



UvA-DARE (Digital Academic Repository)

'Episodic Threat Conditioning': a novel approach to simultaneously measure Pavlovian threat conditioning and episodic memory

de Vries, O.T.; Duken, S.B.; Dzinalija, N.; Kindt, M.; van Ast, V.A.

DOI

[10.1016/j.brat.2025.104882](https://doi.org/10.1016/j.brat.2025.104882)

Publication date

2025

Document Version

Final published version

Published in

Behaviour Research and Therapy

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

Citation for published version (APA):

de Vries, O. T., Duken, S. B., Dzinalija, N., Kindt, M., & van Ast, V. A. (2025). 'Episodic Threat Conditioning': a novel approach to simultaneously measure Pavlovian threat conditioning and episodic memory. *Behaviour Research and Therapy*, 195, Article 104882. <https://doi.org/10.1016/j.brat.2025.104882>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



‘Episodic Threat Conditioning’: a novel approach to simultaneously measure Pavlovian threat conditioning and episodic memory[☆]

Olivier T. de Vries^{a,b,*}, Sascha B. Duken^{a,b}, Nadža Džinalija^c, Merel Kindt^{a,b},
Vanessa A. van Ast^{a,b,**}

^a Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands

^b Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, the Netherlands

^c Free University Medical Center, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Associative memory
Threat conditioning
Pupil dilation
fEMG

ABSTRACT

The powerful ways future behavior and cognition can be affected by emotional events are typically studied either by means of Pavlovian conditioning or episodic memory paradigms, which both rest on the idea that associations are formed between distinct stimuli experienced closely together in space or time. However, due to their incompatible methods, little is known about how physiological read-outs of Pavlovian conditioning and episodic memory work in concert to affect behavior and other cognitive processes, how they relate to each other, or whether they reflect the activity of distinct associative processes in the first place. To shed light on such questions, a paradigm is needed that can assess both conditioned psychophysiology and episodic memory. We reasoned that multimodal stimuli, consisting of congruent image-sound combinations, have the potential to serve as potent unconditioned stimuli in a paradigm where participants encode a large number ‘mini-conditioning events’. Measuring both pupil dilation and facial electromyography, we found that psychophysiological responses to 20 unique aversive USs – but not positive USs – transferred to arbitrary predictors already after a single paired presentation in this novel paradigm. Real-life emotional aversive and stressful events are likely to involve both Pavlovian conditioning and episodic memory processes. The ‘Episodic Threat Conditioning’ paradigm enables their simultaneous assessment, thereby providing an opportunity to gain more holistic insight into how different expressions of memory interact in mental health and disease.

1. Introduction

Emotional experiences have a special status in memory and can thereby profoundly change the course of our lives. By means of both declarative and non-declarative memory processes, they strongly influence current and future behavior (LaBar & Cabeza, 2006; Squire, 2004). The multiple avenues through which emotion exerts its effects on memory are, however, studied in relative isolation by separate research groups using distinct techniques and measures. As a result, our understanding of how learning and memory are modulated by emotion is fractured across research areas, predominantly *Pavlovian conditioning* and *episodic memory*. The former measures anticipatory physiological

responses to learned predictors of significant events (Rescorla, 1987), whereas the latter studies the accuracy and subjective experience of conscious recollections (Tulving, 1983). The separation of research on these phenomena is typically justified by assuming that they hinge on distinct neural systems that are by and large independent from each other in how they function (Squire, 2004), but the degree to which this is in fact true remains contested. Moreover, even if the associative mechanisms that underly the effects found in Pavlovian conditioning and episodic memory studies are fully distinct, their conjoint activity may still have important interaction effects on measurable outputs that are currently precluded from investigation. Indeed, a recent review has argued for the importance of a holistic study of emotional learning and

[☆] All materials and individual stimulus data can be accessed through the following link: https://osf.io/ca3t6/?view_only=851b7df58159440aaa61576e33aafd35.

* Corresponding author. Faculty of Social and Behavioral Science, Department of Clinical Psychology, Nieuwe Achtergracht 129, 1018 WS, Amsterdam, the Netherlands.

** Corresponding author. Faculty of Social and Behavioral Science, Department of Clinical Psychology, Nieuwe Achtergracht 129, 1018 WS, Amsterdam, the Netherlands.

E-mail addresses: o.t.devries@uva.nl (O.T. de Vries), v.a.vanAst@uva.nl (V.A. van Ast).

<https://doi.org/10.1016/j.brat.2025.104882>

Received 11 October 2023; Received in revised form 20 July 2025; Accepted 30 September 2025

Available online 30 September 2025

0005-7967/© 2025 Published by Elsevier Ltd.

memory expressions, especially in regard to maladaptive instances (Dunsmoor & Kroes, 2019). Here, we present a new method (dubbed “Episodic threat conditioning”) that allows for simultaneous assessment of Pavlovian conditioning and episodic memory, and has its own niche of applications that complements existing paradigms that have been developed towards the same end.

Pavlovian conditioning is one of the oldest and most widely used procedures in experimental psychology and affective neuroscience. It consists in the repeated presentation of a neutral stimulus (conditioned stimulus; CS) followed by a biologically significant event (unconditioned stimulus; US), such that the CS comes to elicit physiological responses (conditioned response; CR) in anticipation of the US it predicts (Rescorla, 1987). Learning from experience which cues in the environment signal the onset of threat or reward clearly provides a large evolutionary advantage. As such, it is perhaps unsurprising that the neural mechanisms that support Pavlovian conditioning are considered to be highly preserved across species (LeDoux & Pine, 2016), and that the field is still flourishing more than a century after its conception (Vervliet & Boddez, 2020). What drives most of the current research interest today, however, is its potential to explain the etiology and maintenance of anxiety disorders in terms of behavior and neurobiology (Kindt, 2014).

Within this framework, feelings of fear and anxiety are a part of initially adaptive CRs (Frijda, 1986) that only become maladaptive when they persist despite subsequent disconfirmatory evidence regarding the CS-US relationship, or generalize to similar cues (Duits et al., 2015). Such stubborn persistence (Craske et al., 2008) and overgeneralization (Lissek et al., 2014) of emotional responses to related but harmless signals are key targets for clinical interventions that Pavlovian conditioning allows us to quantify, meaning it provides an valuable model by which treatments can be discovered or further optimized.

Although the simplicity of Pavlovian conditioning can be seen as one of its main strengths, there are worries that the basic procedure is lacking in several ways that limit its external validity and clinical applications (Beckers et al., 2013; Vervliet & Boddez, 2020). First, the vast majority of Pavlovian conditioning studies have used mild electric shocks as the US and geometric shapes as CSs: one that predicts the US and one that does not. One important limitation therefore concerns the role of higher-level cognitive processing in predictive learning and retrieval, which is particularly relevant to theories of associative learning (Mitchell et al., 2009), but virtually untestable using only a handful of stimuli. Simply put, when participants learn only one or two CS-US pairings, with few exceptions, they will have perfect memory for each CS and its associated outcome. This issue is partly addressed by the category conditioning paradigm developed by Dunsmoor and colleagues (Dunsmoor et al., 2012). By employing a large number of CSs, each of which belongs to a semantic category that predicts shock or one that does not, it enables testing whether later recognition of a CS is affected by its association with the reinforced category. The category conditioning paradigm has yielded important insights into the effects of reinforcement on CS recognition (Dunsmoor & Murphy, 2015), and has been employed as a building block for a variety of studies into emotional memory. However, it is less well suited for studies into the effects of emotion on the encoding and retrieval of unique events. For any CS, participants can report the associated US based on whether it belongs to the reinforced or unreinforced semantic category, without having to retrieve the specific memory of a particular CS being followed by a particular aversive outcome. For a conditioning paradigm to probe the encoding and retrieval of specific, episodic-like events, a CS should 1) not have any perceptual or semantic features from which its association with a US can be inferred, and 2) be associated to a unique US.

