Behaviour Research and Therapy 156 (2022) 104158

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Threat belief determines the degree of costly safety behavior: Assessing rule-based generalization of safety behavior with a dimensional measure of avoidance

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

[a] Department of Psychology (Biological Psychology, Clinical Psychology, and Psychotherapy), University of Würzburg, Marcusstraße 9-11, 97070, Würzburg, Germany
[b] Department of Psychology, Educational Sciences, and Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062, PA, Rotterdam, the Netherlands
[c] Translational Psychotherapy, Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 49b, 91052, Erlangen, Germany

ARTICLE INFO

Keywords:
Fear conditioning
Safety behavior
Fear generalization
Rule-based generalization
Threat beliefs
Extinction

ABSTRACT

Excessive generalization of safety behavior to innocuous stimuli that resemble a feared stimulus is oftentimes pathological especially with inflicted impairments. Safety behavior is conventionally assessed dichotomously, requiring multiple presentations of each test stimulus for assessing the proportion of safety behavior executed. Thus, the generalization gradient confounds with ongoing extinction learning during non-reinforced test trials. The present study employed a recently developed dimensional measure of avoidance to examine the extent of safety behavior generalization. We found that a dimensional measure of avoidance was able to assess the generalization gradients of safety behavior even when each test stimulus was presented once, thus minimizing the effect of ongoing extinction learning. Of equal importance is whether higher-order cognitive processes shape generalization of safety behavior. We found a range of distinct generalization gradients in safety behavior, which were highly consistent with participants' verbally reported relational rules. This rule-based generalization parallels to how clinically anxious individuals develop different threat beliefs after trauma exposure, and models how these distinct threat beliefs determine the extent of safety behavior engagement.

1. Introduction

Stimuli that signal threat evoke fear, which in turn motivates avoidance. Avoidance is often used as an umbrella term for distinct defensive responses, including safety behavior. Safety behavior is initiated in the presence of a fear-related stimulus or situation that aims to prevent an aversive outcome (Krypotos, Vervliet, & Engelhard, 2018; Pittig, Wong, Glück, & Boschet, 2020). Thus, safety behavior is regarded as an adaptive behavior given it prevents harm. Safety behavior in anxiety-related disorders is, however, often maladaptive given that it persists in the absence of realistic threat, impairs daily functioning, and preserves pathological threat beliefs (Mendlowicz & Stein, 2000; Olatunji, Cisler, & Tolin, 2007; Salkovskis, 1991). For instance, an individual with clinical social anxiety may have developed fear of small talk after experiencing negative feedback in one. Subsequently, this individual may speak quietly to avoid appearing anxious, therefore preventing an unrealistic perceived threat, such as being negatively criticized (Kim, 2005). Clinically, safety behavior rarely confines to a single learned fear stimulus, but also generalizes to a broad range of stimuli that resemble the original fear-related stimulus (e.g., McManus, Sacadura, & Clark, 2008; Wells et al., 1995). For example, a socially anxious individual may perform similar safety behavior (e.g., speaking quietly) in other feared situations, such as public speeches or job interviews. Excessive generalization of safety behavior to innocuous stimuli or situations is considered a pathological feature in anxiety-related disorders when it becomes unnecessary and causes impairments.

The generalization of fear and safety behavior can be examined in a highly controlled environment in the laboratory via fear and avoidance conditioning models. This well-established model usually combines both Pavlovian fear conditioning and avoidance conditioning. During Pavlovian fear conditioning, conditioned fear is acquired to a formerly neutral conditioned stimulus (CS+) after repeated pairings with an aversive unconditioned stimulus (US). In a following avoidance conditioning phase, performing a designated response during CS+ presentation prevents the upcoming US. This response is regarded as US-

* Corresponding author. 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.brat.2022.104158
Received 10 November 2021; Received in revised form 20 June 2022; Accepted 30 June 2022
Available online 5 July 2022
0005-7967/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

avoidance given it allows one to avoid US occurrence, but does not necessarily terminate the CS (see Krypotos, Effting, Kindt, & Beckers, 2015; Pittig et al., 2020).

Only a few laboratory studies investigated how newly acquired safety behaviors generalize to other stimuli. For example, van Meurs, Wiggert, Wicker, and Lissek (2014) pioneered a laboratory paradigm to examine generalization of costly, fear-related US-avoidance. After differential Pavlovian and avoidance conditioning to a circle CS+ and another safe circle stimulus (CS-) of different sizes, participants were presented with generalization stimuli (GSs) of different sizes intermediate of the CSs during a generalization test. The generalization test consisted of alternating Pavlovian and avoidance trials: the former measured conditioned fear to the stimuli whereas the latter measured US-avoidance responses to the stimuli. Of note, executing US-avoidance resulted in the omission of a competing reward, thus rendering US-avoidance costly. This manipulation arguably taps into the pathological domain of US-avoidance, as it is often costly in anxiety-related disorders (Pittig, Boschet, Glück, & Schneider, 2021). Results showed a generalization decrement from the CS+ to GSs more perceptually dissimilar to the CS+ in both conditioned fear and US-avoidance. In addition, conditioned fear to the GSs positively correlated with US-avoidance to the same stimulus, suggesting that generalization of Pavlovian fear is positively associated with the generalization of US-avoidance. Other studies also found that generalization gradient of US-avoidance as a function of perceptual similarity to the CS+ (Arnaudova, Krypotos, Effting, Kindt, & Beckers, 2017; Hunt, Cooper, Hartnell, & Lissek, 2019; Lommen, Engelhard, & van den Hout, 2010). US-avoidance also generalizes to novel GSs that conceptually resemble the CS+ (Boyle, Roche, Dymond, & Hermans, 2016; Dymond et al., 2011, 2014). Collectively, laboratory studies have shown that US-avoidance can generalize via a perceptual or a conceptual pathway.

A common feature of the aforementioned studies is that US-avoidance was measured dichotomously, that is, US-avoidance could either be executed or not. Therefore, the same GS had to be presented multiple times to examine the generalization gradient of US-avoidance: the proportion of US-avoidance across trials indicated the degree of US-avoidance generalization. One caveat is that most stimuli were presented under extinction during the generalization test, thereby the proportion of US-avoidance over multiple GS trials may be confounded with ongoing extinction learning, potentially reducing the overall amount of US-avoidance (see also Pittig & Wong, 2021). Although some studies (Boyle et al., 2016; Dymond et al., 2011; van Meurs et al., 2014) reduced extinction learning by continuing to reinforce the CS+ in test (steady-state generalization testing; Blough, 1969; Honig & Urcuioli, 1981), this also induced additional safety learning to the GSs in test (i.e., artificially reducing the breadth of generalization). Thereby, examining the generalization of US-avoidance by relying on the proportion of US-avoidance over multiple trials may confound with extinction learning or additional safety learning, biasing the generalization gradient in test.

One way to minimize the aforementioned problem is the usage of a non-dichotomous measurement for US-avoidance. We have recently developed a dimensional assessment for US-avoidance, which participants could engage in US-avoidance on a continuous scale (from 0% to 100%) on each trial (Wong & Pittig, 2021). The extent of US-avoidance engagement was negatively proportionate to US occurrence and the amount of competing reward. For instance, if a participant chose a 70% of US-avoidance during CS+ presentation, there would be a 70% chance of preventing an upcoming US at the cost of missing out 70% of the competing reward. This dimensional assessment of US-avoidance was developed to measure partial engagement in costly safety behavior, which is commonly seen in anxiety-related disorders (see Krypotos et al., 2018; Telch & Lancaster, 2012). However, it also seems fit to measure the generalization of US-avoidance given it does not require the need of averaging across US-avoidance responses across multiple test trials, thereby minimizing the confounding effect of extinction learning to the test stimuli.

