The degree of safety behaviors to a safety stimulus predicts development of threat beliefs

Alex H.K. Wong [a,*], Eva A.M. van Dis [b], Andre Pittig [c], Muriel A. Hagenaars [b], Iris M. Engelhard [b]

[a] *Department of Psychology, Educational Sciences, and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062, PA, Rotterdam, the Netherlands*
[b] *Department of Clinical Psychology, Utrecht University, PO Box 80140, 3508, TC, Utrecht, the Netherlands*
[c] *Translational Psychotherapy, Institute of Psychology, University of Goettingen, Goettingen, Germany*

A B S T R A C T

Safety behaviors are behavioral responses that aim to prevent or minimize an imminent threat when confronting a feared stimulus. Despite its adaptive purpose, preliminary evidence suggests that unnecessary safety behaviors to a safety stimulus induce threat beliefs to it. By allowing participants to engage in safety behaviors dimensionally, this study tested whether the degree of safety behaviors to a safety stimulus predicts the subsequent level of threat expectancies to it. To this end, participants first acquired safety behaviors to a threat-related stimulus (A). Safety behaviors then became available only for one safety stimulus (C), but not to another safety stimulus (B). After engaging in safety behaviors to C, participants exhibited greater threat expectancies to C compared to B, albeit with a small effect size. Importantly, the degree of safety behaviors predicted an increase in threat expectancies. The current findings suggest that safety behaviors to safety stimuli are linked to the development of threat beliefs.

1. Introduction

Low-cost safety behaviors refer to behaviors aimed at preventing or minimizing actual or anticipated harm of a perceived threat with minimal effort (e.g., wearing a seatbelt). Safety behaviors can be modelled via a fear and avoidance conditioning framework. In this well-established framework, a previously neutral conditioned stimulus (CS) is paired repeatedly with an aversive unconditioned stimulus (US). Consequently, presentation of the CS alone is able to evoke conditioned fear. In a subsequent avoidance conditioning procedure, executing a designated response (e.g., pressing the spacebar key) during CS presentation prevents US presentation (de Houwer & Hughes, 2020). This response is referred to as "US-avoidance" given that it prevents the US but does not necessarily terminate CS presentation (Krypotos, Vervliet, & Engelhard, 2018; Pittig, Wong, Glück, & Boschet, 2020).

Recent empirical evidence examined whether safety behaviors to a safety stimulus can increase threat expectancy to it when the behavior is no longer available (Engelhard, van Uijen, van Seters, & Velu, 2015; van Dis, Krypotos, Zondervan-Zwijnenburg, Tinga, & Engelhard, 2022; Xia, Dymond, Lloyd, & Vervliet, 2017). In these studies, after acquiring safety behaviors to a threat-related CS, safety behaviors were then made available only to a safety stimulus. Results showed that after engaging in safety behaviors to this safety stimulus, threat expectancy to it increased when safety behaviors became unavailable (Engelhard et al., 2015; van

Dis et al., 2022; Xia et al., 2017). These studies suggested that engaging in unnecessary safety behaviors to a learned safety stimulus may later induce threat belief to it, suggesting a novel pathway of the formation of maladaptive (albeit mild) threat beliefs.

Even though previous studies demonstrated a role of safety behaviors in increasing threat expectancy to a safety stimulus, they were limited by a dichotomous manipulation of safety behaviors (Engelhard et al., 2015; van Dis et al., 2022; Xia et al., 2017). That is, safety behaviors were either executed or not. This dichotomous manipulation is arguably less sensitive for modelling safety behaviors, as safety behaviors can often be engaged in to a certain degree (Krypotos et al., 2018; Telch & Lancaster, 2012), and are not necessarily maladaptive (Hofmann & Hay, 2018). To illustrate, a socially anxious person may avoid eye contact to a certain extent during a conversation. While this partial engagement in safety behaviors could reduce chances of the onset of a perceived aversive outcome (e.g., rejection) to a certain degree, this person could still appear to be attentive in conversation. Thereby, allowing participants to engage in safety behaviors along a continuum is arguably a more ecologically valid measure for safety behaviors. With regard to this, we have recently developed a protocol for assessing avoidance on a continuum in human fear and avoidance conditioning (Wong & Pittig, 2022), in which one can engage in avoidance to various degrees; the greater the extent of avoidance, the less likely of an aversive outcome onset.

Besides providing a more ecologically valid measure, assessing avoidance on a continuum provided additional strengths and novel approaches for the current study. It can more sensitively examine whether the extent of safety behaviors to a safety stimulus would later predict the level of threat expectancy to it. This is of interest as it is thought that participants might have used their behavior as information (Gangemi et al., 2012; van den Hout et al., 2014). That is, safety behaviors to a threat stimulus previously prevented US-occurrences, so when participants apply safety behaviours to a safety stimulus, they might believe that the safety stimulus would now otherwise be followed by a US. Therefore, assessing avoidance on a continuum provides insights into whether safety behaviors predict development of threat beliefs to a safety stimulus (i.e., the greater the engagement in safety behaviors to a safety stimulus, the stronger the threat belief to it). Relatedly, past studies (e.g., Engelhard et al., 2015; van Dis et al., 2022; Xia et al., 2017) only included participants who engaged in safety behaviors during most safety stimulus trials. This precluded the opportunity to explore an alternative explanation for the increase in threat expectancy to a safety stimulus: the mere availability of safety behaviors during safety stimulus presentation may induce an increase in threat expectancy to it. Thereby, the current study would retain all participants regardless of their engagement of safety behaviors to the safety stimulus, thereby providing some potential insights into this alternative explanation. Furthermore, a dimensional representation of avoidance arguable provides higher statistical power over a dichotomous measure, as the latter is often transformed into proportion scores (e.g., the proportion of avoided trials), a categorical variable. Key limitations of categorical outcomes include artificially limiting variation in responses between participants, limited sensitivity, and problematic for some widely use statistical analyses, such as ANOVA (Jaeger, 2008).

Therefore, the aim of the current study was two-fold. We first sought to replicate findings on increased threat expectancy to a safety stimulus due to prior safety behaviors engagement to it (Engelhard et al., 2015; van Dis et al., 2022; Xia et al., 2017). Second, we examined whether the degree of safety behaviors to a safety stimulus would predict the magnitude of threat expectancy to it.

2. Method

Preregistration and data of this study can be found at the Open Science Framework (https://osf.io/5ceza/).