A second limitation of traditional Pavlovian conditioning is that it strongly relies on just a handful of established simple USs. Shocks and loud sounds are highly effective negative reinforcers, but lack some of the key features that may characterize an associative memory that is formed following an aversive experience in real life. Specifically, they

lack the perceptual richness and semantic depth that allows for an association with a CS to be elaborated on in a meaningful and personally relevant way (Craik & Lockhart, 1972; Tulving, 1983). This hinders the application of the Pavlovian conditioning paradigm to many research questions regarding emotional associative memories of higher complexity. An important consequence is that it is also insufficient as an analogue for several of the clinical phenomena that experimental psychopathology aims to understand in terms of associative learning. This is clearly illustrated by intrusive memories, which refers to the involuntary reliving of traumatic events, and is a transdiagnostic symptom that may contribute to a variety of psychiatric disorders (Brewin et al., 2010). The development and subsequent triggering of intrusive memories is thought to obey the principles of Pavlovian conditioning, meaning that involuntary reliving of a US in the mind’s eye, when triggered by an associated CS, can be considered a CR (Ehlers & Clark, 2000; Franke et al., 2021; Visser, 2020). This insight has led to the development of the *conditioned-intrusion paradigm* (Wegerer et al., 2013), in which aversive film clips are used as USs. Not only do studies using this paradigm yield results similar to standard Pavlovian conditioning on measures such as skin conductance (Ney et al., 2022), they also enable tests of relived perceptions in response to a CS (e.g. Franke et al., 2022; Miedl et al., 2020; Rattel et al., 2019), and of voluntary memory for various features of the aversive films that serve as USs. However, even though the use of movies would overcome the issue of complexity, such studies still make use of a relatively small number of CSs and USs, thereby precluding a declarative test of CS – US associations and instead often focusing on declarative memory for features of the USs. The possibility to include CS-US associative memory tests is essential to align tests of Pavlovian and episodic learning, as they both rest on the idea that associations are formed between distinct stimuli experienced closely together in space or time.

In sum, in order to study how both physiological and declarative expressions of associative memory for an emotional event emerge and change as a function of various manipulations, it is necessary to encode a large number of unique associations, consisting of both unique CSs and USs. A paradigm that simultaneously measures both enables new research into their interactions, as well as long-standing questions as to whether they result from one or multiple memory systems. However, combining physiological and declarative measures of associative memory in a single paradigm requires dealing with a methodological conflict: Pavlovian conditioning is typically achieved by pairing one or two predictors with a salient outcome over many repetitions, whereas studies of episodic memory rely on the presentation of many stimuli, ideally with as few repetitions as the design allows. As such, there is a balance to strike between having sufficient stimuli for subsequent episodic tests of CS-US declarative associations, while still allowing for psychophysiological conditioned responses to develop. Here, we aimed to address this problem by leveraging the effects associated with multimodal stimulus presentation on physiology and memory. To reconcile our demands for US-complexity with the need for multiple USs, we reasoned that movieclips would be too long, and reverted to another solution, employing complex visual images along with congruent sounds. Researchers in various subfields of psychology have argued that congruent images and sounds make for particularly potent aversive stimuli (e.g. Polo et al., 2024; Čosić et al., 2016), but this insight remains to be applied to Pavlovian conditioning. Moreover, multisensory congruence has been demonstrated to enhance stimulus-driven attention (Ee et al., 2009) and recognition memory (Packard & Soto-Faraco, 2025), which are critical factors when participants are to encode many episodic CS-US associations within only few repeated presentations. Finally, involving multiple sensory inputs has the additional benefit of increasing the degree to which a real-world episodic experience is emulated in the lab, thereby increasing the ecological validity of obtained results.

We collected 20 negative, neutral, and positive combinations of congruent images and sounds (60 in total). Physiological responses and

ratings of subjective experience in response to these US stimuli to-be were measured in a separate sample (see supplement 1). For the present experiment, each of the multimodal stimuli (US) was preceded by a unique predictor (CS), and all CS-US pairs were presented three times in total during distinct learning blocks. Unconditioned responses and acquisition of conditioned responses over repeated presentations were measured using both pupil dilation and facial electromyography (fEMG). We chose these physiological indices over self-report indices as we are interested in automatic measures of conditioned responding that lie beyond conscious control. More specifically, we chose pupil dilation as the primary measure of autonomic arousal as it has been shown not to habituate over repeated presentations (Leuchs et al., 2019; Snowden et al., 2016), and to restabilize at a baseline level relatively quickly after stimulus presentation (Korn et al., 2017). For the purpose of conditioning with many CS-US pairs, these are substantial practical advantages over more traditional measures of autonomic arousal such as skin conductance. We additionally included fEMG (corrugator and zygomatics) as a physiological measure of valence, which may represent a dimension of emotion that is distinct and independent of autonomic arousal (Lang et al., 1998), and can be employed without having to mix bursts of noise with the audio of our multi-modal stimuli, as would be required for the fear-potentiated startle, a more traditional index of conditioned physiological valence (Lang et al., 1990).

The goal with this paradigm is to provide a flexible experimental building block that measures both physiological and declarative expressions of emotional associative memory. For the former, there is a clear validity criterion that mirrors traditional Pavlovian conditioning, namely the acquisition of conditioned responses following exposure to aversive and appetitive CS-US associations. We predicted these effects to be consistent with the current literature: Pupil diameter, taken to reflect autonomic arousal, should increase in response to negative and positive US stimuli, and across learning blocks this effect should transfer to their associated CSs. For fEMG, taken to reflect valence, we expect the corrugator to respond to negative USs and the zygomaticus to positive USs, and again a transfer of both these effects to associated CSs. Whereas there is a clear assessment of validity concerning the physiological/conditioning aspect of our paradigm, this is not the case for the declarative memory component. There is currently no established 'main effect' of emotion on associative memory with a consistent direction. Rather, whether emotion enhances, impairs, or does not affect associative memory at all, seems to depend on a large set of interacting moderators that remain an active area of research (e.g. Bisby et al., 2018; de Montpellier & Talmi, 2023; Henson et al., 2016; Madan et al., 2017, 2020). Simply getting associative memory performance within a range of decent variance, such that an effect of emotion can be detected in the first place, may then seem like a more pragmatic criterion to establish the utility of this paradigm. However, given the sensitivity of declarative memory to design elements that can vary greatly with each research question, such as the number of repeated exposures (Hintzman & Block, 1971) and the interval between encoding and testing (Ebbinghaus, 1885; Murre & Dros, 2015), that, too, is insufficient. As such, we only consider acquisition of conditioned responses to be necessary towards demonstrating the utility of the episodic conditioning paradigm. Declarative associative memory is tested for exploratory purposes, and critical design elements that need to be considered when adapting this paradigm are detailed in the discussion.

2. Methods

2.1. Participants

Sixty-one participants were recruited via the University of Amsterdam's online system. Exclusion criteria were physical or neurological illness, having received treatment for any psychological disorder that is recognized by the DSM-5 in the past year (American Psychiatric Association, 2013), having experienced at least one severely traumatic

episode (assessed by asking if they had personally experienced or witnessed any life-threatening situations), relatively high usage of alcohol (consumption above 21 units a week) or other recreational drugs (more than once a week). Four participants aborted the experiment due to the aversiveness of the stimuli, and one was in the end excluded due to technical failures, resulting in a total sample of 56 (mean age = 21.3, SD = 3.4, women = 25). This study was approved by the local ethics committee of the University of Amsterdam.

2.2. Materials

2.2.1. Unconditioned stimuli

2.2.1.1. Images. We sought to include a total of sixty images, equally distributed over three affective conditions: negative, neutral, and positive. Some of the negative images were chosen from the Nencki Affective Picture System (NAPS), which consists of high-quality images, characterized along the dimensions of arousal and valence by subjective participant ratings (Marchewka et al., 2014). Key selection criteria were high arousal and negative valence ratings, but also whether they could be sensibly presented simultaneously with a corresponding sound. The remaining negative images were found elsewhere on the internet. Then, 20 neutral and 20 positive images were picked to each match one of the negative images in terms of their semantic content. For example, the negative image of a fatal car crash was matched with a neutral image of a car driving on a street, as well as a positive image of a retro Volkswagen camper van under a starry sky. All online searches for images were filtered to only show results that fall under a 'creative commons license' for free use and distribution by third parties.

Two modifications were made to each image. First, they were cropped to a resolution of 600 x 600 pixels. Second, to minimize the impact of low-level features on pupil dilation, luminance was equalized across all images. This was achieved using the programming language Python and the 'scikit-image' processing module by 1) converting all images from RGB to HSV space, 2) computing the grand mean of the set of images' value channels (V), which corresponds to luminance, and 3) scaling the V channel of each image such that its mean value matches the grand mean.