The current study, therefore, sought to examine the generalization of costly US-avoidance using the newly developed dimensional measure of avoidance. Participants were first trained in a single-cue conditioning procedure, in which a single CS paired with an US was presented multiple times in the training phases. After acquiring conditioned fear and costly US-avoidance to the CS+, a range of GSs that varied in similarity to the CS+ along a perceptual dimension were presented. Conditioned fear, and most importantly, costly US-avoidance to these GSs were assessed on a single trial.

This study also took advantage of recent developments in the literature to examine the effect of explicit reasoning processes on fear generalization. Recent work has found that participants reported inferring and adopting different relational rules (rules that are inferred based on the relation of physical differences between stimuli) in stimulus generalization, namely rule-based generalization (Lee, Hayes, & Lovibond, 2018; Wong & Lovibond, 2017, 2018). For instance, some participants reported responding based on perceptual similarity between the GSs and the CS+ (e.g., similarity rule) whereas other participants reported responding less to stimuli towards one end of the stimulus dimension while responding more to stimuli towards the other extreme end of the dimension (e.g., linear rule). Importantly, the reported rules were highly consistent with the shape of the individual generalization gradients. These results thereby suggest that inferred rules guide fear generalization, parallel to how clinically anxious individuals form different threat beliefs after trauma exposure. Thus, we categorized participants into different groups in accordance with their reported rules. We expected to replicate findings of a high consistency between the reported rules and the shape of Pavlovian generalization gradients, but more importantly, to examine whether the gradients of US-avoidance generalization also align with the reported rules. Summarized, the aim of the current study was two-fold: first, we sought to examine the generalization of US-avoidance with each GS presented once, with a newly developed dimensional measurement of avoidance. Second, we sought to examine whether the different relational rules (similarity or linear) align with the generalization gradient of US-avoidance and conditioned fear.

2. Method

Preregistration is available at https://osf.io/a79r8 and the data are available at https://osf.io/62qxy/.

2.1. Participants

Undergraduate students or residents from Würzburg were recruited as participants and received either partial course credit or 9€ for participation. We used the same recruitment procedure as in a previous study (Wong, Glück, Boschet, & Engelke, 2020), in which we stopped recruitment until all rule groups had at least 20 participants. This recruitment strategy allowed appropriate sample size in each rule group (e.g., similarity, linear; see *Questionnaire Coding* in *Results*). This led to a total recruitment of 83 participants. The Ethics Committee of the Institute of Psychology at the University of Würzburg approved the study in accordance to the Declaration of Helsinki.

2.2. Apparatus and materials

The stimulus dimension consisted of nine yellow squares with black outline (9.5 cm × 9.5 cm) containing a black dot that varied in position from the left to the right. This stimulus dimension was employed to minimize intensity biases (for details, see Lee et al., 2018; Wong & Lovibond, 2017). These stimuli were labelled A (with the dot at the left-most) to I (with the dot at the right-most), by adjusting the dot position by 0.8 cm from one stimulus to the next (see Fig. 1A). Stimulus E, the stimulus with the black dot in the centre of the box, served as the

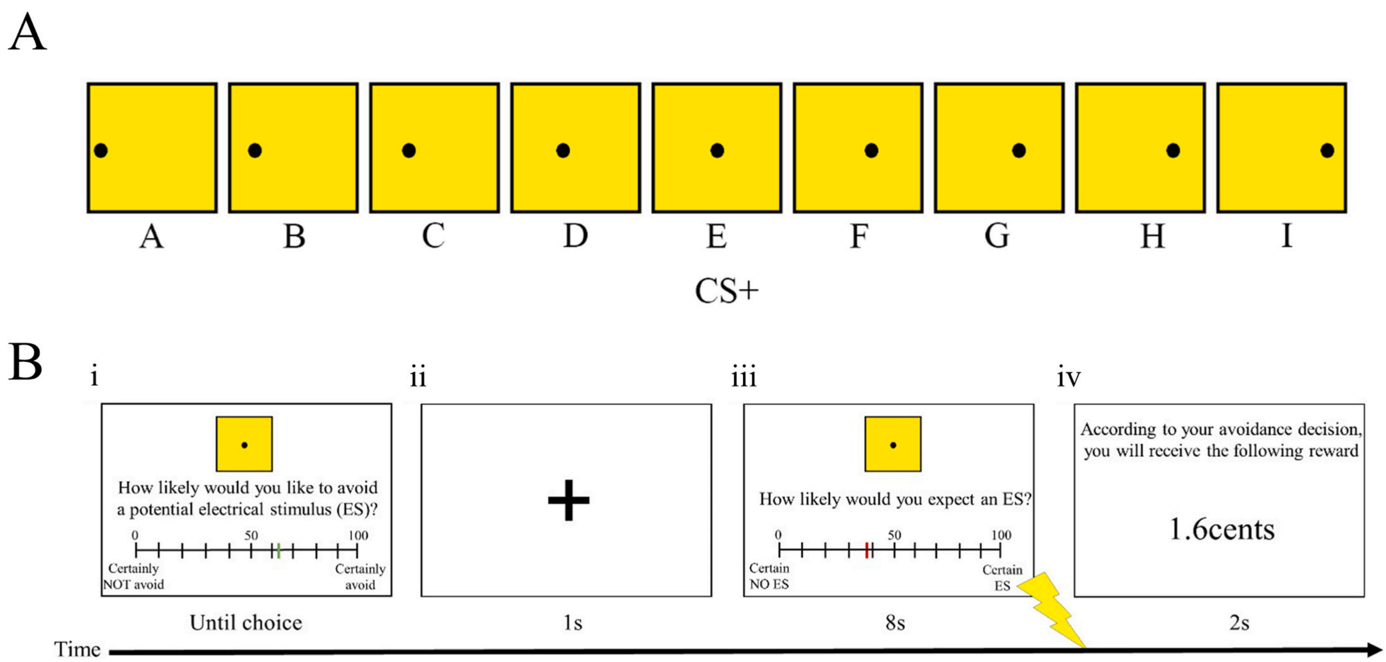

Fig. 1. (A) Stimulus dimension: Stimulus E served as the CS+. The stimulus labels (A–I) were not presented in the experiment. (B) Example of the trial structure in the US-avoidance-acquisition phase and the US-avoidance-reward. (i) Participants were prompted to indicate their US-avoidance ratings. (ii) A fixation cross appeared in the centre of the screen for 1 s (iii) The CS was presented again alongside an US expectancy scale for 8 s. Participants were prompted to indicate their US expectancy ratings. An electrical US would be administered immediately after CS offset depending on the US-avoidance made. (iv) The reward feedback appeared on the screen for 2 s, noted that this feedback occurred in the US-avoidance-reward phase.

CS+. All stimuli were individually presented in the centre of a white screen.

A computer equipped with Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, Version 20.1) presented all stimuli and recorded the US-avoidance ratings and US expectancy ratings. Another computer with BrainVision Recorder (Brain Products GmbH, Gliching, Germany) recorded the skin conductance via two Ag/AgCl electrodes at a sampling rate of 1000Hz. A DS7A Digitimer stimulator generated an electric stimulation US. The US consisted of 125 pulses separated by 5 ms (i.e., US duration of 625 ms), delivered through a bar electrode attached to participants' wrist.