2.1. Participants

Data collection was carried out at two sites, University of Erlangen-Nuremberg in Germany and Utrecht University in the Netherlands. Undergraduate students from each university were recruited and received partial course credit for participation. For a power calculation, we employed a data-based simulation (Kumle, Võ, & Draschkow, 2021) based on a dataset from Wong and Pittig (2022, Experiment 1), revealing that a sample size of 50 achieves 86% power to detect an expected effect size of b = 0.097 (see https://osf.io/5ceza/ for the preregistration). We recruited a total of 60 participants to account for attrition rates due to technical difficulties or different exclusion criteria. Thirty participants were recruited at University of Erlangen-Nuremberg and the other 30 were recruited at Utrecht University. This study was approved by the Ethics Committees of both universities in accordance with the Declaration of Helsinki.

2.2. Apparatus and materials

Three squares with different colors (blue, yellow, and pink) served as visual stimuli presented in the experiment (Engelhard et al., 2015). We used Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, Version 20.1) to present all experimental instructions, visual stimuli, and recorded all self-reported ratings. A DS7A Digitimer stimulator generated an electric US with a duration of 500 ms. Skin conductance was measured via two Ag/AgCl electrodes at a sampling rate of 1000 Hz. Specifically, BrainVision Recorder was used to measure skin conductance at University of Erlangen-Nuremberg whereas Biosemi was used at Utrecht University. Although different hardware was used in the two different universities, all raw data were processed with BrainVision Analyser via the same pipeline (see *Scoring and analysis*).

US-avoidance and US expectancies were measured throughout the experiment. The US-avoidance scale ranged from 0% (Certainly NOT avoid) to 100% (Certainly avoid) with 1% steps of change. The selected avoidance responses were positively proportional to the chance of US omission. For instance, an avoidance response of 80% would lead to an 80% chance of US omission, if a US would have followed the stimulus. (see Wong & Pittig, 2022). The reinforcement rate of the CS+ was negatively mapped to the US-avoidance scale so that US-avoidance was proportional to threat level. This reduced confounding effects on US-avoidance to the safety stimulus of interest that might have caused by US-avoidance that was out of proportion to threat level (e.g., ineffective US-avoidance if it was mapped lower than the reinforcement rate might have artificially reduced US-avoidance to the safety stimulus). The US-avoidance scale was presented with a question "To what extent do you want to avoid a potential electric pulse? Please indicate your degree of avoidance on the scale below". The US expectancy scale ranged from 0% (Certain NO electric stimulation), 50% (Uncertain), to 100% (Certain electric stimulation) with 1% steps of change. The US expectancy scale was presented with a question "To what extent do you expect an electric pulse after this image? Please indicate your expectancy of electric pulse on the scale below."

Two psychometric questionnaires were used: the Depression Anxiety Stress Scale-21 (DASS-21; Lovibond & Lovibond, 1995) that discriminates between three different constructs, namely depression, anxiety, and stress; the Intolerance of Uncertainty scale (IU scales; Freeston, Rhéaume, Letarte, Dugas, & Ladouceur, 1994) that measures cognitive, emotional, and behavioral responses to uncertainty (see Carleton, Norton, & Asmundson, 2007).

2.3. Procedure

After providing written informed consent, participants at University of Erlangen-Nuremberg filled in the German version of DASS-21 (Nilges & Essau, 2015) and the German version of IU scale (Gerlach, Andor, & Patzelt, 2008)[1]; participants at Utrecht University filled in the Dutch version of DASS-21 (de Beurs, Van Dyck, Marquenie, Lange, & Blonk, 2001) and the Dutch version of IU scale (de Bruin, Rassin, van der Heiden, & Muris, 2006). Next, skin conductance electrodes filled with isotonic gel were attached to participants' non-dominant hand. For participants recruited at University of Erlangen-Nuremberg, the electrodes were attached to the hypothenar muscles, whereas the electrodes were attached to the index and middle fingers on participants recruited at Utrecht University.[2] US electrodes were also attached to the wrist of the same hand.

A US workup procedure was carried out, in which participants could set their own individual US intensity at a level that was perceived as "definitely uncomfortable but not painful". Immediately after, the conditioning task was carried out. The conditioning task consisted of five phases: *Practice, Pavlovian acquisition, Safety behavior acquisition, Safety behavior shift*, and *Test* (cf. Engelhard et al., 2015; see Table 1).

---

[1] The German version of IU scale consists of 18 questions whereas the Dutch version consists of 27 questions. Thereby, the IU scores in the German sample were upscaled.

[2] Although the SCR electrodes were attached to different sites of the non-dominant hand between Erlangen and Utrecht, the SCR data did not differ in most of the phases. See Supplementary Materials for more details.

Table 1

+ indicates US presentation; - indicates US omission; * indicates safety behaviors availability; + in brackets indicates US presentation depending on safety behaviors; Number in parentheses indicates the number of trials per trial type.

Practice	Pavlovian acquisition	Safety behavior acquisition	Safety behavior shift	Test
A- (1)	A+ (3)	A*[+] (6)	A+ (3)	A- (4)
B- (1)	A- (1)	A+ (1)	A- (1)	B- (4)
C- (1)	B- (2)	B- (1)	B- (4)	C- (4)
	C- (2)	C- (1)	C*- (4)	

2.3.1. Practice

The blue, yellow, and pink squares served as stimuli A, B, and C, respectively (counterbalanced across participants). Prior to this phase, participants received verbal instructions that some squares would be presented on screen alongside a US expectancy scale. Participants were explicitly informed that no electric US would be delivered in this phase and that the purpose of this phase was for them to familiarize with the US expectancy scale. Stimuli A, B, and C were each presented once.

2.3.2. Pavlovian acquisition

Before this phase began, participants were informed that some squares might be followed by an electric US, whereas some might not (cf. Mertens, Boddez, Krypotos, & Engelhard, 2021). Stimulus A was presented for 4 trials and was reinforced by an electric US at a 75% rate, whereas stimuli B and C were presented for 2 trials each, which were never reinforced. On each trial, a stimulus was presented alongside a US expectancy scale for 8 s. An electric US was administered immediately after stimulus offset for reinforced A trials. The presentation order was pseudo-randomized so that the first and last A trials were always reinforced, and that the same trial type was not presented more than 2 times in a row. The intertrial intervals (ITIs) were randomized between 15 s and 18 s, which were also applied to all the following phases.

