Journal of Anxiety Disorders 103 (2024) 102847

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

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

Unconditioned stimulus devaluation decreases the generalization of costly safety behaviors

Alex H.K. Wong *, Minita Franzen, Matthias J. Wieser

Department of Psychology, Educational Sciences, and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
US devaluation
Safety behaviors
Generalization
Fear and avoidance conditioning
Trait anxiety
Intolerance of uncertainty

ABSTRACT

Safety behaviors are often maladaptive in clinical anxiety as they typically persist without realistic threat and cause various impairments. In the laboratory, safety behaviors are modelled by responses to a conditioned stimulus (CS) that reduce the occurrence of an expected aversive unconditioned stimulus (US). Preliminary evidence suggests that US devaluation, a procedure that decreases US aversiveness, devalues the threat value of the CS and thus diminishes safety behaviors to the CS. This study (n = 78) aimed to extend this finding and examined whether US-devaluation can reduce the generalization of safety behaviors to various stimuli. After acquiring safety behaviors to CSs of different categories, the US predicted by one CS category was devalued. In test, participants showed a selective reduction in safety behaviors to novel stimuli of the devalued CS category, reflecting a decrease in generalization of safety behaviors. Trait anxiety was associated with persistent generalized safety behaviors to novel stimuli of the devalued category. We discuss how US devaluation may improve treatment outcome but also the challenges of clinical translation.

1. Introduction

Safety behaviors are a type of avoidance responses that minimize or even prevent (the onset of) a perceived threat. Although an individual uses safety behaviors with the idea that they induce relief by mitigating the perceived threat, these behaviors often maintain or exacerbate anxiety by preventing the individual from learning that the feared outcomes are unlikely to occur. Safety behaviors are commonly observed in clinical anxiety (Blakey & Abramowitz, 2016; Mendlowicz & Stein, 2000; Olatunji et al., 2007; Pittig et al., 2020). For instance, an individual with social anxiety may over-rehearse a conversation before engaging in social situations. The absence of the perceived threat (e.g., being judged negatively) may then be attributed to the excessive rehearsal thus preventing the correction of one's maladaptive threat belief. This pattern has been empirically demonstrated in laboratory studies (i.e., protection from extinction; Lovibond et al., 2009; Pittig, 2019; Rattel et al., 2017).

In the laboratory, safety behaviors are typically modelled in a fear and avoidance conditioning framework. In this framework, an initially neutral conditioned stimulus (CS) is first repeatedly paired with an aversive unconditioned stimulus (US), thus the CS+ (i.e., a CS that is reinforced by a US) comes to evoke conditioned fear. In a following avoidance conditioning phase, executing a designated response (e.g., pressing a button), during CS+ presentation prevents the upcoming US. To model maladaptive safety behaviors that inflict impairments in clinical anxiety, empirical studies often incorporated tangible (e.g., financial cost; Wong & Pittig, 2022a) or intangible cost such as physical cost (Glogan et al., 2020) in which safety behaviors that could more effectively prevent a US required more physical effort to execute. Furthermore, safety behavior acquired to the CS+ also generalizes to other stimuli that are conceptually related to the CS+ . Empirical studies also showed that safety behaviors generalized to novel stimuli that belonged to the same category of the CS+ (e.g., Dymond et al., 2011, 2014; Kloos et al., 2022). For instance, after acquiring conditioned fear to bird exemplars (e.g., penguin, sparrow), participants showed generalized conditioned fear to other novel exemplars of the same category (e.g., pigeon, emu; Dunsmoor & Murphy, 2014; Wong & Beckers, 2021). These studies provide empirical evidence that safety behaviors spread to other novel innocuous stimuli, providing an explanation why clinically anxious individuals use safety behaviors to a wide range of stimuli or situations.

There has been an increase in laboratory studies examining methods for reducing maladaptive safety behaviors (e.g., see Dymond, 2019; Ball & Gunaydin, 2022; Scheveneels et al., 2021, for reviews) given its

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

https://doi.org/10.1016/j.janxdis.2024.102847
Received 15 September 2023; Received in revised form 9 January 2024; Accepted 19 February 2024
Available online 21 February 2024

clinical importance. However, the literature as of now is scarce in how US devaluation, which refers to decreasing the US aversiveness, can effectively reduce conditioned fear and safety behaviors. Traditionally, US devaluation was carried out by having participants directly experiencing a decrease in US intensity. After US devaluation, empirical studies showed a reduction in conditioned fear to the CS (Hosoba et al., 2001) or to a higher-order CS (a CS that predicts the onset of the CS+, Davey & Mckenna, 1983). More recently, studies employed various US devaluation procedures, such as imagery rescripting (e.g., Dibbets et al., 2012; Woelk et al., 2022), which involved mentally rescripting the US being less aversive and more positive before confronting the CS+ in test again. Participants showed a reduction in conditioned fear to the CS+ after imagery rescripting (Dibbets et al., 2012; Woelk et al., 2022). These laboratory studies align with clinical studies which suggested that imagery rescripting can effectively reduce symptoms of anxiety-related disorders (Kip et al., 2023; Morina et al., 2017). Flores et al., (2018, 2020) extended these findings to safety behaviors. In their studies, participants first acquired safety behaviors to a CS+ that signaled a noise US to the right ear and a different CS+ that signaled a noise US to the left ear. The noise US to the right ear was then devalued. In a following avoidance test, participants showed limited safety behaviors to the rightCS+ but not to the leftCS+ , thus providing evidence that US devaluation reduces safety behaviors to the CS+ .

However, Flores et al., (2018, 2020) only examined whether US devaluation reduced safety behaviors specifically to the CS+ . It is of clinical importance to further test whether reduction of safety behaviors will effectively generalize to other novel stimuli that resemble the devalued CS+ . This is especially important as a wide range of stimuli resembling the feared stimulus evoke maladaptive generalized safety behaviors in clinical anxiety due to excessive fear generalization (e.g., Kaczkurkin et al., 2017; Lissek et al., 2010, 2014). Indeed, excessive fear generalization is thought to be a transdiagnostic mechanism underlying anxiety-related disorders (e.g., panic disorder, specific phobias) and trauma- and stressor-related disorders (Cooper et al., 2022; Fraunfelter et al., 2022; Sep et al., 2019). Preliminary evidence also suggests that individuals with anxiety-related disorders showed maladaptive safety behaviors characterized by persistent costly safety behaviors in the absence of threat (Pittig et al., 2021) and unnecessary safety behaviors to safety stimuli (De Kleine et al., 2023), which interferes with the effectiveness of exposure-based treatments (e.g., Helbig-Lang & Petermann, 2010; Wells et al., 1995). Taken together, a laboratory model that examines whether US devaluation can decrease the spread of safety behaviors provides insights in improving treatment outcome.

The current study thus examined whether US devaluation reduced safety behaviors to novel (but innocuous) generalization stimuli (GSs) that shared the same category membership with the CS+ . In other words, this study provides a laboratory model to examine whether US-devaluation decreases the generalization of maladaptive safety behaviors. To this end, participants first acquired conditioned fear to two CS+s: a rightCS+ that predicted a right noise US and a leftCS+ that predicted a left noise US. They then acquired safety behaviors to the two CS+s that mitigated the chances of receiving the noise USs. Additionally, executing safety behaviors led to a loss in financial reward. This rendered safety behaviors costly, modelling maladaptive safety behaviors in clinical anxiety (Krypotos et al., 2018; Pittig et al., 2020). In a following *US devaluation* phase, the intensity of the right noise US was decreased to a neutral level. In a final *Generalization test* phase, costly safety behaviors to rightGSs and leftGSs, generalization stimuli that belonged to the same category of the rightCS+ and leftCS+ , respectively, were assessed. We expected a lesser extent in costly safety behaviors to the rightGSs (i.e., generalization stimuli related to the *devalued* rightCS+s) compared to the leftGSs (i.e., generalization stimuli related to the *non-devalued* leftCS+s).

This study also explored whether risk factors of clinical anxiety would have any effect on the reduction of costly safety behaviors generalization after US devaluation. Two risk factors were assessed in this study, trait anxiety and intolerance of uncertainty. Trait anxiety refers to a stable trait that is characterized by a strong tendency to react negatively across situations in general. It has been agreed as a risk factor for clinical anxiety (e.g., Chambers et al., 2004; Gershuny & Sher, 1998; Jorm et al., 2000). Intolerance of uncertainty refers to an incapacity to tolerate situations with high level of outcome ambiguity (e.g., Carleton et al., 2012; Freeston et al., 1994; Sexton et al., 2003). Empirical evidence suggests these risk factors are associated with persistent safety behaviors to an extinguished CS+ or to a safety CS- (e.g., Vervliet & Indekeu, 2015; Zuj et al., 2020; also see Wong et al., 2023 for a review), thus it is speculated that these risk factors would also be associated with persistent safety behaviors to innocuous GSs related to a devalued CS+ .