2.3. *Procedure*

After providing written informed consent, participants filled in the German version of DASS-21 (Lovibond & Lovibond, 1995; Nilges & Essau, 2015), which validly measures and discriminates three different constructs: depression, anxiety, and stress. Next, US electrodes were attached to participants' wrist on their non-dominant hand. Two skin conductance electrodes filled with isotonic gel were attached to the hypothenar muscles on the palm on the same hand. Participants were then led through an US intensity calibration, in which the US intensity was gradually increased starting from 0.2 mA until a level of US intensity that was perceived as 'definitely unpleasant but not painful'. Importantly, we carried out a reward matching procedure immediately after US workup procedure to match the US intensity with the amount of competing reward. This means that the amount of the competing reward would neither be too low that artificially increase the degree of US-avoidance nor too high that would artificially induce an opposite pattern. This workup procedure was highly similar to Wong and Pittig (2021, 2022), entailing a series of questions "Are you willing to tolerate the selected level of electrical stimulation if you are given €_?", with the amount of reward ranging from 5 to 31 cents in odd numbers (i.e., 5 cents, 7 cents … 29 cents, 31 cents) presented in a randomized order. This means, a total of 14 questions were presented individually. Participants had to answer either "Yes" or "No" to these questions. The amount of competing reward was the level between the highest amount that received a "No" and the lowest amount that received a "Yes". For instance, if an individual participant was unwilling to tolerate an US when given 5 to 17 cents, but was willing to do so from 19 cents onwards, the amount in between (18 cents) would be used as the maximum competing reward per trial.

This experiment consisted of seven phases: Habituation, Pavlovian fear acquisition, US-avoidance acquisition, US-avoidance-reward, Pavlovian generalization test 1, US-avoidance generalization test and Pavlovian generalization test 2 (see Table 1). Noted that competing reward was only included in US-avoidance-reward and US-avoidance generalization test.

Habituation (US electrodes connected). Participants were instructed that some pictures would appear on the screen, which might or might not be followed by an US. They were informed to learn the relationship between the pictures and the US (cf. Mertens, Boddez, Krypotos, & Engelhard, 2021). Participants were asked to indicate their US expectancy during CS presentations using their dominant hand. The US expectancy visual analogue scale ranged from 0 to 100 in which 0 indicates certain no US and 100 indicates certain US. The habituation phase consisted of four trials of non-reinforced CS+. This phase served to assess baseline responding to the CS+, to compare whether fear had been acquired to the CS+ in the following Pavlovian fear acquisition phase. The CS was presented alongside the US expectancy scale for 8 s. The inter-trial intervals (ITIs) varied between 15 and 18 s and were the same in all the following phases.

Pavlovian fear acquisition (US electrodes connected). Eight trials of CS+ were presented in this phase. Participants were prompted to indicate their US expectancy ratings during the 8 s of CS presentation. The CS+ was reinforced at a 75% rate (i.e., 6 out of 8 trials). The presentation order was pseudo-randomized so that the first and the last trial of CS+ were always reinforced, and the non-reinforced trials were never presented consecutively.

US-avoidance acquisition (US electrodes connected). Before this phase started, participants were informed that they could prevent a potential US that followed the pictures, by indicating their avoidance ratings at

Table 1

A to I indicate the different stimuli across the stimulus dimension; + indicates US presentation; - indicates US omission; * indicates the availability of US-avoidance; + in parentheses indicates that US presentation depends on US-avoidance; € in parentheses indicates competing reward depending on US-avoidance; Number in parentheses indicates the number of trials.

Habituation	Pavlovian fear acquisition	US-avoidance acquisition	US-avoidance-reward	Pavlovian generalization test 1	US-avoidance generalization test	Pavlovian generalization test 2
E− (4)	E+ (6) E− (2)	E* (+) (8)	E* (+, €) (8)	[A-I]- (1)	[A-I]*- (€) (2)	C- (1) E− (1) G- (1) I- (1)

the US-avoidance scale presented alongside the CS (see Fig. 1B for the trial structure). The US-avoidance ratings were negatively proportionate to the chance of receiving an US. For example, an US-avoidance rating of 80% would render a 20% chance of US administration if it would have followed the CS (i.e., reinforced CS+ trials). The CS+ and the US-avoidance scale were presented on the screen until choice. After US-avoidance had been made, a 1 s fixation cross appeared, followed by a CS presented alongside an US expectancy scale for 8 s. Participants were again prompted to indicate their US expectancy ratings. This phase consisted of eight CS+ trials. Six of these CS+ trials were potentially followed by an US depending on the US-avoidance ratings made, whereas two of these CS+ trials were not followed by an US regardless of US-avoidance ratings made. The two non-reinforced CS+ trials were included to follow the 75% reinforcement rate in Pavlovian fear acquisition.

US-avoidance-reward (US electrodes connected). Before this phase started, participants were informed that each trial would be accompanied by a maximum amount of monetary reward. This amount was chosen based on the reward-matching procedure prior to the conditioning task. The amount of reward per trial was, however, inversely proportionate to US-avoidance. For example, an US-avoidance response of 60% would result in a gain of 40% of the reward. Participants were instructed that all reward gained would be paid at the end of the experiment. The trial structure was identical to the previous phase, with the exception that participants received a reward feedback for 2 s at the end of each trial (see Fig. 1B).

Pavlovian generalization test 1 (US electrodes disconnected). The experiment was paused and the experimenter disconnected the US electrodes. The experimenter then proceeded to explain the US electrodes were disconnected to examine whether the disconnection would induce a change in physiological responding, thereby setting up the cover story for disconnecting the US electrodes. In fact, this procedure was established to prevent extinction learning during this phase, which reduces the probability of participants modifying their response strategy due to extinction learning (see Lee et al., 2018; Wong & Lovibond, 2017). Given the US electrodes were disconnected, participants were informed that US-avoidance was unnecessary and hence its availability was removed. They were also informed to continue providing their US expectancy ratings, assuming hypothetically that it was still possible for them to receive an US. All nine stimuli along the stimulus dimension (stimuli A to I) were presented once each in a randomized order. Each stimulus was presented for 8 s and participants were prompted to indicate their US expectancy ratings.

US-avoidance generalization test (US electrodes reconnected). The experiment was paused and the experimenter reconnected the US electrodes. Participants were informed that it was again possible to receive an US, and they were free to decide their US-avoidance ratings for each stimulus. The trial structure was identical to US-avoidance-reward. This phase consisted of two blocks: each block consisted of all nine stimuli along the dimension (stimuli A to I), each presented once. None of the stimuli were reinforced regardless of US-avoidance made.

Pavlovian generalization test 2 (US electrodes connected). Participants were informed that the availability of US-avoidance was removed. Four selected stimuli (C, E, G, & I) were presented once each in a randomized order alongside the US expectancy scale for 8 s. Given that US electrodes were connected in this phase, anticipatory fear to the test stimuli indicated by skin conductance could be analysed. Only four stimuli instead of the whole stimulus dimension was presented to minimize the effect of ongoing extinction in this phase. Noted that in our previous work (Lee et al., 2018; Wong & Lovibond, 2017), participants came up with a Right linear rule (i.e., stimuli on the right end of the stimulus dimension were more likely to predict an US) but not with a Left linear rule (i.e., stimuli on the left end of the stimulus dimension were more likely to predict an US). Thus, we included stimulus I, but not stimulus A, to capture peak responding in participants entertaining a Right linear rule. None of the stimuli were reinforced in this phase.

