Biological Psychology 168 (2022) 108249

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Avoiding a feared stimulus: Modelling costly avoidance of learnt fear in a sensory preconditioning paradigm

Alex H.K. Wong [a,*], Andre Pittig [a,b]

[a] Department of Psychology (Biological Psychology, Clinical Psychology, and Psychotherapy), University of Würzburg, Würzburg, Germany
[b] Translational Psychotherapy, Department of Psychology, Friedrich–Alexander University Erlangen–Nürnberg (FAU), Erlangen, Germany

ARTICLE INFO

Keywords:
Avoidance
Fear conditioning
Higher-order conditioning
Costly avoidance
Threat anticipation
Anxiety

ABSTRACT

Avoidance of learnt fear prevents the onset of a feared stimulus and the threat that follows. In anxiety-related disorders, it turns pathological given its cost and persistence in the absence of realistic threat. The current study examined the acquisition of costly avoidance of learnt fear in healthy individuals (n = 45), via a sensory preconditioning paradigm. Two neutral preconditioning stimuli (PSs) were paired with two neutral conditioned stimuli (CSs). One CS then came to predict an aversive outcome whereas the other CS came to predict safety. In test, participants engaged in stronger avoidance to the PS associated with the fear-related CS than the PS associated with the safety-related CS. Of note, executing behavioral avoidance led to missing out a competing reward, thus rendering avoidance costly. The results also provide preliminary evidence that threat anticipation and a negative change in valence play a role in driving costly avoidance of learnt fear. Future studies should examine how avoidance of learnt fear maintains pathological anxiety.

1. Introduction

Behavioral avoidance is typically adaptive given that it prevents one from harm. Behavioral avoidance in anxiety-related disorders is, however, oftentimes considered maladaptive given that it persists in the absence of threat, and is linked to impairments in daily functioning (Mendlowicz & Stein, 2000; Olatunji, Cisler, & Tolin, 2007). One type of behavioral avoidance is safety behavior, which aims to prevent threat when executed while confronting a threat-signalling stimulus. For instance, a clinically socially anxious individual may actively avoid pauses when giving a public speech to avoid appearing anxious, thus reducing the perceived threat of getting negatively criticized. Another type of behavioral avoidance is avoidance of learnt fear, which refers to behavioral responses that prevent the occurrence of a threat-signalling stimulus in the first place, and ultimately the threat that follows. For instance, the aforementioned individual may avoid attending a conference to avoid the feared situation (public speech) and the potential threat that follows (getting negatively evaluated).

Both types of behavioral avoidance can be modelled in the laboratory via a fear and avoidance conditioning protocol, which combines both Pavlovian fear acquisition and avoidance learning. During Pavlovian fear acquisition, an initially neutral conditioned stimulus (CS) is repeatedly paired with an aversive unconditioned stimulus (US). In a following avoidance learning phase, executing a designated response during CS presentation effectively prevents US occurrence (i.e., US-avoidance), paralleling safety behavior when confronting a warning signal. The laboratory examination of avoidance of learnt fear entails behavioral responses that prevent the occurrence of a CS itself, and the subsequent US it potentially signals (i.e., CS-avoidance). Safety behavior has been extensively examined in the laboratory in the past decade (e.g., Lovibond, Mitchell, Minard, Brady, & Menzies, 2009; Engelhard, van Uijen, van Seters, & Velu, 2015; Flores, López, Vervliet, & Cobos, 2018; Rattel, Miedl, Blechert, & Wilhelm, 2017; Pittig, 2019; Pittig & Wong, 2021; Vervliet & Indekeu, 2015; Vervliet, Lange, & Milad, 2017; Wong & Pittig, 2021), however, laboratory studies on avoidance of learnt fear are relatively scarce.

Higher-order conditioning paradigm is suitable for examining CS-avoidance (Wong, Wirth, & Pittig, under review). This laboratory model entails that not only a CS+ evokes conditioned fear, but also a preceding higher-order CS that signals the CS+, despite this higher-order CS never being directly associated with the US (Gewirtz & Davis, 2000). This paradigm provides great clinical value as it allows the investigation of how a stimulus that predicts a feared stimulus but does not directly signal threat comes to evoke fear and avoidance responses,

* Correspondence to: Lehrstuhl für Psychologie I, Marcusstraße 9-11, Würzburg, Germany.
E-mail address: hon.wong@uni-wuerzburg.de (A.H.K. Wong).

https://doi.org/10.1016/j.biopsycho.2021.108249
Received 6 August 2021; Received in revised form 23 December 2021; Accepted 27 December 2021
Available online 29 December 2021

explaining why clinically anxious individuals avoid stimuli or situations that merely signal a feared stimulus. The association between the higher-order CS and the CS+ can be acquired before CS-US pairings (sensory preconditioning) or after CS-US pairings (second-order conditioning). Both sensory preconditioning and second-order conditioning paradigms are viable for examining CS-avoidance. Indeed, animal studies showed heightened avoidance responses to a higher-order CS in a sensory preconditioning paradigm (Davis & Thompson, 1969; Hoffeld, Kendall, Thompson, & Brogden, 1960) and in a second-order conditioning paradigm (Tabone & de Belle, 2011; Topál & Csányi, 1999), indicating an acquisition of CS-avoidance in animals. Of note, avoidance generalization, via either a perceptual or conceptual pathway, also accounts for avoidance responses to stimuli not directly associated with threat (e.g., Boyle, Roche, Dymond, & Hermans, 2016; Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015; Klein, Berger, Vervliet, & Shechner, 2021a; Lemmens, Beckers, Dibbets, Kang, & Smeets, 2021; Lommen, Engelhard, & van den Hout, 2010; San Martin, Jacobs, & Vervliet, 2020; van Meurs, Wiggert, Wicker, & Lissek, 2014). However, higher-order conditioning is distinct from generalization (e.g., Prewitt, 1967; Rizley & Rescorla, 1972). Specifically, CS-avoidance entails avoidance to a stimulus that signals a CS+, while avoidance generalization occurs to stimuli that resemble a CS+, but do not necessarily signal it. In other words, although both higher-order conditioning and generalization explain how stimuli not directly associated with threat come to evoke avoidance, the underlying mechanisms are distinct from each other. The current study will focus on the relatively understudied CS-avoidance. The assessment of CS-avoidance acquisition in humans via higher-order conditioning is relatively scarce. Similar to animal studies, some higher-order conditioning studies showed that a higher-order CS that signalled a fear-related CS+ evoked avoidance responses (Cho & Mitchell, 1971; Declercq & De Houwer, 2009; Malloy & Levis, 1988). A common issue of these studies is that the acquired avoidance could not be properly justified as CS-avoidance. Avoidance responses in some of these studies (Cho & Mitchell, 1971; Declercq & De Houwer, 2009) were initially acquired during CS presentation to prevent an aversive US, thus were learned as *US-avoidance responses*. Similarly, Malloy & Levis (1988) allowed avoidance responses to be executed during both higher-order CS and CS+ presentations, thus confounding CS-avoidance and US-avoidance. To address this issue, a recent study examined the acquisition of CS-avoidance by confining the acquisition of avoidance responses during higher-order CS presentations (Klein, Berger, Vervliet, & Shechner, 2021b). In this study, stimuli were presented in a serial manner: participants were first presented with a higher-order CS that was followed by a first-order CS, which was directly followed by an electric US. Another non-reinforced first-order CS (CS-) was also presented. In a following avoidance conditioning phase, pressing a designated key during the higher-order CS prevented the first-order CS and the following US, whereas the CS- was unreinforced regardless of key presses. Participants exhibited more frequent avoidance responses to the higher-order CS compared to the CS-. This pattern persisted even after extinction learning to the first-order CS.

Although Klein et al. (2021b) nicely demonstrated the acquisition of CS-avoidance in humans, such avoidance required minimal effort or cost (e.g., merely pressing a key). Such so called low-cost avoidance has been criticized to be unable to tap into the pathological domain of anxiety-related disorders (Krypotos, Vervliet, & Engelhard, 2018; Pittig, Wong, Glück, & Boschet, 2020). Indeed, behavioral avoidance in anxiety-related disorders often bears a cost, for instance, individuals with PTSD reported taking a lengthy detour to avoid places where traumatic reminders are likely to be encountered (Corrigan, Samuelson, Fridlund, & Thomé, 2007). In light of this, some laboratory studies incorporated a competing reward to behavioral avoidance (e.g., Claes, Karos, Meulders, Crombez, & Vlaeyen, 2014; Pittig, 2019; Pittig & Wong, 2021; Rattel et al., 2017; Wong & Pittig, 2021), that is, incorporating an appetitive outcome for disengaging from avoidance responses. In other words, executing the avoidance response would lead to the omission of a reward, rendering avoidance costly, thus more closely modelling pathological avoidance in anxiety-related disorders. In fact, preliminary evidence has shown that individuals with anxiety-related disorders showed elevated costly, but not low-cost avoidance in a human avoidance conditioning study (Pittig, Boschet, Glück, & Schneider, 2020).