2.2.1.2. Sounds. For each of the images a corresponding sound was found on <https://freesound.org>, an online repository of sound files under a creative commons license. All sound clips were entirely non-verbal, thus avoiding potential effects of different languages spoken by participants. The sound files were clipped to a duration of 4 s in the digital audio editing program 'Audacity'. Moreover, to avoid confounding by physiological responses to loud sounds, the maximum amplitude of each file was adjusted to not exceed 72 dB in our lab setting.

2.2.2. Conditioned stimuli

Sixty images of objects were selected from the Bank of Standardized stimuli (BOSS; Brodeur et al., 2010) to use as CSs. The luminance of each CS was set to the same value as the USs, by means of the same procedure described above.

2.2.3. Equipment and software

The experiment was performed using Presentation software version 23.0 (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Sounds were played through Sennheiser HD 25-1 II headphones. Pupil dilation was measured using a Tobii Pro Nano eye tracker at a sampling frequency of 60 Hz, and fEMG data was recorded in VSRPP, a program for physiological measurement developed by the Technical Support Group Psychology at the University of Amsterdam.

2.3. Experimental design and procedure

Participants first signed informed consent. Due to the aversive nature of the stimulus material, it was emphasized that participants were free to abort the experiment at all times without giving reasons. Next, participants filled in several trait questionnaires¹ that were not analyzed here. Then, sensors for physiological measurements were attached after which the participant was asked to rest their head in a chin-rest, positioned 120 cm away from the computer screen. For exploratory purposes, we also measured electrodermal activity and electrocardiography. These data were not analyzed here. After calibrating the eye-tracker the participants were asked to put on the headphones so the experiment could begin. Both just before and right after the experiment, two affective state questionnaires² were presented that were also not analyzed. During the encoding task (Fig. 1), participants encoded 60 associative memories, each consisting of a CS and a US. Each CS was shown for 4 s, and immediately followed by 4 s of US presentation. Participants were instructed to use each combination of stimuli to imagine a story in which they either feature as the main character or observer of an event (Van

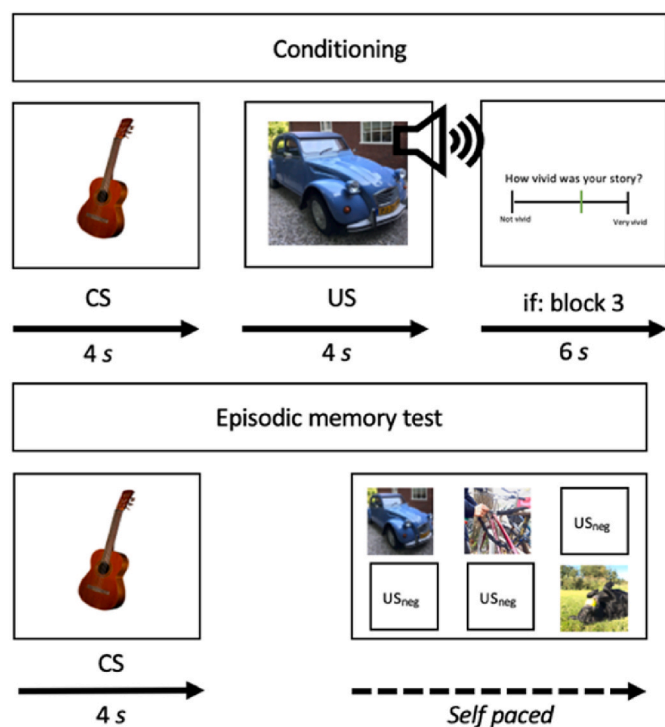


Fig. 1. Schematic overview of the two tasks. During the conditioning phase, participants were presented 60 CS-US pairs. Each US was a unique image combined with a congruent sound that was presented through headphones, and could be negative, neutral, or positive. The figure depicts an example of a neutral US. After all CS-US pairs were presented, they were presented again in a second and third block. In block three, participants would indicate the vividness of the story they imagined using each stimulus pair. During the episodic memory test, each CS would be presented followed by its paired associate and 5 other previously seen USs (2 of each valence in total). Participants used the numpad on the keyboard to select the correctly paired US.

¹ Trait section of the State-Trait Anxiety Inventory (Spielberger et al., 1983), Beck's Depression Inventory (Beck, A. T., Steer, R. A., & Brown, 1996), Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014), and the Spontaneous Use of Imagery Scale (Nelis et al., 2014).

² State section of the State-Trait Anxiety Inventory (Spielberger et al., 1983) and the Positive and Negative Affect Schedule (Watson et al., 1988).

Ast et al., 2013). After all 60 pairs had been presented there was a 1 min break after which they were repeated in a second and third learning block. In the final learning block, participants indicated the vividness of each imagined story on a visual analogue scale (VAS) ranging from 0 to 100 ('not vivid' to 'extremely vivid'). Trials were followed by an ITI that randomly varied from 10 to 14 s, with an average of 12 s. Each participant was shown uniquely randomized CS-US combinations, and the task was programmed so that no more than two USs of the same affective condition would be presented consecutively.

Following encoding, participants completed an associative recognition test for the CS-US associations. Each CS was shown for 4 s, after which six USs (two from each affective condition) were presented from which the associated stimulus had to be selected using the numpad (Fig. 1). The total duration of the experiment was ~75 min.

2.4. Data pre-processing

2.4.1. Pupil

Samples taken 100 ms before and after missing values were also set to missing, since samples near moments when the eye-tracker lost the pupil from its view are not reliable. Missing values were then filled in by linear interpolation, after which a third order bandpass filter (0.01–6 Hz) was applied (Fink et al., 2021). Pupil responses to each CS and US were computed by subtracting the mean pupil dilation from 500 ms before until onset from the mean pupil dilation during the last 2 s of stimulus presentation. Trials were discarded if either the responses to a stimulus or its corresponding baseline were computed based on over 50 % missing values. Participants were excluded from pupil analyses altogether if over 50 % of trials from any affective condition in any of the learning blocks were discarded. These criteria led to the exclusion of 9 participants and a further 1115 trials, leaving a sample of 47 participants and 7345 trials for pupil analysis.

2.4.2. fEMG

Electrodes for fEMG were connected to a custom-made bipolar EMG amplifier with an input resistance of 1G Ω , an amplification factor of 5200, and a bandwidth of 5–1000Hz (6dB/oct). The raw data were sampled at 1000 Hz. The raw data was offline filtered with a 20Hz high-pass filter (Bostel, 2001), a 50Hz notch filter, and a 100Hz notch filter (all 4th order). Then, the data was rectified and integrated using a digital contour follower with a time constant of 25ms. Finally, the data was down-sampled to a rate of 4 Hz. Due to faulty sensors, two participants had to be excluded from corrugator analyses and three from zygomaticus analyses, leaving samples of 54 and 53 participants, respectively.

2.5. Data analysis

All analyses were conducted in R version 4.0.3 (R Team, 2021). We used multilevel linear modeling, performed with the package 'lmer4' (Bates et al., 2015) for each measure separately, and p-values were extracted using 'lmerTest' (Kuznetsova et al., 2017). Intercepts were allowed to vary across participants to account for the nested structure of the data and prevent inflation of the type-1 statistical error rate (Gelman & Hill, 2007). In multilevel regressions, it is often also recommended to include random slopes for the different experimental conditions (Wright, 1998). However, when we did include random slopes, the six main models (three physiological measures x responses to two stimulus types) converged with different sets of parameter inclusions, leading to differences in the exact interpretation of each model. For the sake of consistency across analyses, we opted to report intercept-only models. The main hypothesis that threat conditioning would occur during learning was tested for each physiological measure separately by modelling CS responses as a 3 x 3 linear interaction between Block (1, 2, and 3) and Condition (negative, neutral, positive), representing the valence of the paired US. For measures where calling the base R ANOVA function on the fitted model showed that the interaction between

Condition and Block explained significant variance, we used the R package ‘emmeans’ (Russell et al., 2023) to assess divergence between conditions for each block separately. If conditioning has occurred, responses to predictors of emotional stimuli should increase over repeated presentations of CS-US pairs relative to predictors of neutral stimuli. Where such a pattern was found, we aggregated the data both across participants and stimuli to calculate the proportion that contributed to the effect. Test-retest reliability was estimated for each measure by computing the correlations between stimulus responses in block 1 and 2, and in block 1 and 3, using the package ‘correlation’ for R (Makowski et al., 2019).