After the conditioning task, participants were asked to fill in a 2-page questionnaire (see Supplementary Materials). On the first page, the US expectancy ratings to stimuli A and I that the individual participant had made during Pavlovian generalization test1 (when the US electrodes were disconnected). Participants were asked to explain why such ratings were made, and to write down in detail any strategies or rules of responding they used. The second page was administered only after participants had completed the first page. This page consisted of four statements, each describing a potential relationship between the stimuli and the US in terms of relational rules (right linear, left linear, similarity, and no rule). Participants had to rate how much they endorsed each statement on a scale of 0–100, with 0 being *false* and 100 being *true*. If none of the given statements matched an individual participant's perceived relationship between the stimuli and the US, they were asked to write down their own description in the 'Other' section.

2.4. Scoring and analysis

Only skin conductance recorded during the 8 s stimulus presentation (i.e., when participants were prompted to indicated their US expectancy ratings) were included for analysis. We applied a 1 Hz high-pass filter and a notch filter (50Hz) on the data. We then calculated the SCRs by finding the difference between the peak responding and the corresponding trough in the interval of 1 s after stimulus onset to stimulus offset. We square root transformed the SCRs to reduce skewness (Boucsein et al., 2012). Of note, we did not analyse the SCR data in *Pavlovian generalization test 1* because the US electrodes were disconnected, thus no anticipatory fear responses could be elicited.

We conducted linear mixed models for all analyses. US expectancy ratings and SCRs served as cognitive and physiological indicators of the acquisition of conditioned fear to the CS+. Thus, these two measures served as the dependent variable, whereas Trial (the last two trials in Habituation vs the last two trials of Pavlovian fear acquisition) and Group (all rule groups, see *Questionnaire coding* in Results) served as the fixed effect. To examine the acquisition of costly US-avoidance, Phase (US-avoidance acquisition vs US-avoidance-reward) and Group served as the fixed effects.

For all test phases, US expectancy ratings, SCRs, or US-avoidance served as the dependent variable. Two orthogonal polynomial trend repeated measures contrasts across all test stimuli served as fixed effects to assess the shape of the generalization gradients. Specifically, a linear trend repeated measures contrast assessed linear gradients (represented by Stimulus) whereas a quadratic trend repeated measures contrast assessed curvature gradients (represented by $Stimulus^2$). After categorizing participants into groups according to their reported rules, Group served as another fixed effect. Specifically, the interactions between the polynomial trends across test stimuli and all rule groups were assessed to evaluate any group differences in the shapes of generalization gradients. Follow-up analyses then compared the gradients between each rule group.

In addition, Block (Block 1 vs Block 2) served as an additional fixed effect in the *US-avoidance generalization test* phase. This assessed whether the US-avoidance gradients decreased across blocks (i.e., caused by extinction learning). Follow-up analyses then examined the US-avoidance gradients in each block, assessing the formation of US-avoidance generalization gradients even if all stimuli were presented once each. Although US expectancy ratings and SCRs were measured as anticipatory fear responses *after* US-avoidance responses had been executed, we did not include these analyses in the main text (see Supplementary Materials for these analyses).

For all these linear mixed models, participants served as a random effect. In addition, although not pre-registered, the models examining US-avoidance generalization also included random slopes of participants affected by Block, to account for by individual differences in US-avoidance to the test stimuli between the two blocks in the US-avoidance generalization phase (e.g., individual differences in extinction learning)[1]. The degree of significance was reported with Satterthwaite approximation for degrees of freedom (Satterthwaite, 1941). For all follow-up simple analyses, we applied Bonferroni corrections to control for family-wise error rate. All analyses were carried out using R (R Core Team, 2020) with lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

3. Results

Statistical analyses were restricted to participants who had acquired conditioned fear to the CS+. This means, participants who indicated higher US expectancy ratings in the last two trials of Pavlovian fear acquisition compared to the last two trials of Habituation were included in data analyses. This inclusion criterion was pre-registered. Six participants were excluded based on this acquisition criterion. In addition, data from one participant were not measured due to technical problem. Thus, a total of 7 participants were excluded, leading to a final sample of 76 participants.

Questionnaire coding: Two raters, who were blind to the data, categorized participants into different rule groups based on participants' responses to the post-experimental questionnaires via a two-step procedure. The raters first classified participants into rule groups according to their responses to the open-ended section of the questionnaire. If the reported rules were ambiguous, the raters would consult the close-ended section of the questionnaire and categorized participants according to the rule they most strongly endorsed. The two raters reached a substantial agreement as indicated by Cohen's Kappa (k = 0.74; $p < .001$). Discrepancies in rule group classification was resolved via discussion. Similar to previous studies (Wong & Lovibond, 2017, 2018), three distinct rule groups were identified, namely Similarity (n = 27), Right linear (n = 20), and No rule (n = 29). Participants in the Similarity group reported higher US expectancies when the stimulus was more perceptually similar to the CS+. Participants in the Right linear group reported higher US expectancies when the dot was more towards the extreme right end. Participants in the No rule group reported not being able to identify any clear rules. Table 2 shows the demographic data for each group. No significant differences emerged.

[1] Results remained similar when only participants served as a random effect.

Table 2
Demographic and DASS-21 data. Means (standard deviations).

	Similarity group (*n* = 27)	Right linear group (*n* = 20)	No rule group (n = 29)	*F* or χ^2	*p*
Age	23.48 (3.10)	22.40 (2.35)	23.59 (2.80)	0.54	.583
Sex - Female	22 (81.5%)	14 (70.0%)	19 (65.5%)	1.86	.395
US intensity (mA)	1.06 (0.56)	1.06 (0.38)	1.17 (0.50)	0.44	.647
DASS 21-Anxiety	3.33 (4.22)	3.80 (4.49)	3.52 (3.41)	0.08	.925
DASS 21-Depression	5.11 (6.89)	5.70 (4.69)	6.21(7.20)	0.20	.821
DASS 21-Stress	8.96 (8.58)	8.10 (7.24)	9.10 (7.18)	0.11	.900

3.1. Habituation and pavlovian fear acquisition

US expectancy ratings and SCRs. Fig. 2 shows the responses in the first 4 phases. We examined whether participants acquired conditioned fear to the CS+ by comparing responding to the last two trials of Pavlovian fear acquisition to the last two trials of Habituation. In brief, participants exhibited higher US expectancy ratings to the CS+ at the end of Pavlovian fear acquisition compared to the end of Habituation averaged across all rule groups, $F(1, 76) = 219.21$, $p < .001$. There was no evidence for any group differences in the acquisition of US expectancies to the CS+, $F(2, 76) = 0.20$, $p = .821$. In contrast, although participants showed a descriptive increase in SCRs to the CS+ at the end of Pavlovian fear acquisition than the end of Habituation averaged across all rule groups, this increase did not reach significance, $F(1, 76) = 3.47$, $p = .066$.