Therefore, in the current study, we examined the acquisition of costly CS-avoidance with a higher-order conditioning paradigm. Specifically, we employed a sensory preconditioning paradigm because it may provide insights into how stimuli or situations that are linked to a CS prior to trauma exposure may subsequently evoke behavioral avoidance after trauma exposure. The current study also took advantage of a recently developed dimensional measure of avoidance (Wong & Pittig, 2021). For this non-dichotomous measure, participants could engage in avoidance on a continuous scale from 0% to 100%. The extent of avoidance engagement was negatively proportional to the outcome occurrence and the amount of competing reward. Indeed, avoidance in anxiety-related disorders can be engaged in to a certain extent (Krypotos et al., 2018; Telch & Lancaster, 2012), and is often executed to balance threat at a subjectively acceptable level while limiting the cost of avoidance (cf. Schlund et al., 2016). To summarize, the current study aimed to assess the acquisition of costly CS-avoidance in a sensory preconditioning paradigm with a dimensional measure of avoidance. Our main hypothesis was an increase in costly avoidance to the higher-order CS after the first-order CS had acquired threat value.

2. Method

2.1. Participants

Students or residents from Würzburg were recruited as participants and were compensated by either 9€ or partial course credit. In addition, participants received extra monetary reward depending on their overall avoidance ratings throughout the experiment. A total of 50 participants were recruited. We carried out data analyses with linear mixed models (see *Scoring and analysis*), and did not carry out a power analysis for two reasons. First, there is no agreement in the various methods for power detection in linear mixed models, whereas the different methods (e.g., different types of variance-covariance structures for errors, different estimation methods) for power analyses could produce different power calculations (Bahçecitapar, 2018). Second, there is little to no prior data for appropriate evaluation of the power calculation, especially for interactions (Mathieu, Aguinis, Culpepper, & Chen, 2012). This is critical as we are primarily interested in the interaction between Phase (Baseline costly avoidance vs Test) and Stimulus type (PS1 vs PS2; see *Scoring and analysis*). Therefore, we followed a recent study that examined low-cost CS-avoidance (Klein et al., 2021b), which recommended a sample of at least 40 participants. We recruited a total of 50 participants to account for attrition rate due to exclusion criteria and technical difficulties. This study was approved by the Ethics Committee of the Institute of Psychology at the University of Würzburg in accordance with the Declaration of Helsinki.

2.2. Apparatus and materials

Four standardized 2D black and white drawings from Snodgrass and Vanderwart (1980; dog, apple, bicycle, & bed) and four geometric shapes with different colors (orange triangle, purple hexagon, red circle, & green square) were employed as visual stimuli presented in the experiment. All stimuli were individually presented in the centre of a white screen.

A computer equipped with Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, Version 20.1) presented all visual stimuli and recorded all self-reported ratings. Another computer with BrainVision Recorder (Brain Products, GmbH, Gliching, Germany) measured the skin conductance via two Ag/AgCl electrodes at a sampling rate of 1000

Hz. A DS7A Digitimer stimulator generated an electric US which consisted of 125 pulses separated by 5ms (i.e., US duration of 625 ms).

2.3. Procedure

After providing written informed consent, participants filled in the German version of DASS-21 (Lovibond & Lovibond, 1995; Nilges & Essau, 2015) and the German version of Intolerance of Uncertainty scale (Freeston, Rhéaume, Letarte, Dugas, & Ladouceur, 1994; Gerlach, Andor, & Patzelt, 2008). The DASS-21 measures and discriminates three different constructs: depression, anxiety, and stress, whereas the short German version of Intolerance of Uncertainty scale (IU-18) measures constructs that involve intolerance of uncertainty and worry (see Carleton, Norton, & Asmundson, 2007). Next, we attached skin conductance electrodes filled with isotonic gel to the hypothenar muscles on the palm of participants' non-dominant hand. US electrodes were also attached to participants' wrist on the same hand.

Participants were led through a US workup procedure, in which an US intensity of 0.2 mA was gradually increased until it reached a level that was individually perceived as 'definitely unpleasant but not painful'. Immediately after US calibration, we carried out a reward-matching procedure. This procedure was highly similar to our recent study (Wong & Pittig, 2021), entailing a series of questions "Are you willing to tolerate the selected level of electric stimulation if you are given €_?", with the amount of reward ranging from 5 to 31 cents in odd numbers (i.e., 5 cents, 7 cents,…, 31 cents) presented in a randomized order. A total of 14 questions were presented individually. Participants were prompted to answer either 'Yes' or 'No' to each of these questions. The amount of competing reward between the highest amount that received a 'No' and the lowest amount that received a 'Yes' was selected as the competing incentive for CS-avoidance disengagement. For instance, if an individual participant was unwilling to tolerate an US when given 5 to 21 cents, but was willing to tolerate it when given 23 cents or more, the amount in between (22 cents) would be chosen as the maximum amount of competing reward per trial. This individually calibrated competing reward was presumably neither too high such that would artificially decrease the degree of CS-avoidance nor too low such that would induce an opposite pattern (see Schlund et al., 2016; Wong & Pittig, 2021).

Before the experiment started, four black and white drawings and four colored geometric shapes were presented individually in a randomized order. Participants were prompted to indicate valence and arousal ratings to each of these stimuli using a visual analog scale (valence: -50 = unpleasant, 0 = neutral, and 50 = pleasant; arousal: 0 = not aroused at all and 100 = highly aroused). The scales were presented below the visual stimuli.

The conditioning task consisted of four phases: Preconditioning, Baseline costly avoidance, Pavlovian fear acquisition, and Test (see Table 1).

2.3.1. Preconditioning

Only two out of the four black and white drawings, dog and apple, served as the preconditioning stimuli (PSs), PS1 and PS2. Similarly, only two out of the four colored geometric shapes, an orange circle and a purple hexagon, served as CS1 and CS2 (see Fig. 1A). Each PS was presented eight times, in which PS1 was always followed by CS1, whereas PS2 was always followed by CS2. Each PS was presented alongside a bidirectional CS expectancy visual analog scale, in which the right end of the scale indicates 100% CS1 expectancy, whereas the left end of the scale indicates 100% CS2 expectancy, and the middle of the scale indicates neither presentation of the CSs. The CS expectancy scale was presented alongside a question "How likely would you expect the following outcomes?". After CS expectancy was indicated, the visual analog scale disappeared and the PS remained on screen for 8 s. Immediately after PS offset, the corresponding CS was presented for 4 s. The intertrial intervals (ITIs) were 4 s. The presentation order was pseudo-randomized in a way so that the same trial type would not be presented more than two times in a row. This pseudo-randomization of presentation order was applied to all the following phases. The CSs were counterbalanced across all participants.

2.3.2. Baseline costly avoidance

Prior to this phase, participants were instructed that they could avoid the outcome that followed the PSs, by indicating their avoidance ratings at the avoidance scale presented alongside the PS. The avoidance ratings were negatively proportional to the chance of CS presentation. For instance, an avoidance rating of 70% would result in a 30% chance of CS presentation, or a 70% chance of CS omission for that particular trial. Participants were also informed that whenever avoidance was available, each trial would come with a competing reward. The amount of competing reward was, however, inversely proportional to the avoidance made. For instance, an avoidance response of 70% would result in a gain of 30% of the maximum reward. Participants were instructed that all rewards gained throughout the task would be paid at the end of the experiment. The PS and the avoidance scale were presented on screen until choice. After avoidance ratings had been indicated, a 1 s fixation cross appeared, followed by the CS expectancy scale. After CS expectancy ratings were indicated, the PS remained on screen for 8 s. Depending on the avoidance ratings made, either the corresponding CS or a white blank screen was presented for 4 s. Reward feedback was then presented for 2 s. The ITI was randomized between 11 and 15 s, and the same was applied to all the following phases. See Fig. 1B for an example of trial structure in this phase.

2.3.3. Pavlovian fear acquisition

Prior to this phase, participants were informed that avoidance responses and the competing rewards were temporarily removed. The two CSs were presented eight times each. CS1 was reinforced by an electric US at a 75% rate (i.e., 6 out of 8 trials) whereas CS2 was never reinforced. Each CS was first presented alongside an US expectancy scale, in which the left end of the scale indicates 0% US expectancy, whereas the right end of the scale indicates 100% US expectancy, and the middle indicates 50% US expectancy. The US expectancy scale was presented alongside a question asking "How likely would you expect an electric stimulation?". After US expectancy ratings had been made, the CS remained on the screen for 8 s, which was then followed by an US depending on the trial type. The presentation order was pseudo-randomized so that the first and the last trial of CS1 were always reinforced.

Table 1
PS indicates preconditioning stimuli; CS indicates conditioned stimuli; + indicates US presentation; - indicates US omission; * indicates the availability of avoidance; CS and € in brackets indicate the presentation of a CS and a competing reward, respectively, depend on avoidance; Number in parentheses indicates the number of trials.

