Behavioural Brain Research 470 (2024) 115078

Contents lists available at ScienceDirect

Behavioural Brain Research

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

The generalization of threat beliefs to novel safety stimuli induced by safety behaviors

Alex H.K. Wong[a,*], Andre Pittig[b], Iris M. Engelhard[c]

[a] Department of Psychology, Educational Sciences, and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam 3062 PA, the Netherlands
[b] Translational Psychotherapy, Institute of Psychology, Georg-August-Universität Göttingen, Kurze-Geismar-Straße 1, Göttingen 37073, Germany
[c] Department of Clinical Psychology, Utrecht University, P.O. Box 80140, Utrecht 3508 TC, the Netherlands

ARTICLE INFO

Keywords:
Safety behaviors
Threat belief
Fear generalization
Fear and avoidance conditioning
Anxiety

ABSTRACT

Safety behaviors are responses that can reduce or even prevent an expected threat. Moreover, empirical studies have shown that using safety behaviors to a learnt safety stimulus can induce threat beliefs to it. No research so far has examined whether threat beliefs induced this way generalize to other novel stimuli related to the safety stimulus. Using a fear and avoidance conditioning model, the current study (n=116) examined whether threat beliefs induced by safety behaviors generalize to other novel generalization stimuli (GSs). Participants first acquired safety behaviors to a threat predicting conditioned stimulus (CSthreat). Safety behaviors could then be performed in response to one safe stimulus (CSsafeShift) but not to another (CSsafe). In a following generalization test, participants showed a significant but small increase in threat expectancies to GSs related to CSsafeShift compared to GSs related to CSsafe. Interestingly, the degree of safety behaviors used to the CSsafeShift predicted the subsequent increase in generalized threat expectancies, and this link was elevated in trait anxious individuals. The findings suggest that threat beliefs induced by unnecessary safety behaviors generalize to other related stimuli. This study provides a potential explanation for the root of threat belief acquisition to a wide range of stimuli or situations.

1. Introduction

Safety behaviors are behavioral responses that are typically performed when encountering threatening objects or situations and can mitigate or prevent the onset of an expected imminent threat. For instance, upon hearing the fire alarm, safety behaviors include running out of the building to minimize chances of perishing in a fire. Therefore, safety behaviors performed in high threat situations are typically considered adaptive for survival. In clinical anxiety, safety behaviors are not always adaptive, especially if they maintain maladaptive and unrealistic threat beliefs [7]. For instance, a client with social anxiety may stay on the edge of a social group to limit social engagement, thus mitigating the perceived threat of social rejection, despite the fact that the perceived threat rarely occurs. Using safety behaviors thus precludes one from disconfirming their maladaptive threat beliefs (e.g., the absence of social rejection is attributed to safety behaviors) and thus interferes with the effectiveness of exposure-based therapies [14,55].

Given the clinical importance of understanding how safety behaviors and threat beliefs interact, safety behaviors are examined empirically in highly controlled fear and avoidance conditioning models. In this framework, a previously neutral conditioned stimulus (CS) is repeatedly paired with a biologically aversive unconditioned stimulus (US). Participants then learn that performing a designated response during CS presentation effectively mitigates US onset, thus modelling the acquisition of safety behaviors.[1] Empirical studies have shown the maladaptive characteristics of such safety behaviors, including the persistence of safety behaviors without realistic threat (e.g., [10,12,34,38,52]), how persistent safety behaviors protect one from correcting maladaptive beliefs [28,35,39], and that safety behaviors after fear extinction can lead to a return of fear [51]. Preliminary evidence suggests that maladaptive safety behaviors are enhanced in individuals with clinical anxiety [36] or individuals at risk (e.g., [10,54,63]).

Another characteristic of safety behaviors is that using safety behaviors to safety cues can subsequently increase threat expectancies to them, even when participants had learnt that these cues were safe. In a controlled laboratory study [9], participants first acquired safety

[*] Correspondence to: Erasmus School of Social and Behavioural Sciences, Burgemeester Oudlaan 50, Rotterdam 3062 PA, the Netherlands.
E-mail address: h.k.wong@essb.eur.nl (A.H.K. Wong).

[1] Noted that while the use of safety behaviors reduces or even prevents US onset, it does not terminate CS presentation (see [21]).

https://doi.org/10.1016/j.bbr.2024.115078
Received 29 January 2024; Received in revised form 8 April 2024; Accepted 30 May 2024
Available online 31 May 2024
0166-4328/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

behaviors to a threat-related CS+ that effectively prevents an imminent US. In a following phase, the availability of safety behaviors was shifted to a safety-related CS-. In test, participants showed an increase in threat expectancy to the CS- when safety behaviors became unavailable, even though participants had previously learnt that CS- signaled the absence of a US. This pattern suggests that unnecessary safety behaviors to a learnt safety cue induces an increase in threat belief to it. This suggests a potential pathway for the formation of maladaptive beliefs in clinical anxiety. This finding has been replicated in other laboratory studies [50,61,62].

Yet, an unexplored question is whether the formation of threat beliefs induced by safety behaviors would generalize to other novel cues. Empirical studies have shown that fear acquired to a threat-related CS generalizes to novel stimuli that perceptually or conceptually resemble the original threat stimulus (e.g., [6,8,18,1]). It may thus be possible that threat beliefs induced by safety behaviors may generalize beyond the original stimulus to which safety behaviors were performed. The current study aimed to address this question by examining whether this acquired threat belief would generalize to other novel generalization stimuli (GSs). This is of clinical importance as it examines a potential pathway of the formation and generalization of threat beliefs.

In addition, we have found that the degree of safety behaviors to a safety cue determined the magnitude of increase in threat expectancy to it [61]. This provides preliminary evidence that the actual use of safety behaviors induces maladaptive threat belief to a safety cue (e.g., using behavior as information), as opposed to the mere availability of safety behaviors to a safety cue that would increase threat expectancies to it [44]. Therefore, we aimed to explore whether the level of safety behavior usage to a safety cue also predicts the level of generalized threat expectancies to novel cues that resemble the safety cue.

This study also explored whether risk factors of clinical anxiety would have any effect on the magnitude of generalized threat beliefs induced by safety behaviors. Three risk factors were assessed in this study, intolerance of uncertainty, trait anxiety, and (low) distress tolerance. Intolerance of uncertainty refers to an individual's disposition to experience distress or anxiety in the face of ambiguity [4]. Trait anxiety refers to an individual's disposition to experience heightened level of distress and anxiety when reacting to stressful situations [47]. (Low) distress tolerance refers to a low level of capacity to endure negative emotional states [43]. These three factors have been widely accepted to contribute to a greater vulnerability to clinical anxiety (e.g., [5,42,48]). In laboratory studies, these factors have been linked to enhanced fear generalization [32,40,58] and increased safety behaviors (e.g., [24,53,63]). Based on these findings, the current study explored whether these risk factors are associated with stronger safety behaviors to the safety cue and subsequently greater generalized threat beliefs.

In sum, the current study aimed to examine whether using safety behaviors to a learnt safety cue would increase threat expectancies to novel cues that conceptually resemble it. A second aim was to examine whether the degree of safety behaviors usage to the safety cue would be associated with the strength of fear generalization to novel cues associated with it. We further explored the impact of individual risk factors on unnecessary safety behaviors and the generalized threat beliefs induced by safety behaviors.

2. Method

2.1. Participants

Psychology undergraduates at Erasmus University Rotterdam were recruited and compensated with partial course credits for participation. Using the dataset from a similar study from our lab ([61]; dataset available at https://osf.io/5ceza/), a data-based simulation analysis [22] suggested that 110 participants provided 90.5 % power to detect a difference in US expectancy ratings to the two safety stimuli in test with an effect size of b = 1.49. A total of 120 participants were recruited to account for data exclusion due to participants not reaching the acquisition criterion or other technical issues. This study was approved by the Ethics Committee of the Erasmus School of Social and Behavioural Science (ETH2223–0321) in accordance to the Declaration of Helsinki.