2.3.3. Safety behaviors acquisition

Before this phase started, participants were instructed that they had the opportunity to prevent a US that potentially followed the squares (see exact instructions in Supplementary Materials). This opportunity to engage in US-avoidance was signalled by the presentation of a US-avoidance scale. In this phase, 6 trials of stimulus A (A* trials) were presented alongside an avoidance scale, whereas one trial for stimuli A, B, and C were presented without the US-avoidance scale. The trial structure for A* trials consisted of stimulus A presented alongside a US-avoidance scale for 5 s, followed by an 8 s presentation of the same stimulus alongside a US expectancy scale. US administration depended on the US-avoidance response made, and if presented, was administered immediately after stimulus offset. For A, B, and C trials, stimulus was presented alongside the US expectancy scale for 8 s. Stimulus A was reinforced by a US as a reminder that A without opportunity to engage in US-avoidance still led to a US, whereas B and C trials were not reinforced (see Table 1).

2.3.4. Safety behavior shift

This phase immediately followed the previous phase. Each stimulus was presented for 4 trials. Trial structure was the same as the previous phase, except that only C trials were presented with the US-avoidance scale. Stimulus A trials were reinforced at a 75% rate, whereas none of the other stimuli were reinforced. The purpose of this phase was to shift US-avoidance availability from A trials to C trials, so that the use of safety behaviors to C might increase US expectancy to it in the following *Test* phase.

2.3.5. Test

This phase started immediately after the previous phase. Each stimulus was presented for 4 trials. For all trials, there was no opportunity to engage in US-avoidance, and none of them were reinforced by an electric US.

2.4. Scoring and analysis

Skin conductance measured during the 8 s of CS presentation was analyzed. We used BrainVision Analyzer to process raw SCR data. We applied a 1 Hz low-pass filter to remove high-frequency noise and a 50 Hz notch filter to the SCR data. The SCR was calculated by subtracting the averaged skin conductance level 2 s prior to CS onset from the peak response 1 s after CS onset to CS offset (see Pineles, Orr, & Orr, 2009). All negative SCRs or SCRs lower than 0.02 μS were scored as zero. All SCRs were then square-root transformed to reduce skewness (Boucsein et al., 2012). The processing of SCR data was done by an independent research assistant blinded to the trial types.

Most data were analyzed within a linear mixed model framework. The analyses were separated into three sections: manipulation check, main hypotheses, and exploratory analyses. All analyses were preregistered on OSF (https://osf.io/5ceza/).

2.4.1. Manipulation check

2.4.1.1. Pavlovian acquisition. We first checked whether participants acquired differential conditioned fear to the reinforced stimulus (A) and the non-reinforced stimuli (B & C), as indexed by US expectancy ratings and SCRs. To this end, US expectancy ratings or SCRs served as dependent variable, whereas Stimulus type (A, B, & C) served as a fixed factor. We applied two orthogonal contrasts to this model. The first contrast compared responding to the last 2 trials of A with all two trials of B and C, examining the acquisition of differential responding to a threat-related stimulus and to safety-related stimuli (the first 2 trials of A were not included due to the lack of excitatory learning on early trials). The second contrast examined whether there was any difference in inhibitory learning to B and C, so that differences in conditioned fear in *Test* could be attributed to the manipulation of *Safety behavior shift* rather than a pre-existing difference in safety learning.

2.4.1.2. Safety behavior acquisition. To examine whether participants acquired US-avoidance to A* trials, we provided the magnitude of US-avoidance averaged across all A* trials. In addition, to check whether US-avoidance engagement reduced conditioned fear to A*, US expectancy ratings or SCRs served as dependent variable, whereas Stimulus type (A*, A, B, & C) served as a fixed factor in a linear mixed model framework. Two orthogonal contrasts were employed. The first contrast assessed whether responding to A would be stronger than those averaged across A*, B, and C trials. This contrast evaluated whether conditioned fear to an unavoidable threat stimulus (A) would be stronger than an avoidable threat stimulus (A*) and the two safety stimuli (B & C). The second contrast assessed whether the reduction of conditioned fear to A* was comparable to that of B and C.

2.4.1.3. Safety behavior shift. To examine whether participants engaged in US-avoidance to a safety stimulus (C) when possible, we provided the magnitude of US-avoidance averaged across all trials of C*. Furthermore, to assess whether differential conditioned fear to stimuli changed once US-avoidance availability shifted to C trials, US expectancy ratings or SCRs served as dependent variable, whereas Stimulus type (A, B, and C) and Trial (a linear trend repeated measures across trials) served as fixed factors in a linear mixed model framework. Similarly, two

Table 2
Demographic data and questionnaire data. Means (standard deviation).

	Mean (Standard deviation)			
	Whole sample (n = 54)	Subsample from Erlangen (n = 28)	Subsample from Utrecht (n = 26)	Difference between subsamples (*p*)
Age	22.07 (3.41)	21.64(4.25)	22.54 (2.16)	*.339*
Sex – Female	39 (72.22%)	23 (82.14%)	16 (61.54%)	*.091*
DASS21-Anxiety (0-42)	4.63 (3.84)	4.43 (3.28)	4.85 (4.42)	*.694*
DASS21-Depression (0-42)	5.41 (5.41)	6.00 (5.50)	4.77 (5.34)	*.408*
DASS21-Stress (0-42)	10.04 (6.45)	9.64 (6.69)	10.46 (6.28)	*.646*
IU (27–135)	63.05 (17.18)	60.91 (17.74)	65.35 (16.59)	*.348*

orthogonal contrasts were applied to this model. The first contrast compared responding to A with responding averaged across B and C, which evaluated whether conditioned fear to A persisted once US-avoidance to it became unavailable. The second contrast compared responding to B with C, testing whether responding to C would differ from those to B once US-avoidance availability shifted from A to C trials.

2.4.2. Main hypotheses

2.4.2.1. Test. To examine whether conditioned fear to the test stimuli changed after *Safety behavior shift*, US expectancy ratings or SCRs served as dependent variable, whereas Stimulus type and Trial served as fixed factors. Two orthogonal contrasts were employed to this model. The first contrast examined whether differential conditioned fear to A and the safety stimuli (B & C) persisted in *Test*, whereas the second contrast examined whether responding to C would be greater than B once US-avoidance availability to the former was removed in *Test*.