Preconditioning	Baseline costly avoidance	Pavlovian fear acquisition	Test
PS1 → CS1 (8)	PS1* [CS1, €] (8)	CS1+ (6)	PS1*- [€] (4)
PS2 → CS2 (8)	PS2* [CS2, €] (8)	CS1- (2)	PS2*- [€] (4)
		CS2- (8)	

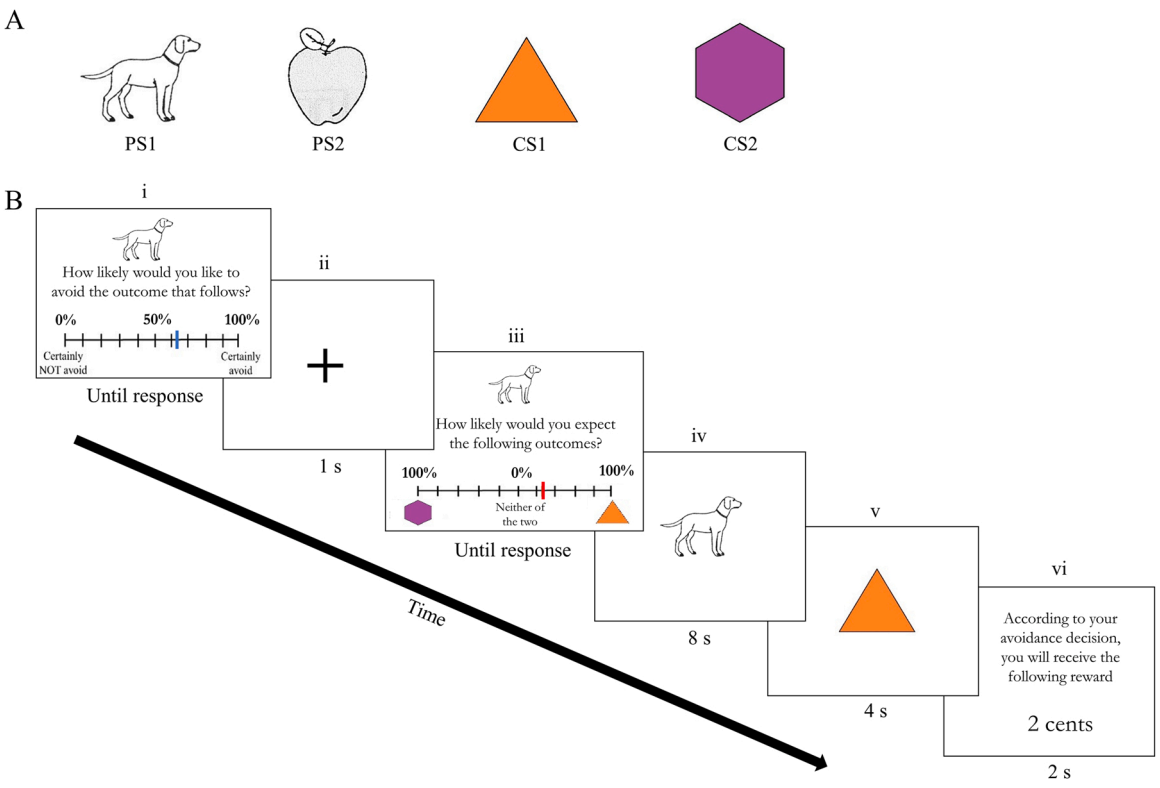

Fig. 1. A. Stimuli used in the experiment. The black and white drawings always served as PSs, whereas the color geometric figures always served as CSs. B. Example of the trial structure during the Baseline costly avoidance phase. (i) Participants had to indicate their CS-avoidance. (ii) A fixation cross appeared for 1 s. (iii) The PS appeared again with the CS expectancy scale until response. (iv) The CS expectancy scale was removed, and the PS was presented alone for 8 s. (v) The corresponding CS or a blank screen was presented for 4 s depending on CS-avoidance made. In this example a CS was presented. (vi) The reward feedback was presented for 2 s.

2.3.4. Test

Participants were informed that avoidance was available again, and they could choose to prevent the outcomes that potentially followed the PSs. Participants were also reminded that their avoidance ratings determined the magnitude of reward on each trial. The two PSs were presented four times each. On each trial, the PS was presented alongside the avoidance scale. After avoidance ratings had been made, a 1 s fixation cross appeared and participants were then prompted to indicate their CS expectancy. Upon responding, a 1 s fixation cross appeared again and participants were prompted to indicate their US expectancy. When all ratings were made, the PS remained on screen for 8 s. None of the stimuli in this phase were reinforced by a CS nor an US. In other words, a blank screen appeared for 4 s immediately after PS offset, followed by a reward feedback of 2 s.

After the experiment, participants were prompted to indicate their valence and arousal ratings to each of the PSs and the CSs. However, we only reported the change in valence ratings in the main text (see Supplementary Materials for the analyses regarding change in arousal ratings[1]).

2.4. Scoring and analysis

Skin conductance prior to Pavlovian fear acquisition was not analysed due to the absence of anticipatory fear. Only skin conductance recorded during the 8 s of stimulus presentations were analysed (i.e., the PSs in Test and the CSs in Pavlovian fear acquisition). However, for the sake of transparency, SCRs in all phases are depicted in Fig. 2. A 50 Hz notch filter and a 1 Hz high-pass filter that removes high-frequency noise were applied to the skin conductance data. Next, we calculated the SCRs by finding the difference between the maximum response and the corresponding trough in the interval of 1 s after stimulus onset to stimulus offset. This was done by a two-step procedure. First, we used BrainVision Analyzer to automatically detect the maximum and minimum responses in each stimulus interval. We then checked if the minimum response was located at the trough corresponding to the maximum response (e.g., a stimulus interval with multiple peaks); if not, we manually adjusted it accordingly. If no responses were detected, it would be scored as zero. At last, the SCR data was then square root transformed to reduce skewness (Boucsein et al., 2012).

We analysed all data within a linear mixed model framework. The analyses were separated into three parts: manipulation check, main hypotheses, and exploratory analyses.

2.4.1. Manipulation check

We first analysed whether participants acquired the association between the PSs and the CSs in the preconditioning phase. Thus, CS expectancy ratings served as dependent variable, whereas Stimulus type (PS1 vs PS2) and a linear trend repeated measures across trials (Trial) served as fixed effects. Similarly, to check whether participants acquired differential conditioned fear to the CSs in Pavlovian fear acquisition, US expectancy ratings or SCRs served as dependent variable, whereas CS type (CS1 vs CS2) and Trial (a linear trend repeated measures across trials) served as fixed effects. The interactions of these fixed effects were of primary interest to evaluate the development of differential responding to the PSs or the differential responding to the CSs.

We also compared baseline avoidance to the PSs in Baseline costly avoidance. Given that we hypothesized no differences in baseline avoidance to both PSs, we used a Bayesian approach to support the absence of an effect (Kruschke, 2015). In this particular model, we obtained the 95% highest density intervals (HDIs), which contain the most

[1] We had no a priori expectations for arousal ratings, thus we moved related analyses to the Supplementary Materials.

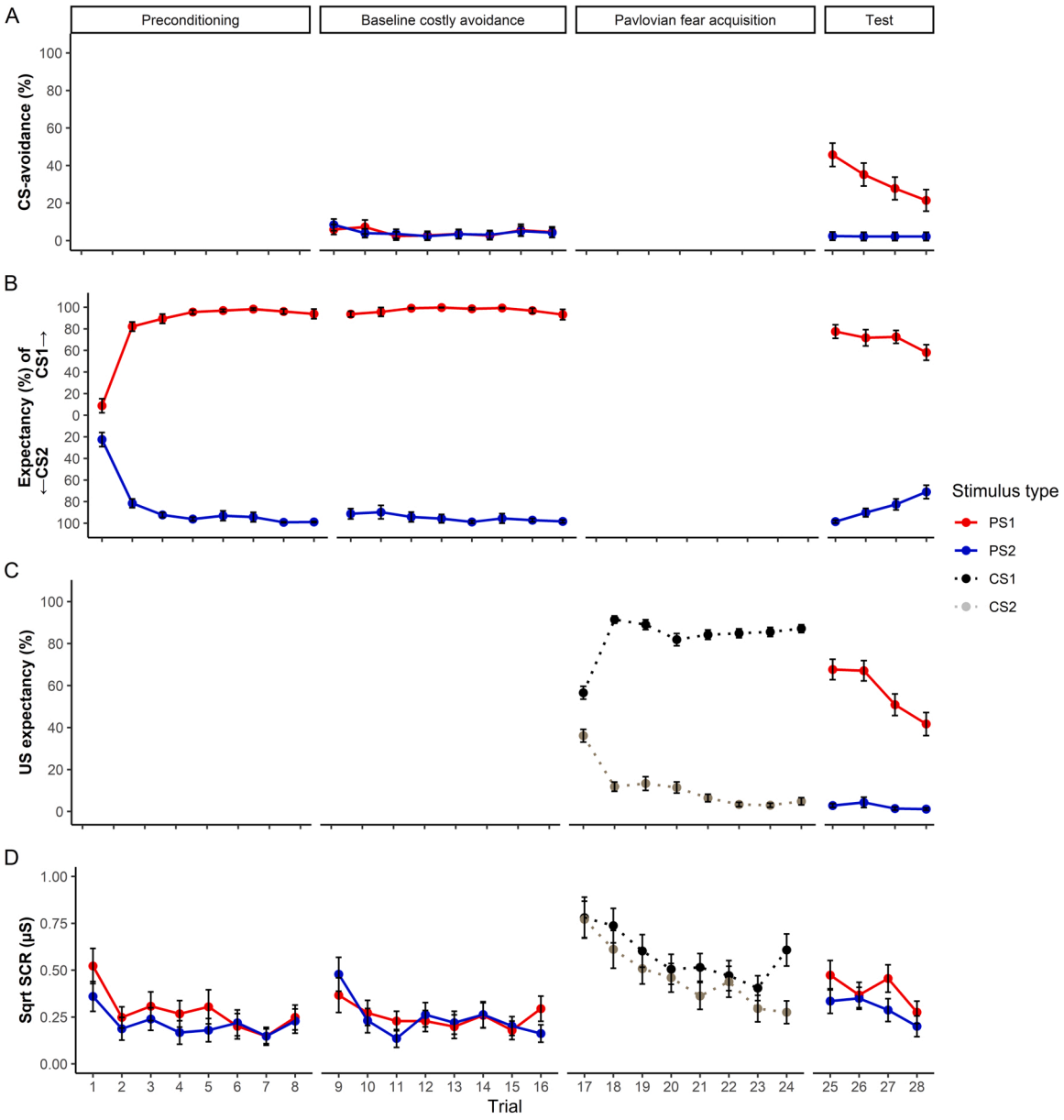

Fig. 2. (A) CS-avoidance, (B) CS expectancy ratings, (C) US expectancy ratings, and (D) square-root SCRs across all phases. See the color version of this figure online.

credible values, paralleled with the 95% confidence interval in frequentist analyses. We calculated the posterior distribution (obtained via Markov Chain Monte Carlo) that falls under the area around the null value, namely the Region Of Practical Equivalence (ROPE). We then calculated the percentage of HDIs that fell into areas under ROPE; the higher this value is, the more likely it indicates an absence of an effect (Kruschke, 2015; Kruschke & Liddell, 2018). To this end, avoidance served as dependent variable, whereas Stimulus type (PS1 vs PS2) and Trial (linear trend repeated measures across trials) served as fixed effects in this model.