2. Method

The pre-registration and data of this study can be found at Open Science Framework (Pre-registration: https://osf.io/gvzyn; Data: https://osf.io/84pza/).

2.1. Participants

Undergraduates from Erasmus University Rotterdam were recruited for the study in exchange for partial course credits. Participants were also informed that they would receive a voucher of up to 5€ depending on their performance in the experiment. However, all participants received a 5€ voucher regardless. Using the dataset from Flores et al. (2018), a data-based simulation power analysis (Kumle et al., 2021) revealed that 70 participants provided 94.5% power to detect the effect of US devaluation of costly safety behaviors of the smallest effect size of interest (see Kumle et al., 2021) in *Test* (see https://osf.io/gvzyn). A total of 80 participants were recruited to account for attrition rate such as participants not reaching the acquisition criteria or technical issues. This study was approved by the Ethics Committee of the Erasmus School of Social and Behavioural Sciences (ETH2122–0453) in accordance to the Declaration of Helsinki.

2.2. Apparatus and Materials

Twenty-four standardized black-and-white drawings (Snodgrass & Vanderwart, 1980) served as the conditioned stimuli (CSs) or the generalization stimuli (GSs). These drawings were picked from three categories: mammal, fruit, and vehicle (**mammal**: bear, cow, deer, dog, horse, lion, rabbit, and sheep; **fruit**: apple, banana, cherry, lemon, orange, pear, pineapple, and strawberry; **vehicle**: bicycle, bus, car, helicopter, motorbike, plane, train, and truck; see Fig. 1). The exemplars that served as CSs and GSs were counterbalanced across participants.

Two noise USs were delivered via a headphone (Audio-Technica ATH-M40x) connected to an amplifier (Beyerdynamic A-20). One noise US was delivered to the right ear (right noise US) whereas the other noise US was delivered to the left ear (left noise US). The noise USs were a 1000 ms beep of 44100 Hz. Skin conductance was measured via a pair of Ag/AgCl electrodes attached to the hypothenar muscles of participants' non-dominant hand. Skin conductance was recorded at a 1000 Hz sampling rate by a Biopac MP150 system equipped with a EDA100 amplifier.

US expectancy ratings and safety behaviors were measured via visual analogue scales (VAS) presented on the screen. The US expectancy VAS was a bidirectional scale in which the right end of the scale indicates 100% expectancy for the right noise US, the left end of the scale indicates 100% expectancy for the left noise US, and the middle of the scale indicates neither USs. Similarly, the safety behaviors VAS was a bidirectional scale in which the right end of the scale indicates 100% avoidance to the right noise US, the left end of the scale indicates 100% avoidance to the left noise US, and the middle of the scale indicates 0% prevention of neither USs. Of note, this continuous avoidance VAS has been shown to sensitively measure safety behaviors to GSs even when

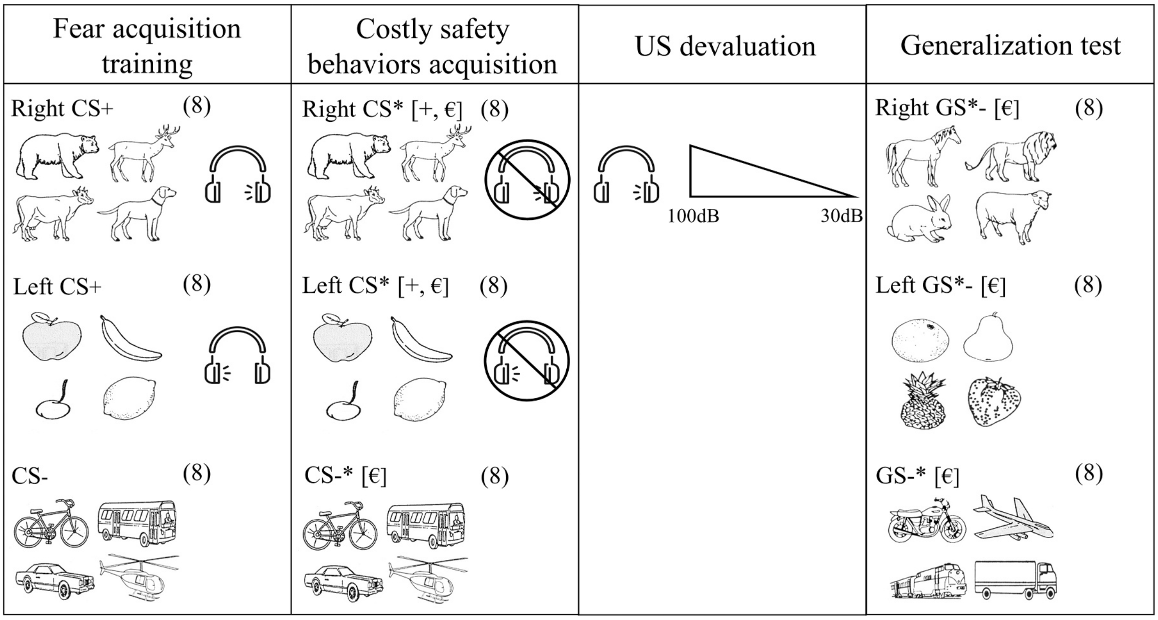

Fig. 1. Experimental design. CS indicates conditioned stimuli. Each CS represents 4 exemplars from one of the three categories (mammal, fruit, vehicle). Each exemplar was presented twice in each phase. + means US presentation (either right noise US or left noise US); - indicates US omission; * indicates safety behaviors availability; + and € in brackets indicate US presentation and a competing reward, respectively, depending on safety behaviors and CS type. GS indicates generalization stimuli. Each GS represent 4 novel exemplars from one of each category. Each exemplar was presented twice. Number in parentheses indicates the number of trials per trial type.

each GS is presented for one trial (see Wong & Pittig, 2022b; Wong & Lee, Engelke et al., 2023 for detailed arguments). Both VASs were presented at the bottom of the screen along with the CSs or the GSs. A computer equipped with Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, Version 20.1) presented all visual stimuli and VASs and recorded all self-reported ratings.

Trait anxiety was assessed by the trait version of Depression Anxiety Stress Scale-21 (DASS-21; Lovibond & Lovibond, 1995; Lovibond, 1998), which measures and discriminates between depression, anxiety, and stress. The DASS-21 anxiety subscale has great internal consistency with a Cronbach's alpha of 0.87 (Antony et al., 1998). The subscale of DASS also showed selective stability; anxiety assessed by the DASS anxiety subscale remained the best predictor for anxiety after 8 years over the other subscales (Lovibond, 1998). Trait anxiety was scored by multiplying the sum of scores of the anxiety subscales by two, so that the scores of trait anxiety become comparable to the original 42-item DASS and its cut-offs (Lovibond & Lovibond, 1995). Intolerance of uncertainty was assessed by the Intolerance of Uncertainty scales (IUS; Buhr & Dugas, 2002). IUS shows great internal consistency with a Cronbach's alpha of 0.94. It also appears to show great stability within a 5-week period (Buhr & Dugas, 2002). Intolerance of uncertainty was scored as sum of scores of IUS.

2.3. Procedure

After providing informed consent, participants filled in the DASS-21 and IUS questionnaires. Skin conductance electrodes filled with isotonic gel were attached. A US familiarization procedure was then carried out, in which the US was presented for five trials, starting from 30 dB to 100 dB. The US was delivered to the right and the left ear on alternative trials, so that a 100 dB right noise US and a 100dB left noise US were delivered on the fourth and last trial, respectively. If participants reported that the USs were too aversive, the US intensity would be reduced to 95 dB. In the final sample ($n = 78$), 9 participants requested to reduce the US intensity to 95 dB. On a scale of 1 to 10 in which 1 indicated not at all unpleasant whereas 10 indicated highly unpleasant, the averaged unpleasantness ratings for the 100 dB noise ($M=7.9$, $SD=0.38$) did not differ from those to the 95 dB noise ($M=7.9$, $SD=0.33$).