The second model examined whether the degree of US-avoidance engagement during *Safety behavior shift* predicted conditioned fear in *Test*. To this end, we carried out robust regression models to assess whether the individual degree of US-avoidance to C on the last trial of *Safety behavior shift* predicted the magnitude of US expectancy ratings or SCRs to the same stimulus on the first trial of *Test*. Robust regression linear models were conducted to minimize the influence of outliers by assigning lower weights to outliers in an iterative manner (Koller & Stahel, 2011).

2.4.3. Exploratory analyses

To exploratively examine whether the increase in US expectancy to C in test was long-lived or not, we compared US expectancy or SCRs to C with B averaged across the last two test trials.

We also exploratorily examined whether trait anxiety or intolerance of uncertainty would be associated with an increase in safety behaviors engagement to a safety stimulus (C*) during *Safety behavior shift*. To this end, US-avoidance served as dependent variable, whereas trait anxiety or intolerance of uncertainty served as a continuous fixed factor. Furthermore, we examined whether these factors modulate how US-avoidance engagement to C* predicts conditioned fear or threat expectancies to it. Thereby, US expectancy ratings or SCRs on the first C trial of *Test* served as dependent variable, whereas US-avoidance to the last C* trial of *Safety behavior shift* and trait anxiety or intolerance of uncertainty served as predictors in robust regression models.

For all the linear mixed models, participants served as a random effect. All the main effects and higher-order interactions were analyzed in separate models (Hayes, Glynn, & Huge, 2012). The degree of significance was reported with Satterthwaite approximation for degrees of freedom (Satterthwaite, 1941). All statistical analyses were conducted in R (R core team, 2022), with *lmer* package for frequentist linear mixed models (Bates, Mächler, Bolker, & Walker, 2015) and *robustbase* package for robust regression analyses (Mächler et al., 2023). Effect sizes were reported as partial-R^2 with *r2glmm* package (Jaeger, 2017).

3. Results

Statistical analyses were restricted to participants who had 1) acquired differential US expectancies to the threat and safety stimuli in *Pavlovian acquisition* (as indexed by higher US expectancies averaged across A trials than US expectancies averaged across B and C trials), and 2) acquired US-avoidance to the threat stimulus in *Safety behaviors acquisition* (as indexed by 50% or higher US-avoidance averaged across all A* trials; see preregistration). The first criterion was defined as higher US expectancy ratings to A than that averaged across B and C on the last trial of *Pavlovian acquisition*. The second criterion was defined as engaging in at least an average of 50% US-avoidance to all A* trials in *Safety behaviors acquisition*. Six participants were excluded based on these criteria, leaving a total of 54 participants in the final sample (see Table 2 for descriptive statistics for the final sample). Of note, SCRs for 4 participants were not recorded due to technical issues. Nonetheless, these participants were retained in the sample for analyses of US expectancy ratings and US-avoidance responses.

3.1. Manipulation check

3.1.1. Pavlovian acquisition

US expectancy and SCRs. Fig. 1 shows the US expectancy ratings to stimuli across the experiment. Averaged US expectancy ratings to the last 2 trials of A+ were higher than the ratings averaged across all B- and C- trials in *Pavlovian acquisition*, bStimulus type(A vs B&C)[3] = 12.40, SE = 0.73, *p* < .001, R^2 = 0.45, indicating successful fear acquisition. Furthermore, there was no evidence that inhibitory learning to B-differed from that to C- across acquisition trials, bStimulus type(B vs C) = 0.86, SE = 1.27, *p* = .499, R^2 = 0.001.

Fig. 1 shows the SCRs to stimuli across the experiment. Similar to the US expectancy data, averaged SCRs to the last 2 trials of A+ were stronger than those averaged across B- and C- trials, bStimulus type(A vs B&C) = 0.029, SE = 0.0085, *p* < .001, R^2 = 0.031. There was no evidence that inhibitory learning in SCRs differed between B- and C-, bStimulus type(B vs C) = 0.025, SE = 0.015, *p* = .083, R^2 = 0.008.

3.1.2. Safety behavior acquisition

US-avoidance. Averaged across trials, participants exhibited an average of 88.06% (SD = 11.90) of US-avoidance engagement to A* trials, suggesting that participants acquired the response - US omission contingency.

US expectancy and SCRs. Participants exhibited higher US expectancy to A+ compared to A*, B-, and C- trials across *Safety behavior acquisition*, bStimulus type(A vs A*&B&C) = 11.66, SE = 0.90, *p* < .001, R^2 = 0.24. Furthermore, US expectancy ratings averaged across A* trials were higher than B and C trials, bStimulus type(A* vs B&C) = 13.84, SE = 0.88, *p* < .001, R^2 = 0.32. We carried out additional analyses to check whether responding to C differed from B. There was no evidence that US expectancy to C differed from B, b$_{Stimulus type(B vs C)}$ = 0.94, SE = 2.27, *p* = .678, R^2 < 0.001. For the SCR data, responding to A+ were stronger

---

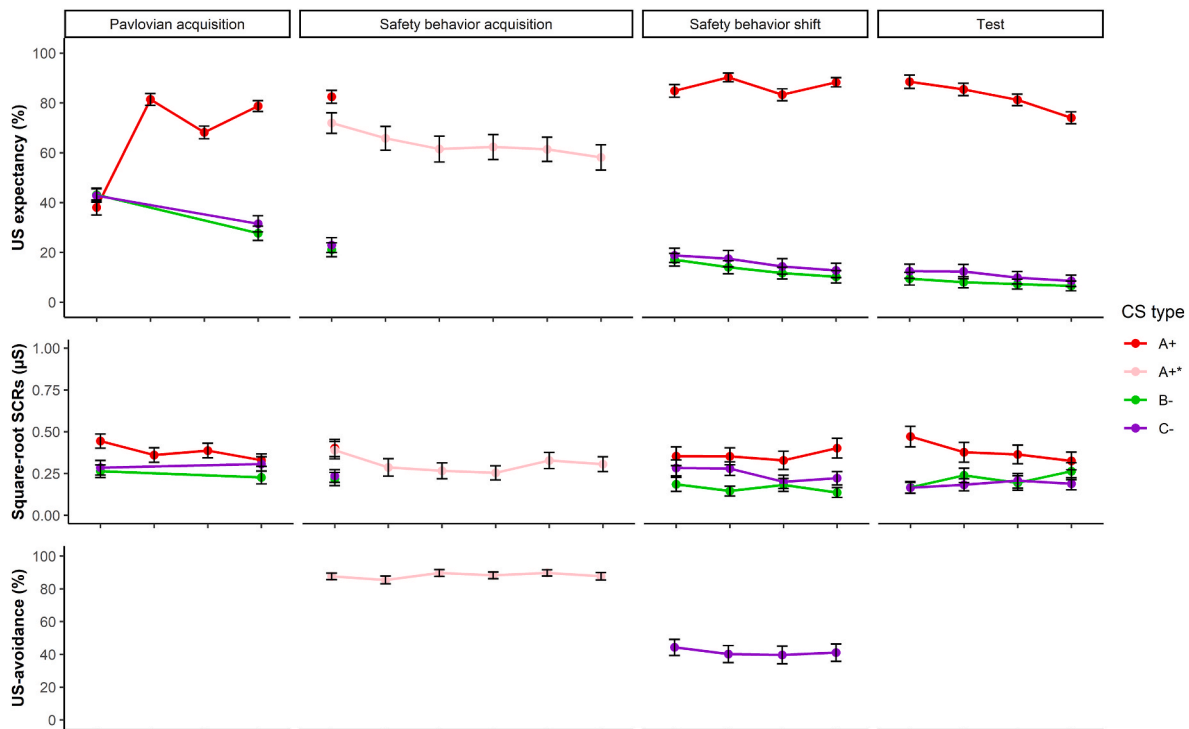[3] b stands for unstandardized regression weight.