2.4.2. Main hypotheses

Our primary interest was to examine the acquisition of CS-avoidance. To this end, we first analysed whether participants exhibited differential costly avoidance to the PSs (i.e., stronger avoidance to PS1 compared to PS2) in Test, and whether the extent of costly avoidance decreased across trials (i.e., extinction learning). More importantly, to further confirm the acquisition of costly avoidance, we compared avoidance responses on the first trial of Test to the last trial of Baseline costly avoidance. The remaining trials in Test were not included to preclude the ongoing effect of extinction learning. Thus, avoidance served as dependent variable, whereas Stimulus type (PS1 vs PS2) and Trial (last trial of Baseline costly avoidance vs first trial of Test) served as fixed effects. The interaction of these fixed effects were evaluated to examine the differential increase in avoidance response to the PSs before and after fear acquisition to the CSs. Furthermore, we assessed the changes in expectancy ratings or SCRs to the PSs in Test. To this end, CS expectancy ratings, US expectancy ratings, or SCRs served as dependent variable, whereas Stimulus type and Trial served as fixed effects.

Second, we tested whether an increase in CS-avoidance would predict a decrease in conditioned fear, especially to PS1. These analyses were only carried out on the first trial of Test to prevent the confounding effect of extinction learning. Thus, US expectancy ratings or SCRs served as dependent variable, whereas Avoidance and Stimulus type (PS1 vs

Table 2
Demographic data for the final sample.

	Mean (Standard deviation)
Age	26.62 (7.48)
Sex – Female	31 (68.89%)
US intensity (mA)	0.97 (0.55)
DASS21-Anxiety (0-42)	3.51 (5.48)
DASS21-Depression (0-42)	5.87 (6.20)
DASS21-Stress (0-42)	9.24 (7.85)
IU-18 (18-90)	41.09 (12.18)

Table 3
95% highest density intervals (HDI) for each fixed effect in the models of Baseline costly avoidance. ROPE: Region of Practical Equivalence.

Parameters	HDI low	HDI high	ROPE%
Stimulus type	-1.29	1.47	100%
Trial	-0.53	0.08	100%
Stimulus type* Trial	-0.54	0.69	100%

PS2) served as fixed effect.

Third, we assessed the change in valence ratings to the PSs and CSs before and after the experiment. Thus, valence ratings served as dependent variable, whereas Phase (*pre-* and *post*-experiment), Trial type (PSs vs CSs), and Threat type (PS1 & CS1 vs PS2 & CS2) served as fixed effects. Interactions of these fixed effects were evaluated to examine the changes in valence to all stimuli before and after the experiment.

2.4.3. *Exploratory analyses*

Finally, we carried out two exploratory analyses. First, we examined whether the change in valence ratings to both PSs (i.e., the valence ratings after the experiment minus the valence ratings before the experiment for each PS) would be associated with the extent of CS-avoidance engagement on the first trial in Test. Thus, avoidance served as dependent variable, whereas the change in valence ratings to each PS after the experiment served as fixed effect. Second, we examined whether higher trait anxiety or intolerance of uncertainty would be associated with an increase in differential CS-avoidance to the PSs (i.e., difference in avoidance between the PSs) on the first trial in Test. To this end, we carried out robust linear regression analyses with differential avoidance as dependent variable, whereas trait anxiety or intolerance of uncertainty served as a predictor. Robust regression linear analyses were carried out to minimize the influence of outliers by reweighting least squares estimation in an iterative manner (Koller & Stahel, 2011; see also Field & Wilcox, 2017).

For all the aforementioned linear mixed models, participants served as a random effect. The degree of significance was reported with Satterthwaite approximation for degrees of freedom (Satterthwaite, 1941). Importantly, the main effects and higher-order interactions were analysed in separate models (Hayes, Glynn, & Huge, 2012). All analyses were carried out using R (R Core Team, 2021), with *lmer* package for frequentist models (Bates, Mächler, Bolker, & Walker, 2015), *rstanarm* (Goodrich, Gabry, Ali, & Brilleman, 2020) and *bayestestR* (Makowski, Ben-Shachar, & Lüdecke, 2019) for the Bayesian models, and *robustbase* for the robust regression analyses (Maechler et al., 2021). Effect sizes for the fixed effects in the frequentist models were reported as partial-R^2 (see Jaeger, Edwards, Das, & Sen, 2017) with *r2glmm* package (Jaeger, 2017).

3. Results

We restricted analyses to participants who had 1) acquired differential CS expectancy ratings to the PSs and 2) acquired differential US expectancy ratings to the CSs. The first criterion was defined by an average difference of at least 100 points in CS expectancies on the last four trials of Preconditioning (maximum difference was 200 points in CS expectancies); the second criterion was defined by an average difference of at least 50 points in the last four trials of Pavlovian fear acquisition (maximum difference was 100 points in US expectancies), same as in our prior study (Wong, Glück, Boschet, & Engelke, 2020). Three participants were excluded based on these criteria. Two additional participants were excluded for requesting to reduce the US intensity during the experiment. In sum, 5 participants were excluded, leaving a total of 45 participants (see Table 2 for descriptive statistics for the final sample) for

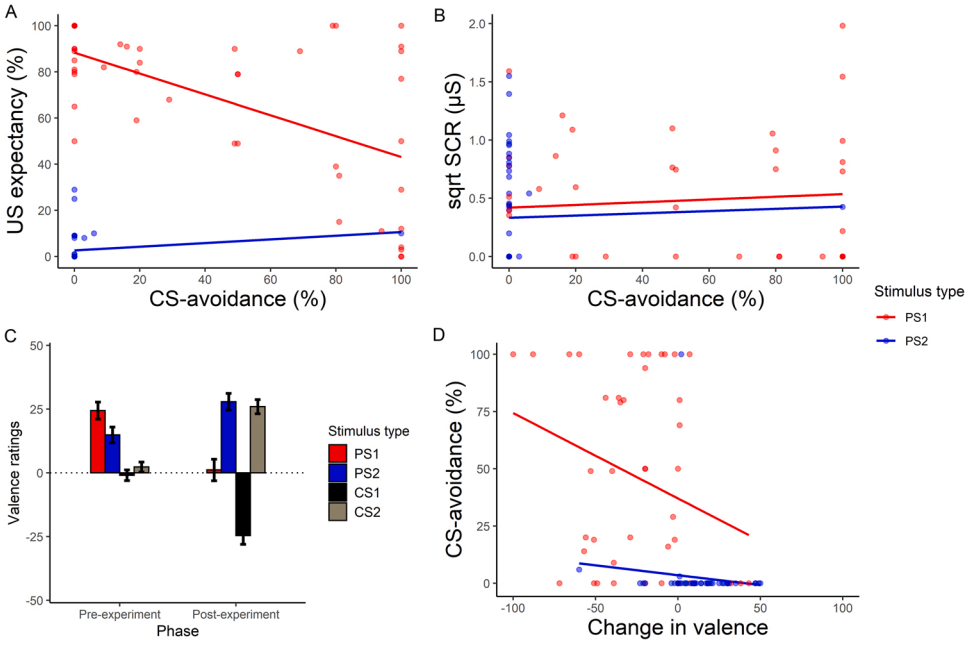

Fig. 3. Top panel. Relationship between CS-avoidance and the subsequent conditioned fear as indicated by (A) US-expectancy ratings and (B) SCRs on the first trial of Test. Bottom Panel. Association of the change in valence and CS-avoidance. (C) The comparison of valence ratings to the PSs and CSs *pre-* and *post*-experiment. (D) The relationship between the change in valence to the PSs and CS-avoidance on the first trial of Test. Negative value on the x-axis indicates a negative change in valence, whereas a positive value indicates an opposite pattern. Red dots represent responding to PS1 whereas blue dots represent responding to PS2. Darker color indicates more overlapping data points. The lines represent the line of best fit for each PS for visual aid. See the color version of this figure online.

statistical analyses[2].

3.1. Manipulation check

3.1.1. Preconditioning

3.1.1.1. CS expectancy. Fig. 2A shows CS expectancy ratings to the PSs . Participants exhibited a rapid development of differential CS expectancies to the PSs across trials, confirmed by a significant interaction between Stimulus type and Trial, $b_{Stimulus\ type*Trial} = 965.10$, $SE = 62.31$, $p < .001$, $R^2 = 0.30$. This suggests that participants acquired the correct PS-CS contingencies.