2.2. Apparatus and materials

A total of twelve standardized 2-D black-and-white drawings [45] from three categories (mammal, fruit, and vehicle) served as the conditioned stimuli (CSs) and generalization stimuli (GSs). The mammal stimuli included bear, cow, dog, and sheep; the fruit stimuli included apple, pear, pineapple, and strawberry; the vehicle stimuli included bus, plane, train, and truck. For each category, two drawings served as CSs and two drawings served as GSs.

US expectancy ratings were measured via a visual analog scale ranging from 0 % to 100 % with steps of 1 %, 0 % indicates certain absence of a US whereas 100 % indicates certain presence of a US. Likewise, safety behaviors were assessed by a visual analog scale ranging from 0 % to 100 % with steps of 1 %; 0 % indicates certainly not avoid a US whereas 100 % indicates certainly avoid a US. Safety behaviors were assessed via this continuum, as it has been suggested to more sensitively measure the use of safety behavior to different extents [59] and shown to overcome common limitations when assessing generalization of safety behaviors (see [57,60]). All visual stimuli and scales were presented via Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, Version 20.1).

Skin conductance was measured via a pair of Ag/AgCl electrodes attached to the hypothenar muscles of participants' non-dominant hand. Skin conductance was measured at a 1000 Hz sampling rate by a Biopac MP150 system equipped with a EDA100 amplifier. The electric US consisted of a train of electric pulses amounting to a duration of 500 ms. The electric US was generated by a DS7A Digitimer stimulator, delivered via a bar-electrode attached to the wrist of participants' non-dominant hand.

Three psychometric questionnaires were used to assess various individual differences of interest. First, a trait version of Depression Anxiety Stress Scale-21 (DASS-21; [25,27]) was used to assess trait anxiety. Intolerance of uncertainty scale (IUS; [11]) was used to assess intolerance of uncertainty. Distress tolerance scale was used to assess distress tolerance (DTS; Simon & Gaher, 2005).

2.3. Procedure

After providing informed consent, participants filled in the DASS-21, IUS, and DTS questionnaires. The experimenter then attached the SCR electrodes filled with isotonic gel and the electric US electrodes on participants' non-dominant hand. A US workup procedure was then carried out. Participants first sampled a US with an intensity of 0.2 mA; the US intensity was gradually increased in a stepwise manner until it reached a level that was perceived as 'definitely unpleasant but not painful'. Immediately after the workup procedure, the conditioning task was carried out. The conditioning task consisted of five phases: *Practice, Fear acquisition training, Safety behavior acquisition training, Safety behavior shift, and Generalization test* (see Fig. 1).

2.3.1. Practice

Before this phase began, participants were informed that some geometric shapes would appear on screen along with a US expectancy scale. This phase allowed participants to familiarize with the US expectancy scale. Participants were explicitly informed that no US would be delivered in this phase. Three colored squares were presented. On each trial, the colored square was presented with a US expectancy scale for 8 s. Afterwards, participants were presented with the US expectancy ratings they made (e.g., a text of "your expectancy for an electric pulse is 70 %" if participants indicated a US expectancy of 70 %), reassuring that the US expectancy scale worked as intended. The intertrial intervals were

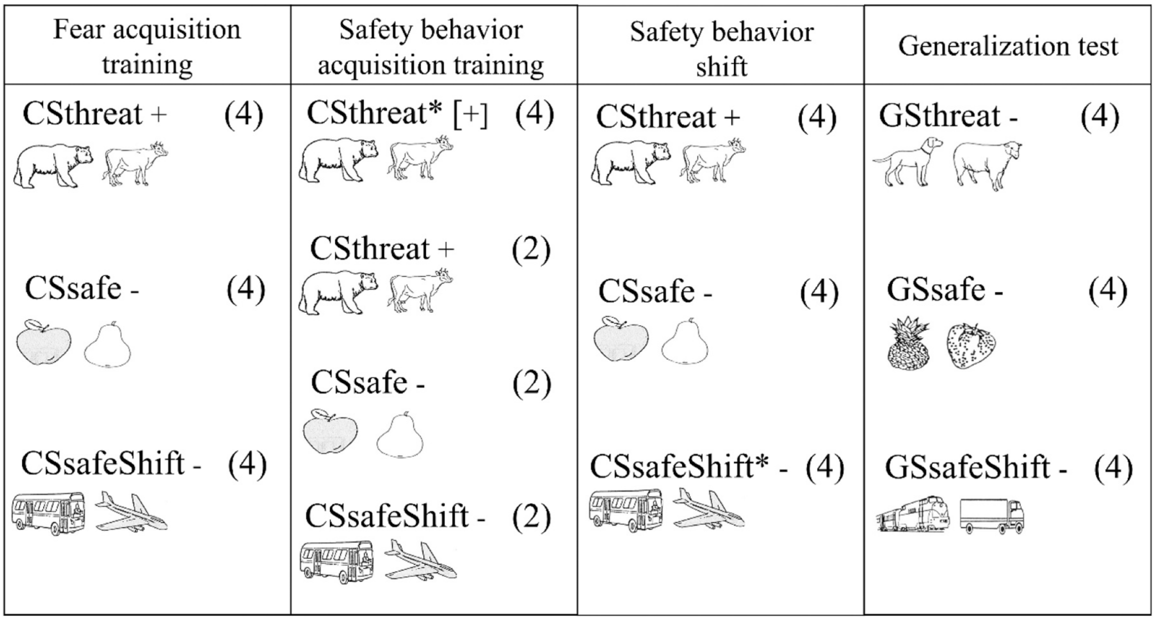

Fig. 1. Experimental design. CS indicates conditioned stimuli. Each CS represents 2 exemplars from one of the three categories (mammal, fruit, vehicle). + represents electric US presentation; - represents electric US omission; * indicates safety behavior availability; + in brackets indicates electric US presentation depending on safety behavior. GS indicates GS stimuli. Number in parentheses indicates the number of trials per trial type.

4 s.

2.3.2. *Fear acquisition training*

Before this phase started, participants were verbally informed that different drawings would appear on screen, which might or might not signal a US (the exact instructions for the experiment can be seen in the Supplementary Materials). In this phase, one category served as CSthreat (e.g., mammal), one category served as CSsafe (e.g., fruit), and one category served as CSsafeShift (e.g., vehicle). This phase consisted of two blocks. In each block, two drawings from each category were presented once (e.g., a bear and a cow drawing were presented once each for CSthreat trials), totaling to six trials in each block. All CSthreat trials were reinforced by an electric US, whereas CSsafe and CSsafeShift trials were never reinforced. Each CS was presented at the center of the screen with the US expectancy scale for 8 s. Participants were prompted to indicate their US expectancy ratings on each CS trial. The presentation order of the CSs was pseudo-randomized so that the same CS type would not occur more than twice in a row between blocks. The intertrial intervals were randomized between 11 and 15 s, which were applied to all the following phases. The CS categories were counterbalanced across participants.

2.3.3. *Safety behavior acquisition training*

Participants were informed that they could probabilistically prevent the US onset that potentially followed the CSs. This opportunity to mitigate the US onset was signaled by the presentation of a safety behavior scale. This phase consisted of two blocks. In each block, an avoidable CSthreat (CSthreat*) was presented for two trials: CSthreat was presented with a safety behavior scale for 5 s, followed by an 8 s presentation of the same CS along with a US expectancy scale. On each CSthreat* trial, participants had to make a safety behavior response on the scale; their safety behavior response probabilistically determined whether a US would be delivered or not. For instance, a safety behavior response of 70 % would lead to a 70 % chance of US prevention. If a US is delivered, it would be delivered immediately after CS offset. CSthreat, CSsafe, and CSsafeShift were presented for one trial each with only the US-expectancy scale for 8 s. CSthreat trial was reinforced by a US to remind participants that CSthreat without an opportunity to use safety behavior would still lead to a US. CSsafe and CSsafeShift trials were never reinforced.

2.3.4. *Safety behavior shift*

This phase followed seamlessly from the previous phase. This phase consisted of two blocks. In each block, each CS type was presented for two trials. CSthreat and CSsafe trials were presented along with a US expectancy scale for 8 s, in which CSthreat was reinforced by a US while CSsafe was not reinforced. CSsafeShift* trials were presented with a safety behavior scale for 5 s, followed by the same stimulus presented with a US expectancy scale for 8 s. Regardless of safety behavior, CSsafeShift* trials were not reinforced. This phase was critical as safety behavior availability was shifted from CSthreat trials to CSsafeShift trials, so that any increase in responding to novel stimuli related to CSsafeShift could be attributed to the use of safety behavior to CSsafeShift.