Next, a reward-matching procedure was carried out. This procedure aimed to identify a reward that is neither too high nor too low to minimize floor or ceiling effects on safety behaviors (see Schlund et al., 2016; Wong & Pittig, 2022a). Participants were asked a series of questions "Are you willing to tolerate the noise (either to your right or left ear) if you are given €_?" in which the amount of reward ranged from 5cents to 31cents in odd numbers (5cents, 7cents … 29cents, 31cents). This amounted to a total of 14 questions, presented in a randomized order. Participants had to answer either 'Yes' or 'No' to each question. The reward between the highest amount that received a 'No' and the lowest amount that received a 'Yes' was selected as the competing reward for safety behaviors. For instance, a participant who was only willing to tolerate the noise US starting from 27cents (i.e., answering 'No' from 5 to 25cents but answering 'Yes' from 27cents onward) would have 26cents selected as the maximum amount of competing reward per trial.

The fear and avoidance conditioning task that followed consisted of four consecutive phases: *Fear acquisition training*, *Costly safety behaviors acquisition*, *US devaluation*, and *Generalization test* (see Fig. 1).

2.3.1. Fear acquisition training

This phase comprised two blocks, with four black-and-white drawings from each category (mammal, fruit, and vehicle) presented once each in each block. CSs from one category (e.g., mammal) signaled a right noise US at a 75% reinforcement rate (rightCS+), CSs from another category (e.g., fruit) signaled a left noise US at a 75% reinforcement rate (leftCS+), CSs from the remaining category (e.g., vehicle) were not reinforced (CS-). Each CS was presented at the center of the screen along with the bidirectional US expectancy VAS for 8 s (see Fig. 2 A). Participants indicated their US expectancy ratings on each trial. The presentation order of the CSs was pseudo-randomized so that the same trial type would not occur more than twice in a row. The intertrial intervals (ITIs) were randomized between 11 and 15 s. The same pseudo-randomization and ITIs were applied to all the following phases. The CSs were counterbalanced across participants.

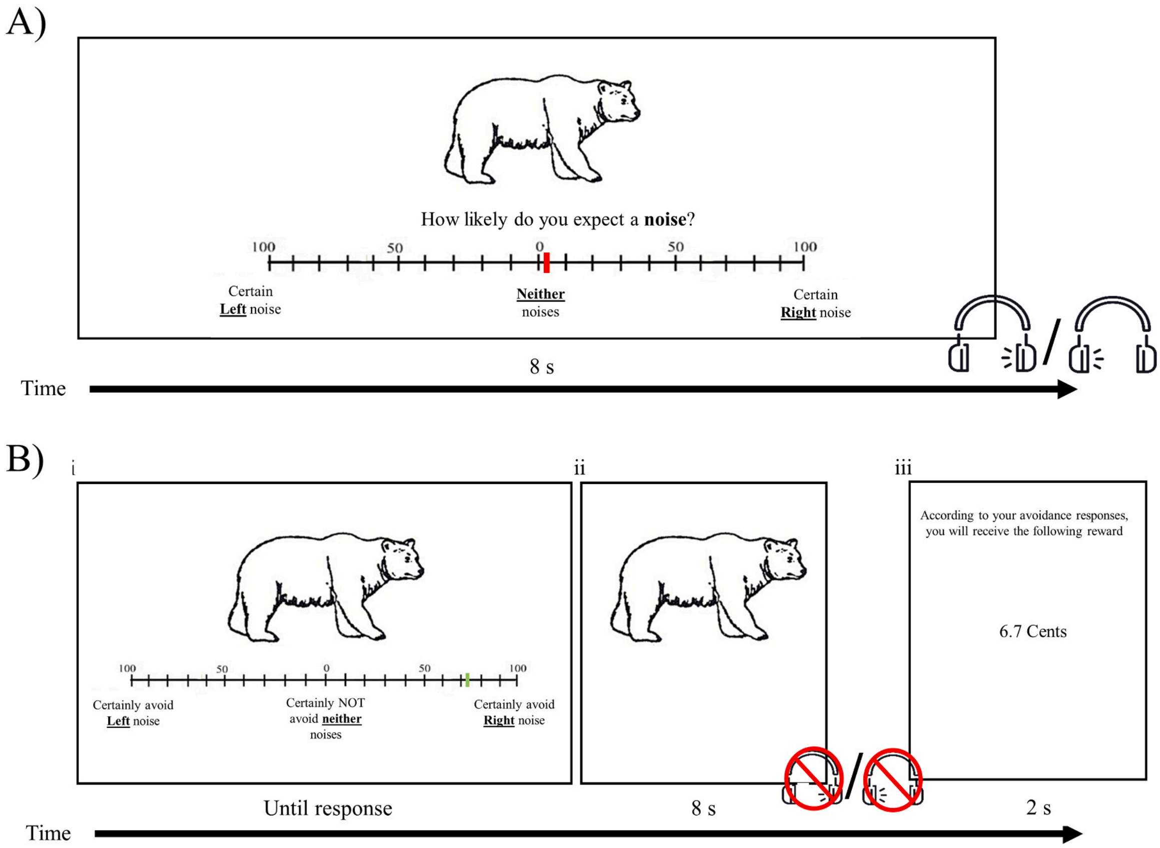

Fig. 2. (A) Trial structure during *Fear acquisition training*. The CS was presented along with the bidirectional US expectancy scale for 8 s; participants were prompted to indicate their US expectancies. Immediately after CS offset, a right noise US, a left noise US, or no US was delivered depending on the CS type. (B) Trial structure during *Costly safety behaviors acquisition* and *Generalization test*. (i) Participants were prompted to indicate their engagement in safety behaviors. (ii) After response, the CS or the GS remained on screen for 8 s. Immediately after CS/GS offset, a right noise US or a left noise US might be administered depending on the degree of safety behaviors engagement and trial type during *Costly safety behaviors acquisition*, whereas none of the stimuli were reinforced during *Generalization test*. (iii) A reward feedback appeared on screen for 2 s.

2.3.2. Costly safety behaviors acquisition

Participants were instructed that they could now avoid the noise USs signaled by the CSs by indicating their safety behaviors on the bidirectional safety behaviors VAS (see Fig. 2B). The degree of safety behaviors engagement probabilistically determined the chance of US omission. For instance, an 85% engagement in safety behaviors would lead to an 85% chance of US omission if the CS would have signaled that particular US. Additionally, each trial came with a competing reward that was individually selected via the reward-matching procedure. The amount of the competing reward was, however, inversely proportional to the degree of safety behaviors engagement on each trial. For instance, an 85% safety behaviors engagement would lead to a gain of 15% of the maximum reward for that trial. Participants were instructed that all reward gained through the experiment would be paid off at the end of the task. Participants were also informed that inaccurate safety behaviors would not effectively prevent US presentation. For example, indicating 100% prevention of a *right noise US* on a *leftCS+* trial would still result in a *left noise US* presentation.

This phase comprised two blocks, with each of the four CSs from each category presented once in each block. On each trial, the CS would be presented along with the safety behaviors VAS until response. After that, the CS remained on screen for 8 s. Immediately after CS offset, participants might receive a right noise US, a left noise US, or nothing depending on the trial type, the degree of safety behaviors engagement, and the accuracy of safety behaviors. A reward feedback would then be presented for 2 s

2.3.3. US devaluation

In this phase, the intensity of the right noise US was decreased in a stepwise procedure from 100 dB/95 dB to 30 dB across four trials (100/95→77→54→30). The devalued right noise US (at 30 dB) was then repeated twice, totaling 6 right noise US trials. On each trial, participants rated the aversiveness of the right noise US on a VAS ranging from 0 to 100, with 0 indicating *Not at all aversive* and 100 indicating *Very aversive*.

2.3.4. Generalization test

Participants were instructed that they could use costly safety behaviors again. In this phase, novel GSs that shared the same category membership of the CSs were present (rightGS, leftGS, GS-). This phase comprised two blocks, with four GSs from each category presented once in each block, totaling 12 trials per block. Each GS was presented along with the bidirectional safety behaviors VAS until response. The same GS then remained on screen for 8 s, followed by a reward feedback for 2 s. None of the GSs were reinforced regardless of trial type and safety behaviors.

2.4. Scoring and analysis

2.4.1. SCRs

Only skin conductance recorded during the 8 s CS presentation during *Fear acquisition training* was included for analysis. A 50 Hz notch filter and a 1 Hz low-pass filter were applied to the SCR data. The SCRs were obtained by the trough-to-peak method 1 s after CS onset till CS offset. The SCRs were then square root transformed to reduce skewness

(Boucsein et al., 2012). The processing of SCRs was done by research assistants blinded to the trial types.

2.4.2. US expectancy ratings and safety behaviors

Due to the bidirectional VASs, US expectancy or safety behaviors to the rightCS+ /rightGS ranged from + 1 to + 100 whereas the same measure for the leftCS+ /leftGS ranged from − 1 to − 100. To make these measures comparable between stimuli, the absolute values of responding to the leftCS+ /GS were taken. Furthermore, inaccurate safety behaviors throughout the entire experiment (e.g., using safety behaviors to left noise US during rightCS+ presentations) were removed and treated as missing data.