3.1.2. Baseline costly avoidance

3.1.2.1. CS-avoidance. Fig. 2B shows the CS-avoidance responses to the PSs. Participants showed similarly low levels of costly CS-avoidance to both PSs. There was no evidence that differential CS-avoidance developed across trials, $b_{Stimulus\ type*Trial} = 5.04$, $SE = 18.69$, $p = .787$, $R^2 < 0.001$, neither was there any evidence of differential CS-avoidance to the PSs averaged across trials, $b_{Stimulus\ type} = 0.056$, $SE = 0.70$, $p = .937$, $R^2 < 0.001$. The Bayesian model further confirmed an absence of these effects: 100% of the HDIs depicting the interaction between Stimulus type and Trial fell within ROPE, the same also applies for the HDIs depicting the main effect of Stimulus type (see Table 3). This means that baseline costly CS-avoidance to both PSs did not differ from each other.

3.1.2.2. CS expectancy. Participants continued to exhibit strong differential CS expectancies to the PSs averaged across trials, confirmed by a significant main effect of Stimulus type, $b_{Stimulus\ type} = 192.06$, $SE = 1.60$, $p < .001$, $R^2 = 0.96$. The maintenance of such strong differential responding was due to two reasons. First, participants had already acquired the PS-CS contingencies in the previous phase. Second, participants showed very low levels of CS-avoidance, thus strongly expecting the CSs to follow the corresponding PSs. No other effects reached significance (smallest $p = .098$, largest $R^2 = 0.005$).

In sum, participants exhibited similarly low levels of baseline costly CS-avoidance to both PSs. Strong differential CS expectancies to the PSs was observed due to little to no costly CS-avoidance in this phase.

3.1.3. Pavlovian fear acquisition

3.1.3.1. US expectancy and SCRs. Fig. 2C shows the US expectancy ratings to the CSs. Participants developed an increase in US expectancy ratings to CS1 and a decrease in US expectancy ratings to CS2 across acquisition trials, confirmed by a significant interaction between CS type and Trial, $b_{CS\ type*Trial} = 345.45$, $SE = 33.01$, $p < .001$, $R^2 = 0.16$.

For the SCRs (Fig. 2D), we did not observe any development in differential responding to the CSs, $b_{Stimulus\ type*Trial} = 1.54$, $SE = 0.97$, $p = .113$, $R^2 = 0.004$. In fact, responding to both CSs decreased across trials, $b_{Trial} = -3.15$, $SE = 0.49$, $p < .001$, $R^2 = 0.061$, presumably due to habituation to the stimuli. Averaged across all trials, participants exhibited stronger responding to CS1 compared to CS2, confirmed by a main effect of CS type, $b_{CS\ type} = 0.11$, $SE = 0.037$, $p = .003$, $R^2 = 0.014$.

In sum, participants acquired differential conditioned fear to the CSs in both US expectancies and SCRs measures.

3.2. Main hypotheses

3.2.1. Test

3.2.1.1. CS-avoidance. As hypothesized, participants exhibited stronger CS-avoidance to PS1 than to PS2. While CS-avoidance to PS1 decreased across test trials, CS-avoidance to PS2 maintained relatively stable across test trials (Fig. 2A). This pattern was confirmed by a significant interaction between Stimulus type and Trial, $b_{Stimulus\ type*Trial} = -169.14$, $SE = 45.51$ $p < .001$, $R^2 = 0.033$.

3.2.1.2. CS expectancy. Participants exhibited strong differential CS expectancies to the PSs (Fig. 2B), however, the magnitude of differential CS expectancies declined across test trials due to extinction learning, $b_{Stimulus\ type*Trial} = -312.16$, $SE = 74.87$, $p < .001$, $R^2 = 0.050$. Although participants engaged in CS-avoidance to PS1 more frequently than to PS2, differential CS expectancies to the PSs remained significant, $b_{Stimulus\ type} = 155.52$, $SE = 4.06$, $p < .001$, $R^2 = 0.82$. This pattern was presumably due to participants not fully engaging in CS-avoidance, thus leaving some residual CS expectancies.

3.2.1.3. US expectancy and SCRs. Participants exhibited higher US expectancy ratings to PS1 compared to PS2, however, while US expectancies to PS1 declined across test trials, US expectancies to PS2 remained steadily at a low level (Fig. 2C). This pattern was confirmed by a significant interaction between Stimulus type and Trial, $b_{Stimulus\ type*Trial} = -182.67$, $SE = 44.11$, $p < .001$, $R^2 = 0.045$. The initial higher level of US expectancies to PS1 was presumably reflecting residual fear because of partial engagement in CS-avoidance; US expectancies declined across test trials due to extinction learning. In contrast, US expectancies to PS2 were low across Test despite the low level of CS-avoidance to it.

For the SCRs measure (Fig. 2D), participants exhibited stronger SCRs to PS1 than to PS2 averaged across test trials, $b_{Stimulus\ type} = 0.10$, $SE = 0.037$, $p = .007$, $R^2 = 0.018$. Unlike US expectancies, there was a general decrease in SCRs to the PSs across test trials, $b_{Trial} = -1.03$, $SE = 0.35$, $p = .003$, $R^2 = 0.022$. No differences for the decline in responding were observed between the PSs, $b_{Stimulus\ type*Trial} = -0.08$, $SE = 0.69$, $p = .904$, $R^2 < 0.001$.

In sum, participants exhibited stronger costly CS-avoidance to PS1 than PS2, however, CS-avoidance to PS1 declined across trials due to extinction learning. As a result, CS expectancies to PS1 decreased across trial due to extinction of CS occurrence. Similarly, conditioned fear to PS1 declined due to extinction of US occurrence, as indicated by a decrease in US expectancy ratings and SCRs across test trials.

3.2.2. Cross-phase analysis: comparing CS-avoidance between Baseline costly avoidance and Test

When comparing CS-avoidance to the PSs on the last trial of Baseline costly avoidance to the first trial of Test, we observed an increase in CS-avoidance to PS1, whereas CS-avoidance to PS2 remained at a low level across the two phases. This pattern was supported by a significant interaction between Stimulus type and Trial, $b_{=Stimulus\ type*Trial} = 43.04$, $SE = 6.36$, $p < .001$, $R^2 = 0.18$.

3.2.3. CS-avoidance predicting US expectancies or SCRs

On the first trial of Test, an increase in CS-avoidance decreased US expectancy ratings to PS1 more than PS2 (Fig. 3A), confirmed by a significant interaction between Stimulus type and Avoidance, $b_{Stimulus\ type*Avoidance} = 0.53$, $SE = 0.20$, $p = .010$, $R^2 = 0.074$. In contrast, CS-avoidance on the first test trial had no differential predictive value on SCRs to the PSs, $b_{Stimulus\ type*Avoidance} = 0.0014$, $SE = 0.0041$, $p = .724$, $R^2 = 0.001$ (Fig. 3B).

3.2.4. Valence ratings

Fig. 3C shows the valence ratings to the PSs and the CSs before and after the experiment. Only interactions involving Phase were reported here given our primary interest was the change in valence ratings to the stimuli post-experiment. Averaged across Stimulus type, we observed a decrease in valence ratings to the threat-relevant stimuli (PS1 & CS1)

[2] All statistical analyses remain similar when including all participants.

post-experiment but an increase in valence ratings to safety-relevant stimuli (PS2 & CS2) post-experiment, confirmed by a significant interaction between Phase and Threat type, $b_{Phase*Threat\ type} = -41.82$, $SE = 4.70$, $p < .001$, $R^2 = 0.20$. This 2-way interaction between Phase and Threat type, however, did not further interact with Stimulus type (PSs vs CSs), $b_{Phase*Threat\ type*Stimulus\ type} = 11.07$, $SE = 8.35$, $p = .186$, $R^2 = 0.005$, suggesting no evidence of any differences between the change in differential valence ratings to the PSs and the CSs. No other interactions reached significance (lowest $p = .333$, largest $R^2 = 0.003$).

3.3. Exploratory analyses

3.3.1. Association between a change in valence and CS-avoidance

Change in valence ratings to the PSs was negatively associated with CS-avoidance to the PSs in Test (Fig. 3D), supported by a main effect of Change in valence ratings, $b_{Change\ in\ valence} = -0.58$, $SE = 0.11$, $p < .001$, $R^2 = 0.25$. This indicates that a more negative change in valence ratings to the PSs was associated with a larger increase in CS-avoidance. However, there was no evidence indicating that the negative association between change in valence and CS-avoidance differed between PS1 and PS2, $b_{Change\ in\ valence*Stimulus\ type} = -0.38$, $SE = 0.26$, $p = .154$, $R^2 = 0.023$.

3.3.2. Association between risk factors and CS-avoidance

In regard to the association between individual risk factors and CS-avoidance, the DASS-Anxiety scores ranged from 0 to 28 ($M = 3.51$, $SD = 5.48$), whereas Intolerance of uncertainty ranged from 19 to 70 ($M = 41.08$, $SD = 12.18$). None of these risk factors were significantly associated with differential CS-avoidance to the PSs (lowest $p = .223$).