Fig. 1. US expectancy (top panel), square-root SCRs (middle panel), and US-avoidance (bottom panel) across all phases. Error bars indicate standard error of the mean. See the color version of this figure online. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

compared to A*, B-, and C- trials across *Safety behavior acquisition*, bStimulus type(A vs A*&B&C) = 0.037, SE = 0.011, $p < .001$, R^2 = 0.025. Similarly, responding to A* trials was stronger than to B- and C-trials, bStimulus type(A* vs B&C) = 0.026, SE = 0.010, $p = .012$, R^2 = 0.013. Similar to US expectancy ratings, there was no evidence that SCRs to C differed from B, $b_{\text{Stimulus type(B vs C)}}$ = 0.0093, SE = 0.027, $p = .730$, $R^2 < 0.001$.

3.1.3. Safety behavior shift

US-avoidance. When US-avoidance availability shifted to C trials, participants utilized the entire scale of avoidance (0%–100%), with an average of 41.29% (SD = 37.92) of US-avoidance engagement, suggesting that participants did engage in US-avoidance to a learnt safety stimulus when avoidance became available.

US expectancy and SCRs. Participants exhibited higher US expectancy ratings to A+ trials compared to B- and C* trials, whereas ratings to the latter decreased gradually across trials, ratings to the former showed an irregular pattern. This pattern was supported by a significant interaction between Stimulus type and Trial, bStimulus type(A vs B&C*) *Trial = 24.07, SE = 10.28, $p = .020$, R^2 = 0.008. Despite US expectancy ratings to C* were higher than to B- averaged across trials, this difference did not reach significance, bStimulus type(B vs C*) = 1.31, SE = 0.70, $p = .064$, R^2 = 0.005. There was no evidence that the decrease in US expectancy ratings differed between B- and C*, bStimulus type(B vs C*)*Trial = 2.37, SE = 17.80, $p = .894$, $R^2 < 0.001$.

With regard to the SCR data, participants exhibited stronger responding to A+ compared to B- and C*, bStimulus type(A vs B&C*) = 0.052, SE = 0.0071, $p < .001$, R^2 = 0.08. Participants also exhibited stronger responding to C* than to B- averaged across trials, bStimulus type(B vs C*) = 0.042, SE = 0.012, $p < .001$, R^2 = 0.019. No other effects reached significance (smallest $p = .104$).

In sum, participants acquired stronger conditioned fear to A+ compared to B- and C-, as indexed by US expectancy ratings and SCRs. The acquisition of US-avoidance on A* trials led to a decrease in US expectancy ratings, but this pattern was not observed in SCR data. When

US-avoidance availability shifted to C, SCRs to it increased but not US expectancy ratings.

3.2. Main hypotheses

3.2.1. Test

US expectancy and SCRs. With regard to the first contrast comparing the threat-related stimulus to safety-related stimuli, participants exhibited higher US expectancy ratings to A- compared to B- and C-; ratings to A- declined more rapidly across trials compared to B- and C-. This pattern was supported by a significant interaction between Stimulus type and Trial, bStimulus type(A vs B&C)*Trial = −34.05, SE = 9.09, $p < .001$. R^2 = 0.02. For the second contrast comparing responding to the two safety stimuli, there was no evidence that the decline in US expectancies across trials differed between B- and C-, bStimulus type(B vs C)*Trial = −6.64, SE = 15.75, $p = .673$, $R^2 < 0.001$. However, more importantly, US expectancy ratings to C- were higher than that to B- averaged across test trials, bStimulus type(B vs C) = 1.49, SE = 0.63, $p = .018$, R^2 = 0.008. The effect size was, however, small.[4] Although there were no significant differences in US expectancy ratings to B and C during the acquisition phases, the expectancy ratings to C were descriptively higher than to B. Therefore, we added the difference in US expectancy ratings to B and C averaged across *Pavlovian acquisition* and *Safety behavior acquisition* as a covariate in *Test*. Results showed no evidence that the prior descriptive differences between B and C contributed to the same difference observed in *Test*, $b_{\text{Covariate(B vs C)}}$ = 0.059, SE = 0.043, $p = .175$, R^2 = 0.003. Therefore, there was no

---

[4] We carried out an additional analysis to compare whether data collected at Utrecht University were systematically different from those collected at Erlangen University, with regard to the main hypotheses. The was a three-way interaction involving Stimulus type, Trial, and Site in the SCR data. However, no follow-up analyses reached significance (see Supplementary Materials for more details).

evidence that the descriptive pre-existing differential US expectancy ratings to B and C had a confounding effect in *Test*.

Similar to the US expectancy data, SCRs to A- were stronger compared to B- and C-, whereas responding to A- showed a more rapid decrease across trials than to B- and C-. This pattern was supported by an interaction between Stimulus type and Trial, bStimulus type(A vs B&C)*Trial $= -0.59$, SE $= 0.19$, $p = .002$, $R^2 = 0.016$. However, there was no evidence that SCRs differed between B- and C- in *Test*, (smallest $p = .246$).