All analyses were conducted with linear mixed models. In addition, when a null difference was expected (e.g., no difference in responding to the rightCS+ and the leftCS+ during *Fear acquisition training*), a Bayesian approach was used to support the absence of an effect (Kruschke, 2015). In these Bayesian models, the 95% highest density intervals (HDIs) that contained the most credible values were calculated. Then, a Markov Chain Monte Carlo method was used to calculate the posterior distribution that fell under the area of null value, namely the Region Of Practical Equivalence (ROPE). The percentage of HDIs that fell under ROPE was obtained; a higher percentage reflects a higher likelihood of an absence of an effect. (Kruschke, 2015; Kruschke & Liddell, 2018). All analyses were carried out using R (R Core Team, 2023), with *lmer* package for the frequentist linear mixed models (Bates et al., 2015) and *bayetestR* (Makowski et al., 2019) for the Bayesian models. The degree of significance was reported with Satterthwaite approximation for degrees of freedom (Satterthwaite, 1941). The effect sizes in the frequentist models were reported as partial-R^2 (Jaeger, Edwards et al., 2017) with *r2glmm* package (Jaeger, 2017). The analyses were separated into three parts: manipulation check, main hypotheses, and exploratory analyses. All analyses were pre-registered (see https://osf.io/gvzyn).

2.4.3. Manipulation check

2.4.3.1. The acquisition of conditioned fear and costly safety behaviors. Two orthogonal contrasts were employed to the two acquisition phases. The first contrast examined whether participants successfully acquired differential responding to the CSs. To this end, responding averaged across the rightCS+ and the leftCS+ was compared to CS-. The second contrast examined whether there were any differences in responding between the two threat-related CSs, by comparing responses to the rightCS+ with the leftCS+ . For these two contrasts, US expectancy ratings or SCRs (during *Fear acquisition training)* or safety behaviors (during *Costly safety behaviors acquisition)* served as a continuous dependent variable, whereas CS type (rightCS+, leftCS+, & CS-) and Block served as fixed effects. Given that we expected no differences in responding between the rightCS+ and the leftCS+ , we also carried out Bayesian models to support the absence of an effect. Participants served as a random effect; it was the only random effect served in all the following linear mixed models.

2.4.3.2. US devaluation. To check whether US devaluation was successful, aversiveness ratings to the right noise US on the last trial (30 dB) were compared to the first trial (100 dB/95 dB) during this phase. To this end, aversiveness ratings served as a continuous dependent variable, whereas Trial (first trial & last trial) served as a fixed effect.

2.4.4. Main hypotheses

2.4.4.1. Generalization Test. Two non-orthogonal planned contrasts were applied. The first contrast examined whether costly safety behaviors selectively generalized to novel GSs belonging to the same category of the threat-related leftCS+ . To this end, costly safety behaviors to the leftGS was compared with that to GS- that belonged to the safety category CS-. Noted that this contrast deviated from the pre-registered analysis, as the current contrast could more sensitively detect whether generalized costly safety behaviors occurred. The second contrast directly compared whether US devaluation decreased generalized costly safety behaviors by comparing costly safety behaviors to (the devalued) rightGS with (the non-devalued) leftGS. For these two contrasts, safety behaviors served as a continuous dependent variable, whereas GS type (rightGS, leftGS, & GS-) and Block served as fixed effects. Noted that the two contrasts were non-orthogonal to each other, thus the reported p-values were Bonferroni-corrected.

2.4.4.2. Cross-phase analyses. Two non-orthogonal planned contrasts were applied. The first contrast examined whether generalization decrement (e.g., a decrease in responding from CS+ to GS of the same category) took place. Costly safety behaviors to the leftCS+ and CS- on the last block of *Costly safety behaviors acquisition* were compared to the leftGS and GS- on the first block of *Generalization test*. A second contrast directly compared whether generalized costly safety behaviors to the rightGS decreased after US devaluation while accounting for generalization decrement. This contrast compared costly safety behaviors to the rightCS+ /rightGS with the leftCS+ /leftGS. For these two contrasts, safety behaviors served as a continuous dependent variable, whereas Trial type (rightCS+/rightGS, leftCS+/leftGS, & CS-/GS-) and Phase served as fixed effects. The reported p-values were Bonferroni-corrected.

2.4.5. Exploratory analyses

Exploratory analyses were included to further explore the effect of trait anxiety or intolerance of uncertainty on responding during *Fear acquisition training, Costly safety behaviors acquisition,* and *Generalization test*. To this end, the aforementioned contrasts and models described in each phase were employed again, with trait anxiety or intolerance of uncertainty added as a continuous variable.

3. Results

Data exclusion criteria were pre-registered at https://osf.io/gvzyn. Overall, two participants were excluded due to not acquiring the correct CS-US contingencies during *Fear acquisition training*. This led to a final sample of 78 participants (M_{age}=21.2, SD=2.8; $Gender_{women}$=72%). In addition, SCRs data from 6 participants were excluded due to technical problem; however, behavioral data from these 6 participants were included for analyses. In sum, 78 participants were included for the analysis of behavioral data, whereas 72 participants were included for SCR analyses. A total of 6.4% of safety behaviors were inaccurate and thus treated as missing data.

3.1. Manipulation check

3.1.1. The acquisition of conditioned fear and costly safety behaviors

Two orthogonal contrasts were applied (Contrast 1: rightCS+ & leftCS+ vs CS-; Contrast 2: rightCS+ vs leftCS+). In brief, participants showed stronger responding to the threat-related CSs compared to the CS- in US expectancy ratings and SCRs during *Fear acquisition training* (Fig. 3A & 3B; all $ps \leq .001$) and costly safety behaviors during *Costly safety behaviors acquisition* (Fig. 3C; $p < .001$). Participants also showed no differences in responding to the two threat-related CSs in all measures (all $ps \geq .268$). Bayesian models further supported these null effects as ≥ 89.77% HDIs fell under the area of ROPE (see Supplementary Materials for the full analysis).

3.1.2. US devaluation

Averaged aversive ratings to the right noise US on the first trial (100 dB or 95 dB) was 66.6 (SD=24.3), while averaged aversive ratings to the devalued right noise US on the last trial (30 dB) was 4.4 (SD=8.4). The decrease in aversiveness was significant, bTrial= −62.16, $p < .001$,