4. Discussion

Using a sensory preconditioning paradigm, the current study examined the acquisition of costly CS-avoidance to a preconditioning stimulus (PS1) that signalled a fear-related CS (CS1). A key finding was that participants exhibited a selective increase in costly avoidance to a PS that signalled a fear-related CS, but not to a PS (PS2) that signalled a safety-related CS (CS2). This differential avoidance to the PSs indicated the acquisition of costly CS-avoidance despite PS1 had no direct association with the US. Importantly, the differential CS-avoidance to the PSs in test was not due to any pre-existing differences in baseline avoidance to the PSs, given that costly avoidance to both stimuli did not differ from each other prior to Pavlovian fear acquisition. Thus, the current study provides preliminary evidence of the acquisition of costly CS-avoidance by means of sensory preconditioning.

Another major finding was that on the first test trial, an increase in CS-avoidance was associated with a decrease in US expectancy ratings. This pattern indicated that participants adjusted their US expectancies accordingly after engaging in various degrees of CS-avoidance, suggesting that threat anticipation plays a role in CS-avoidance. Furthermore, the decrease in CS-avoidance to PS1 across test trials indicated extinction learning. This decrease in CS-avoidance aligned with a general decrease in SCRs to the PSs in test, and a decrease in US expectancies specifically to PS1 in test. These patterns aligned with the findings of Wong and Pittig (2020), in which a decrease in avoidant decision to novel exemplars of the same category of a fear-related CS coincided with a decrease in threat expectancies. These patterns suggested that US anticipation and prevention may be a potential mechanism driving CS-avoidance, despite PS1 had no direct association with an US.

The aforementioned pattern aligned with and expand on the notions of the Expectancy model (Lovibond, 2006). This model puts forward the idea that behavioral avoidance is executed based on two assumptions: avoidance is executed when one has a propositional belief that a stimulus signals threat, and such avoidance is effective in preventing threat. Thus, a propositional belief that a PS signals a CS which in turns predicts a threatening US may lead to behavioral avoidance to the PS, with the ultimate aim of US prevention. The alignment between the decrease in CS-avoidance and US expectancies in test due to extinction aligns with this model, given that avoidance became unnecessary if the US was absent after PS presentation. The Expectancy model also predicts a decrease in anticipatory fear after avoidance has been executed, which aligns with the negative relationship between CS-avoidance and the subsequent US expectancy ratings. Interestingly, participants still exhibited stronger US expectancy ratings and SCRs to PS1 compared to PS2 in test, despite engaging in stronger avoidance to PS1. On face value, this pattern was in contrast to the prediction of the Expectancy model, which suggests a decrease in anticipatory fear after engaging in avoidance. However, this differential fear to the PSs was due to the partial engagement in CS-avoidance, thus leaving some residual fear to the PSs.

Besides threat anticipation and its prevention, CS-avoidance can also be driven by other potential factors. In fact, Klein et al. (2021b) provided preliminary evidence that CS-avoidance persisted after extinguishing the CS-US contingency, suggesting other non-threat-related factors may drive CS-avoidance. In fact, negative valence alone is thought to be sufficient to guide behavioral avoidance (Chen & Bargh, 1999; Hans Phaf, Mohr, Rotteveel, & Wicherts, 2014; Krieglmeyer, Deutsch, De Houwer, & De Raedt, 2010), even when an US is not anticipated (see referential account; Baeyens, Eelen, & Crombez, 1995; De Houwer, Thomas, & Baeyens, 2001). The current study provides preliminary support that a negative change in valence plays a role in driving CS-avoidance. First, participants exhibited a decrease in valence ratings to PS1 after the experiment, consistent to preliminary evidence in the literature (Yu, Lang, Birbaumer, & Kotchoubey, 2014). Second, a greater increase in negative valence to the PSs was associated with a stronger engagement in costly CS-avoidance. However, it should be noted that the current results do not entirely support the notion that negative valence *per se* can drive avoidance, given that the absolute valence to PS1 was merely neutral after CS1 had gained threat value. Instead, this pattern suggests that a negative *change* in valence is associated with avoidance responses.

One may argue that avoidance to PS1 could be seen as avoidance generalization. Indeed, the current paradigm was somewhat similar to studies examining symbolic generalization of avoidance employing a matching-to-sample (MTS) task (Dymond, Schlund, Roche, De Houwer, & Freegard, 2012; Dymond, Schlund, Roche, & Whelan, 2014; Dymond et al., 2011). In a typical MTS task, participants learn to classify neutral stimuli into two artificial categories. Following the MTS task, a stimulus from one category is paired with an aversive US (CS+), whereas a stimulus from the other category is not paired with an US (CS-). Participants then acquire differential US-avoidance to the CSs. Interestingly, US-avoidance selectively generalizes to stimuli of the same category of the CS+, despite these stimuli have no direct association with the US. These MTS studies seemingly share similar procedures with sensory preconditioning, specifically, the MTS task parallels with the preconditioning stage in which neutral stimuli were paired together. However, the stimulus relationship in sensory preconditioning entails that one stimulus signals the onset of another stimulus (i.e., a PS signals CS onset), whereas stimulus relationship in MTS has no inherent predictive value (i.e., stimuli merely belong to the same artificial category). Furthermore, avoidance is first acquired to the CS+ that prevents an US in the MTS studies, effectively rendering it as US-avoidance. Thus avoidance to stimuli of the same artificial category of CS+ is seen as generalization of US-avoidance. In contrast, the current study confined avoidance to a PS that prevented CS onset, thus rendering avoidance as CS-avoidance. To summarize, avoidance in the current sensory preconditioning paradigm can be seen as avoidance responses to stimuli that *signal* a feared stimulus (CS-avoidance), whereas avoidance in MTS studies can be seen as avoidance responses to stimuli that conceptually *resemble* a feared stimulus (generalization of US-avoidance).

We did not find any associations between individual risk factors and costly CS-avoidance, contrary to findings of positive associations

between risk factors and avoidance (e.g., Hunt, Cooper, Hartnell, & Lissek, 2020; Pittig, Schulz, Craske, & Alpers, 2014; San Martin et al., 2020). One potential reason for this null effect is that we had limited variance in trait anxiety and intolerance of uncertainty, given that we did not selectively recruit participants high and low in these risk factors. One may further argue that the finding of costly avoidance acquisition in the entire healthy sample, but not selectively in individuals at risk, may suggest that the competing reward was insufficient to encourage disengagement from CS-avoidance. However, this is unlikely due to two reasons. First, we individually calibrated the amount of competing reward so that it was at an optimal level to create a conflict between approach and avoidance (see Schlund et al., 2016). Second, unlike studies employing a dichotomous measure of avoidance in which engaging in costly avoidance forfeits the competing reward entirely (e.g., Pittig et al., 2014; Rattel et al., 2017; van Meurs et al., 2014), participants in the current study could engage in avoidance to a certain extent while retaining a portion of the competing reward. Thus, this encourages healthy participants to partially engage in costly avoidance. In fact, combining these two aspects, Wong and Pittig (2021) found that healthy individuals acquired costly avoidance. Therefore, we interpreted that the acquisition of costly CS-avoidance in a healthy sample was not due to an inadequate level of competing reward.

From a clinical perspective, costly CS-avoidance models the acquisition of maladaptive avoidance of learnt fear to stimuli associated with a fear-related stimulus. This closely parallels with individuals with PTSD taking a lengthy detour to avoid situations where trauma reminders can be encountered (Corrigan et al. 2007), or individuals with needle phobia avoiding medical centres where injections are likely to be encountered (e.g., Katz, 1974; Kleinknecht & Lenz, 1989). Importantly, alongside avoidance generalization (Boyle et al. 2016; Lommen, et al., 2010; van Meurs et al. 2014), the current findings help to explain why a broad range of stimuli and situations are avoided after trauma exposure. Specifically, CS-avoidance accounts for avoidance responses to stimuli that signal the presence of feared stimuli, whereas avoidance generalization accounts for avoidance responses to stimuli that resemble the feared stimuli. This laboratory model can also be expanded to study the nuances of compulsive behaviors in OCD, for instance, repetitively tapping on a table to prevent an intrusive thought (CS). Recent studies in the literature suggest that these compulsive behaviors are inflexible, goal-insensitive habitual behaviors (Gillan et al. 2014, 2015), thus future studies can examine whether these compulsive behaviors persist even when costs of these behaviors gradually increase (see also Glück, Zwosta, Wolfensteller, Ruge, & Pittig, 2021).

Given the lack of studies in avoidance of learnt fear, it remains unclear whether avoidance of learnt fear shares similar pathological features with safety behavior. For instance, avoidance of learnt fear to a stimulus may generalize to other similar stimuli; a constant engagement in avoidance of learnt fear may result in "protection from extinction" (Lovibond et al., 2009; Pittig, 2019; Rattel et al., 2017), given that the absence of an US would be attributed to the execution of avoidance of learnt fear. Thus, future studies are required to further examine the pathological features of avoidance of learnt fear. Furthermore, treatments for reducing avoidance of learnt fear may address different mechanisms underlying avoidance of learnt fear. Exposure-based treatments focusing on violating threat expectancy may effectively reduce avoidance of learnt fear. However, if negative valence is sufficient to evoke avoidance of learnt fear and persists after threat expectancy violation, additional interventions are required. For example, counterconditioning, a procedure that reduces negative valence of a stimulus, may effectively reduce avoidance of learnt fear.