2.3.5. *Generalization test*

This phase followed seamlessly from the previous phase and consisted of two blocks. In each block, novel GSs from each of the CS category were presented. GSthreat, GSsafe, and GSsafeShift trials were presented twice each block. All GS trials were presented with a US expectancy scale for 8 s, and none of the GSs were reinforced by a US.

2.4. *Scoring and analyses*

Only skin conductance measured during the 8 s of CS/GS presentation was analyzed (i.e., the 8 s window when US expectancy ratings were prompted). Using BrainVision Analyzer, we applied a 1 Hz low-pass filter to remove high frequency noises and a 50 Hz notch filter to the SCR data. The SCRs were obtained by identifying the peak responding and its corresponding trough 1 s after CS/GS onset till CS/GS offset. We then square root transformed the SCRs to reduce skewness [46]. The SCR data processing was carried out by research assistants blinded to the trial types.

All analyses were carried out with linear mixed models and robust regression models. The analyses were separated into three parts: *Manipulation check*, *Main hypotheses*, and *Exploratory analyses*. All analyses were pre-registered on Open Science Framework (https://osf.io/auvpy).

2.4.1. Manipulation check

2.4.1.1. Fear acquisition training. We applied two orthogonal contrasts for the models in this phase. The first contrast examined whether conditioned fear was successfully acquired to the threat-related CS, as reflected by stronger responding to CSthreat compared to the safety-related CSsafe and CSsafeShift. Therefore, responding to the CSthreat was compared to responding averaged across CSsafe and CSsafeShift. The second contrast assessed whether differences in responding to the two safety-related CSs already occurred during acquisition. Therefore, responding to CSsafe was compared with CSsafeShift. For these two contrasts, US expectancy ratings or SCRs served as the dependent variable, whereas CS type (CSthreat, CSsafe, & CSsafeShift), Block (Block1 & Block 2), and their interaction served as fixed effects. Participants served as the only random effect in the linear mixed models for all analyses.

2.4.1.2. Safety behavior acquisition training. To examine whether safety behavior to CSthreat* trials was acquired, we provided the descriptive magnitude of safety behavior averaged across all CSthreat* trials. Regarding US expectancy ratings and SCRs, we examined whether the reinforced CSthreat trials would evoke stronger responding to avoidable CSthreat* trials and the safety-related CS trials. Thus, responding to CSthreat was compared to responding averaged across CSthreat*, CSsafe, and CSsafeShift. We also added a non-preregistered orthogonal contrast. This contrast examined whether responding to the two safety-related CS trials already occurred during acquisition to check for potential differences before safety behavior shift. Thus, responding to CSsafe was compared with CSsafeShift. To this end, US expectancy ratings or SCRs served as the dependent variable whereas CS type (CSthreat, CSthreat*, CSsafe, & CSsafeShift), Block, and their interaction served as fixed effects.

2.4.2. Main hypotheses

2.4.2.1. Safety behavior shift. The descriptive magnitude of safety behavior to CSsafeShift* trials was provided to assess whether participants used safety behavior to these trials. Furthermore, to assess whether conditioned fear to the CSs changed once safety behavior availability shifted to CSsafeShift trials, two orthogonal contrasts were applied to the models in this phase. The first contrast compared responding to CSthreat trials with responding averaged across CSsafe and CSsafeShift* trials; this contrast assessed whether conditioned fear persisted to reinforced CSthreat trials. The second contrast compared responding to CSsafe with CSsafeShift*, examining whether conditioned fear to these two CSs already differed once safety behavior availability was shifted to CSsafeShift. To this end, US expectancy or SCRs served as dependent variable, whereas CS type and Block served as fixed factors.

2.4.2.2. Generalization test. To assess fear generalization to the GSs and whether generalized fear to CSsafeShift increased due to safety behavior to CSsafeShift trials in the previous phase, two orthogonal contrasts were applied. The first contrast examined whether participants exhibited fear generalization, as indexed by stronger responding to GSthreat compared to responding averaged across GSsafe and GSsafeShift trials. More importantly, the second contrast examined one of the main hypotheses, examining whether generalized fear to GSsafeShift was stronger than GSsafe. To reduce confounded extinction learning, only the first block of *Generalization test* was analyzed, as preregistered. To this end, US expectancy or SCRs served as dependent variable whereas GS type served as the fixed effect.

2.4.2.3. Robust regression models. Robust regression models examined whether the degree of safety behavior to CSsafeShift trials during *Safety behavior shift* predicted generalized fear to GSsafeShift in *Generalization test*. To this end, safety behavior to the CSsafeShift trials on the last block of *Safety behavior shift* predicted the degree of generalized fear, as reflected by US expectancy ratings or SCRs, to the GSsafeShift trials on the first block of *Generalization test*.

2.4.3. Exploratory analyses

We explored whether the expected increase in US expectancy to GSsafeShift was not merely due to negative generalization decrement (i.e., a decrease in fear inhibition to safety-related GSs due to stimulus generalization). To this end, we compared US expectancies to CSsafe and CSsafeShift on the last block of *Fear acquisition training* with US expectancies to GSsafe and GSsafeShift on the first block of *Generalization test*. Note that we did not use US expectancy ratings to the safety-related CSs during *Safety behavior shift* given that the use of safety behavior might have already modulated US expectancy ratings to CSsafeshift (see Expectancy model; [26]).

We also explored whether individual differences such as trait anxiety, intolerance of uncertainty, or distress tolerance would have any effect on 1) the degree of safety behavior to CSsafeShift trials during *safety behavior shift*, 2) degree of fear generalization to the GSs during *Generalization test*, and 3) modulating the relationship between safety behavior to CSsafeShift and generalized fear to GSsafeShift. To this end, these individual differences were added into the aforementioned models in *Main hypotheses* as continuous variables.

In all the linear mixed models, the main effects and higher-order interactions were analyzed in separate models [13]. The degree of significance was reported with Satterthwaite approximation for degrees of freedom [41]. All analyses were conducted in R (R core team, 2022), with *lmer* package for linear mixed models [2] and *robust* package for robust regression analyses [30]. The effect sizes in the frequentist models were reported as partial-R^2 [17] with *r2glmm* package [16]. Furthermore, as we expected null differences in responding between CSsafe and CSsafeShift during the two acquisition phases, we used a Bayesian approach to support the absence of an effect [19]. In these Bayesian models, we obtained the 95 % highest density intervals (HDIs) that contained the most credible values (analog to 95 % confidence interval in frequentist analyses). We then calculated the posterior distribution via Markov Chain Monte Carlo that fell under the area of the null value, namely the Region Of Practical Equivalence (ROPE). The percentage of HDIs that fell under ROPE was calculated; the higher this percentage was, the more likely it reflected an absence of an effect [19, 20].

3. Results

We only included participants who acquired greater US expectancy ratings to CSthreat than US expectancy ratings averaged across CSsafe and CSsafeShift on the last block of *Fear acquisition training*. Four participants were excluded from this criterion, leaving 116 participants in the final sample. In addition, SCR data for eleven participants were not recorded due to technical issues, thus they were treated as missing data in the dataset. However, their behavioral data were retained for analyses (i.e., n = 116 for behavioral data, n = 105 for SCR data). The descriptive statistics for the sample can be seen in Table 1. All these exclusion

Table 1
Descriptive statistics for the sample. DTS = Distress tolerance scale; IUS = Intolerance of uncertainty scale; DASS21 = Depression Anxiety Stress scale −21.

	Mean (SD)
Gender (Women/Men/Other/Prefer not to disclose)	80/32/3/1
Age	21.09 (2.56)
US intensity (mA)	1.29 (0.49)
DTS (1−5)	3.33 (0.82)
IUS (27−135)	61.89 (17.24)
DASS21-Depression (0−42)	6.34 (6.22)
DASS21-Anxiety (0−42)	10.03 (6.96)
DASS21-Stress (0−42)	12.48 (6.85)

criteria were preregistered (see https://osf.io/auvpy). The dataset is available at https://osf.io/z2dgh/.