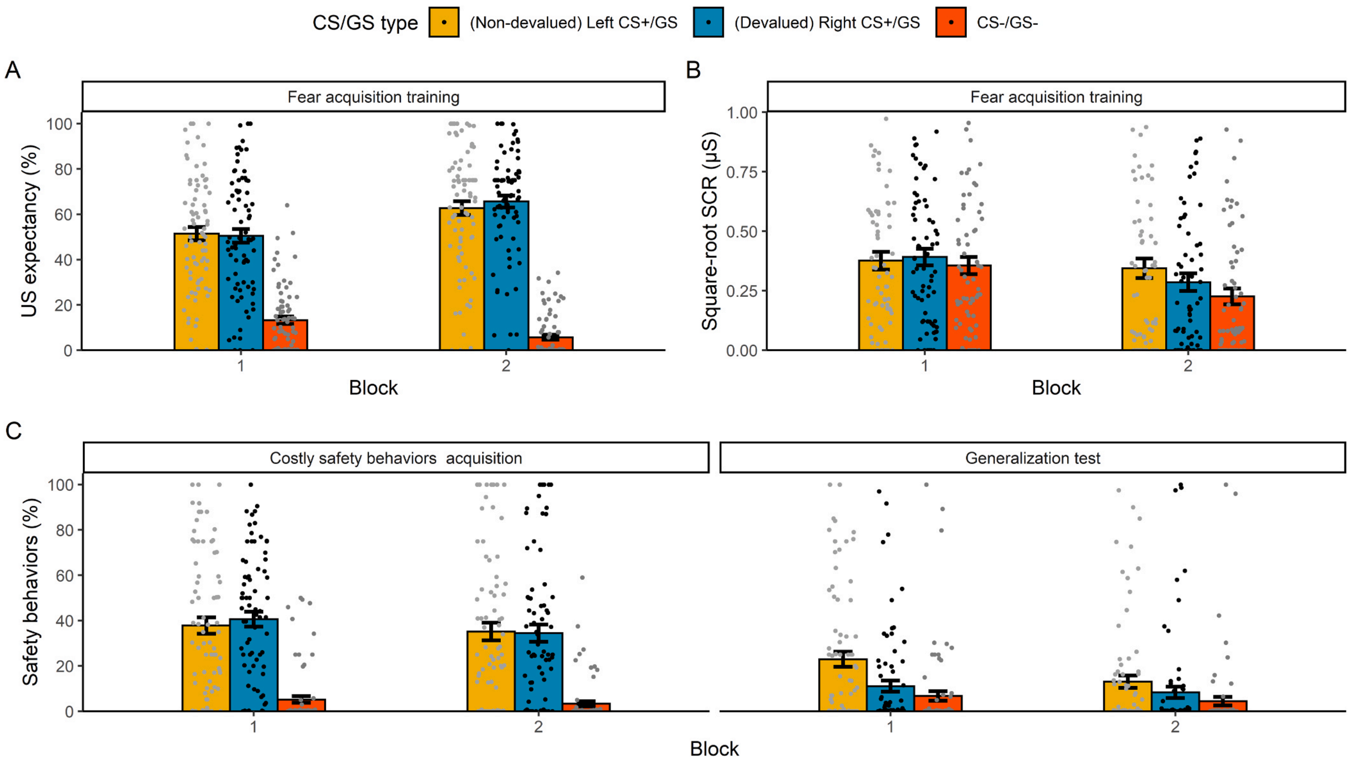

Fig. 3. US expectancy ratings (A) and SCRs (B) during *Fear acquisition training.* Costly safety behaviors during *Costly safety behaviors acquisition* and *Generalization test* (C). Error bars indicate standard error of the mean. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

$R^2 = 0.75$. This suggests that the manipulation of US devaluation was successful.

3.2. Main hypotheses

3.2.1. Generalization test

Data in *Generalization test* were analyzed in linear mixed models with GS type and Block as fixed effects. Fig. 3C shows costly safety behaviors to the GSs during *Generalization test*. Two non-orthogonal contrasts were used: the first contrast (leftGS vs GS-) tested whether participants showed generalized costly safety behaviors. Participants exhibited stronger costly safety behaviors to the leftGS compared to GS-; this differential difference decreased across blocks presumably due to ongoing extinction learning. This pattern was supported by a significant interaction between GS type and Block, bGStype(leftGS vs GS-)*Block= 3.95, *Bonferroni-corrected* $p < .001$, $R^2 = 0.009$.

The second contrast assessed whether US devaluation selectively reduced generalized safety behaviors to GSs that belonged to the same category of the devalued CS+ (i.e., rightGS). Participants exhibited less costly safety behaviors to (the devalued) rightGS compared to (the non-devalued) leftGS, but this difference decreased over blocks presumably due to greater extinction learning to the leftGS. This pattern was supported by a significant interaction between GS type and Block, bGStype (rightGS vs leftGS)*Block= 3.81, *Bonferroni-corrected* $p < .001$, $R^2 = 0.006$.

3.2.2. Cross-phase analyses

The first contrast compared generalization decrement in costly safety behaviors to the leftGS and GS-. Generalization decrement was observed to both the leftGS and GS- (i.e., decreased responding from the leftCS+ to the leftGS but an increase in responding from CS- to GS-). This pattern was supported by a significant interaction between GS type and Phase, bGStype(leftCS+/leftGS vs CS-/GS-)*Phase= 8.91, *Bonferroni-corrected* $p < .001$, $R^2 = 0.021$. Importantly, the second contrast directly compared the change in responding to the rightGS and the leftGS when transiting from acquisition to test. There was seemingly a greater generalization decrement to the rightGS compared to the leftGS, bGS type(leftCS+/leftGS vs rightCS+/rightGS)*Phase= 5.79, *Bonferroni-corrected* $p = .001$, $R^2 = 0.007$. The apparent greater decrease in costly safety behaviors to the rightGS in the transition across the two phases was presumably due to a combination of generalization decrement and an effect of US devaluation. Follow-up analyses revealed that on the last block of *Costly safety behaviors acquisition,* there was no evidence that costly safety behaviors differed between the rightCS+ and the leftCS+ , bGStype(leftCS+ vs rightCS+)= 0.12, *Bonferroni-corrected* $p > .999$, $R^2 < 0.001$ (100% HDIs fell under the area of ROPE in the Bayesian model). However, after US devaluation, costly safety behaviors were significantly weaker to the rightGS compared to the leftGS, bGStype(leftGS vs rightGS)= 5.91, *Bonferroni-corrected* $p < .001$, $R^2 = 0.015$, further suggesting that the difference in responding to the rightGS and the leftGS was not merely due to generalization decrement.

In sum, participants showed generalized safety behaviors to novel but threat-related leftGS. More importantly, participants showed less generalized costly safety behaviors to the rightGS when compared with the leftGS, suggesting that US devaluation decreased generalized safety behaviors. This difference in generalized safety behaviors to the rightGS and the leftGS was not merely due to generalization decrement.

3.3. Exploratory analyses

Trait anxiety scores ranged from 0 to 26 (M=7.7, SD=7.6) whereas intolerance of uncertainty ranged from 31 to 107 (M=60.9, SD=19.8). We only reported significant effects below. See the Supplementary Materials for the detailed full analyses.

During the acquisition phases, only effects involving intolerance of uncertainty reached significance. Specifically, when controlling for the effect of trait anxiety on the differential US expectancy ratings to the CSs, intolerance of uncertainty was significantly associated with a decrease in differential US expectancy ratings to the CSs averaged across blocks, bCStype(rightCS+ & leftCS+ vs CS-)*IU= −0.10, p = .007, R^2 = 0.005, (see Fig. S1 & S2 in the Supplementary Materials). Bayes factor (BF_{10}) for this contrast was 33.50; according to Jeffreys (1961), a Bayes factor of 10 or larger reflects strong evidence of an effect. However, follow-up analyses showed that this apparent decrease in differential US expectancy ratings to the CSs was driven by a decrease in US expectancies to the two CS+s, bIU= −0.34, p = .036, R^2 = 0.018, but there was no association between intolerance of uncertainty and US expectancies to the CS-, bIU< −0.001, p = .999, R^2 < 0.001. Similarly, during *Costly safety behaviors acquisition* (Fig, 4), intolerance of uncertainty was associated with a decrease in differential costly safety behaviors to the CSs averaged across blocks when controlling for trait anxiety, bCStype (rightCS+ & leftCS+ vs CS-)*IU= −0.11, p = .003, R^2 = 0.005, BF_{10} = 22.66. However, follow-up analyses suggested no evidence that intolerance of uncertainty was associated with changes to safety behaviors in the CS+ or CS- (all ps ≥.125).

During *Generalization test* (Fig. 4), only effects involving trait anxiety reached significance. When directly comparing costly safety behaviors to (the devalued) rightGS with the (non-devalued) leftGS, trait anxiety was significantly associated with a decrease in differential costly safety behaviors to these two GSs averaged across blocks when controlling for the effect of intolerance of uncertainty on differential responding to the same GSs, bGStype(rightGS vs leftGS)*Anxiety= −0.24, *Bonferroni-corrected* p = .045, R^2 = 0.003, BF_{10} = 10.94. There was no evidence that intolerance of uncertainty had any effect on the same contrast, bGS type (right GS vs left GS)*IU = −0.037, SE = 0.039, *Bonferroni-corrected* p = .678, R^2 = 0.001. However, the BF_{10} for this contrast was 5.48, suggesting that there was substantial evidence that an increase in intolerance of uncertainty was associated with impaired differential responding to the rightGS and leftGS.

In sum, exploratory analyses suggested that intolerance of uncertainty was associated with impaired discriminative responding to the CSs, whereas trait anxiety was associated with a persistence in generalized costly safety behaviors to the rightGS after US devaluation.

4. Discussion

Using a fear and avoidance conditioning procedure, the current study examined whether devaluing US aversiveness led to a decrease in costly safety behaviors to novel generalization stimuli. As predicted, participants exhibited stronger costly safety behaviors to GSs that belonged to the same category of a threat-related CS+ (i.e., leftGS) compared to safety-related GS- in *Generalization test*, despite none of these stimuli had a direct history of predicting the presence or absence of a US. This suggested that costly safety behaviors acquired to the threat-related leftCS+ generalized selectively to novel stimuli of the same category. This pattern is consistent with past studies that found generalized safety behaviors to novel stimuli that were conceptually related to the CS+ (e.g., Boyle et al., 2016; Dymond et al., 2014; Kloos et al., 2022).