The current study had some limitations. First, we only counterbalanced the CSs across participants, but did not counterbalance the stimuli serving as PSs and CSs across participants (e.g., the black and white drawings always served as PSs whereas the geometric shapes always served as CSs). Thus, the intrinsic properties of these stimuli might have induced some unspecified effects that potentially affected valence ratings. For instance, there was an observable bias in baseline valence ratings between the PSs and the CSs. Second, we provided explicit instructions about the contingencies of avoidance response-outcome and avoidance response-reward. Such instructions precluded the possibility to examine how costly avoidance is acquired by trial-and-error (c.f. Meulders et al., 2016; Glogan et al., 2020). This might also reduce threat ambiguity, minimizing the effects of individual risk factors on costly CS-avoidance. Third, despite negative valence is widely accepted to be resistant to extinction (see Baeyens et al., 1995; Hermans et al., 2005; Keller, Hennings, & Dunsmoor, 2020), there is evidence suggesting a reduction in negative valence after extinction (Zbozinek, Hermans, Prenoveau, Liao, & Craske, 2015). Therefore, given that we assessed the change in valence after test which was carried out in extinction, negative valence to PS1 might have been attenuated. Future studies can assess the change in valence immediately after fear acquisition. Fourth, we did not pre-register our hypotheses and analyses.

In conclusion, the current study established a laboratory model for the acquisition of costly avoidance of learnt fear, in which participants were more likely to engage in costly avoidance to a higher-order PS that signalled a fear-related stimulus than to a higher-order PS that signalled a safety stimulus. Such costly avoidance is suggested to be motivated by threat anticipation, a negative change in valence, or a combination of both. Future studies are required to further delineate the detrimental role of avoidance of learnt fear in anxiety-related disorders.

Funding

AHKW was supported in part by funds of the Bavarian State Ministry of Science and the Arts and the University of Würzburg to the Graduate School of Life Sciences (GSLS), University of Würzburg. AP was supported by a Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) grant PI1269/2-1 – 389569971.

Declaration of interest

None.

Data availability

The datasets generated and analysed during the current study are available in the Open Science Framework repository, https://osf.io/94uyj/.

Acknowledgement

The authors would like to thank Franziska Wirth and AnnKathrin Evers for their help in data collection and data processing.

References

Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever: on classical conditioning and extinction. *Journal of Psychophysiology, 9*, 127–141.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bahçecitapar, M. K. (2018). Some factors affecting statistical power of approximate tests in the linear mixed model for longitudinal data. *Communications in Statistics – Simulation and Computation, 47*(1), 294–314. https://doi.org/10.1080/03610918.2017.1283699

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

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, G., Dawson, W. T., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Carleton, R. N., Norton, P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety and Disorder, 21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social*

Psychology Bulletin, 25(2), 215–224. https://doi.org/10.1177/0146167299025002007
Cho, D., & Mitchell, D. S. (1971). Stimulus generalization in sensory preconditioning. *Journal of Experimental Psychology, 87*(3), 405–409. https://doi.org/10.1037/h0030528
Claes, N., Karos, K., Meulders, A., Crombez, G., & Vlaeyen, J. W. S. (2014). Competing goals attenuate avoidance behavior in the context of pain. *The Journal of Pain, 15*(11), 1120–1129. https://doi.org/10.1016/j.jpain.2014.08.003
Corrigan, I., Samuelson, K. A. M., Fridlund, B., & Thomé, B. (2007). The meaning of posttraumatic stress-reactions following critical illness or injury and intensive care treatment. *Intensive and Critical Care Nursing, 23*(4), 206–215. https://doi.org/10.1016/j.iccn.2007.01.004
Davis, J. L., & Thompson, R. F. (1969). Sensory preconditioning of cats in a shuttle box avoidance situation. *Psychonomic Science, 15*(3), 142. https://doi.org/10.3758/BF03342397
De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853
Declercq, M., & De Houwer, J. (2009). Transfer of avoidance responding to a sensory preconditioned cue: evidence for the role of S-S and R-S knowledge in avoidance learning. *Learning and Motivation, 40*, 197–208. https://doi.org/10.1016/j.lmot.2008.11.003
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*(5), 561–582. https://doi.org/10.1016/j.beth.2014.10.001
Dymond, S., Schlund, M. W., Roche, B., De Houwer, J., & Freegard, G. P. (2012). Safe from harm: learned, instructed, and symbolic generalization pathways of human threat-avoidance. *PLoS One, 7*(10), Article e47539. https://doi.org/10.1371/journal.pone.0047539
Dymond, S., Schlund, M. W., Roche, B., & Whelan, R. (2014). The spread of fear: symbolic generalization mediates graded threat-avoidance in specific phobia. *Quarterly Journal of Experimental Psychology, 67*(2), 247–259. https://doi.org/10.1080/17470218.2013.800124
Dymond, S., Schlund, M. W., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: symbolic generalization of human avoidance learning. *Behaviour Research and Therapy, 49*(10), 614–621. https://doi.org/10.1016/j.brat.2011.06.007
Engelhard, I. M., van Uijen, S. L., van Seters, N., & Velu, N. (2015). The effects of safety behavior directed towards a safety cue on perceptions of threat. *Behavior Therapy, 46*(5), 604–610. https://doi.org/10.1016/j.beth.2014.12.006
Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.robust013
Flores, A., López, F. J., Vervliet, B., & Cobos, P. L. (2018). Intolerance of uncertainty as a vulnerability factor for excessive and inflexible avoidance behaviour. *Behaviour Research and Therapy, 104*, 34–43. https://doi.org/10.1016/j.brat.2018.02.008
Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why dopeople worry? *Personality and Individual Differences, 17*(6), 791–802. https://doi.org/10.1016/0191-8869(94)90048-5
Gerlach, A. L., Andor, T., & Patzelt, J. (2008). Die Bedeutung von Unsicherheitsintoleranz für die Generalisierte Angststörung Modellüberlegungen und Entwicklung einer deutschen Version der Unsicherheitsintoleranz-Skala. *Zeitschrift für Klinische Psychologie und Psychotherapie, 37*, 190–199. https://doi.org/10.1026/1616-3443.37.3.190
Gewirtz, J. C., & Davis, M. (2000). Using pavlovian higher-order conditioning paradigms to investigate the neuralsubstrates of emotional learning and memory. *Learning & Memory, 7*(5), 257–266. https://doi.org/10.1101/lm.35200
Gillan, C. M., Apergis-Schoute, A. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., … Robbins, T. W. (2015). Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *American Journal of Psychiatry, 172*(3), 284–293. https://doi.org/10.1176/appi.ajp.2014.14040525
Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., … Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry, 75*(8), 631–638. https://doi.org/10.1016/j.biopsych.2013.02.002
Glogan, E., Gatzounis, R., Meulders, M., & Meulders, A. (2020). Generalization of instrumentally acquired pain-related avoidance to novel but similar movements using a robotic arm-reaching paradigm. *Behaviour Research and Therapy, 124*, Article 103525. https://doi.org/10.1016/j.brat.2019.103525
Glück, V. M., Zwosta, K., Wolfensteller, U., Ruge, H., & Pittig, A. (2021). Costly habitual avoidance in reduced by concurrent goal-directed approach in a modified devaluation paradigm. *Behaviour Research and Therapy, 146*, Article 103964. https://doi.org/10.1016/j.brat.2021.103964
Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). "Rstanarm: Bayesian applied regression modeling via Stan." R package version 2.21.1. https://mc-stan.org/rstanarm.
Hans Phaf, H., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology, 5*, 378. https://doi.org/10.3389/fpsyg.2014.00378
Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions regarding the interpretation of regression coefficients and hypothesis tests in linear models with interactions. *Communication Methods and Measures, 6*(1), 1–11. https://doi.org/10.1080/19312458.2012.651415
Hermans, D., Dirikx, R., Vansteenwegen, D., Baeyens, F., van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*(4), 533–551. https://doi.org/10.1016/j.brat.2004.03.013
Hoffeld, D. R., Kendall, S. B., Thompson, R. F., & Brogden, W. G. (1960). Effect of amount of preconditioning training upon the magnitude of sensory preconditioning. *Journal of Experimental Psychology, 59*(3), 198–204. https://doi.org/10.1037/h0048857
Hunt, C., Cooper, S. E., Hartnell, M. P., & Lissek, S. (2020). Anxiety sensitivity and intolerance of uncertainty facilitate associations between generalized Pavlovian fear and maladaptive avoidance decisions. *Journal of Abnormal Psychology, 128*(4), 315–326.
Jaeger, B.C. (2017). R2glmm: computes R squared for mixed (multilevel) models. R package version 0.1, 2.
Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics, 44*, 1086–1105. https://doi.org/10.1080/02664763.2016.1193725
Katz, R. C. (1974). Single session recovery from a hemodialysis phobia: a case study. *Journal of Behavior Therapy and Experimental Psychiatry, 5*(2), 205–206. https://doi.org/10.1016/0005-7916(74)90116-5
Keller, N. E., Hennings, A. C., & Dunsmoor, J. E. (2020). Behavioral and neural processes in counterconditioning: past and future directions. *Behaviour Research and Therapy, 125*, Article 103532. https://doi.org/10.1016/j.brat.2019.103532
Klein, Z., Berger, S., Vervliet, B., & Shechner, T. (2021a). Fear learning, avoidance, and generalization are more context-dependent for adults than adolescents. *Behaviour Research and Therapy, 147*, Article 103993. https://doi.org/10.1016/j.brat.2021.103993
Klein, Z., Berger, S., Vervliet, B., & Shechner, T. (2021b). High avoidance despite low fear of a second-order conditional stimulus. *Behaviour Research and Therapy, 136*, Article 103765. https://doi.org/10.1016/j.brat.2020.103765
Kleinknecht, R. A., & Lenz, J. (1989). Blood/injury fear, fainting and avoidance of medically-related situations: a family correspondence study. *Behaviour Research and Therapy, 27*(5), 537–547. https://doi.org/10.1016/0005-7967(89)90088-0
Koller, M., & Stahel, W. A. (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis, 55*(8), 2504–2515. https://doi.org/10.1016/j.csda.2011.02.014
Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved: Valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science, 21*(4), 607–613. https://doi.org/10.1177/0956797610365131
Kruschke, J. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (second ed.). Academic Press.
Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4
Krypotos, A. M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy, 111*, 99–105. https://doi.org/10.1016/j.brat.2018.10.011
Lemmens, A., Beckers, T., Dibbets, P., Kang, S., & Smeets, T. (2021). Overgeneralization of fear, but not avoidance, following acute stress. *Biological Psychology, 164*, Article 108151. https://doi.org/10.1016/j.biopsycho.2021.108151
Lommen, M. J. J., Engelhard, I. M., & van den Hout, M. (2010). Neuroticism and avoidance of ambiguous stimuli: better safe than sorry? *Personality and Individual Differences, 49*(8), 1001–1006. https://doi.org/10.1016/j.paid.2010.08.012
Lovibond, P. F. (2006). Fear and avoidance: an integrated expectancy model. *Fear and Learning: From Basic Processes to Clinical Implications* (pp. 117–132). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/11474-006
Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (second ed.). Sydney Psychology Foundation.
Lovibond, P. F., Mitchell, C. J., Minard, E., Brady, A., & Menzies, R. G. (2009). Safety behaviours preserve threat beliefs: protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy, 47*(8), 716–720. https://doi.org/10.1016/j.brat.2009.04.013
Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., … Anna di Palma, M. (2021). Robustbase: Basic Robust Statistics. R package version 0.93-98, http://robustbase.r-forge.r-project.org/.
Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541. ⟨https://joss.theoj.org/papers/10.21105/joss.01541⟩
Malloy, P., & Levis, D. J. (1988). A laboratory demonstration of persistent humanavoidance. *Behavior Therapy, 19*(2), 229–241. https://doi.org/10.1016/S0005-7894(88)80045-5
Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 97*(5), 951–966. https://doi.org/10.1037/a0028380
Meulders, A., Franssen, M., Fonteyne, R., & Vlaeyen, J. W. S. (2016). Acquisition and extinction of operant pain-related avoidance behavior using a 3 degrees-of-freedom robotic arm. *Pain, 157*(5), 1094–1104. https://doi.org/10.1097/j.pain.0000000000000483
Mendlowicz, M. V., & Stein, M. B. (2000). Quality of life in individuals with anxiety disorders. *The American Journal of Psychiatry, 157*(5), 669–682. https://doi.org/10.1176/appi.ajp.157.5.669
Nilges, P., & Essau, C. (2015). Die depressions-angst-stress-skalen: Der DASS–einscreeningverfahren nicht nur fur schmerzpatienten [depression, anxiety and stress scales: DASS – a screening procedure not only for pain patients]. *Der Schmerz, 29*(6), 649–657. https://doi.org/10.1007/s00482-015-0019-z