3.1. Manipulation check

3.1.1. Fear acquisition training

Fig. 2 shows the US expectancy ratings and SCRs to stimuli across the experiment. For the first contrast (CSthreat vs CSsafe & CSsafeShift), participants exhibited higher US expectancy ratings to the CSthreat compared to the average of CSsafe and CSsafeShift; this difference was larger in the second block compared to the first block. This pattern was supported by a significant interaction between CS type and Block, bCS type(CSthreat vs CSsafe & CSsafeShift)*Block = 22.89, SE = 0.76, p <.001, R^2 = 0.39. Similarly, we observed stronger SCRs to CSthreat compared to responding averaged across CSsafe and CSsafeShift while this difference was larger in the second block compared to the first block, bCS type(CSthreat vs CSsafe & CSsafeShift)*Block = 0.090, SE = 0.015, p <.001, R^2 = 0.024.

For the second contrast (CSsafe vs CSsafeShift), there was no evidence that neither US expectancy ratings nor SCRs differed between CSsafe and CSsafeShift (smallest p =.204). Bayesian models further supported the absence of an effect in both measures, as ≥ 98 % of HDIs of the interactions and main effect involving CS type fell under ROPE. In sum, participants successfully acquired differential US expectancy ratings and conditioned fear to the CSs during acquisition without any differences in responding to CSsafe and CSsafeShift.

3.1.2. Safety behavior acquisition training

Participants showed an average of 88.35 % (SD = 21.24) safety behavior to CSthreat* trials, suggesting that participants used safety behaviors to a great extent when it could probabilistically prevent US onset. Additionally, only 9.91 % of CSthreat* trials were reinforced by a US, suggesting that the contingencies between a safety behavior response and US omission worked as intended. For the first contrast (CSthreat vs CSthreat*, CSsafe, & CSsafeShift), participants exhibited higher US expectancy ratings to reinforced CSthreat trials, compared to US expectancies averaged across CSthreat*, CSsafe, and CSsafeShift trials, bCS type(CSthreat vs CSthreat* & CSsafe & CSsafeShift) = 17.72, SE = 0.49, p <.001, R^2 = 0.52. This effect did not significantly interact with Block, bCS type(CSthreat vs CSthreat* & CSsafe & CSsafeShift) *Block = 1.22, SE = 0.96, p =.204, R^2 = 0.013. Similarly, participants showed stronger SCRs to the reinforced CSthreat trials when compared to responding across the remaining CS trials, bCS type(CSthreat vs CSthreat* & CSsafe & CSsafeShift) = 0.081, SE = 0.0079, p <.001, R^2 = 0.082, while there was no evidence that this difference differed between blocks, bCS type(CSthreat vs CSthreat* & CSsafe & CSsafeShift)*Block = 0.014, SE = 0.016, p =.382, R^2 = 0.001.

For the second contrast (CSsafe vs CSsafeShift), there was no evidence that participants showed any differences in US expectancy ratings or SCRs between CSsafe and CSsafeShift (smallest p =.317). Bayesian models supported the absence of an effect in both measures, as ≥ 97.26 % of HDIs of main effects fell under ROPE. An exception was that only 61.96 % of HDIs of the CStype*Block interaction fell under ROPE. However, this was due to a descriptively weaker responding to CSsafeShift compared to CSsafe, while this descriptive difference was larger in the second block compared to the first block. If anything, this descriptive difference would have contributed to greater responding to GSsafe compared with GSsafeShift in *Generalization test* (i.e., a pattern opposite of our hypothesis).

In sum, participants successfully acquired the safety behavior – US omission contingency. They continued to show heightened conditioned fear to the reinforced CSthreat trials compared to other CSs and no differences in responding between the safety CSs.

3.2. Main hypotheses

Safety behavior shift. Participants showed an average of 25.35 % (SD = 36.36) safety behavior to CSsafeShift* trials. For the first contrast (CSthreat vs CSsafe & CSsafeShift*), participants continued to show higher US expectancy ratings to the reinforced CSthreat compared to ratings averaged across CSsafe and CSsafeShift*; this difference was larger in Block 2 compared to Block 1. This pattern was supported by a significant interaction between CS type and Block, bCS type(CSthreat vs CSsafe & CSsafeShift*)*Block = −0.71, SE = 0.84, p =.030, R^2 = 0.003. In contrast, although SCRs were stronger to CSthreat compared to averaged responding to the safety-related CSs averaged across blocks, bCS type(CSthreat vs CSsafe & CSsafeShift*) = 0.13, SE = 0.0089, p <.001, R^2 = 0.13, this differential difference did not differ between blocks, bCS type(CSthreat vs CSsafe & CSsafeShift*)*Block = 0.028, SE = 0.018, p =.121, R^2 = 0.002.

For the second contrast (CSsafe vs CSsafeShift*), there was no evidence that US expectancy or SCRs differed between CSsafe and CSsafeShift (smallest p =.132). Bayesian models supported the absence of an effect in both measures, as ≥ 95.23 % of HDIs of the interactions and main effects involving CS type fell under ROPE. Thus, threat expectancy nor SCRs to CSsafeShift* was not elevated compared to CSsafe when safety behavior to CSsafeShift was available.

3.2.1. Generalization test (first test block only)

Two orthogonal contrasts were applied to the US expectancy and SCR data. For the first contrast (GSthreat vs GSsafe & GSsafeShift), participants showed higher US expectancy ratings to GSthreat compared with ratings averaged across GSsafe and GSsafeShift, bGS type(GSthreat vs GSsafe & GSsafeShift) = 20.04, SE = 0.59, p <.001, R^2 = 0.60. Similarly, participants showed stronger SCRs to GSthreat compared with responding averaged across the safety-related GSs, bGS type(GSthreat vs GSsafe & GSsafeShift) = 0.073, SE = 0.011, p <.001, R^2 = 0.05.

Most importantly, for the second contrast (GSsafe vs GSsafeShift), participants exhibited higher US expectancy ratings to GSsafeShift compared to GSsafe, bGS type(GSsafe vs GSsafeShift) = 2.44, SE = 1.02, p =.018, R^2 = 0.007. In contrast, there was no evidence that this difference was observed in the SCR data, bGS type(GSsafe vs GSsafeShift) = −0.0046, SE = 0.020, p =.814, R^2 < 0.001. In sum, participants showed generalized fear to GSthreat as indexed by both US expectancy ratings and SCR. More importantly, participants showed an increase in generalized US expectancies to GSsafeShift (due to the avoidable CSsafeShift* trials during *Safety behavior shift*). We also explored whether this effect was long-lived by including both test blocks (see detailed analyses in the Supplementary Materials). In brief, no effects comparing the differences in responding to GSsafeShift and GSsafe reach significance. This null effect was presumably due to extinction learning, especially in the second test block.

3.2.2. Regression models

Fig. 3A and B show the relationship between safety behavior to CSsafeShift trials on the last block of *Safety behavior shift* and generalized fear to GSsafeShift trials on the first block of *Generalization test*. Stronger use of safety behavior to CSsafeShift was significantly associated with higher US expectancy ratings to GSsafeShift, $b_{\mathrm{Avoidance}}$ = 0.49 SE = 0.060, p <.001. In contrast, there was no evidence that safety behavior to CSsafeShift predicted SCRs to GSsafeShift, $b_{\mathrm{Avoidance}}$ = 0.0018 SE = 0.0013, p =.175. We further explored whether the increase in generalized US expectancy ratings to GSsafeShift was specifically due to the use of safety behavior to CSsafeShift. If so, we should observe limited to no association between the use of safety behavior to CSsafeShift and generalized fear to GSsafe. To this end, we explored whether safety behavior to CSsafeShift in the last block of *Safety behavior shift* was associated with generalized fear to GSsafe in the first block of *Generalization test*. We found no evidence that safety behavior use to CSsafeShift was associated with either US expectancy ratings or SCRs to GSsafe