Prior to the main analyses, physiological response data was converted to within-subject Z-scores for two reasons. First, this means that the beta coefficients of all linear models can be interpreted as an indication of effect size, which for this reason will be reported using the symbol Z rather than β . Note however that this does not indicate a test statistic, but instead refers to the estimated difference relative to a reference condition. For example, an analysis of a difference between conditions that is reported like this: ($Z = 0.5$, $CI_{95} = [0.35, 0.65]$) indicates that the point-estimate of the effect is half a standard deviation (Freedman, 2009). Critically, as Z-scores were computed for each subject individually, this means that an average participant will show an effect of condition that is 0.5 standard deviations, regardless of their exact response scale and distribution shape. A second advantage of Z-transforming physiological data within participants is that it normalizes responses of individuals whose response distributions may vary widely, thereby potentially causing biased sample means. Particularly on the fEMG measures we noticed that several participants showed responses in a much higher range than most (100 %–1500 % change relative to baseline compared to 0 %–20 %). These large responses only occurred in specific valence conditions (i.e., the corrugator in the negative condition and the zygomaticus in the positive condition), which indicates that these values do not reflect artefacts but expected affective expressions. However, if analyzed without transforming the data, such values would inflate condition effects and response means for individual stimuli. Stimulus-specific effect sizes (Z-score) and probability of superiority (Ruscio, 2008), are listed in sheet 2 of supplement 2 for each measure and learning block, and represent their effectiveness at evoking physiological responses to neutral predictors following learning. If for example these values are [$Z_{\text{block2}} = 0.25$ and $P_{\text{sup_block2}} = 65\%$], this means that the expected response to a predictor (CS) paired to this particular stimulus (US) in the second block is 0.25 standard deviations large, and the probability that it will be higher than that to a random CS that is paired to a neutral stimulus is 65 %.

Performance on the memory test that immediately followed the three learning blocks was analyzed to test whether declarative memory for CS-US associations differed across affective conditions. The effect of emotional condition on memory was tested by fitting a multilevel logistic regression with random intercepts for participants to the binary memory response variable. The overall effect of Condition was tested using the base-R anova function to compare the model including Condition to a null-model of only random participant intercepts. Additionally, we used multilevel regression paired with ANOVA to test the effect of reaction times as an alternative measure of memory strength (e.g. Meyer et al., 2017).

3. Results

3.1. Vividness of encoded episodes

Participants failed to enter their vividness ratings within 6 s on 7.9 % of trials. An ANOVA of vividness ratings revealed a significant effect of Condition ($F_{2,3203} = 69.25$, $p < 0.001$, n participants = 55). Relative to neutral CS-US pairs (mean vividness = 59.14, $sd = 30.40$), vividness of negative trials was lower (difference = -8.45 , $CI_{95} = [-5.69, -11.21]$, $t = 7.17$, $p < 0.001$) whereas that of positive trials was higher (difference

= 5.30, $CI_{95} = [2.54, 8.06]$, $t = 4.50$, $p < 0.001$). Only 7 participants reported average vividness values below the mid-point on the VAS (50), suggesting that the majority was largely able to follow our encoding instructions to imagine vivid stories involving themselves and the stimulus pairs. Moreover, a regression analysis further showed that there was no effect of trial-number ($\beta = 0.006$, $p = 0.904$), indicating that willingness and/or ability to comply with task instructions was not affected by the relatively large number of trials.

3.2. Unconditioned responding and test-retest reliability

3.2.1. Pupil

We found significant main effects of Condition ($F_{2,7072} = 266.78$, $p < 0.001$) and Block ($F_{2,7072} = 106.37$, $p < 0.001$) on pupil dilation to the unconditioned stimuli, as well as an interaction between Condition and Block ($F_{4,7072} = 95.35$, $p < 0.001$). This indicates that pupil dilation differed both when averaged across learning blocks and affective conditions, and that the effect of condition varies across learning blocks. To further disentangle the interaction, pairwise comparisons of marginal means showed that pupil responses to negative USs were higher than to neutral USs in all three learning blocks, starting in block 1 at a difference of $Z = 0.52$ ($CI_{95} = [0.41, 0.63]$, $p < 0.001$), then $Z = 0.54$ ($CI_{95} = [0.43, 0.63]$, $p < 0.001$) in block 2, and finally at a relatively smaller difference of $Z = 0.30$ ($CI_{95} = [0.19, 0.42]$, $p < 0.001$). Pupil responses to positive USs were smaller than to neutral USs in all blocks, although for block 2 the difference was only marginally significant. For block 1 we found a difference of $Z = 0.24$ ($CI_{95} = [0.13, 0.35]$, $p < 0.001$), for block 2 of $Z = 0.11$ ($CI_{95} = [-0.005, 0.22]$, $p = 0.064$), and $Z = 0.14$ ($CI_{95} = [0.03, 0.25]$, $p = 0.011$) for block 3. Test-retest reliability between pupil responses in block 1 and block 2 was $\rho_{2091} = 0.26$ ($CI_{95} = [0.22, 0.30]$, $p < 0.001$), and $\rho_{2091} = 0.23$ ($CI_{95} = [0.19, 0.27]$, $p < 0.001$) between block 1 and block 3. These results suggest that especially the negative stimuli are consistently potent in eliciting pupil responses. However, unlike in the rating experiment described in supplement 1, pupil responses to positive stimuli were significantly lower than to neutral stimuli. This difference is relatively small compared to the effect of negative stimuli, and this inconsistency may be explained by the larger statistical power of the main experiment. Despite the Block \times Condition interaction, which indicates a degree of habituation, the effect of Condition remained significant over repeated presentations.

3.2.2. Corrugator

The ANOVA of corrugator responses revealed a main effect of Condition ($F_{2,9691} = 439.42$, $p < 0.001$) and an interaction between Condition and Block $F_{4,9691} = 5.36$, $p < 0.001$), but no main effect of Block ($F_{2,9691} = 0.82$, $p = 0.438$). Thus, the corrugator was sensitive to Condition, and the effect differed across learning blocks. Pairwise comparisons of marginal means showed that corrugator responded to negative and positive USs in the same way as the pupil: compared to neutral USs, responses to negative USs were higher and responses to positive USs were lower in all three learning blocks. In block 1, the difference between negative and neutral USs was $Z = 0.64$ ($CI_{95} = [0.54, 0.73]$, $p < 0.001$), then $Z = 0.50$ ($CI_{95} = [0.41, 0.60]$, $p < 0.001$) in block 2, and $Z = 0.46$ ($CI_{95} = [0.37, 0.56]$, $p = 0.016$) in block 3. The difference between positive and neutral stimuli in block 1 was $Z = -0.18$ ($CI_{95} = [-0.28, -0.08]$, $p < 0.001$), then $Z = -0.11$ ($CI_{95} = [-0.20, -0.01]$, $p = 0.029$) in block 2, and $Z = -0.11$ ($CI_{95} = [-0.21, -0.02]$, $p = 0.016$) in block 3. Test-retest reliability between corrugator responses in block 1 and block 2 was $\rho_{3222} = 0.22$ ($CI_{95} = [0.19, 0.26]$, $p < 0.001$), and $\rho_{2091} = 0.15$ ($CI_{95} = [0.12, 0.19]$, $p < 0.001$) between block 1 and block 3. These findings show that the effect of condition decreases, but remains statistically significant across the three learning blocks.

3.2.3. Zygomaticus

A main effect of Condition ($F_{2,9477} = 42.91$, $p < 0.001$) shows that the zygomaticus was sensitive to US condition. There was no main effect of

Block ($F_{2,9477} = 0.26, p = 0.769$), nor an interaction between Condition and Block ($F_{4,9477} = 0.26, p = 0.081$), indicating that the effect of Condition on zygomaticus responses remained stable across learning blocks. Comparisons of marginal Condition means, collapsed across Block, revealed that both negative ($Z = 0.12, CI_{95} = [0.07, 0.19], p < 0.001$) and positive USs ($Z = 0.23, CI_{95} = [0.41, 0.60], p < 0.001$) evoked larger responses than neutral USs, and responses to positive USs were significantly higher than to negative USs ($Z = 0.10, CI_{95} = [0.04, 0.16], p = 0.001$). Test-retest reliability between zygomaticus responses in block 1 and block 2 was $\rho_{3222} = 0.11$ ($CI_{95} = [0.07, 0.14], p < 0.001$), and $\rho_{2091} = 0.05$ ($CI_{95} = [0.02, 0.09], p = 0.002$) between block 1 and block 3. Overall, this shows that although trial level zygomaticus responses are less reliable across blocks as they are for the pupil and corrugator, they can still be partly predicted from earlier presentations.