A key finding is that after devaluing the aversiveness of the right noise US, participants exhibited limited generalized costly safety behaviors to the rightGS (novel stimuli of the same category of the devalued rightCS+) compared to the leftGS (novel stimuli of the same category of the non-devalued leftCS+) in *Generalization test*. Two additional patterns further supported that US devaluation decreases costly safety behaviors generalization. First, participants showed similar levels of costly safety behaviors to both threat-related CS+s during *Safety behaviors acquisition*, indicating no pre-existing difference in responding between the two CS+s. However, after devaluing the aversiveness of the right noise US, costly safety behaviors reduced selectively to the rightGS, a set of novel stimuli that belonged to the same category of the (now) devalued rightCS+ . Second, the cross-phase analysis suggests that the decrease in costly safety behaviors to the rightGS from training to test was not solely due to generalization decrement (i.e., a decrease in responding due to generalization). The decrease in costly safety behaviors from the rightCS+ to rightGS was significantly greater than from the leftCS+ to the leftGS; this significant decrease in responding to the rightGS was largely attributed to US devaluation, assuming that the levels of generalization decrement for the rightGS and leftGS were similar. Thus, the current finding suggests that US devaluation reduces

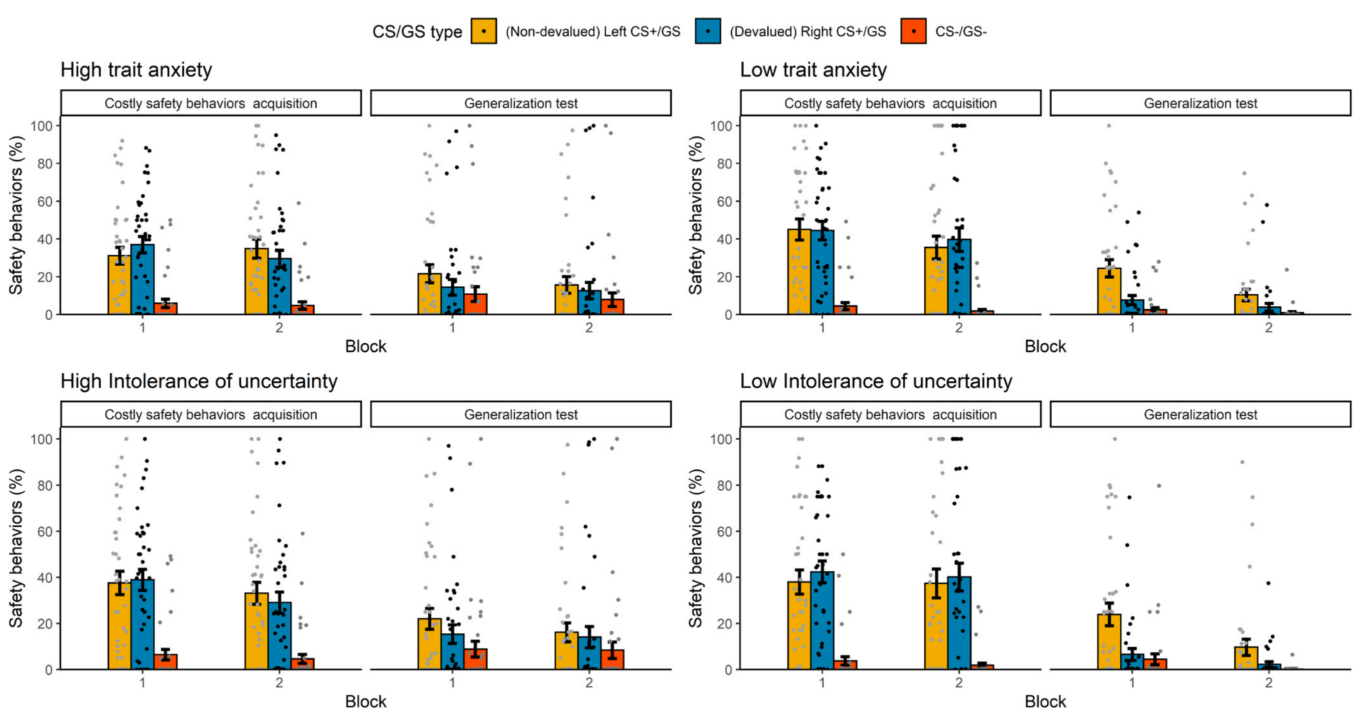

Fig. 4. The effect of trait anxiety (Top panel) and intolerance of uncertainty (Bottom panel) on costly safety behaviors during *Costly safety behaviors acquisition* and *Generalization test*. Trait anxiety and intolerance of uncertainty were median split for descriptive purpose. Error bars indicate standard error of the mean. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

generalized costly safety behaviors to novel stimuli that categorically resemble a devalued CS+ .

4.1. Theoretical and methodological considerations

The reduction in costly safety behaviors generalization after US devaluation is consistent with cognitive accounts of conditioning. First, these accounts proposed that after repeated CS-US pairings, a propositional belief that CS+ is threatening is formed (e.g., De Houwer, 2009; Mitchell et al., 2009). After US devaluation, a neutral US is expected to follow the CS+ , thus threat belief to the CS+ is reevaluated and hence markedly reduced (e.g., Davey, 1989, 1992). Second, safety behaviors to the CS+ is thought to involve mental processes that compare the outcomes of using safety behaviors or not (Lovibond, 2006). If threat belief to the CS+ is established, one is likely to engage in safety behaviors to prevent US onset. Likewise, this model explains how safety behaviors generalize to novel exemplars that belong to the same category of CS+ : a threat is expected upon GS presentation due to categorical generalization, thus driving one to engage in safety behaviors. These cognitive accounts jointly explain how US devaluation selectively decreases generalized safety behaviors to the rightGS (i.e., the devalued threat category): the rightGS evokes generalized threat beliefs similar to the rightCS+ due to categorical generalization. Therefore, devaluing the right noise US not only attenuated threat belief to the rightCS+ , but also to the rightGS (i.e., participants expected a non-aversive US during rightGS presentation), thus leading to a reduction in costly safety behaviors to the rightGS. In contrast, threat belief to the leftCS+ was not devalued (and thus the generalized threat beliefs to the leftGS), hence participants exhibited stronger generalized costly safety behaviors to the leftGS compared to the devalued rightGS.

Most laboratory studies (e.g., Cameron et al., 2015; van Meurs et al., 2014) that examined generalization of safety behaviors typically measured safety behaviors dichotomously (i.e., to use safety behaviors or not). Thus, generalized safety behaviors were measured by averaging multiple trials of the same GS. Considering that the GSs are typically presented under extinction, some studies partially reinforced CS+ during test (e.g., Boyle et al., 2016; Hunt et al., 2019). However, this may create new safety learning to the GSs, thus artificially reducing generalized responses in test (see Wong & Pittig, 2022b). Another way to minimize the confounding effects of ongoing extinction learning is to limit the number of test trials. This was made possible in the current study by assessing safety behaviors on a continuum, which was capable to reflect generalized responses even when each GS was presented only once. However, extinction learning was observed despite the limited number of test trials, as evident in the reduced responding in the last block in *Generalization test*.

4.2. Risk factors

This study also explored how risk factors of clinical anxiety, such as trait anxiety and intolerance of uncertainty, had any effect on the acquisition and generalization data. Trait anxiety was associated with persistent generalized costly safety behaviors even when the expected outcome had been devalued, while controlling for intolerance of uncertainty. This finding is consistent with preliminary evidence highlighting trait anxiety is linked to persistent avoidance (Pittig et al., 2014; Wong & Pittig, 2023). On the other hand, intolerance of uncertainty was associated with impaired acquisition of discriminative fear and safety behaviors to the CSs, independently of trait anxiety. Although these patterns were apparently consistent with some previous findings (e.g., Kanen et al., 2021; Sjouwerman et al., 2020), they were not driven by an increase in responding to the safety CS-. For the generalization data, despite the frequentist model suggested no evidence that intolerance of uncertainty was associated with persistent generalized costly safety behaviors after US devaluation, the Bayes model suggested “substantial” evidence for this effect. There are two potential reasons to account for this discrepancy between the two analyses. First, we used a quite conservative method to correct for the non-orthogonal contrasts in the frequentist model. Second, the sample size might be too small to detect any potential effect of intolerance of uncertainty in persistent generalized safety behaviors (c.f. Zuj et al., 2020). Taken together, despite research on individual differences in safety behaviors is still emerging (see Wong et al., 2023), replication is required with sufficient sample size (Morriss et al., 2021).