Given that US expectancy ratings to C were descriptively greater than to B during *Safety behavior shift*, we carried out an exploratory analysis to examine whether this difference was greater in *Test* compared to *Safety behavior shift*. Results showed no evidence that the difference in US expectancy ratings to C compared to B was greater in *Test* compared to *Safety behavior shift*, bStimulus type(B vs C)*Phase $= 0.37$, SE $= 1.19$, $p = .756$, $R^2 < 0.001$.

3.2.2. US-avoidance predicting conditioned fear

Fig. 2A and B show whether US-avoidance to the last C* trial of *Safety behavior shift* predicted US expectancy and SCRs to it on the first test trial, respectively. Higher degrees of US-avoidance to C* on the last trial of *Safety behavior shift* was positively associated with higher US expectancy ratings to C- on the first trial of *Test*, $\beta_{Avoidance} = 0.14$ SE $= 0.053$, $p = .010$ ($r = 0.31$). In contrast, we found no evidence that the degree of US-avoidance engagement to C predicted the level of SCRs to it, $\beta_{Avoidance} = 0.044$, SE $= 0.11$, $p = .704$ ($r = -0.0074$). To further examine whether this predictive relationship was confounded by the descriptive difference to B and C during the two acquisition phases, we added the same covariate described above in the regression model. Results showed no evidence that the covariate predicted expectancy to C in *Test*, $\beta_{Covariate(B vs C)} = 0.070$, SE $= 0.058$, $p = .233$, whereas US-avoidance to C* still significantly predicted expectancy to C in *Test*, $\beta_{Avoidance} = 0.16$ SE $= 0.056$, $p = .005$.

Given that US expectancy ratings to C* were descriptively greater than that to B-, we ran an exploratory robust regression to examine whether there was already an association between US-avoidance and US expectancy ratings or SCRs to C* during *Safety behavior shift* (see Fig. 2C & D). In contrast to the direct comparison in US expectancies between C* and B, US-avoidance to C* significantly predicted the level of US expectancy ratings to it, $\beta_{Avoidance} = 0.21$, SE $= 0.079$, $p = .011$ ($r = 0.43$). In contrast, there was no evidence that US-avoidance to C* was associated with the magnitude of SCRs to C, $\beta_{Avoidance} = 0.017$, SE $= 0.14$, $p = .901$ ($r = 0.14$).

Furthermore, given that US expectancy ratings to C* was descriptively greater than those to B during *Safety behavior shift*, we explored whether this descriptive difference was associated with the same comparison observed in *Test*. Results showed no evidence that the differential US expectancy ratings during *Safety behavior shift* predicted the same difference in *Test*, $\beta_{USexpectancy(B vs C)} = 0.025$, SE $= 0.022$, $p = .252$ ($r = 0.11$). This suggests that the differential US expectancies to B and C in *Test* were likely due to previous US-avoidance engagement to C*, but not merely a "residual effect" carried over from *Safety behavior shift*.

3.3. Exploratory analyses

We explored whether the difference in threat expectancy and SCRs to the two safety stimuli were long-lived. To this end, we compared responding to these two stimuli on the last 2 test trials. US expectancy to C- were indeed significantly higher than to B-, bStimulus type(B vs C) $= 2.30$, SE $= 0.85$, $p = .008$, $R^2 = 0.016$, but the effect size was small. In contrast, there was no evidence for any differences in SCRs to these two stimuli, bStimulus type(B vs C) $= -0.03$, SE $= 0.04$, $p = .389$, $R^2 = 0.003$. No other effects reached significance (smallest $p = .241$).

With regard to inter-individual trait factors, there was no evidence that trait anxiety or intolerance of uncertainty was associated with increased US-avoidance to C* during *Safety behavior shift*, and neither associated with increased threat expectancy or SCRs to C- in *Test* (see Supplementary Materials for the full analysis). We also explored whether there was an association in US-avoidance engagement to A* during *Safety behavior acquisition* and to C* during *Safety behavior shift* via a robust regression model. Results showed no evidence that US-avoidance to A* was associated with US-avoidance to C*, $\beta_{Avoidance} =$
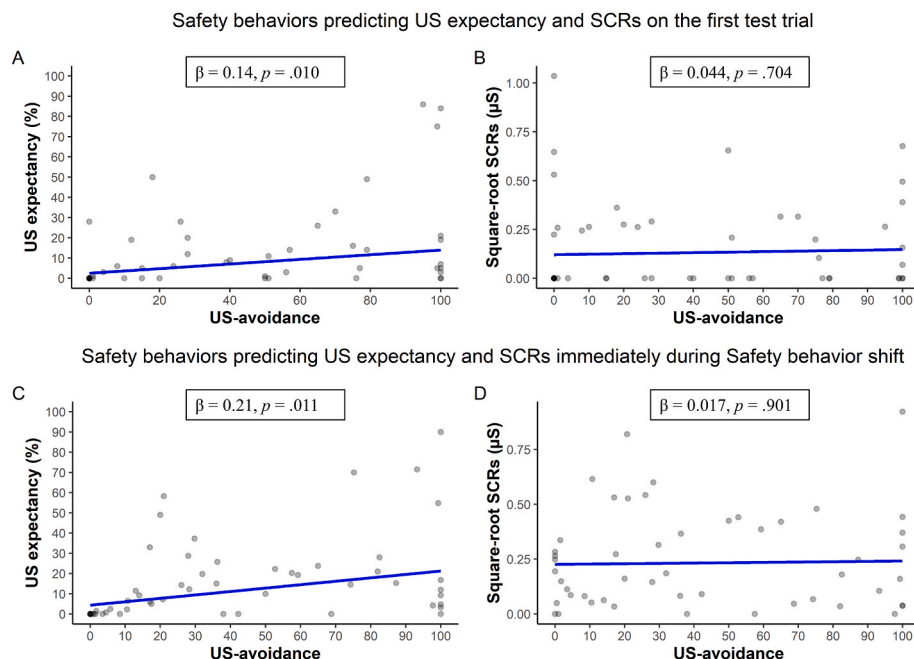

Fig. 2. Relationship between US-avoidance and conditioned fear. Top panel: US-avoidance to the last C* trial in *Safety behavior shift* predicts a) US expectancy ratings and b) square-root SCRs to the same stimulus on the first trial in *Test*. Bottom panel: US-avoidance averaged across C* trials in *Safety behavior shift* predicts c) US expectancy ratings and d) square-root SCRs to the same stimulus averaged across the same phase. Darker color indicates more overlapping data points. The lines represent the robust line of best fit for visual aid. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

0.28 SE = 0.15, p = .067, suggesting no evidence that participants who more strongly engaged in US-avoidance to A* during *Safety behavior acquisition* tended to more strongly avoid C*.