3.3. Acquisition of conditioned responses

3.3.1. Pupil

There was a significant interaction between Condition and Block ($F_{4,7336} = 7.71, p < 0.001$), indicating that there is an effect of stimulus valence on pupil CRs that differs across learning blocks. There was no significant difference in pupil responses evoked by CS_{neg} as compared to CS_{neu} when all CS-US pairs were first presented in block 1 ($Z = 0.03, CI_{95} = [-0.08, 0.14], p = 0.806$), but these conditions did diverge in both block 2 ($Z = 0.39, CI_{95} = [0.27, 0.50], p < 0.001$), and block 3 ($Z = 0.20, CI_{95} = [0.08, 0.31], p < 0.001$). Pupil responses to CS_{pos} however, did not differ from CS_{neu} in any of the learning blocks (all $p > 0.177$). These results demonstrate that, after learning the CS-US associations, predictors paired to negative USs came to evoke conditioned pupil responses, whereas those paired to positive USs did not. The average response to CS_{neg} was higher than to CS_{neu} for 39 out of 46 participants (84.8 %) in block 2, and 34 out of 46 (73.9 %) in block 3. All 20 negative USs on average elicited larger conditioned responses than the mean of neutral USs in block 2, and 18 (90 %) in block 3. Together, these findings show that the negative stimuli, but not the positive stimuli are potent USs for measuring conditioned responses on pupil dilation.

3.3.2. Corrugator

There was a significant interaction between Condition and Block ($F_{4,9710} = 4.30, p = 0.002$), indicating that there is an effect of stimulus valence on corrugator CRs that differs across learning blocks. Contrary to expectations, responses to CS_{neg} were already higher than to CS_{neu} in block 1, prior to learning the associations ($Z = 0.12, CI_{95} = [0.02, 0.23], p = 0.012$). However, this effect became larger in block 2 ($Z = 0.20, CI_{95} = [0.10, 0.30], p < 0.001$), and then even more in block 3 ($Z = 0.28, CI_{95} = [0.19, 0.39], p < 0.001$). Corrugator responses to predictors of positive stimuli did not differ from those that predicted neutral stimuli in any of the learning blocks (all $p > 0.151$). Thus, corrugator responses to CS_{neg} increased relative to CS- following repeated presentation of the CS-US pairs, whereas responses to CS_{pos} did not. The average corrugator response to CS_{neg} was higher than to CS_{neu} for 34 out of 54 participants (63.0 %) in block 2, and for 37 (68.5 %) in block 3. Of the 20 negative stimuli, 18 on average elicited larger CRs than the mean of neutral stimuli in both block 2 and block 3. Similar to the pupil, the corrugator is an effective measure of aversive conditioning to the negative stimuli, but shows no appetitive conditioning to the positive stimuli. It must be noted that there was already a difference between CS_{neg} and CS_{neu} in the first block, before the CS-US pairs had ever been presented. The interaction between condition and block indicates that this difference grew larger following the first learning block, consistent with the pattern expected of Pavlovian conditioning effects.

3.3.3. Zygomaticus

An ANOVA showed that there was no interaction effect of Condition and Block on CRs of the zygomaticus ($\eta^2 = 0.000, CI_{90} = [0.000, 0.001], F_{4,9531} = 0.52, p = 0.720$), thus providing no evidence for the occurrence

of conditioning in either the negative or positive condition.

Conditioned and unconditioned responses for all physiological measures and valence conditions are shown in Fig. 2.

3.4. Memory test

There was no effect of Condition on associative recognition accuracy ($F = 0.25, p = 0.775$), meaning that there was no difference in how well participants remembered the associations between CSs paired with negative, neutral, or positive USs. However, this analysis was however most likely hindered by high performance close to ceiling. The average proportion of correctly recognized associations were for 0.91, 0.90, and 0.91 for the negative, neutral, and positive pairings, respectively. Moreover, 42 out of 58 participants achieved a perfect score in at least one affective condition, and 23 achieved a perfect score overall. Excluding participants with a perfect score did not yield another result ($F = 0.25, p = 0.777$). We also analyzed reaction times, which as a continuous and unbounded variable may be more sensitive to an effect of affective condition on memory strength. Yet, an ANOVA with Condition and Accuracy (correct or incorrect memory) as interacting predictors only revealed a significant effect of Accuracy ($F_{1,3064} = 251.15, p < 0.001$), indicating that participants entered correct answers faster than incorrect answers. There was no main effect of Condition ($F_{2,3424} = 2.78, p = 0.062$), nor an interaction between Condition and Accuracy ($F_{2,3427} = 3.477, p = 0.081$).

4. Discussion

The present research aimed to develop a new paradigm to simultaneously measure Pavlovian conditioning and episodic memory using multimodal affective stimuli. Using a relatively large number of unique CS-US episodic 'mini-conditioning events', we demonstrated that randomly paired predictors (CS) of negative stimuli (US) triggered conditioned responses in the form of increased pupil dilation and corrugator activity already after a single block of associative learning. We found no such conditioned responses to predictors of positive stimuli on the zygomaticus or any other of the physiological indices. There was no effect of emotion condition on declarative associative memory. This test may have been hindered by high overall performance, but excluding participants with a perfect score did not change the result. More important, however, is that the mere quantity of employed stimuli does allow for meaningful declarative tests of associative memory, given some adaptations are made to the episodic conditioning paradigm. Options such as a reduction in learning blocks, or extending the time interval between learning and testing, are discussed below. Together, these results indicate that congruent image-sound combinations can be used as potent emotional USs in studies that require simultaneous assessment of physiological and declarative expressions of emotional associative memory. Thereby, the episodic conditioning paradigm opens the door to new research questions where they intersect.

Although 'episodic conditioning' was successful for negative USs, it was not for positive USs, which is a fairly typical pattern of results for studies investigating both aversive and appetitive conditioning in humans (e.g. Exner et al., 2021; Hermann et al., 2000; Stussi et al., 2018). One reason for this may be that the unconditioned responses to positive stimuli were on all measures substantially smaller than those to negative stimuli were on the pupil and corrugator. Conditioned responses tend to be lower than their corresponding unconditioned responses, making positive CRs harder to detect in the present study. It is also possible that we did not find positive conditioning because the positive stimuli did not require anticipatory physiological responses to be approached in the same way that the negative stimuli did to be avoided (Krypotos et al., 2014). Since the difference between negative and neutral stimuli was much larger than that between positive and neutral stimuli on each measure except the zygomaticus, the CSs of neutral and positive stimuli may have simply clustered in affective space