4.3. Clinical implications

The present findings modeled the reduction of maladaptive safety behaviors to novel stimuli that resembled the feared stimulus. For example, the current findings suggest that devaluing the aversiveness of a perceived threat (e.g., getting judged negatively) can reduce maladaptive safety behaviors (e.g., over-rehearsal of a speech) across various social situations (e.g., public speech, a group conversation). At face value, US devaluation may seem unfeasible (nor ethical) for clinical translation as it involves directly exposing the feared outcome or trauma (e.g., the US) to clients. However, there are interventions that utilize US devaluation that are more feasible for clinical translation. For instance, imagery rescripting has been found to effectively reduce fear in laboratory studies (e.g., Dibbets et al., 2012; Woelk et al., 2022) and also symptoms in clinical anxiety (e.g., Frets et al., 2014; Norton & Abbott, 2016) and PTSD (Morina et al., 2017). Of note, in the current study, the same US was used for training and devaluation for all participants for practical reasons. However, in a clinical context, US devaluation requires recalling a US that is individually meaningful and disorder-specific (Kip et al., 2023; Morina et al., 2017). Nonetheless, there are still challenges on translating the current findings to a clinical context. For instance, it remains unclear whether devaluing a single US is sufficient to decrease safety behaviors to the CS or GSs if multiple feared outcomes are expected. For instance, a socially anxious individual may avoid social interaction because of multiple feared outcomes, such as social rejection and getting humiliated. It is unknown whether merely devaluing one of the feared outcomes (e.g., social rejection) can effectively reduce safety behaviors to this or other similar situations. Furthermore, US devaluation may not be an effective intervention for disorders that do not involve a specific feared outcome, such as generalized anxiety disorder (but see Ovanessian et al., 2019). In sum, the current findings provide a laboratory model suggesting that US devaluation can decrease maladaptive generalized safety behaviors. Future studies are warranted for addressing the aforementioned challenges in translating laboratory findings into clinical practice.

4.4. Limitations

One limitation of the current study is that the US devaluation procedure and costly safety behaviors generalization were assessed on the same day. As a result, it remains unclear whether US devaluation has a long-term effect in reducing the generalization of safety behaviors, which can more relevantly apply to a clinical context. A second limitation is that the CSs and the GSs were too simple and hence not necessarily ecologically valid, unlike the acquisition and generalization of complex fear memories in real life (Beckers et al., 2013). Future studies can incorporate complex stimuli with multimodal input (e.g., both visual and auditory), for instance, using a trauma film as both the CS and US (Holmes & Bourne, 2008), or asking participants to imagine the US via mental imagery (Mertens et al., 2020).

4.5. Conclusion

In conclusion, the current study has demonstrated that devaluing US aversiveness consequently led to a decrease in generalized costly safety behaviors to novel stimuli that shared the same category membership as the devalued CS+ . This provides a laboratory model for examining interventions that utilize US devaluation to reduce the spread of maladaptive safety behaviors in clinical anxiety. Exploratory analyses suggest that trait anxiety, but not intolerance of uncertainty, was associated with a persistence in generalized costly safety behaviors to the devalued GSs. Future studies can examine whether US devaluation remains effective in reducing (the generalization of) maladaptive safety behaviors in the long-term in multiple-day paradigms.

Ethical approval

This study was approved by the Ethics Committee of the Erasmus School of Social and Behavioural Sciences (ETH2122–0453) in accordance to the Declaration of Helsinki.

Funding

This study was not funded by any external parties.

CRediT authorship contribution statement

Minita Franzen: Conceptualization, Writing – review & editing. **Alex H.K. Wong:** Data curation, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – original draft. **Matthias J. Wieser:** Conceptualization, 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.

We further confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

Data Availability

I have shared the link to my data on OSF.

Acknowledgement

The authors would like to thank Celina Steffens, Flora Steger, Gina Lehmann, and Henry Sticher for their help in data collection and processing.

Preregistration

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

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.janxdis.2024.102847.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

Ball, T. M., & Gunaydin, L. A. (2022). Measuring maladaptive avoidance: from animal models to clinical anxiety. *Neuropsychopharmacology, 47*(5), 978–986. https://doi.org/10.1038/s41386-021-01263-4

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

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

Blakey, S. M., & Abramowitz, J. S. (2016). The effects of safety behaviors during exposure therapy for anxiety: critical analysis from an inhibitory learning perspective. *Clinical Psychology Review, 49*, 1–15. https://doi.org/10.1016/j.cpr.2016.07.002

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

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

Buhr, K., & Dugas, M. J. (2002). The intolerance of uncertainty scale: psychometric properties of the English version. *Behaviour Research and Therapy, 40*(8), 931–945. https://doi.org/10.1016/S0005-7967(01)00092-4

Cameron, G., Schlund, M. W., & Dymond, S. (2015). Generalization of socially transmitted and instructed avoidance. *Froniters in Behavioral Neuroscience, 9*. https://doi.org/10.3389/fnbeh.2015.00159

Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., & Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *Journal of Anxiety Disorders, 26*(3), 468–479. https://doi.org/10.1016/j.janxdis.2012.01.011

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

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

Davey, G. C. L. (1989). UCS revaluation and conditioning models of acquired fears. *Behaviour Research and Therapy, 27*(5), 521–528. https://doi.org/10.1016/0005-7967(89)90086-7

Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: a review and synthesis of the literature. *Advances in Behaviour Research and Therapy, 14*(1), 29–66. https://doi.org/10.1016/0146-6402(92)90010-L

Davey, G. C. L., & McKenna, I. (1983). The Effects of postconditioning revaluation of CS $_1$ and UCS following pavlovian second-order electrodermal conditioning in humans. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 35B*(2), 125–133. https://doi.org/10.1080/14640748308400899

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Kleine, R. A., Hutschemaekers, M. H. M., Hendriks, G. J., Kampman, M., Papalini, S., Van Minnen, A., & Vervliet, B. (2023). Impaired action-safety learning and excessive relief during avoidance in patients with anxiety disorders. *Journal of Anxiety Disorders, 96*, Article 102698. https://doi.org/10.1016/j.janxdis.2023.102698

Dibbets, P., Poort, H., & Arntz, A. (2012). Adding imagery rescripting during extinction leads to less ABA renewal. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(1), 614–624. https://doi.org/10.1016/j.jbtep.2011.08.006

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized, 1821 *Psychological Science, 25*(9), 1816–1821. https://doi.org/10.1177/0956797614535401.

Dymond, S. (2019). Overcoming avoidance in anxiety disorders: the contributions of Pavlovian and operant avoidance extinction methods. *Neuroscience & Biobehavioral Reviews, 98*, 61–70. https://doi.org/10.1016/j.neubiorev.2019.01.007

Dymond, S., Schlund, M. W., Roche, B., & Whelan, R. (2014). The spread of fear: symbolic generalization mediates graded threat-avoidance in specific phobia. *Quarterly Journal of Experimental Psychology, 67*(2), 247–259. https://doi.org/10.1080/17470218.2013.800124

Dymond, S., Schlund, M. W., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: Symbolic generalization of human avoidance learning. *Behaviour Research and Therapy, 49*(10), 614–621. https://doi.org/10.1016/j.brat.2011.06.007

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

Flores, A., López, F. J., Vervliet, B., & Cobos, P. L. (2020). Prospective intolerance of uncertainty is associated with maladaptive temporal distribution of avoidance responses: An extension of Flores, López, Vervliet, and Cobos (2018). *Journal of Behavior Therapy and Experimental Psychiatry, 68*, Article 101527. https://doi.org/10.1016/j.jbtep.2019.101527

Fraunfelter, L., Gerdes, A. B. M., & Alpers, G. W. (2022). Fear one, fear them all: a systematic review and meta-analysis of fear generalization in pathological anxiety. *Neuroscience & Biobehavioral Reviews, 139*, Article 104707. https://doi.org/10.1016/j.neubiorev.2022.104707

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

Frets, Petra G., Kevenaar, C., & van der Heiden, C. (2014). Imagery rescripting as a stand-alone treatment for patients with social phobia: a case series. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 160–169. https://doi.org/10.1016/j.jbtep.2013.09.006

Gershuny, B. S., & Sher, K. J. (1998). The relation between personality and anxiety: Findings from a 3-year prospective study. *Journal of Abnormal Psychology, 107*(2), 252–262. https://doi.org/10.1037/0021-843X.107.2.252

Glogan, E., Gatzounis, R., Meulders, M., & Meulders, A. (2020). Generalization of instrumentally acquired pain-related avoidance to novel but similar movements using a robotic arm-reaching paradigm. *Behaviour Research and Therapy, 124*, Article 103525. https://doi.org/10.1016/j.brat.2019.103525

Helbig-Lang, S., & Petermann, F. (2010). Tolerate or eliminate? a systematic review on the effects of safety behaviors across anxiety disorders. *Clinical Psychology: Science and Practice, 17*(3), 218–233. https://doi.org/10.1111/j.1468-2850.2010.01213.x

Holmes, E. A., & Bourne, C. (2008). Inducing and modulating intrusive emotional memories: a review of the trauma film paradigm. *Acta Psychologica, 127*(3), 553–566. https://doi.org/10.1016/j.actpsy.2007.11.002

Hosoba, T., Iwanaga, M., & Seiwa, H. (2001). The effect of UCS inflation and deflation procedures on 'fear' conditioning. *Behaviour Research and Therapy, 39*(4), 465–475. https://doi.org/10.1016/S0005-7967(00)00025-5

Hunt, C., Cooper, S. E., Hartnell, M. P., & Lissek, S. (2019). Anxiety sensitivity and intolerance of uncertainty facilitate associations between generalized Pavlovian fear and maladaptive avoidance decisions. *Journal of Abnormal Psychology, 128*(4), 315–326. https://doi.org/10.1037/abn0000422

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

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics, 44*, 1086–1105. https://doi.org/10.1080/02664763.2016.1193725

Jeffreys, H. (1961). *Theory of Probability* (third ed.). Oxford University Press.