4. Discussion

Using a dimensional manipulation of avoidance, the current study sought to replicate the findings of whether safety behaviors to a safety stimulus increases threat expectancy to it when safety behaviors are unavailable (Engelhard et al., 2015; van Dis et al., 2022; Xia et al., 2017). More importantly, we examined whether the degree of safety behaviors engagement to the same stimulus predicts the degree of threat expectancy to it.

Before *Safety behavior Shift,* threat expectancies to B and C did not differ. However, threat expectancy to C was higher than to B *after* participants had the opportunity to engage in safety behaviors to C. This suggested that the opportunity to engage in safety behaviors increases threat expectancy to a learnt safety stimulus, replicating past findings (Engelhard et al., 2015; van Dis et al., 2022; Xia et al., 2017). This increase in threat expectancy to C was unlikely due to fear generalization from A, assuming that fear generalization from A to C was similar to that from A to B, but rather the manipulation of shifting safety behaviors from A to C. However, it should be noted that the effect size in the current study was small compared to past findings. This small effect size was presumably due to the inclusion of participants that exhibited limited safety behaviors to C during *Safety behavior shift*, whereas the previous studies excluded such participants.

More importantly, the degree of safety behaviors to C significantly predicted the magnitude of threat expectancy to it when safety behaviors became unavailable. An interesting finding was that the degree of safety behaviors to C during *Safety behavior shift* also predicted threat expectancy to it in the same phase, despite the difference in threat expectancy ratings between B and C in the same phase did not reach significance. These two patterns suggest that engaging in safety behaviors to a safety stimulus increases one's threat expectancy to it. These patterns support two accounts: behavior as information (Gangemi, et al., 2012; van den Hout et al., 2014) and cognitive dissonance (e.g., Festinger, 1957; Harmon-Jones, Harmon-Jones, & Levy, 2015). The former account puts forward the idea that one infers the likelihood of threat based on the intensity of behavior, whereas the latter presumes one would resolve an unpleasant state of cognitive dissonance by adjusting threat beliefs to match the prior action. Nonetheless, both accounts predict a positive link between the degree of safety behaviors to a safety stimulus and subsequent threat expectancy to it. However, this study did not aim to disentangle between these two possible accounts. Furthermore, it remains unclear in the literature whether the actual engagement in safety behaviors or the mere availability of safety behaviors imposes this detrimental effect (e.g., Kemp, Blakey, Wolitzky-Taylor, Sy, & Deacon, 2019; Sloan & Telch, 2002). The current findings provide preliminary support for the former account, as the degree of safety behaviors engagement directly predicted the increase in threat expectancy to a safety stimulus.

Threat expectancy ratings were larger to C compared to B only in *Test* when safety behaviors became unavailable. On face value, this pattern suggests that the unavailability of safety behaviors played a major role in inducing differential threat expectancies to the two safety stimuli. However, this differential threat expectancies to the two safety stimuli did not differ significantly between the two phases, thus, there was no strong evidence indicating that this pattern was only observed when safety behaviors became unavailable. An alternative explanation is that the differential threat expectancies to the safety stimuli in *Test* was carried over by the descriptive differential responding to the same stimuli during *Safety behavior shift*, rather than being attributed to safety behaviors to C. However, there was no evidence to support this claim as the differential threat expectancies to the safety stimuli in the two phases were not significantly associated with each other. In sum, the current findings suggest that engaging in safety behaviors to a safety stimulus induces an increase in threat beliefs to it, in addition to the degree of safety behaviors engagement determining the magnitude of threat beliefs to the safety stimulus.

During *Safety behavior shift*, participants still engaged in safety behaviors to C* to a significant extent even though safety behaviors were unnecessary. This was presumably due to the minimal cost of executing safety behaviors. We have previously found an increase in safety behaviors to a CS- when safety behaviors were low-cost compared to costly safety behaviors (Wong & Pittig, 2022). This increase in unnecessary safety behaviors were interpreted as participants employing a "why not strategy" (e.g., I know safety behaviors are unnecessary to a safety signal, but if safety behaviors cost nothing, then why not?).

Despite the differential threat expectancies to B and C during *Safety behavior shift* didn't reach significance, skin conductance responses to C were greater than B immediately after safety behaviors were engaged in *Safety behavior shift*, however this effect was not observed in US expectancy. On face value, this suggests that conditioned fear as indexed by skin conductance responses to a safety stimulus *immediately* increased after engaging in safety behaviors to it. However, skin conductance responses also reflect non-fear related responses, such as orienting responses induced by the sudden shift in safety behaviors availability to a safety stimulus. Therefore, it remains unclear whether the increase in skin conductance responses to C could be interpreted as an immediate increase in conditioned fear. Future studies can include measurements that can more sensitively measure fear, such as eyeblink startle responses, to delineate the increase in responding to a safety stimulus right after engaging in safety behaviors to it.

Regarding clinical implications, evidence has showed that clinically anxious individuals often engage in faulty reasoning, like emotional reasoning (Arntz, Rauner, & van den Hout, 1995; Engelhard, Macklin, McNally, van den Hout, & Arntz, 2001; Mansell & Clark, 1999) and behavior as information (Gangemi, Mancini, & van den Hout, 2012; van den Hout et al., 2014). Thereby, although the extent of safety behaviors engagement to a safety stimulus is significantly lower than to a threat-related stimulus, the former may inevitably expand the scope of threat beliefs to a wide range of stimuli that pose no actual harm. Furthermore, the safety behaviors being modelled in the current study are so-called low-cost safety behaviors, which are behaviors that require minimal effort. These low-cost safety behaviors are often subtle and can easily go unnoticed, for instance, an individual with panic disorder may carry pills along or an individual with social anxiety disorder may hold arms stiffly at sides to prevent trembling. Extra effort is required to notice these safety behaviors and prevent their engagement when confronting innocuous stimuli or situations. Future studies can examine whether the increase in threat beliefs to a safety stimulus would generalize to other innocuous stimuli, or whether unnecessary safety behaviors would turn habitual.