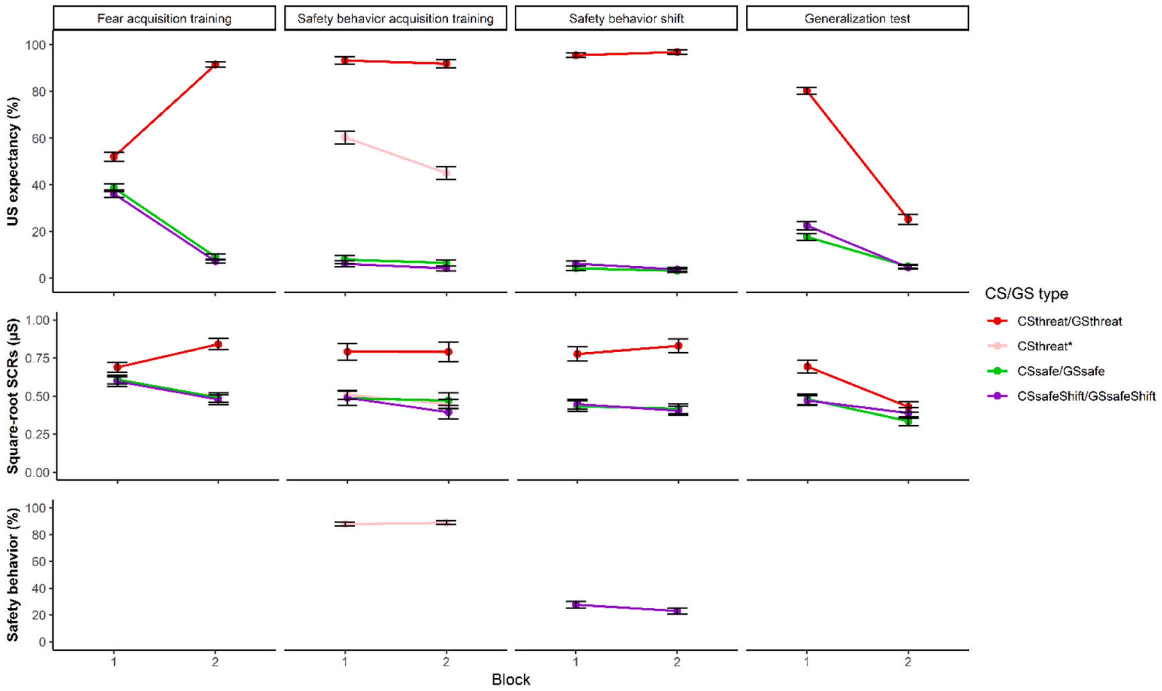

Fig. 2. US expectancy (top panel), square-root SCRs (middle panel), and safety behavior (bottom panel) across all phases. Error bars indicate standard error of the mean. See the color version of this figure online.

(smallest p =.249).

3.3. Exploratory analyses

3.3.1. Cross-phase analysis to account for negative generalization decrement

When comparing US expectancies to CSsafe and CSsafeShift on the last block of *Fear acquisition training,* there was an increase in US expectancy ratings to *both* GSsafe and GSsafeShift on the first block of *Generalization test*, bPhase = 11.94, SE = 1.13, p <.001, R^2 = 0.094. This increase was presumably due to a negative generalization decrement (i.e., a decrease in fear inhibition to safety-related GSs due to stimulus generalization). The increase in US expectancy ratings from CSsafeShift to GSsafeShift was significantly greater than those from CSsafe to GSsafe, supported by a significant interaction between Phase and GS type, bGS type(CSsafe/GSsafe vs CSsafeShift/GSsafeShift)*Phase = 6.69, SE = 2.25, p =.003, R^2 = 0.08. This pattern suggests that the increase in expectancy to GSsafeShift was not only due to negative generalization decrement, assuming that the magnitude of negative generalization decrement was the same for GSsafe and GSsafeShift. In contrast, no effects in the SCR data reached significance (smallest p =.667).

3.3.2. Individual differences in safety behavior and conditioned fear

During *Safety behavior shift*, there was no evidence that trait anxiety, intolerance of uncertainty, nor distress tolerance modulated safety behavior to CSsafeShift when controlling for each other (smallest p =.391). Similarly, during *Generalization test,* there was no evidence that either predisposition factors had any effect on the differential US expectancy ratings or SCRs between the threat-related GSs and the safety-related GSs, nor any differences in responding between the two safety-related GSs (smallest p =.052).

Regarding the relation between safety behavior to CSsafeShift and US expectancy ratings to GSsafeShift, there was no evidence that neither intolerance of uncertainty nor distress tolerance modulated this relation (smallest p =.371). Trait anxiety did positively modulate this relation when controlling for the other two traits (Fig. 3C & D): the higher trait anxiety was, safety behavior to CSsafeShift more strongly predicted US expectancy to GSsafeShift, $b_{Avoidance*Anxiety}$ = 0.051 SE = 0.018, p =.006. However, the variance inflation factors (VIFs) for this model ranged from 6.7 to 32.3, indicating a severe multicollinearity issue. Therefore, we mean-centered all the continuous trait variables, which markedly reduced the VIFs for the interaction terms (largest VIF = 2.3; see [15]). In this revised model, trait anxiety still positively modulated the association between safety behavior to CSsafeShift and US expectancy to GSsafeShift, $b_{Avoidance*Anxiety}$ = 0.36 SE = 0.13, p =.006. For the SCR data, there was no evidence that any individual trait modulated the relation between safety behavior to CSsafeShift and responding to GSsafeShift (smallest p =.208).[2]

4. Discussion

The current study sought to examine whether the formation of threat expectancy to safety cues induced by safety behavior generalizes to related but novel generalization stimuli. In addition, we examined whether the degree of safety behaviors to the safety cue would predict the degree of generalized threat responses to it.

During the first block of *Generalization test*, participants showed higher US expectancy ratings to GSsafeShift compared to GSsafe, in which the former belonged to the same category of CSsafeShift, a safety cue that participants could use safety behaviors to during *Safety behavior shift*. This pattern expands on past findings that not only engaging in safety behaviors to a safety cue induces maladaptive threat belief to it (e.g., [9,50,62]), but this maladaptive belief also generalizes to other novel stimuli that conceptually resemble the avoided safety cue. This pattern was unlikely due to a pre-existing bias in responding between CSsafe and CSsafeShift, as the Bayesian models supported the absence of differences in responding between these two safety-related CSs during the acquisition phases.

[2] We have also examined the effect of gender on generalized threat beliefs due to the use of safety behaviors. Detailed analyses can be found in the Supplementary Materials.