3.2. US-avoidance acquisition and US-avoidance-reward

US-avoidance. Participants exhibited high levels of US-avoidance to the CS+ immediately in the US-avoidance acquisition, presumably due to the verbal instructions entailing the effectiveness of US-avoidance in US prevention. Averaged across all rule groups, participants exhibited lower levels of US-avoidance to the CS+ in the US-avoidance-reward phase compared to the US-avoidance acquisition phase, $F(1, 76) = 94.16$, $p < .001$. This decrease in US-avoidance between phases was due to the introduction of a reward competing with US-avoidance. The interaction between Phase and Group did not reach significance, $F(2, 76) = 0.71$, $p = .497$, suggesting no evidence for any group differences in the reduction of US-avoidance when a competing reward was introduced.

US expectancy ratings and SCRs. Due to high levels of US-avoidance, participants showed a relatively low level of US expectancy ratings to the CS+ in the US-avoidance acquisition phase. Furthermore, there was a significant increase in US expectancy ratings to the CS+ when transiting from US-avoidance acquisition to US-avoidance-reward, $F(1, 76) = 73.43$, $p < .001$. This pattern was attributed to a decrease in US-avoidance due to the introduction of a competing reward (see Pittig, 2019; Pittig & Wong, 2021). The SCRs showed a similar pattern: participants exhibited a significant increase in responding to the CS+ in the transition from US-avoidance acquisition to US-avoidance-reward, $F(1, 76) = 6.40$, $p = .013$. There was no evidence for any group differences in the increase in US expectancies or SCRs to the CS+ when a competing reward was introduced (smallest $p = .503$).

Collectively, an introduction of a competing reward reduced levels of US-avoidance. Alongside the decrease in US-avoidance, participants exhibited higher US expectancy ratings and stronger SCRs to the CS+ with no group differences. Of note, the increase in SCRs did not purely reflect an increase in anticipatory fear, but also reflected an increase in general arousal due to anticipating a competing reward (e.g., Hulsman et al., 2020; Le et al., 2019).

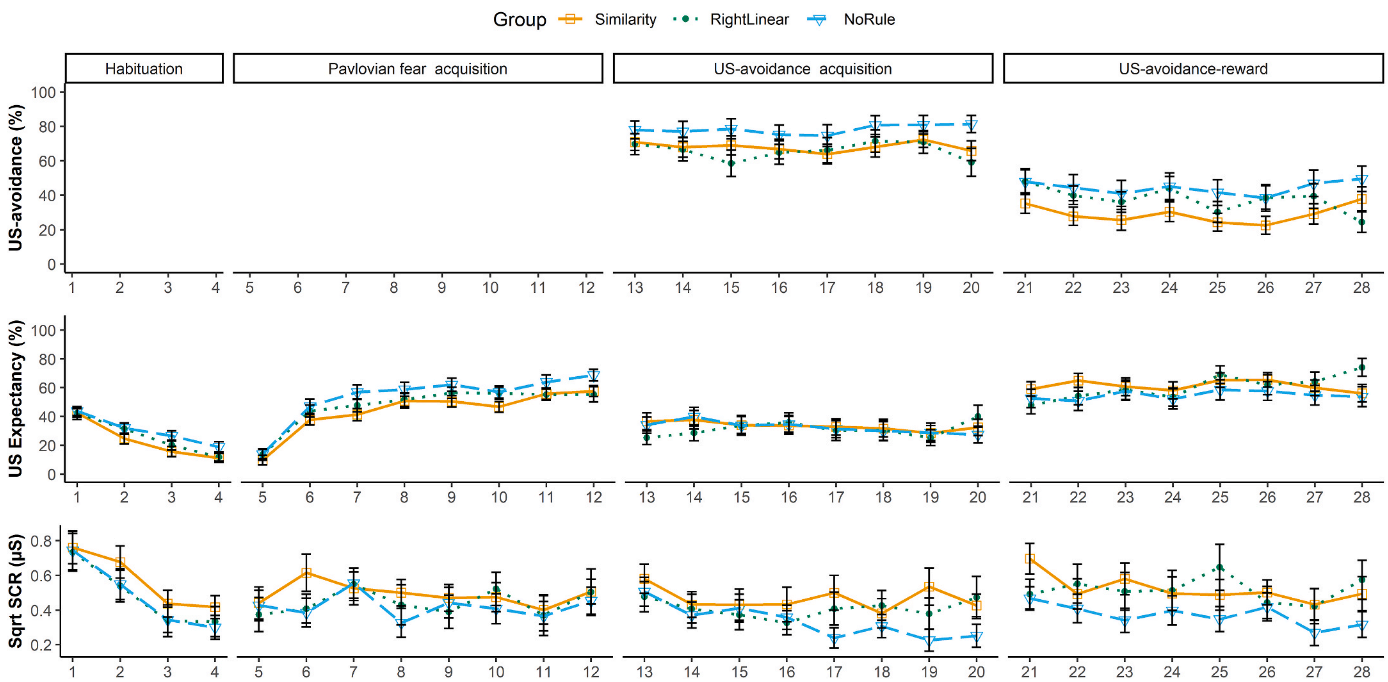

Fig. 2. US-avoidance (top panel), US expectancy ratings (middle panel), and square-root SCRs in the first four phases. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3. *Pavlovian generalization test 1*

US expectancy ratings. Fig. 3A shows the generalization gradients of US expectancy ratings for each rule group. Only interactions of interest are reported here given we were primarily interested in the distinct generalization patterns in each rule group. We observed significant interactions between Group and Stimulus, $F(2, 608) = 22.91$, $p < .001$, and between Group and Stimulus^2 $F(2, 608) = 9.16$, $p < .001$.