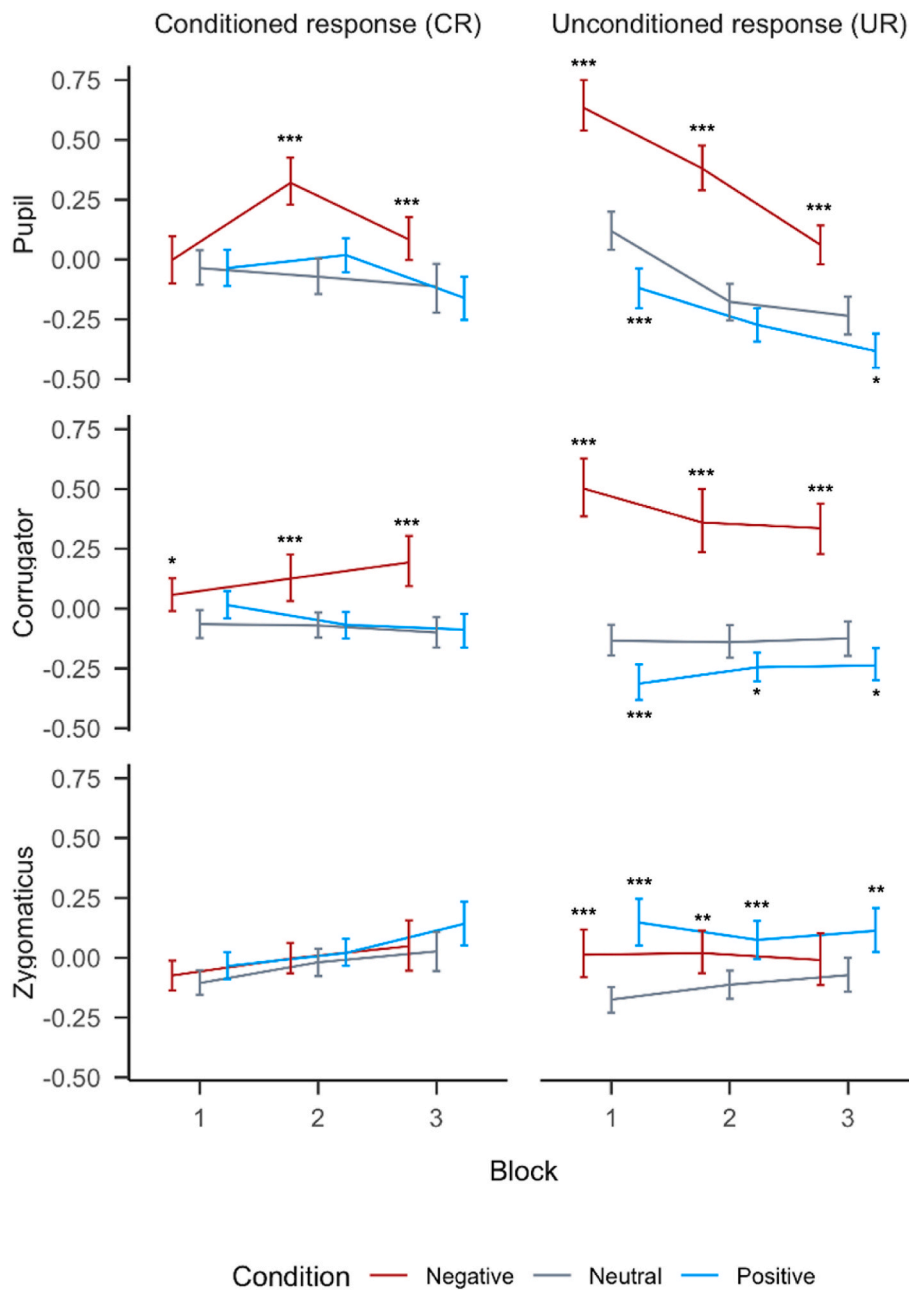


Fig. 2. Conditioned (left panel) and unconditioned responses (right panel) for each physiological measure (pupil, corrugator, and zygomaticus), each affective condition (negative, neutral positive), and each learning block. Data was Z-transformed and aggregated across participants, meaning that 'Z-score' is the unit for each physiological measure shown here. Error bars indicate confidence intervals. Asterisks indicate significant differences relative to the neutral condition (* = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.001$).

as safety signals (Sangha et al., 2020), in which case the presence of the successful negative conditioning trials thwarted the detection of a positive conditioning effect. Regardless, we only found evidence that the negative multimodal stimuli used here are effective as USs, hence specifically the name 'Episodic Threat Conditioning'.

This study provides a chart of subjective and physiological measures of emotion in response to 60 unique multi-modal stimuli (see supplement 1), as well as to arbitrarily paired predictors, that can be utilized in a broad range of research areas. Where they are uniquely useful, however, is in studies aimed at understanding the relation between Pavlovian threat conditioning and episodic memory, as well as how they interact to produce other outcomes of interest. One straightforward example is the question of whether conditioned responding can exist in the absence of memory for the CS-US association. This has previously been

investigated in brain damaged patients (Buchanan, 1995), which are rare and yield results that may not generalize to healthy individuals, and through subliminal stimulus presentation (Helmstetter, 2019; Raio et al., 2012), which comes with the downsides of weakened US impact and methodological issues that may bias results towards false positives (Mertens & Engelhard, 2020). The episodic threat conditioning paradigm enables the direct presentation of many stimulus pairs that trigger strong CRs and URs in healthy participants, which can be related to the presence of declarative memory associations, thus providing a new approach to address this fundamental question of human learning. As another example of how this paradigm can be used to integrate ideas from Pavlovian threat conditioning and episodic memory, one study from our lab used it to investigate whether threat affects the ability to draw connections between distinct memories through a shared element,

also known as associative inference (de Vries et al., 2022). We found that associative inference was impaired by threat, but that this effect was not moderated by intensity of emotional responses to the USs. Another study, incorporating the episodic conditioning paradigm in a very similar design, found that conditioned responses spread to arbitrary stimuli that were previously associated with a CS, and that this was likely facilitated by an episodic memory integration mechanism (de Vries et al., 2025). Importantly, the pattern of URs and CRs to negative stimuli during acquisition in both these studies was the same as in the present study, providing evidence for the robustness of the episodic threat conditioning effect.

When making use of the episodic threat conditioning paradigm and adjusting it to the needs of a new study, it is important to be aware of several design choices that influence both Pavlovian conditioning and episodic memory, and may need to be balanced carefully. As also evidenced by the ceiling effect of associative memory in the present study, episodic memories are stronger following repeated stimulus presentations (Hintzman & Block, 1971) and when the interval between encoding and testing is short (Ebbinghaus, 1885; Murre & Dros, 2015). Repetition of CS-US stimulus pairs is however essential to measure conditioned responses. Here, as well as in our previous studies using this paradigm (de Vries et al., 2022, 2025), we found significant conditioned responses already after presenting each 'mini-conditioning event' one time, meaning that for the purpose of measuring acquisition, the third block is redundant and could probably be omitted to bring episodic memory performance immediately following encoding into a workable range. This would have the added benefit of keeping the memories more episodic in nature, since that is thought to decrease as a function of repeated experience (O'Reilly et al., 2014). Alternatively, when more than two CS-US presentations are required for other reasons, increasing the interval between encoding and testing would likely accomplish the same. Another factor that affects episodic memory is the cognitive demand of the encoding task (Underwood, 1978). Since the paradigm was only effective at inducing aversive conditioning effects, future studies will likely only use the negative and neutral USs, thus decreasing the number of associations to be learned. This may enhance acquisition and have the possibly unwanted effect of further increasing episodic memory strength (Underwood, 1978), in which case can be is even more important to take measures that reduce memory performance. It is furthermore important to note that the encoding instructions here are likely to be key to the paradigm's efficiency. Letting participants use the CS-US pairs to imagine stories that centrally involve them recruits elaborative processes that are beneficial to memory (Craig & Lockhart, 1972). It is possible that without this specific kind of engagement, acquisition of conditioned responses will require more than two learning blocks, or may even not be possible at all. Note, however, that as we instructed participants that they could also imagine stories in which they witnessed an event, it is possible that conditioning is more driven by first-person relative to third-person narratives. Whether this is actually the case is an interesting avenue for future studies.

Future studies employing the episodic conditioning approach should carefully consider what dependent variables are of interest for their specific study. We developed this paradigm using primarily measures of pupil size and fEMG for reasons specified in the introduction, most importantly their resistance to habituation. However, studies that make use of fewer stimuli and/or repetitions at encoding could also consider using more traditional measures such as SCR and startle. We also did not index self-report measures such as conditioned distress, since we were mostly interested in psychophysiological indices of autonomic arousal. Moreover, we reasoned that online ratings were likely to interfere with both pupil measurement and participants' ability to imagine coherent stories encompassing both CS and US, which may be essential for forming a measurable association between the two. However, for future studies with a clinical interest, online ratings such as conditioned distress or US expectancy could be required. When implementing such measurements, we recommend considering the effect this

may have on pupil assessment and episodic encoding. Note that implementing traditional US-expectancy ratings is not possible, given that there are many different USs that such a rating could refer to, though it is possible to ask more generally if the participant anticipates any of the aversive outcome stimuli.

One final recommendation regarding the use of this paradigm concerns the statistical analyses. Although every research question demands a critical evaluation of what statistical models are most appropriate, we would like to emphasize that a key feature of this paradigm is the encoding of multiple distinct events, consisting of unique stimulus combinations. Therefore, unless there is a clear reason not to, the default recommendation is to fit hierarchical models rather than aggregating data within participants and thereby discarding trial-level variance. For the analysis of physiological data a standard multilevel linear regression is suitable, whereas for the associative recognition test used here, a multilevel logistic regression that fits the true distribution of binary data is recommended. Standard works on hierarchical regression typically recommend fitting the maximum random effects structure to avoid inflation of the false-positive rate (Barr et al., 2013; Gelman & Hill, 2007), while others note that this may reduce statistical power (Matuschek et al., 2017). Additionally, issues with model convergence can pose a more pragmatic challenge when fitting hierarchical regressions with random slopes. Here we reported six main analyses (CRs and URs for three physiological measures), and since there was no consistency in which random effects structure converged, we opted to report the intercept-only models so that each result has the same exact interpretation. Notably, the converging maximum effect structure for each analysis yielded similar SEs and p-values, which in some cases were more statistically significant rather than less. This suggests that the false-positive rate was not inflated by our decision to report intercept-only models. With that being said, we urge users of this paradigm to consider and explore the matter of random effects structures for each study independently.