One limitation of the current study was the limited effects in the SCR data. The increase in responding to stimulus C in *Test* and the predictive relation between safety behaviors and responding were only observed in US expectancies, but not in SCRs. It is important to note that the null effects in SCRs was not due to a failure of fear acquisition, given that participants exhibited stronger SCRs to the threat-related CS compared to the safety-related CSs in *Pavlovian acquisition*. The limited findings in SCRs were potentially due to its large individual variability (Lykken & Venables, 1971), resulting in less statistical power for within-subject comparisons.

In conclusion, the current study replicated findings that executing safety behaviors to a safety stimulus increases threat expectancy to it. A key novel finding was that the degree of safety behaviors engagement was positively associated with the level of threat expectancy to a safety stimulus, suggesting that the actual engagement of safety behaviors but not the mere availability of safety behaviors induced maladaptive threat beliefs.

CRediT authorship contribution statement

Alex H.K. Wong: Conceptualization, Formal analysis, Methodology, Software, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. **Eva A.M. van Dis:** Conceptualization, Data curation, Methodology, Writing – review & editing. **Andre Pittig:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Muriel A. Hagenaars:** Conceptualization, Methodology, Writing – review & editing. **Iris M. Engelhard:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Data availability

I have shared my link to the data at https://osf.io/5ceza/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2023.104423.

References

Arntz, A., Rauner, M., & van den Hout, M. (1995). "If I feel anxious, there must be danger": Ex-consequentia reasoning in inferring danger in anxiety disorders. *Behaviour Research and Therapy, 33*(8), 917–925. https://doi.org/10.1016/0005-7967(95)00032-S

de Beurs, E., Van Dyck, R., Marquenie, L. A., Lange, A., & Blonk, R. W. B. (2001). De DASS: Een vragenlijst voor het meten van depressie, angst en stress. *Gedragstherapie, 34*, 35–53.

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

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

de Bruin, G. O., Rassin, E., van der Heiden, C., & Muris, P. (2006). Psychometric properties of a Dutch version of the intolerance of uncertainty scale. *Netherlands Journal of Psychology, 62*, 87–92. https://doi.org/10.1007/BF03061055

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders, 21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

van Dis, E. A. M., Krypotos, A.-M., Zondervan-Zwijnenburg, M. A. J., Tinga, A. M., & Engelhard, I. M. (2022). Safety behaviors toward innocuous stimuli can maintain or increase threat beliefs. *Behaviour Research and Therapy, 156*, Article 104142. https://doi.org/10.1016/j.brat.2022.104142

Engelhard, I. M., Macklin, M. L., McNally, R. J., van den Hout, M., & Arntz, A. (2001). Emotion- and intrusion-based reasoning in Vietnam veterans with and without chronic posttraumatic stress disorder. *Behaviour Research and Therapy, 39*(11), 1339–1348. https://doi.org/10.1016/S0005-7967(00)00101-7

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*, 604–610. https://doi.org/10.1016/j.beth.2014.12.006

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford: Stanford Univ. Press.

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences, 17*(6), 791–802. https://doi.org/10.1016/0191-8869(94)90048-5

Gangemi, A., Mancini, F., & van den Hout, M. (2012). Behavior as information: "If I avoid, then there must be a danger". *Journal of Behavior Therapy and Experimental Psychiatry, 43*(4), 1032–1038. https://doi.org/10.1016/j.jbtep.2012.04.005

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

Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An action-based model of cognitive-dissonance processes. *Current Directions in Psychological Science, 24*(3), 183–189. https://doi.org/10.1177/0963721414566449

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

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

van den Hout, M., Gangemi, A., Mancini, F., Engelhard, I. M., Rijkeboer, M. M., van Dams, M., et al. (2014). Behavior as information about threat in anxiety disorders: A comparison of patients with anxiety disorders and non-anxious controls. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(4), 489–495. https://doi.org/10.1016/j.jbtep.2014.07.002

de Houwer, & Hughes, S. J. (2020). *The psychology of learning: An introduction from a functional-cognitive perspective.* MIT Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, B. C. (2017). R2glmm: computes R squared for mixed (multilevel) models. *R package version 0., 1, 2.*

Kemp, J. J., Blakey, S. M., Wolitzky-Taylor, K. B., Sy, J. T., & Deacon, B. J. (2019). The effect of safety behavior availability versus utilization on inhibitory learning during exposure. *Cognitive Behaviour Therapy, 48*(6), 517–528. https://doi.org/10.1080/16506073.2019.1574312

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

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

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales (2nd ed.).* Sydney Psychology Foundation.

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology, 8*(5), 656–672. https://doi.org/10.1111/j.1469-8986.1971.tb00501.x

Mächler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., et al. (2023). *robustbase: Basic Robust Statistics. R package version 0.99-0.* http://robustbase.r-forge.r-project.org/.

Mansell, W., & Clark, D. M. (1999). How do I appear to others? Social anxiety and processing of the observable self. *Behaviour Research and Therapy, 37*(5), 419–434. https://doi.org/10.1016/S0005-7967(98)00148-X

Mertens, G., Boddez, Y., Krypotos, A.-M., & Engelhard, I. M. (2021). Human fear conditioning is moderated by stimulus contingency instructions. *Biological Psychology, 158*, Article 107994. https://doi.org/10.1016/j.biopsycho.2020.107994

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]. *Schmerz, Der, 29*(6), 649–657. https://doi.org/10.1007/s00482-015-0019-z

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

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

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309–316. https://doi.org/10.1007/BF02288586

Sloan, T., & Telch, M. J. (2002). The effects of safety-seeking behavior and guided threat reappraisal on fear reduction during exposure: An experimental investigation. *Behaviour Research and Therapy, 40*, 235–251. https://doi.org/10.1016/S0005-7967(01)00007-9

Telch, M. J., & Lancaster, C. L. (2012). Is there room for safety behaviors in exposure therapy for anxiety-related disorders?. In *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.

Wong, A. H. K., & Pittig, A. (2022). A dimensional measure of safety behavior: A non-dichotomous assessment of costly avoidance in human fear conditioning. *Psychological Research, 86*, 312–330. https://doi.org/10.1007/s00426-021-01490-w

Xia, W., Dymond, S., Lloyd, K., & Vervliet, B. (2017). Partial reinforcement of avoidance and resistance to extinction in humans. *Behaviour Research and Therapy, 96*, 79–89. https://doi.org/10.1016/j.brat.2017.04.002