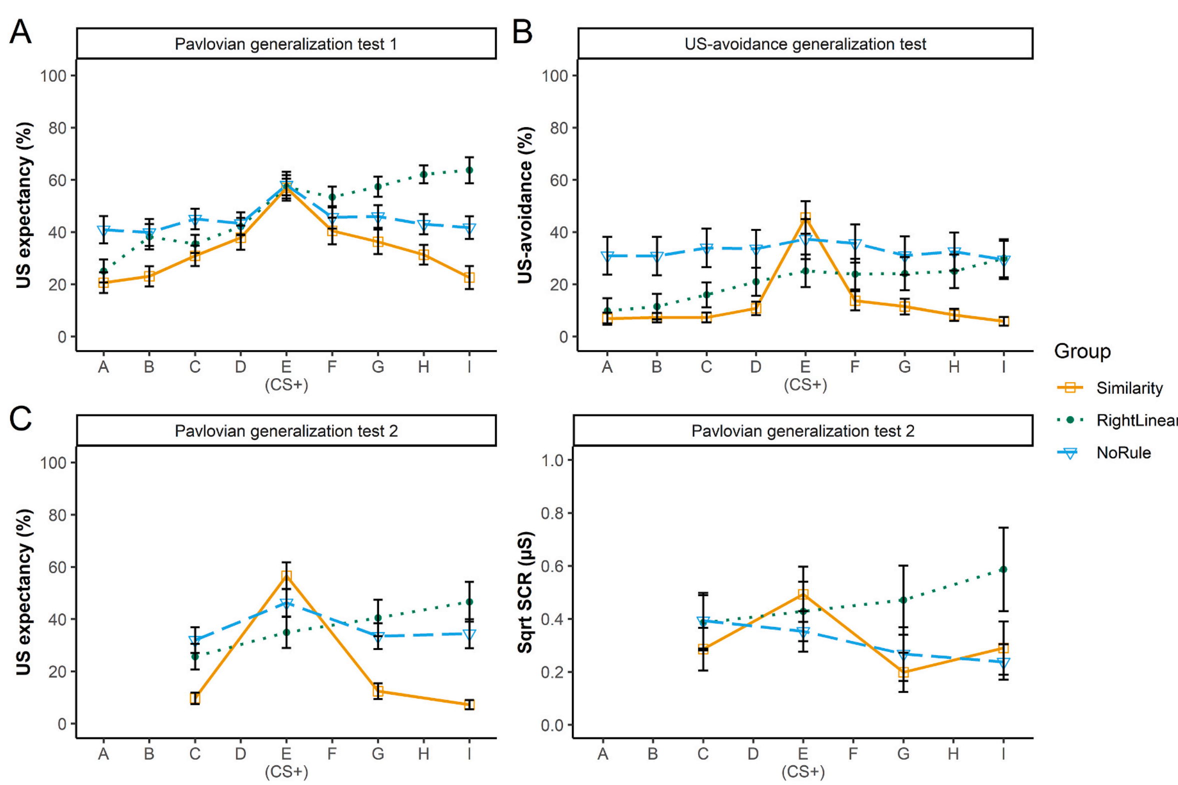

Fig. 3. Responding to test stimuli in the different test phases for each rule group. (A) Mean US expectancy ratings in Pavlovian generalization test 1. Noted that no SCRs data were analysed in this phase due to the disconnection of US electrodes, thus no anticipatory fear responses could be measured (B) Mean US-avoidance in US-avoidance generalization test. (C) Mean US expectancy ratings (left panel) and square-root SCRs (right panel) in Pavlovian generalization test 2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

This suggests that the generalization gradients differed between rule groups in terms of linearity and curvature, respectively. Follow-up analyses revealed that the gradients in all rule groups differed from each other substantially. Compared to the No rule group, the Right linear group exhibited a more linear gradient, indicated by a significant interaction in linear trend across stimuli between these two groups $b_{Group(Right\ linear\ vs\ No\ rule)*Stimulus} = 297.53$, SE = 46.75, $p < .001$, whereas the Similarity group exhibited a more curved gradient, $b_{Group(Similarity\ vs\ No\ rule)*Stimulus^2} = -150.67$, SE = 43.02, $p = .003$. Compared to the Right linear group, the Similarity group revealed a more curved gradient, $b^2_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = -181.16$, SE = 47.46, $p < .001$, but a less linear gradient, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = -265.09$, SE = 47.46, $p < .001$.

In sum, the three rule groups revealed distinct generalization gradients: visually inspecting, the Similarity group exhibited a sharp unimodal gradient; the Right linear group showed a linear gradient with increasing responding to test stimuli on the right end of the stimulus dimension; the No rule group exhibited a flat unimodal gradient.

3.4. US-avoidance generalization test

US-avoidance. Fig. 3B shows the generalization gradients of US-avoidance in the entire US-avoidance generalization test for each rule group. We observed a significant interaction between Group and Stimulus, $F(2, 1292) = 16.11$, $p < .001$, and between Group and Stimulus^2, $F(2, 1292) = 11.95$, $p < .001$, suggesting that the shape of the gradients differed between groups in terms of linearity and curvature, respectively. Follow-up analyses revealed that the Right linear group exhibited a more linear gradient than the No rule group, $b_{Group(Right\ linear\ vs\ No\ rule)*Stimulus} = 231.11$, SE = 43.35, $p < .001$, whereas the Similarity group exhibited a more curved gradient than the No rule group, $b_{Group(Similarity\ vs\ No\ rule)*Stimulus^2} = -164.91$, SE = 39.88, $p < .001$. Furthermore, the Right linear group showed a more linear gradient than the Similarity group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = 206.40$, SE = 44.00, $p < .001$, whereas the Similarity group showed a more curved gradient than the Right linear group, $b^2_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = -186.72$, SE = 44.00, $p < .001$.

There was a significant decrease in US-avoidance from the first block to the second block of US-avoidance generalization, supported by a main effect of Block, $F(1, 76) = 37.21$, $p < .001$. This presumed extinction significantly changed the group differences in gradient's linearity between blocks, $F(2, 1216) = 16.11$, $p < .001$, but did not significantly change the group differences in gradient's curvature between blocks, $F(2, 1292) = 0.44$, $p = .642$. However, follow-up analyses revealed that none of the groups differed with each other in terms of linearity between blocks (smallest $p = .140$).

3.5. US-avoidance generalization test block 1

Given that we hypothesized that US-avoidance gradients could be observed even when each test stimulus was presented once, we analysed the group differences in generalization of US-avoidance in each US-avoidance generalization test block.

US-avoidance. In the first block of US-avoidance generalization test (Fig. 4), we observed a significant interaction between Group and Stimulus, $F(2, 608) = 14.75$, $p < .001$, and between Group and Stimulus^2 $F(2, 608) = 7.17$, $p < .001$, suggesting some group differences in linearity and curvature of gradients of US-avoidance, respectively.[2] Visually inspecting the gradients, the Similarity group showed a unimodal gradient, the Right linear group showed a linear gradient whereas the No rule group exhibited a flat gradient. Follow-up analyses confirmed that these gradients in the rule groups differed from each other: the Right linear group exhibited a more linear gradient than the No rule group, $b_{Group(Right\ linear\ vs\ No\ rule)*Stimulus} = 230.18$, SE = 45.12, $p < .001$, whereas the Similarity group showed a more curved gradient than the No rule group, $b^2_{Group(Similarity\ vs\ No\ rule)*Stimulus} = -123.78$, SE = 4152, $p = .018$. In addition, the Right linear group had a more linear gradient than the Similarity group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = 205.61$, SE = 45.80, $p < .001$, whereas the Similarity group had a more curved gradient than the Right linear group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus^2} = -158.46$, SE = 45.80, $p = .003$.

3.6. US-avoidance generalization test block 2

US-avoidance. In the second block of US-avoidance generalization test (Fig. 4), we observed a significant interaction between Group and Stimulus, $F(2, 608) = 4.67$, $p = .010$, and between Group and Stimulus^2, $F(2, 608) = 7.56$, $p < .001$, suggesting group differences in linear-based gradients and in curvature-based gradients, respectively. Follow-up analyses revealed that compared to the No rule group, the Right linear group showed a more linear gradient, $b_{Group(Right\ linear\ vs\ No\ rule)*Stimulus} = 96.65$, SE = 33.66, $p = .025$, whereas the Similarity group showed a more curved gradient than the No rule group, $b_{Group(Similarity\ vs\ No\ rule)*Stimulus^2} = -109.44$, SE = 30.97, $p = .003$. Compared to the Right linear group, the Similarity group revealed a more curved gradient, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus^2} = -105.61$, SE = 34.17, $p = .013$, however the Right linear group did not show a more linear gradient than the Similarity group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = 86.29$, SE = 34.17, $p = .071$.

Collectively, the generalization gradients of US-avoidance in each rule group were distinct from each other. Aligning with participants' reported rules, the Similarity group showed a sharp peaked gradient, the Right linear group showed a linear gradient, whereas the No rule group showed a relatively flat gradient. Group differences in the generalization of US-avoidance were less robust in the second block compared to the first block, presumably due to partial extinction learning.

3.7. Pavlovian generalization test 2

US expectancy ratings and SCRs. For the US expectancy ratings (Fig. 3C), there was a significant interaction between Group and Stimulus, $F(2, 228) = 8.902$, $p < .001$, and between Group and Stimulus^2, $F(2, 228) = 8.90$, $p < .001$, suggesting that the shape of the gradients differed between groups in terms of linearity and curvature, respectively.