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (2000). Predicting anxiety and depression from personality: Is there a synergistic effect of neuroticism and extraversion? *Journal of Abnormal Psychology, 109*(1), 145–149. https://doi.org/10.1037/0021-843X.109.1.145

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., Sponheim, S. R., & Lissek, S. (2017). Neural Substrates of Overgeneralized Conditioned Fear in PTSD. *American Journal of Psychiatry, 174*(2), 125–134. https://doi.org/10.1176/appi.ajp.2016.15121549

Kanen, J. W., Arntz, F. E., Yellowlees, R., Christmas, D. M., Price, A., Apergis-Schoute, A. M., Sahakian, B. J., Cardinal, R. N., & Robbins, T. W. (2021). Effect of Tryptophan Depletion on Conditioned Threat Memory Expression: Role of Intolerance of Uncertainty. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(5), 590–598. https://doi.org/10.1016/j.bpsc.2020.12.012

Kip, A., Schoppe, L., Arntz, A., & Morina, N. (2023). Efficacy of imagery rescripting in treating mental disorders associated with aversive memories – An updated meta-analysis. *Journal of Anxiety Disorders, 99*, Article 102772. https://doi.org/10.1016/j.janxdis.2023.102772

Kloos, T., van Vliet, C., Riecke, J., & Meulders, A. (2022). Indoor or outdoor? generalization of costly pain-related avoidance behavior to conceptually related contexts. *The Journal of Pain, 23*(4), 657–668. https://doi.org/10.1016/j.jpain.2021.10.010

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A tuotial with R, JAGS, and Stan* (Second ed..,). Academic Press,.

Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Krypotos, A.-M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy, 111*, 99–105. https://doi.org/10.1016/j.brat.2018.10.011

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*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry, 75*(11), 909–915. https://doi.org/10.1016/j.biopsych.2013.07.025

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *Armerican Journal of Psychiatry, 167*(1), 47–55. https://doi.org/10.1176/appi.ajp.2009.09030410

Lovibond, P. (2006). 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* (pp. 117–132). American Psychological Association. https://doi.org/10.1037/11474-006.

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology, 107*(3), 520–526. https://doi.org/10.1037/0021-843X.107.3.520

Lovibond, P. F., Mitchell, C. J., Minard, E., Brady, A., & Menzies, R. G. (2009). Safety behaviours preserve threat beliefs: Protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy, 47*(8), 716–720. https://doi.org/10.1016/j.brat.2009.04.013

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541. https://joss.theoj.org/papers/10.21105/joss.01541.

Mendlowicz, M. V., & Stein, M. B. (2000). Quality of life in individuals with anxiety disorders. *American Journal of Psychiatry, 157*(5), 669–682. https://doi.org/10.1176/appi.ajp.157.5.669

Mertens, G., Krypotos, A.-M., & Engelhard, I. M. (2020). A review on mental imagery in fear conditioning research 100 years since the 'Little Albert' study. *Behaviour Research and Therapy, 126*, Article 103556. https://doi.org/10.1016/j.brat.2020.103556

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*(2), 183–198. https://doi.org/10.1017/S0140525×09000855

Morina, N., Lancee, J., & Arntz, A. (2017). Imagery rescripting as a clinical intervention for aversive memories: a meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 55*, 6–15. https://doi.org/10.1016/j.jbtep.2016.11.003

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

Norton, A. R., & Abbott, M. J. (2016). The efficacy of imagery rescripting compared to cognitive restructuring for social anxiety disorder. *Journal of Anxiety Disorders, 40*, 18–28. https://doi.org/10.1016/j.janxdis.2016.03.009

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

Ovanessian, M. M., Koerner, N., Antony, M. M., & Dugas, M. J. (2019). A preliminary test of the therapeutic potential of written exposure with rescripting for generalized anxiety disorder. *Journal of Experimental Psychopathology, 10*(2). https://doi.org/10.1177/2043808719841529

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

Pittig, A., Schulz, A. R., Craske, M. G., & Alpers, G. W. (2014). Acquisition of behavioral avoidance: task-irrelevant conditioned stimuli trigger costly decisions. *Journal of Abnormal Psychology, 123*(2), 314–329. https://doi.org/10.1037/a0036136

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

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

R Core Team. (2023). *R: A Language And Environment For Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

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

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

Scheveneels, S., Boddez, Y., & Hermans, D. (2021). Predicting clinical outcomes via human fear conditioning: a narrative review. *Behaviour Research and Therapy, 142*, Article 103870. https://doi.org/10.1016/j.brat.2021.103870

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

Sep, M. S. C., Steenmeijer, A., & Kennis, M. (2019). The relation between anxious personality traits and fear generalization in healthy subjects: a systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 107*, 320–328. https://doi.org/10.1016/j.neubiorev.2019.09.029

Sexton, K. A., Norton, P. J., Walker, J. R., & Norton, G. R. (2003). Hierarchical model of generalized and specific vulnerabilities in anxiety. *Cognitive Behaviour Therapy, 32*(2), 82–94. https://doi.org/10.1080/16506070302321

Sjouwerman, R., Scharfenort, R., & Lonsdorf, T. B. (2020). Individual differences in fear acquisition: multivariate analyses of different emotional negativity scales, physiological responding, subjective measures, and neural activation. *Scientific Reports, 10*(1), Article 15283. https://doi.org/10.1038/s41598-020-72007-5

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of*

Experimental Psychology: Human Learning and Memory, 6(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

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

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

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

Woelk, M., Krans, J., Raes, F., Vervliet, B., & Hagenaars, M. A. (2022). Imagery rescripting versus extinction: distinct and combined effects on expectancy and revaluation learning. *Clinical Psychological Science, 10*(4), 622–639. https://doi.org/10.1177/21677026211055169

Wong, A. H. K., & Beckers, T. (2021). Trait anxiety is associated with reduced typicality asymmetry in fear generalization. *Behaviour Research and Therapy, 138*, Article 103802. https://doi.org/10.1016/j.brat.2021.103802

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

Wong, A. H. K., & Pittig, A. (2022b). Threat belief determines the degree of costly safety behaviors: assessing rule-based generalization of safety behaviors with a dimensional measure of avoidance. *Behaviour Research and Therapy, 156*, Article 104158. https://doi.org/10.1016/j.brat.2022.104158

Wong, A. H. K., & Pittig, A. (2023). Reducing avoidance of learnt fear: extinction of an imminent threat signal partly decreases costly avoidance to a distal threat signal, 204380872311751 *Journal of Experimental Psychopathology, 14*(2). https://doi.org/10.1177/20438087231175147.

Wong, A. H. K., Lee, J. C., Engelke, P., & Pittig, A. (2023). Reduction of costly safety behaviors after extinction with a generalization stimulus is determined by individual differences in generalization rules. *Behaviour Research and Therapy, 160*, Article 104233. https://doi.org/10.1016/j.brat.2022.104233

Wong, A. H. K., Aslanidou, A., Malbec, M., Pittig, A., Wieser, M. J., & Andreatta, M. (2023). A systematic review of the inter-individual differences in avoidance learning. *Collabra: Psychology, 9*(1), Article 77856. https://doi.org/10.1525/collabra.77856

Zuj, D. V., Xia, W., Lloyd, K., Vervliet, B., & Dymond, S. (2020). Negative reinforcement rate and persistent avoidance following response-prevention extinction. *Behaviour Research and Therapy, 133*, Article 103711. https://doi.org/10.1016/j.brat.2020.103711