With this study, we presented a new experimental paradigm that can be used as a building block for studies aimed at understanding how threat conditioning and episodic memory processes interact to facilitate the prediction of aversive outcomes. Although we have emphasized the various ways these fields and their methods may complement each other, the result of this fusion is best thought of not as more, but rather as being different than the sum of its parts. The large number of CS-US pair and the instruction to imagine oneself outside the context of the lab likely requires the recruitment of brain areas and networks beyond those involved in traditional Pavlovian conditioning tasks, particularly prefrontal areas and the medial temporal lobe (Rugg & Vilberg, 2013). Given that these areas are known to be highly interconnected with the amygdala and other structures that are considered part of the threat learning network (Fastenrath et al., 2014; Ledoux & Daw, 2018), interactions at the neural level may give rise to new behavioral effects, as well as interfere with previously established findings. Similarly, effects of emotion on episodic associative memory may differ from those established in the literature when using this paradigm due to its sequential as opposed to simultaneous (e.g. Bisby et al., 2018; Madan et al., 2017) presentation of stimuli, and the acquisition of predictive value that follows. This could trigger the engagement of the defensive network when a CS is presented during a memory test, which may interact with the mechanisms of declarative retrieval. The episodic threat conditioning paradigm thus provides a way to blur the perhaps artificial lines between cognitive and emotional memory processes (Pessoa et al., 2022), and study the mechanisms underlying the prediction and recollection of aversive events with stimuli that are more ecologically valid and meaningful to participants.

CRedit authorship contribution statement

Olivier T. de Vries: Writing – original draft, Visualization, Software, Project administration, Methodology, Formal analysis, Data curation,

Conceptualization. **Sascha B. Duken:** Methodology, Formal analysis, Nadza Džinalija, Investigation. **Nadža Džinalija:** Investigation. **Merel Kindt:** Writing – review & editing, Supervision, Conceptualization. **Vanessa A. van Ast:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vanessa A. van Ast reports financial support was provided by Dutch Research Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.brat.2025.104882>.

Data availability

Data will be made available on request.

References

- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. <https://thuvienso.hoasen.edu.vn/handle/123456789/9853>.
- Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014). Assessing vividness of mental imagery: The plymouth sensory imagery questionnaire. *British Journal of Psychology*, 105(4), 547–563. <https://doi.org/10.1111/bjop.12050>
- 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.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II manual* (p. 2pages). San Antonio, TX: The Psychological Corporation.
- Beckers, T., Krypotos, A. M., Boddez, Y., Eftting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology*, 92(1), 90–96. <https://doi.org/10.1016/j.biopsycho.2011.12.015>
- Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, 147(2), 243–256. <https://doi.org/10.1037/xge0000356>
- Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, 38, 22–34. <https://doi.org/10.1111/1469-8986.3810022>
- Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychological Review*, 117(1), 210–232. <https://doi.org/10.1037/a0018113>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, 5(5). <https://doi.org/10.1371/journal.pone.0010773>
- Čosić, K., Popović, S., Kukuljica, D., Dropuljić, B., Ivanec, D., & Tonković, M. (2016). Multimodal analysis of startle type responses. *Computer Methods and Programs in Biomedicine*, 129, 186–202. <https://doi.org/10.1016/j.cmpb.2016.01.002>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684. <https://doi.org/10.4324/9781315440446>
- Craske, M. G., Kirscanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy*, 46(1), 5–27. <https://doi.org/10.1016/j.brat.2007.10.003>
- de Montpellier, E., & Talmi, D. (2023). Are multiple types of associative memory differently impacted by emotion? *Cognition & Emotion*, 0(0), 1–24. <https://doi.org/10.1080/02699931.2023.2279182>
- de Vries, O. T., Grasman, R. P. P., Kindt, M., & van Ast, V. A. (2022). Threat learning impairs subsequent associative inference. *Scientific Reports*, 12(1), 1–13. <https://doi.org/10.1038/s41598-022-21471-2>
- de Vries, O. T., Kindt, M., & Van Ast, V. A. (2025). Integration of conditioned threat with pre-existing memories. *Learning & Memory*, 32(1), Article a054019. <https://doi.org/10.1101/lm.054019.124>
- Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., ... Baas. (2015). UPDATED META-ANALYSIS OF CLASSICAL FEAR CONDITIONING IN THE ANXIETY DISORDERS: Review: Updated Meta-Analysis of Fear Conditioning in Anxiety Disorders. *Depression and Anxiety*, 32(4), 239–253. <https://doi.org/10.1002/da.22353>.
- Dunsmoor, J. E., & Koes, M. C. W. (2019). Episodic memory and Pavlovian conditioning: Ships passing in the night. *Current Opinion in Behavioral Sciences*, 26, 32–39. <https://doi.org/10.1016/j.cobeha.2018.09.019>
- Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology*, 89(2), 300–305. <https://doi.org/10.1016/j.biopsycho.2011.11.002>
- Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: How humans generalize fear. *Trends in Cognitive Sciences*, 19(2), 73–77. <https://doi.org/10.1016/j.tics.2014.12.003>
- Ebbinghaus, H. (1885). Über das Gedächtnis. <https://home.uni-leipzig.de/wundtbriefe/wwdc/opera/ebbing/memory/Gdaechtfl.htm>.
- Ee, R. van, Boxtel, J. J. A. van, Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, 29(37), 11641–11649. <https://doi.org/10.1523/JNEUROSCI.0873-09.2009>
- Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, 38(4), 319–345. [https://doi.org/10.1016/S0005-7967\(99\)00123-0](https://doi.org/10.1016/S0005-7967(99)00123-0)
- Exner, A., Tapia León, I., Mueller, E. M., & Klucken, T. (2021). Cardiac response in aversive and appetitive olfactory conditioning: Evidence for a valence-independent CS-elicited bradycardia. *Psychophysiology*, 58(11), 1–14. <https://doi.org/10.1111/psyp.13912>
- Fastenrath, M., Coynel, D., Spalek, K., Milnik, A., Gschwind, L., Roozendaal, B., Papasotiropoulos, A., & de Quervain, D. J. F. (2014). Dynamic modulation of amygdala-hippocampal connectivity by emotional arousal. *Journal of Neuroscience*, 34(42), 13935–13947. <https://doi.org/10.1523/JNEUROSCI.0786-14.2014>
- Fink, L., Simola, J., Tavano, A., Lange, E., Wallot, S., & Laeng, B. (2021). From pre-processing to advanced dynamic modeling of pupil data. *PsyArXiv*, 1–56.
- Franke, L. K., Miedl, S. F., Danböck, S. K., Grill, M., Liedgruber, M., Kronbichler, M., Flor, H., & Wilhelm, F. H. (2022). Neuroscientific evidence for pain being a classically conditioned response to trauma- and pain-related cues in humans. *Pain*, 163(11), 2118. <https://doi.org/10.1097/j.pain.0000000000002621>
- Franke, L. K., Rattel, J. A., Miedl, S. F., Danböck, S. K., Bürkner, P. C., & Wilhelm, F. H. (2021). Intrusive memories as conditioned responses to trauma cues: An empirically supported concept? *Behaviour Research and Therapy*, 143(April). <https://doi.org/10.1016/j.brat.2021.103848>
- Freedman, D. A. (2009). *Statistical models: Theory and practice*. Cambridge University Press.
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=QkNuuVf-pBMC&oi=fnd&pg=PR11&dq=frijda+1986+the+emotions&ots=BNK8h-7pPs&sig=GJkTz_RELqhiS1Tudl-uV42p0k8
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Helmstetter, F. J. (2019). Independent of contingency awareness. <https://doi.org/10.1037/a0020263.Classical>.
- Henson, R. N., Campbell, K. L., Davis, S. W., Taylor, J. R., Emery, T., Erzincinlioglu, S., & Kievit, R. A. (2016). Multiple determinants of lifespan memory differences. *Scientific Reports*, 6(1), Article 32527. <https://doi.org/10.1038/srep32527>
- Hermann, C., Ziegler, S., Birbaumer, N., & Flor, H. (2000). Pavlovian aversive and appetitive odor conditioning in humans: Subjective, peripheral, and electrocortical changes. *Experimental Brain Research*, 132(2), 203–215. <https://doi.org/10.1007/s002210000343>
- Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 88(3), 297–306. <https://doi.org/10.1037/h0030907>
- 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>
- 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>
- Krypotos, A. M., Eftting, M., Arnaudova, I., Kindt, M., & Beckers, T. (2014). Avoided by association: Acquisition, extinction, and renewal of avoidance tendencies toward conditioned fear stimuli. *Clinical Psychological Science*, 2(3), 336–343. <https://doi.org/10.1177/2167702613503139>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/JSS.V082.I13>
- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1), 54–64. <https://doi.org/10.1038/nrn1825>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, 97(3), 377–395. <https://doi.org/10.1037/0033-295X.97.3.377>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, 44(12), 1248–1263. [https://doi.org/10.1016/S0006-3223\(98\)00275-3](https://doi.org/10.1016/S0006-3223(98)00275-3)
- Ledoux, J., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, 19(5), 269–282. <https://doi.org/10.1038/nrn.2018.22>
- LeDoux, J. E., & 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>
- 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>