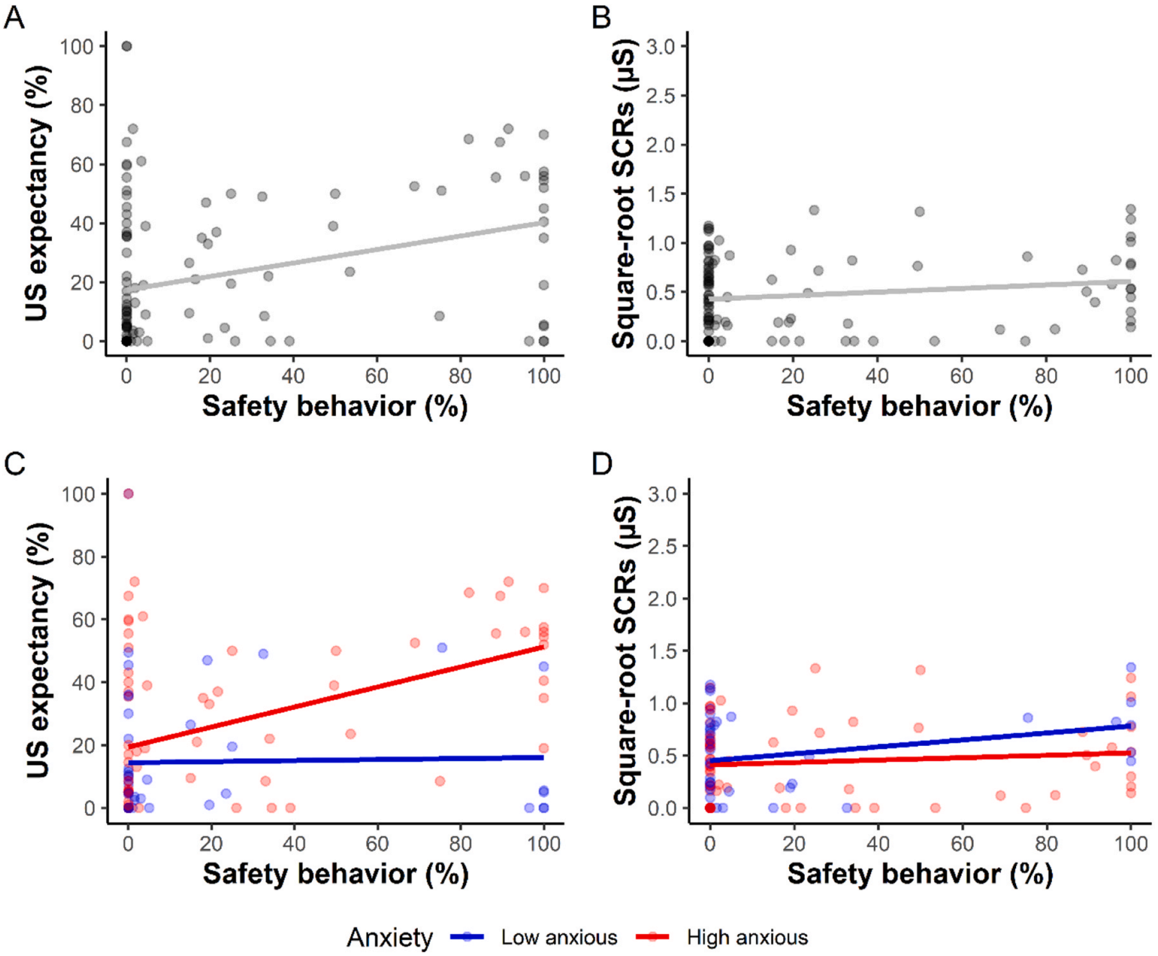

Fig. 3. Top panel). Relationship between safety behavior and conditioned fear. Safety behavior to CSsafeShift* trials on the last block of *Safety behavior shift* predicts generalized US expectancy ratings (A) and square-root SCRs (B) to generalization stimuli on the first block of *Generalization test*. Bottom panel). Trait anxiety modulates the predictive relationship between safety behavior and generalized US expectancy ratings (C) and square-root SCRs (D). Trait anxiety was median split for visual aid. Darker color indicates more overlapping data points.

The increase in US expectancy ratings to GSsafeShift was not merely due to generalization decrement. Generalization decrement refers to a discrepancy in responding between the original trained cue and a novel generalization stimulus [29]. In this context, a generalization decrement was indicated by a general increase in US expectancy ratings to the safety-related GSs compared to the safety-related CSs (i.e., negative generalization decrement: a decrease in inhibitory responses due to generalization). However, a cross-phase analysis revealed that the increase in responding from CSsafeShift to GSsafeShift was greater than from CSsafe to GSsafe, thus suggesting that the increase in threat expectancy from CSsafeShift to GSsafeShift was not merely due to negative generalization decrement, but also due to previous use of safety behaviors to CSsafeShift. Therefore, the current study provides robust evidence that maladaptive threat belief induced by safety behaviors to safety cues generalizes to other related novel cues.

It has been proposed that the mere availability of safety behaviors may suffice to signal threat and thereby form threat beliefs [44]. Therefore, the mere availability of safety behaviors to a safety cue may increase threat expectancy to it even when safety behaviors were not used [51]. However, we found that the individual degree of safety behaviors positively predicted the level of generalized threat expectancies: while participants who used little safety behaviors to the safety cue (CSsafeShift) showed limited generalized threat expectancies (to GSsafeShift), those who used safety behaviors to a greater extent showed stronger generalized threat expectancies. The current findings, along with preliminary evidence from our lab [61], suggest that the degree of safety behavior use to safety stimuli determined the increase in generalized threat beliefs to novel stimuli related to the safety stimuli.

Exploratory analyses showed that when controlling for other individual risk factors, trait anxiety positively moderated the link between safety behaviors to CSsafeShift and generalized threat expectancies to GSsafeShift. This pattern tentatively suggests that highly anxious individuals are more likely to weigh the use of safety behaviors to a safety cue for their evaluation of threat. That means, trait anxious individuals might be more likely to infer generalized threat based on their use of safety behaviors (i.e., behavior as information). This pattern fits with findings that trait anxiety modulates the link between fear and avoidance [37] and clinical evidence that individuals with anxiety-related disorders and obsessive-compulsive disorders tend to infer threat from their behaviors [49]. Therefore, the current study provides an experimental model of how individuals expand their scope of maladaptive (albeit mild) threat beliefs to novel cues that are related to safety cues that one had used safety behaviors to, whereas this pattern was preliminary suggested to be elevated in trait anxious individuals. Although this pattern was not found in intolerance of uncertainty or distress tolerance, it could be due to a power issue (e.g., [33]). Future studies with appropriate sample sizes are required to examine whether these risk factors are associated with an increase in generalized threat expectancies due to the use of safety behaviors.

One advantage of measuring safety behavior on a continuum is that it can more sensitively measure the extent of safety behavior use. It arguably increases face validity as it allows the measure of partial use of safety behavior [21,59]. Preliminary studies [56,57] measuring safety behavior on a continuum also found that individuals at risk of developing clinical anxiety showed excessive safety behavior even in the absence of realistic threat, a pathological feature commonly observed in clinical anxiety. Therefore, these findings provide preliminary support that a continuum measure of safety behavior might have met the diagnostic validity criteria. However, it remains unclear whether this measure of safety behavior is strong in other external validities., for instance, predictive validity. Future research is much required to fully validate this measure on other validity criteria (see [21]).One limitation of the study was that the significant difference in threat expectancies between GSsafe and GSsafeShift was small, compared to previous studies (e.g., [9,62]). This small effect size was presumably due to the current study including all participants despite some of them showed limited to no use of safety behaviors to the safety cue, whereas previous studies (e.g., [9, 62]) excluded these participants. However, including all participants regardless of their use of safety behaviors to the safety cue allowed us to examine whether the degree of safety behaviors use determined the degree of threat expectancies generalization to related cues. Another limitation was that most findings were observed in threat expectancy data but not in SCR data. Noted that the null effect in SCR data was not due to unsuccessful differential acquisition to the CSs. This apparent dissociation between the two measures could be attributed to the three-system model of anxiety [23,31], which proposed that fear is acquired and expressed via three independent systems, including cognitive system such as verbal report of subjective experience, overt behavior, and physiological activity. This model proposes that these three systems can function somewhat autonomously, leading to varying response levels across different measures to the same stimulus. However, this model fails to clearly predict when and why responding across different systems will diverge, or under what conditions responding should converge. An alternative explanation for the apparent discrepancy between the measures in the current study could be SCRs being less sensitive in detecting differences in responding to the safety-related GSs, due to its high inter-individual variability [3].

In conclusion, the current study extends the findings of an increase in threat expectancies to safety cues due to the use of safety behaviors and found that such increase in threat expectancies also generalizes to other related cues. Another key finding was that the degree of safety behaviors use to a safety cue predicted the increase in generalized threat expectancies to related cues, suggesting that the actual use of safety behaviors contributed to the formation of (generalized) maladaptive threat beliefs to other cues. Clinical implications suggest that clinicians should pay attention to minimize subtle safety behaviors to innocuous situations or objects.

Ethical approval

This study was approved by the Ethics Committee of the Erasmus School of Social and Behavioural Sciences (ETH2223–0321) in accordance to the Declaration of Helsinki.

Funding

IM Engelhard was supported by a Vici grant from NWO (grant no. 453–15–005).

CRediT authorship contribution statement

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

Declaration of Competing Interest

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Data Availability

The authors have shared the data via the link: http://osf.io/z2dgh/.

Acknowledgement

The authors would like to thank Felix Teichmann, Sven Eijkelenkamp, Vlada Voskobiinyk, and Denis Torbin for their help in data collection and processing.

Preregistration

This study was preregistered on the Open Science Framework (https://osf.io/auvpy)

Reporting

We reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.bbr.2024.115078.