Olatunji, B. O., Cisler, J. M., & Tolin, D. F. (2007). Quality of life in the anxiety-related disorders: a meta-analytic review. *Clinical Psychology Review, 27*(5), 572–581. https://doi.org/10.1016/j.cpr.2007.01.015

Pittig, A. (2019). Incentive-based extinction of safety behaviors: positive outcomes competing with aversive outcomes trigger fear-opposite action to prevent protection from fear extinction. *Behaviour Research and Therapy, 121*, Article 103463. https://doi.org/10.1016/j.brat.2019.103463

Pittig, A., Boschet, J. M., Glück, V. M., & Schneider, K. (2020). Elevated costly avoidance in anxiety disorders: patients show little downregulation of acquired avoidance in face of competing rewards for approach. *Depression and Anxiety, 38*, 361–371.

Pittig, A., Schulz, A. R., Craske, M. G., & Alpers, G. W. (2014). Acquisition of behavioral avoidance: task-irrelevant conditioned stimuli trigger costly decisions. *Journal of Abnormal Psychology, 123*(2), 314–329.

Pittig, A., & Wong, A. H. K. (2021). Incentive-based, instructed, and social observational extinction of avoidance: fear-opposite actions and their influence on fear extinction. *Behaviour Research and Therapy, 137*, Article 103797. https://doi.org/10.1016/j.brat.2020.103797

Pittig, A., Wong, A. H. K., Glück, V. M., & Boschet, J. M. (2020). Avoidance and its bi-directional relationship with conditioned fear: mechanisms, moderators, and clinical implications. *Behaviour Research and Therapy, 126*, Article 103550. https://doi.org/10.1016/j.brat.2020.103550

Prewitt, E. P. (1967). Number of preconditioning trials in sensory preconditioning using cer training. *Journal of Comparative and Physiological Psychology, 64*(2), 360–362. https://doi.org/10.1037/h0088033

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rattel, J. A., Miedl, S. F., Blechert, J., & Wilhelm, F. H. (2017). Higher threat avoidance costs reduce avoidance behaviour which in turn promotes fear extinction in humans. *Behaviour Research and Therapy, 96*, 37–46. https://doi.org/10.1016/j.brat.2016.12.010

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology, 81*(1), 1–11. https://doi.org/10.1037/h0033333

San Martin, C., Jacobs, B., & Vervliet, B. (2020). Further characterization of relief dynamics in the conditioning and generalization of avoidance: effects of distress tolerance and intolerance of uncertainty. *Behaviour Research and Therapy, 124*, Article 103526. https://doi.org/10.1016/j.brat.2019.103526

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309–316. https://doi.org/10.1007/BF022 88586

Schlund, M. W., Brewer, A. T., Magee, S. K., Richman, D. M., Solomon, S., Ludlum, M., & Dymond, S. (2016). The tipping point: value differences and parallel dorsal-ventral frontal circuits gating human approach-avoidance behavior. *NeuroImage, 136*(1), 94–105. https://doi.org/10.1016/j.neuro image.2016.04.070

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Tabone, C. J., & de Belle, J. S. (2011). Second-order conditioning in Drosophila. *Learning and Memory, 18*(4), 250–253. https://doi.org/10.1101/lm.2035411

Telch, M. J., & Lancaster, C. L. (2012). Is there room for safety behaviors in exposure therapy for anxiety-related disorders?. *Exposure Therapy: Rethinking the Model — Refining the Method* (pp. 313–334). New York, NY, US: Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-3342-2_18

Topál, J., & Csányi, V. (1999). Interactive learning in the paradise fish (Macropodus opercularis): an ethological interpretation of the second-order conditioning paradigm. *Animal Cognition, 2*(4), 197–206. https://doi.org/10.1007/s100710050040

van Meurs, B., Wiggert, N., Wicker, I., & Lissek, S. (2014). Maladaptive behavioural consequences of conditioned fear-generalization: a pronounced, yet sparsely studied, feature of anxiety pathology. *Behaviour Research and Therapy, 57*, 29–37. https://doi.org/10.1016/j.brat.2014.03.009

Vervliet, B., & Indekeu, E. (2015). Low-Cost avoidance behaviors are resistant to fear extinction in humans. *Frontiers in Behavioral Neuroscience, 9*, 351. https://doi.org/10.3389/fnbeh.2015.00351

Vervliet, B. V., Lange, I., & Milad, M. R. (2017). Temporal dynamics of relief in avoidance conditioning and fear extinction: experimental validation and clinical relevance. *Behaviour Research and Therapy, 96*, 66–78. https://doi.org/10.1016/j.brat.2017.04.011

Wong, A. H. K., Glück, V. M., Boschet, J. M., & Engelke, P. (2020). Generalization of extinction with a generalization stimulus is determined by learnt threat beliefs. *Behaviour Research and Therapy, 135*, Article 103755. https://doi.org/10.1016/j.brat.2020.103755

Wong, A. H. K., & Pittig, A. (2020). Costly avoidance triggered by categorical fear generalization. *Behaviour Research and Therapy, 129*, Article 103606. https://doi.org/10.1016/j.brat.2020.103606

Wong, A. H. K., & Pittig, A. (2021). A dimensional measure of safety behavior: a non-dichotomous assessment of costly avoidance in human fear conditioning. *Psychological Research*. https://doi.org/10.1007/s00426-021-01490-w

Wong, A.H.K., Wirth, F.M., & Pittig, A. (under review). Avoidance of learnt fear: models, potential mechanisms, and future directions.

Yu, T., Lang, S., Birbaumer, N., & Kotchoubey, B. (2014). Neural correlates of sensory preconditioning: a preliminary fMRI investigation. *Human Brain Mapping, 35*(4), 1297–1304. https://doi.org/10.1002/hbm.22253

Zbozinek, T. D., Hermans, D., Prenoveau, J. M., Liao, B., & Craske, M. G. (2015). Post-extinction conditional stimulus valence predicts reinstatement fear: relevance for long term outcomes of exposure therapy. *Cognition & Emotion, 29*(4), 654–667. https://doi.org/10.1080/02699931.2014.930421