Latency of skin conductance responses across stimulus modalities. *Psychophysiology*, *56*(4), e13307. https://doi.org/10.1111/psyp.13307

Sperl, M. F. J., Panitz, C., & Hermann, C. (2016). *A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials*. *53*, 1352–1365. https://doi.org/10.1111/psyp.12677

Sperl, M. F. J., Panitz, C., Hermann, C., & Mueller, E. M. (2016). A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials. *Psychophysiology*, *53*(9), 1352–1365. https://doi.org/10.1111/psyp.12677

Sperl, M. F. J., Panitz, C., Rosso, I. M., Dillon, D. G., Kumar, P., Hermann, A., Whitton, A. E., Hermann, C., Pizzagalli, D. A., & Mueller, E. M. (2018). Fear Extinction Recall Modulates Human Frontomedial Theta and Amygdala Activity. *Cerebral Cortex*, *April*, 1–15. https://doi.org/10.1093/cercor/bhx353

Sperl, M. F. J., Wroblewski, A., Mueller, M., Straube, B., & Mueller, E. M. (2021). Learning dynamics of electrophysiological brain signals during human fear conditioning. *NeuroImage*, *226*, 117569. https://doi.org/10.1016/j.neuroimage.2020.117569

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*.

Steinberg, C., Bröckelmann, A., 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. (2012). 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

Stolz, C., Endres, D., & Mueller, E. M. (2019). Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study. *Psychophysiology*, *56*(4). https://doi.org/10.1111/psyp.13308

Subramaniyam, N. (2018). *Pitfalls of Filtering the EEG Signal | Sapien Labs | Neuroscience | Human Brain Diversity Project*. https://sapienlabs.co/pitfalls-of-filtering-the-eeg-signal/

Sung, Y. T., & Wu, J. S. (2018). The Visual Analogue Scale for Rating, Ranking and Paired-Comparison (VAS-RRP): A new technique for psychological measurement. *Behavior Research Methods*, *50*(4), 1694–1715. https://doi.org/10.3758/s13428-018-1041-8

Szucs, D., & Ioannidis, J. P. A. (2019). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. In *bioRxiv* (p. 809715). bioRxiv. https://doi.org/10.1101/809715

Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T.,

& Stark, R. (2011). Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. *Social Cognitive and Affective Neuroscience*, *6*(4), 495–506. https://doi.org/10.1093/scan/nsq070

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*, 1–13. https://doi.org/10.1155/2011/879716

Tesche, C. D., Moses, S. N., Houck, J. M., Martin, T., Hanlon, F. M., Jackson, E., & Kičić, D. (2007). Dynamics of frontal and cerebellar activation during aversive conditioning: An MEG study. *International Congress Series*, *1300*, 437–440. https://doi.org/10.1016/j.ics.2007.02.057

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4), i–109. https://doi.org/10.1037/h0092987

Tierney, T. M., Mellor, S., O'Neill, G. C., Holmes, N., Boto, E., Roberts, G., Hill, R. M., Leggett, J., Bowtell, R., Brookes, M. J., & Barnes, G. R. (2020). Pragmatic spatial sampling for wearable MEG arrays. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-77589-8

Tovote, P., Fadok, J. P., & Lüthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews Neuroscience*, *16*(6), 317–331. https://doi.org/10.1038/nrn3945

Trenado, C., Pedroarena-leal, N., Cif, L., Nitsche, M., & Ruge, D. (2018). *Neural Oscillatory Correlates for Conditioning and Extinction of Fear*. 1–10. https://doi.org/10.3390/biomedicines6020049

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. *PLOS Computational Biology*, *14*(8), e1006243. https://doi.org/10.1371/journal.pcbi.1006243

Tzovara, A., Meyer, S. S., Bonaiuto, J. J., Abivardi, A., Dolan, R. J., Barnes, G. R., & Bach, D. R. (2019). High-precision magnetoencephalography for reconstructing amygdalar and hippocampal oscillations during prediction of safety and threat. *Human Brain Mapping*, *40*(14), 4114–4129. https://doi.org/10.1002/hbm.24689