References

[1] A. Aslanidou, M. Andreatta, A.H.K. Wong, M.J. Wieser, No influence of threat uncertainty on fear generalization, Psychophysiology 61 (1) (2024) e14423, https://doi.org/10.1111/psyp.14423.

[2] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. *67* (1) (2015), https://doi.org/10.18637/jss.v067.i01.

[3] T. Beckers, A.-M. Krypotos, Y. Boddez, M. Effting, M. Kindt, What's wrong with fear conditioning? Biol. Psychol. *92* (1) (2013) 90–96, https://doi.org/10.1016/j.biopsycho.2011.12.015.

[4] R.N. Carleton, The intolerance of uncertainty construct in the context of anxiety disorders: theoretical and practical perspectives, Expert Review of Neurotherapeutics 12 (8) (2012) 937–947, https://doi.org/10.1586/ern.12.82.

[5] J.A. Chambers, K.G. Power, R.C. Durham, The relationship between trait vulnerability and anxiety and depressive diagnoses at long-term follow-up of generalized anxiety disorder, J. Anxiety Disord. *18* (5) (2004) 587–607, https://doi.org/10.1016/j.janxdis.2003.09.001.

[6] S.E. Cooper, E.A.M. van Dis, M.A. Hagenaars, A.M. Krypotos, C.B. Nemeroff, S. Lissek, I.M. Engelhard, J.E. Dunsmoor, A meta-analysis of conditioned fear generalization in anxiety-related disorders, Neuropsychopharmacology *47* (*9*) (2022) 1652–1661, https://doi.org/10.1038/s41386-022-01332-2.

[7] M.G. Craske, K. Kircanski, M. Zelikowsky, J. Mystkowski, N. Chowdhury, A. Baker, Optimizing inhibitory learning during exposure therapy, Behav. Res. Ther. *46* (1) (2008) 5–27, https://doi.org/10.1016/j.brat.2007.10.003.

[8] J.E. Dunsmoor, A. Martin, K.S. LaBar, Role of conceptual knowledge in learning and retention of conditioned fear, Biol. Psychol. *89* (2) (2012) 300–305, https://doi.org/10.1016/j.biopsycho.2011.11.002.

[9] I.M. Engelhard, S.L. van Uijen, N. van Seters, N. Velu, The effects of safety behavior directed towards a safety cue on perceptions of threat, Behav. Ther. *46* (5) (2015) 604–610, https://doi.org/10.1016/j.beth.2014.12.006.

[10] A. Flores, F.J. López, B. Vervliet, P.L. Cobos, Intolerance of uncertainty as a vulnerability factor for excessive and inflexible avoidance behavior, Behav. Res. Ther. *104* (2018) 34–43, https://doi.org/10.1016/j.brat.2018.02.008.

[11] M.H. Freeston, J. Rhéaume, H. Letarte, M.J. Dugas, R. Ladouceur, Why do people worry? Personal. Individ. Differ. *17* (6) (1994) 791–802, https://doi.org/10.1016/0191-8869(94)90048-5.

[12] R. Gatzounis, A. Meulders, Once an avoider always an avoider? Return of pain-related avoidance after extinction with response prevention, J. Pain. *21* (2020) 1224–1235, https://doi.org/10.1016/j.jpain.2020.02.003.

[13] A.F. Hayes, C.J. Glynn, M.E. Huge, Cautions regarding the interpretation of regression coefficients and hypothesis tests in linear models with interactions, Commun. Methods Meas. *6* (1) (2012) 1–11, https://doi.org/10.1080/19312458.2012.651415.

[14] S. Helbig-Lang, F. Petermann, Tolerate or eliminate? A systematic review on the effects of safety behaviors across anxiety disorders, Clin. Psychol.: Sci. Pract. *17* (3) (2010) 218–233, https://doi.org/10.1111/j.1468-2850.2010.01213.x.

[15] D. Iacobucci, M.J. Schneider, D.L. Popovich, G.A. Bakamitsos, Mean centering helps alleviate "micro" but not "macro" multicollinearity, Behav. Res. Methods *48* (2016) 1308–1317, https://doi.org/10.3758/s13428-015-0624-x.

[16] Jaeger, B.C. (2017). R2glmm: computes R squared for mixed (multilevel) models. R. package version 0.1, 2.

[17] B.C. Jaeger, L.J. Edwards, K. Das, P.K. Sen, An R2 statistic for fixed effects. in the generalized linear mixed model, J. Appl. Stat. *44* (2017) 1086–1105, https://doi.org/10.1080/02664763.2016.1193725.

[18] Z. Klein, S. Berger, B. Vervliet, T. Shechner, Fear learning, avoidance, and generalization are more context-dependent for adults than adolescents, Behav. Res. Ther. *147* (2021) 103993, https://doi.org/10.1016/j.brat.2021.103993.

[19] J. Kruschke. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, second ed., Academic Press, 2015.

[20] J.K. Kruschke, T.M. Liddell, The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective, Psychon. Bull. Rev. *25* (1) (2018) 178–206, https://doi.org/10.3758/s13423-016-1221-4.

[21] A.-M. Krypotos, B. Vervliet, I.M. Engelhard, The validity of human avoidance paradigms, Behav. Res. Ther. *111* (2018) 99–105, https://doi.org/10.1016/j.brat.2018.10.011.

[22] L. Kumle, M.L.-H. Võ, D. Draschkow, Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R, Behav. Res. Methods *53* (6) (2021) 2528–2543, https://doi.org/10.3758/s13428-021-01546-0.

[23] P.J. Lang, M.M. Bradley, B.N. Cuthbert, Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology, Biol. Psychiatry *44* (12) (1998) 1248–1263, https://doi.org/10.1016/S0006-3223(98)00275-3.

[24] A. Lemmens, T. Smeets, T. Beckers, P. Dibbets, Avoiding at all costs? An exploration of avoidance costs in a novel Virtual Reality procedure, Learn. Motiv. *73* (2021) 101710, https://doi.org/10.1016/j.lmot.2021.101710.

[25] P.F. Lovibond, Long-term stability of depression, anxiety, and stress syndromes, J. Abnorm. Psychol. *107* (3) (1998) 520–526, https://doi.org/10.1037/0021-843X.107.3.520.

[26] P. Lovibond, Fear and Avoidance: An Integrated Expectancy Model, in: M. G. Craske, D. Hermans, D. Vansteenwegen (Eds.), Fear and learning: From basic processes to clinical implications, American Psychological Association, 2006, pp. 117–132, https://doi.org/10.1037/11474-006.

[27] S.H. Lovibond, P.F. Lovibond. Manual for the Depression Anxiety Stress Scales, 2nd ed., Sydney Psychology Foundation, 1995.

[28] P.F. Lovibond, C.J. Mitchell, E. Minard, A. Brady, R.G. Menzies, Safety behaviours preserve threat beliefs: protection from extinction of human fear conditioning by an avoidance response, Behav. Res. Ther. *47* (8) (2009) 716–720, https://doi.org/10.1016/j.brat.2009.04.013.

[29] N.J. Mackintosh. The Psychology of Animal Learning, Academic Press, 1974.

[30] Maechler M., Rousseeuw P., Croux C., Todorov V., Ruckstuhl A., Salibian-Barrera M., Verbeke T., Koller M., Conceicao E.L., Anna di Palma M. (2023). robustbase: Basic Robust Statistics. R package version 0.99-0, http://robustbase.r-forge.r-project.org/.

[31] I.B. Mauss, M.D. Robinson, Measures of emotion: a review, Cogn. Emot. *23* (2) (2009) 209–237, https://doi.org/10.1080/02699930802204677.

[32] J. Morriss, B. Macdonald, C.M. van Reekum, What is going on around here? Intolerance of uncertainty predicts threat generalization, PLoS ONE 0154494 (2016), https://doi.org/10.1371/journal.pone.0154494.

[33] J. Morriss, D.V. Zuj, G. Mertens, The role of intolerance of uncertainty in classical threat conditioning: Recent developments and directions for future research, International Journal of Psychophysiology 166 (2021) 116–126, https://doi.org/10.1016/j.ijpsycho.2021.05.011.