- Lissek, S., Kaczurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized Anxiety Disorder Is Associated With Overgeneralization of Classically Conditioned Fear. *Biological Psychiatry*, 75(11), 909–915. <https://doi.org/10.1016/j.biopsych.2013.07.025>.
- Madan, C. R., Fujiwara, E., Caplan, J. B., & Sommer, T. (2017). Emotional arousal impairs association-memory: Roles of amygdala and hippocampus. *NeuroImage*, 156 (May), 14–28. <https://doi.org/10.1016/j.neuroimage.2017.04.065>
- Madan, C. R., Knight, A. G., Kensinger, E. A., & Mickley Steinmetz, K. R. (2020). Affect enhances object-background associations: Evidence from behaviour and mathematical modelling. *Cognition & Emotion*, 34(5), 960–969. <https://doi.org/10.1080/02699931.2019.1710110>
- Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdtke, D. (2019). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.02306>
- Marchewka, A., Żurawski, L., Jednoróg, K., & Grabowska, A. (2014). The Nencki affective picture system (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2), 596–610. <https://doi.org/10.3758/s13428-013-0379-1>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neuroscience & Biobehavioral Reviews*, 108 (April 2019), 254–268. <https://doi.org/10.1016/j.neubiorev.2019.11.012>
- Meyer, T., Krans, J., van Ast, V., & Smeets, T. (2017). Visuospatial context learning and configuration learning is associated with analogue traumatic intrusions. *Journal of Behavior Therapy and Experimental Psychiatry*, 54, 120–127. <https://doi.org/10.1016/j.jbtep.2016.07.010>
- Miedl, S. F., Rattel, J. A., Franke, L. K., Blechert, J., Kronbichler, M., Spooemaker, V. I., & Wilhelm, F. H. (2020). Neural processing during fear extinction predicts intrusive memories. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(4), 403–411. <https://doi.org/10.1016/j.bpsc.2019.12.017>
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning: The behavioral and brain sciences. *Behavioral and Brain Sciences*, 32(2), 183–246. <https://doi.org/10.1017/S0140525X09000855>
- Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS One*, 10(7), 1–23. <https://doi.org/10.1371/journal.pone.0120644>
- Nelis, S., Holmes, E. A., Griffith, J. W., & Raes, F. (2014). Mental imagery during daily life: Psychometric evaluation of the spontaneous use of imagery scale (SUIS). *Psychologica Belgica*, 54(1), 19–32. <https://doi.org/10.5334/pb.ag>
- Ney, L. J., Schenker, M., & Lipp, O. V. (2022). Combining the trauma film and fear conditioning paradigms: A theoretical review and meta-analysis with relevance to PTSD. *Behaviour Research and Therapy*, 152, Article 104081. <https://doi.org/10.1016/j.brat.2022.104081>
- O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, 38(6), 1229–1248. <https://doi.org/10.1111/j.1551-6709.2011.01214.x>
- Packard, P. A., & Soto-Faraco, S. (2025). Crossmodal semantic congruence and rarity improve episodic memory. *Memory & Cognition*. <https://doi.org/10.3758/s13421-024-01659-9>
- Pessoa, L., Medina, L., & Desfilis, E. (2022). Refocusing neuroscience: Moving away from mental categories and towards complex behaviours. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 377(1844), Article 20200534. <https://doi.org/10.1098/rstb.2020.0534>
- Polo, E. M., Farabbi, A., Mollura, M., Mainardi, L., & Barbieri, R. (2024). Understanding the role of emotion in decision making process: Using machine learning to analyze physiological responses to visual, auditory, and combined stimulation. *Frontiers in Human Neuroscience*, 17. <https://doi.org/10.3389/fnhum.2023.1286621>
- Raio, C. M., Carmel, D., Carrasco, M., & Phelps, E. A. (2012). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology*, 22(12), R477–R479. <https://doi.org/10.1016/j.cub.2012.04.023>
- Rattel, J. A., Weeger, M., Miedl, S. F., Blechert, J., Grünberger, L. M., Craske, M. G., & Wilhelm, F. H. (2019). Peritraumatic unconditioned and conditioned responding explains sex differences in intrusions after analogue trauma. *Behaviour Research and Therapy*, 116, 19–29. <https://doi.org/10.1016/j.brat.2019.01.009>
- Rescorla, R. A. (1987). Rescorla (1988) Pavlovian conditioning—Its not what you think it is. *Annual Reviews of Neuroscience*, 43(3), 151–160.
- Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, 23(2), 255–260. <https://doi.org/10.1016/j.conb.2012.11.005>
- Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, 13(1), 19–30. <https://doi.org/10.1037/1082-989X.13.1.19>
- Russell, A., Lenth, V., Bolker, B., Buerkner, P., Giné-vázquez, I., Herve, M., Love, J., Singmann, H., & Lenth, M. R. V. (2023). Package 'emmeans' R topics documented, 34 (4), 216–221. <https://doi.org/10.1080/00031305.1980.10483031>>.License
- Sangha, S., Diehl, M. M., Bergstrom, H. C., & Drew, M. R. (2020). Know safety, no fear. *Neuroscience & Biobehavioral Reviews*, 108(February 2019), 218–230. <https://doi.org/10.1016/j.neubiorev.2019.11.006>
- Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology*, 53(8), 1217–1223. <https://doi.org/10.1111/psyp.12668>
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). State-trait-anxiety-inventory for adults self-evaluation questionnaire STAI form Y-1 nad form Y-2. In *Mind garden* (p. 5).
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171–177. <https://doi.org/10.1016/j.nlm.2004.06.005>
- Stussi, Y., Delplanque, S., Coraj, S., Pourtois, G., & Sander, D. (2018). Measuring Pavlovian appetitive conditioning in humans with the postauricular reflex. *Psychophysiology*, 55(8). <https://doi.org/10.1111/psyp.13073>
- Team, R. C. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Tulving, E. (1983). Elements of episodic memory. In *Canadian psychology*, 26New York: Oxford University Press, 3.
- Underwood, B. J. (1978). Recognition memory as a function of. *12(2)*, 89–91.
- Van Ast, V. A., Cornelisse, S., Meeter, M., Joëls, M., & Kindt, M. (2013). Time-dependent effects of cortisol on the contextualization of emotional memories. *Biological Psychiatry*, 74(11), 809–816. <https://doi.org/10.1016/j.biopsych.2013.06.022>
- Vervliet, B., & Boddez, Y. (2020). Memories of 100 years of human fear conditioning research and expectations for its future. *Behaviour Research and Therapy*, 135 (September). <https://doi.org/10.1016/j.brat.2020.103732>
- Visser, R. M. (2020). Why do certain moments haunt us? Conceptualizing intrusive memories as conditioned responses. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(4), 375–376. <https://doi.org/10.1016/j.bpsc.2020.02.007>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Wegerer, M., Blechert, J., Kerschbaum, H., & Wilhelm, F. H. (2013). Relationship between fear conditionability and aversive memories: Evidence from a novel conditioned-intrusion paradigm. *PLoS One*, 8(11), Article e79025. <https://doi.org/10.1371/journal.pone.0079025>