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*. https://doi.org/10.1038/ncomms14637

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*. https://doi.org/10.1177/2331216519832483

VanRullen, R. (2011a). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*(DEC), 1–6. https://doi.org/10.3389/fpsyg.2011.00365

VanRullen, R. (2011b). Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Frontiers in Psychology*, *2*, 365. https://doi.org/10.3389/fpsyg.2011.00365

Västfjäll, D. (2013). Emotional Reactions to Tonal and Noise Components of Environmental Sounds. *Psychology*, *04*(12), 1051–1058. https://doi.org/10.4236/psych.2013.412153

Vaurio, R. (2011). Symptom Checklist-90-Revised. In *Encyclopedia of Clinical Neuropsychology* (pp. 2447–2450). Springer New York. https://doi.org/10.1007/978-0-387-79948-3_2012

Vinograd, M., & Craske, M. G. (2020). History and theoretical underpinnings of exposure therapy. In *Exposure Therapy for Children with Anxiety and OCD* (pp. 3–20). Elsevier. https://doi.org/10.1016/b978-0-12-815915-6.00001-9

Visser, R. M., de Haan, M. I. C., Beemsterboer, T., Haver, P., Kindt, M., & Scholte, H. S. (2016). Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology*, *53*(8), 1117–1127. https://doi.org/10.1111/psyp.12665

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology*, *55*, 8–20. https://doi.org/10.1016/j.psyneuen.2015.01.021

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience*, *16*(4), 388–390. https://doi.org/10.1038/nn.3345

Wagenmakers, E. J., Krypotos, A. M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory and Cognition*, *40*(2), 145–160. https://doi.org/10.3758/s13421-011-0158-0

Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning? *Annual Review of Psychology*, *48*, 573–607. https://doi.org/10.1146/annurev.psych.48.1.573

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1-14. https://doi.org/10.1037/h0069608

Weidemann, G., Satkunarajah, M., & Lovibond, P. F. (2016). I Think, Therefore Eyeblink: The Importance of Contingency Awareness in Conditioning. *Psychological Science*, *27*(4), 467–475. https://doi.org/10.1177/0956797615625973

Weike, A. I., Schupp, H. T., & Hamm, A. O. (2006). *Fear acquisition requires awareness in trace but not delay conditioning*. https://doi.org/10.1111/j.1469-8986.2006.00469.x

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biology*, *13*(4), 1–10. https://doi.org/10.1371/journal.pbio.1002128

Wendt, J., Hufenbach, M. C., König, J., & Hamm, A. O. (2020). Effects of verbal instructions and physical threat removal prior to extinction training on the return of conditioned fear. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-57934-7

Westfall, J., Nichols, T. E., & Yarkoni, T. (2017). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*, 23. https://doi.org/10.12688/wellcomeopenres.10298.2

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data - a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Wiens, S., Katkin, E. S., & Öhman, A. (2003). Effects of trial order and differential conditioning on acquisition of differential shock expectancy and skin conductance conditioning to masked stimuli. *Psychophysiology*, *40*(6), 989–997. https://doi.org/10.1111/1469-8986.00117

Wilcox, R. (2017). Copyright. In *Introduction to Robust Estimation and Hypothesis Testing* (4th ed.). Academic Press. https://doi.org/10.1016/B978-0-12-804733-0.00015-9

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. In *Trends in Hearing* (Vol. 22). SAGE Publications Inc. https://doi.org/10.1177/2331216518800869

Wolpe, J. (1968). Psychotherapy by reciprocal inhibition. *Conditional Reflex : A Pavlovian Journal of Research & Therapy*, *3*(4), 234–240.

https://doi.org/10.1007/BF03000093

Yarkoni, T. (2020). *PsyArXiv Preprints | The Generalizability Crisis*. https://psyarxiv.com/jqw35

Zalta, A. K., & Foa, E. B. (2012). Exposure Therapy: Promoting Emotional Processing of Pathological Anxiety. In W. O'donohue & J. E. Fisher (Eds.), *Cognitive Behavior Therapy Core Principles for Practice* (pp. 75–105). y John Wiley & Sons, Inc. www.wiley.com.