[34] S. Papalini, M. Ashoori, J. Zaman, T. Beckers, B. Vervliet, The role of context in persistent avoidance and the predictive value of relief, Behav. Res. Ther. *138* (2021) 103816, https://doi.org/10.1016/j.brat.2021.103816.

[35] A. Pittig, Incentive-based extinction of safety behaviors: Positive outcomes competing with aversive outcomes trigger fear-opposite action to prevent protection from fear extinction, Behav. Res. Ther. *121* (2019) 103463, https://doi.org/10.1016/j.brat.2019.103463.

[36] A. Pittig, J.M. Boschet, V.M. Glück, K. Schneider, Elevated costly avoidance in anxiety disorders: Patients show little downregulation of acquired avoidance in face of competing rewards for approach, Depress Anxiety *38* (3) (2021) 361–371, https://doi.org/10.1002/da.23119.

[37] A. Pittig, A.R. Schulz, M.G. Craske, G.W. Alpers, Acquisition of behavioral avoidance: Task-irrelevant conditioned stimuli trigger costly decisions, Journal of Abnormal Psychology 123 (2) (2014), https://doi.org/10.1037/a0036136.

[38] A. Pittig, A.H.K. Wong, Incentive-based, instructed, and social observational extinction of avoidance: fear-opposite actions and their influence on fear extinction, Behav. Res. Ther. *137* (2021) 103797, https://doi.org/10.1016/j.brat.2020.103797.

[39] J.A. Rattel, S.F. Miedl, J. Blechert, F.H. Wilhelm, Higher threat avoidance costs reduce avoidance behaviour which in turn promotes fear extinction in humans, Behav. Res. Ther. *96* (2017) 37–46, https://doi.org/10.1016/j.brat.2016.12.010.

[40] C. San Martin, B. Jacobs, B. Vervliet, Further characterization of relief dynamics in the conditioning and generalization of avoidance: Effects of distress tolerance and intolerance of uncertainty, Behav. Res. Ther. *124* (2020) 103526, https://doi.org/10.1016/j.brat.2019.103526.

[41] F.E. Satterthwaite, Synthesis of variance, Psychometrika *6* (5) (1941) 309–316, https://doi.org/10.1007/BF02288586.

[42] K.A. Sexton, P.J. Norton, J.R. Walker, G.R. Norton, Hierarchical model of generalized and specific vulnerabilities in anxiety, Cogn. Behav. Ther. *32* (2) (2003) 82–94, https://doi.org/10.1080/16506070302321.

[43] J.S. Simons, R.M. Gaher, The distress tolerance scale: development and validation of a self-report measure, Motiv. Emot. *29* (2) (2005) 83–102, https://doi.org/10.1007/s11031-005-7955-3.

[44] T. Sloan, M.J. Telch, The effects of safety-seeking behavior and guided threat reappraisal on fear reduction during exposure: an experimental investigation, Behav. Res. Ther. *40* (3) (2002) 235–251, https://doi.org/10.1016/S0005-7967(01)00007-9.

[45] J.G. Snodgrass, M. Vanderwart, A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity, J. Exp. Psychol.: Hum. Learn. Mem. *6* (2) (1980) 174–215, https://doi.org/10.1037/0278-7393.6.2.174.

[46] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Publication recommendations for electrodermal measurements: Publication standards for EDA, Psychophysiology *49* (8) (2012) 1017–1034, https://doi.org/10.1111/j.1469-8986.2012.01384.x.

[47] C.D. Spielberger, Anxiety, cognition and affect: a state-trait perspective, in: A. H. Tuma, J.D. Maser (Eds.), Anxiety and the Anxiety Disorders, Erlbaum, Hillsdale, NJ, 1985, pp. 171–182.

[48] M.J. Telch, K. Jacquin, J.A.J. Smits, M.B. Powers, Emotional responding to hyperventilation as a predictor of agoraphobia status among individuals suffering from panic disorder, J. Behav. Ther. Exp. Psychiatry *34* (2) (2003) 161–170, https://doi.org/10.1016/s0005-7916(03)00037-5.

[49] M. van den Hout, A. Gangemi, F. Mancini, I.M. Engelhard, M.M. Rijkeboer, M. van Dams, I. Klugkist, Behavior as information about threat in anxiety disorders: a comparison of patients with anxiety disorders and non-anxious controls, J. Behav. Ther. Exp. Psychiatry *45* (4) (2014) 489–495, https://doi.org/10.1016/j.jbtep.2014.07.002.

[50] E.A.M. van Dis, A.-M. Krypotos, M.A.J. Zondervan-Zwijnenburg, A.M. Tinga, I. M. Engelhard, Safety behaviors toward innocuous stimuli can maintain or increase threat beliefs, Behav. Res. Ther. *156* (2022) 104142, https://doi.org/10.1016/j.brat.2022.104142.

[51] S.L. Van Uijen, A. Leer, I.M. Engelhard, Safety behavior after extinction causes a return of threat expectancies, Behav. Ther. *49* (2018) 450–458.

[52] B. Vervliet, E. Indekeu, Low-cost avoidance behaviors are resistant to fear extinction in humans, Front. Behav. Neurosci. *9* (2015) 351, https://doi.org/10.3389/fnbeh.2015.00351.

[53] B. Vervliet, I. Lange, M.R. Milad, Temporal dynamics of relief in avoidance conditioning and fear extinction: experimental validation and clinical relevance, Behav. Res. Ther. *96* (2017) 66–78, https://doi.org/10.1016/j.brat.2017.04.011.

[54] S. Wake, C.M. van Reekum, H. Dodd, The effect of social anxiety on the acquisition and extinction of low-cost avoidance, Behav. Res. Ther. *146* (2021) 103967, https://doi.org/10.1016/j.brat.2021.103967.

[55] A. Wells, D.M. Clark, P. Salkovskis, J. Ludgate, A. Hackmann, M. Gelder, Social phobia: the role of in-situation safety behaviors in maintaining anxiety and negative beliefs, Behav. Ther. *26* (1) (1995) 153–161, https://doi.org/10.1016/S0005-7894(05)80088-7.

[56] A.H.K. Wong, M. Franzen, M.J. Wieser, Unconditioned stimulus devaluation decreases the generalization of costly safety behaviors, J. Anxiety Disord. *103* (2024) 102847, https://doi.org/10.1016/j.janxdis.2024.102847.

[57] A.H.K. Wong, J.C. Lee, P. Engelke, A. Pittig, Reduction of costly safety behaviors after extinction with a generalization stimulus is determined by individual differences in generalization rules, Behav. Res. Ther. *160* (2023) 104233, https://doi.org/10.1016/j.brat.2022.104233.

[58] A.H.K. Wong, P.F. Lovibond, Excessive generalization of conditioned fear in trait anxious individuals under ambiguity. Behav. Res. Ther. *107* (2018) 53–63, https://doi.org/10.1016/j.brat.2018.05.012.

[59] A.H.K. Wong, A. Pittig, A dimensional measure of safety behavior: a non-dichotomous assessment of costly avoidance in human fear conditioning, Psychol. Res. *86* (1) (2022) 312–330, https://doi.org/10.1007/s00426-021-01490-w.

[60] A.H.K. Wong, A. Pittig, Threat belief determines the degree of costly safety behavior: assessing rule-base generalization of safety behavior with a dimensional measure of avoidance, Behav. Res. Ther. *156* (2022) 104158, https://doi.org/10.1016/j.brat.2022.104158.

[61] A.H.K. Wong, E.A.M. van Dis, A. Pittig, M.A. Hagenaars, I.M. Engelhard, The degree of safety behaviors to a safety stimulus predicts development of threat beliefs, Behav. Res. Ther. *170* (2023) 104423, https://doi.org/10.1016/j.brat.2023.104423.

[62] W. Xia, S. Dymond, K. Lloyd, B. Vervliet, Partial reinforcement of avoidance and resistance to extinction in humans, Behav. Res. Ther. *96* (2017) 79–89, https://doi.org/10.1016/j.brat.2017.04.002.

[63] D.V. Zuj, W. Xia, K. Lloyd, B. Vervliet, S. Dymond, Negative reinforcement rate and persistent avoidance following response-prevention extinction, Behav. Res. Ther. *133* (2020) 103711, https://doi.org/10.1016/j.brat.2020.103711.