Follow-up analyses revealed that the Similarity group exhibited a more curved gradient than the No rule group, $b_{Group(Similarity\ vs\ No\ rule)*Stimulus^2} = -168.67$, SE = 50.40, $p = .006$. However, there was no evidence that the Right Linear group had a more linear gradient than the No rule group, $b_{Group(Right\ linear\ vs\ No\ rule)*Stimulus} = 144.35$, SE = 54.78, $p = .054$. Furthermore, the Right linear group had a more linear gradient than the Similarity group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus} = 234.23$, SE = 55.60, $p < .001$, whereas the Similarity group had a more curved gradient compared to the Right linear group, $b_{Group(Similarity\ vs\ Right\ linear)*Stimulus^2} = -213.47$, SE = 55.60, $p = .001$.

For the SCRs (Fig. 3D), there were some group differences in the gradients in terms of linearity, $F(2, 228) = 3.15$, $p = .045$, but not in terms of quadratic trend, $F(2, 228) = 0.39$, $p = .676$. However, the follow-up analyses revealed that there was no evidence that the rule groups differed in linearity (smallest $p = .091$).

4. Discussion

With a single-cue conditioning procedure, the current study measured the generalization of safety behavior with a recently developed dimensional measure of avoidance (Wong & Pittig, 2021). This dimensional measure was able to capture costly safety behavior generalization gradients even if each test stimulus was presented once (e.g.,

[2] US-avoidance generalization gradient for each participant can be seen in the Supplementary Materials.

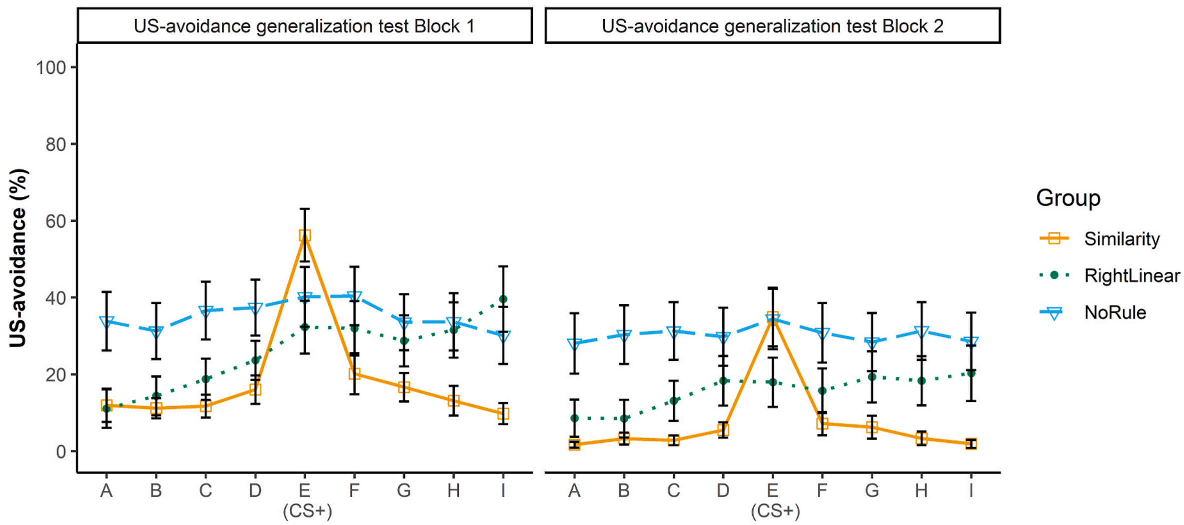

Fig. 4. Generalization gradients of US-avoidance in the first block (left panel) and second block (right panel) of US-avoidance generalization test for each rule group. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

block 1 of US-avoidance generalization test). Importantly, individual generalization gradients varied from each other considerably and were consistent to relational rules (i.e., rule-based generalization). Participants were categorized into three distinct rule groups, namely Similarity, Right linear and No rule. Notably, we demonstrated that both generalization of Pavlovian fear and generalization of safety behavior were highly consistent with these relational rules.

A key finding is that generalization gradients of safety behavior could be assessed even when each test stimulus was presented once when measured in a non-dichotomous manner. Past studies that measured safety behavior dichotomously had to assess the proportion of safety behavior averaged across multiple trials of the same stimulus. This assessment is, however, confounded with ongoing extinction learning to the test stimuli given that none of them were reinforced in test (except the CS+; Blough, 1969; Honig & Urcuioli, 1981). The ability of a dimensional measure to assess generalization gradients of safety behavior without averaging across multiple trials of the same stimulus thus minimizes any confounding effect with extinction learning. This notion is highlighted by the reduction in safety behavior generalization from the first test block to the second test block. This reduction of responding was presumably due to extinction learning, suggesting that extinction occurred rapidly in humans even when the number of non-reinforced test trials was limited. Therefore, this pattern highlights the potential of a dimensional measure of avoidance in minimizing the confounded extinction effect. Although a dimensional measure of avoidance might effectively minimize extinction learning occurring between test blocks, generalized extinction might still occur from trial to trial (given that the GSs are highly similar).

Another advantage of using a dimensional measure is that it arguably provides higher face validity given that safety behavior is not necessarily dichotomous in real life: safety behavior can be partially engaged in to a certain extent (see Krypotos et al., 2018; Telch & Lancaster, 2012). For instance, an individual with social anxiety may avoid eye contact to a greater extent when giving a public speech, but engage in the same safety behavior to a lesser extent when conversing in a small group. Future studies can examine whether a dimensional measure of avoidance is strong in other external validities (e.g., predictive validity, see Vervliet & Raes, 2013). Furthermore, as this measure allowed the partial engagement in safety behavior while being able to retain a portion of the competing reward, it provides great potential for future research to examine the nuances in approach-avoidance conflict.

Another key finding is that the generalization of safety behavior aligns with the self-generated relational rules. Generalization gradients generated by different rules were distinct from each other. More importantly, these gradients of safety behavior were highly consistent with the relational rules. Specifically, participants who reported a similarity rule engaged in safety behavior to a lesser extent when presented with a novel stimulus that only slightly resembled the CS+, whereas participants who reported a right linear rule engaged in safety behavior to a greater extent to stimuli right of the stimulus dimension. Furthermore, participants who failed to identify any relational rules (No rule) exhibited a flat gradient, characterized with a relative absence of generalization decrement. These patterns suggest that relational rules determine how safety behavior generalizes to other stimuli that resemble the CS+. The current findings align with studies that examined safety behavior generalization via higher-order processes. For instance, participants were more likely to engage in safety behavior when presented with novel stimuli that belonged to the same artificial category of the CS+ than those that belonged to the same category of the CS- (Dymond et al., 2011, 2014). Similarly, safety behavior was more frequently engaged in when presented with a novel word that was a synonym of the CS+ (semantic generalization; Boyle et al., 2016). The current findings thus extend the role of higher-order processes in the generalization of safety behavior.

The current study also found that the generalization gradients of fear were consistent with the reported relational rules, consistent with other rule-based generalization studies (Lee et al., 2018; Wong et al., 2020; Wong & Lovibond, 2017). The rule group differences in generalization gradients were, however, observed in the US expectancy ratings, whereas the SCR data partially aligned with the reported rules. Specifically, given that acquisition of SCRs to the CS+ was not established, the limited group differences in SCRs warranted caution in interpreting how inferred relational rules determined SCRs generalization. In sum, self-generated relational rules determine the generalization of both self-reported ratings and costly safety behavior.

The high consistency between the self-generated relational rules and fear as well as safety behavior generalization aligns with the Expectancy model (Lovibond, 2006). This model entails that the acquisition of both conditioned fear and safety behavior to a warning signal are based on one's propositional belief. The current findings extend this notion to novel stimuli that resemble the warning signal. Specifically, how propositional beliefs determine the degree of safety behavior engagement to these novel stimuli. It is, however, largely unknown how participants came up with a variety of relational rules despite receiving highly similar training. In addition, it is unclear whether there are any individual factors that may encourage participant to endorse one rule but not another. Perhaps the endorsement of different rules was caused by the individual differences participants brought to the study, for instance, prior history of learning or different personalities. Future studies can delineate what factors directly or indirectly determine the type of relational rules generated.

Regarding clinical implications, partial engagement in safety behavior can be seen as balancing threat at a subjectively acceptable level and limiting the cost of executing safety behavior (cf. Schlund et al., 2016). Thus, a dimensional measure of US-avoidance captures the nuances of the degree of safety behavior engagement (see Krypotos et al., 2018; Wong & Pittig, 2021). More importantly, the individually inferred relational rules parallel with the formation of different threat beliefs in anxiety-related disorders. Specifically, participants acquired fear in highly similar conditions, but inferred distinct relational rules, which in turn guided the generalization of conditioned fear and costly safety behavior. This may parallel with individuals who experienced similar trauma exposures, but later exhibit safety behavior to different extents when presented with situations or objects that resemble the trauma. For instance, when encountering a dark color Chihuahua, an individual with dog phobia believing that the fur's darkness is positively associated with aggressiveness may engage in safety behavior to a greater extent, whereas another individual believing that the size of a dog is positively associated with aggressiveness may engage in safety behavior to a lesser extent. Furthermore, we incorporated a competing reward to encourage disengagement in safety behavior, thus effectively rendered the execution of safety behavior costly, resembling pathological safety behavior in anxiety-related disorders. Thus, the current findings suggested that the self-generated threat beliefs determine the balance between minimization of perceived threat and cost of safety behavior to novel stimuli that resemble the feared stimulus.

The current study emphasized on the "initial" generalization of costly safety behavior as confounded extinction was minimized, providing a useful testbed for evaluating the interplay between threat appraisal and approach-avoidance conflict with minimum interference from ongoing extinction. Thus, future studies may expand the current design and assess whether individuals at risk of or with clinical anxiety are more likely to exhibit excessive generalized safety behaviors before ongoing extinction learning in test occurs. Future studies may also assess another maladaptive aspect of generalized safety behavior, namely the persistent generalized safety behavior after extinction learning to the GSs (e.g., Hunt et al., 2019; van Meurs et al., 2014). By presenting the same GSs for multiple trials and employing a dimensional measure of avoidance, one can evaluate how the initial generalized safety behavior impacts subsequent extinction learning to the same GS. For example, stronger engagement in safety behavior to a GS likely protects one from extinction learning, further fuelling unnecessary, excessive safety behavior to it.

The current study had some limitations. First, we found only a few significant effects on the skin conductance measure, especially the lack of evidence in SCR acquisition to the CS+ and SCR generalization gradients. The limited findings in the SCRs could be due to its large individual variability (Lykken & Venables, 1971), resulting in less statistical power for group comparison. Second, we employed a single-cue conditioning procedure so that we could assess how excitatory learning to the CS+ generalizes to other GSs without being interfered by the generalization of inhibitory learning to a safety cue (CS-). However, without a CS-, it was difficult to determine whether conditioned fear was acquired to the CS+, or whether the increase in conditioned fear was due to non-associative learning processes (e.g., sensitization). Third, we included only four test stimuli in Pavlovian generalization test 2 to minimize ongoing extinction. This led to an asymmetrical distribution of test stimuli (i.e., the inclusion of stimulus I but not A). Fourth, the stimulus dimension in use had a clearly defined mid-point, thus encouraging the formation of explicit relational rules and overshadowed effects from low-level associative learning processes. Future studies can employ a more continuous stimulus dimension with no clearly defined mid-point (e.g., blue-green dimension, see Lee et al., 2018). Fifth, the non-reinforced CS+ trials in Habituation might have induced latent inhibition, slowing down fear learning to it in the following phase. However, at the end of Pavlovian fear acquisition, threat expectancy ratings reached asymptote that was similar to the reinforcement rate, suggesting that latent inhibition had been minimized.

In conclusion, this study employed a dimensional measure to assess the generalization gradients of costly safety behavior. Impressively, this dimensional measure allowed us to examine the generalization of safety behavior even when each test stimulus was presented once, thus minimizing the confounding effect of ongoing extinction learning to all test stimuli. In addition, participants inferred distinct relational rules, which directly determined to what extent one would engage in safety behavior to different generalization stimuli. In terms of clinical implication, the present work emphasizes that threat beliefs determine the extent of costly safety behavior engagement to various stimuli that resemble the feared stimulus. Thus, the current findings can be extended to identify individual threat belief that could help identify sets of objects or situations that evoke costly safety behavior.

Authors' declaration

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.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from h.k.wong@essb.eur.nl.

CRediT authorship contribution statement

Alex H.K. Wong: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Funding acquisition. **Andre Pittig:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Declaration of competing interest

None.

Acknowledgement

AHKW was supported in part by funds of the Bavarian State Ministry of Science and the Arts and the University of Würzburg to the Graduate School of Life Sciences (GSLS), University of Würzburg. AP was supported by a Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) grant PI1269/2-1 – 389569971. The authors would like to thank Lea Zwilling, Anja Dolling, Lena Bergmann, Annkathrin Evers, Anica Pilz, Tabea Seibel, and Maria Yepez for their help in data collection and processing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104158.

References

Arnaudova, I., Kindt, M., Fanselow, M., & Beckers, T. (2017). Pathways towards the proliferation of avoidance in anxiety and implications for treatment. *Behavior Research and Therapy, 96*, 3–13. https://doi.org/10.1016/j.brat.2017.04.004

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

Blough, D. S. (1969). Generalization gradient shape and summation in steady-state tests. *Journal of the Experimental AAnalysis of Behavior, 12*(1), 91–104. https://doi.org/10.1901/jeab.1969.12-91

Boucsein, W., Fowles, D. C., Grimnes, S., et al. (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 & Emotion, 30*(2), 340–352. https://doi.org/10.1080/02699931.2014.1000831

Core Team, R. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

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. *Behavior Research and Therapy, 49*(10), 614–621. https://doi.org/10.1016/j.brat.2011.06.007

Honig, W. K., & Urcuioli, P. J. (1981). The legacy of guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior, 36*(3), 405–445. https://doi.org/10.1901/jeab.1981.36-405

Hulsman, A. M., Kaldewaij, R., Hashemi, M. M., Zhang, W., Koch, S. B. J., Figner, B., et al. (2020). Individual differences in costly fearful avoidance and the relation to psychophysiology. *Behaviour Research and Therapy, 137*, Article 103788. https://doi.org/10.1016/j.brat.2020.103788

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

Kim, E. J. (2005). The effect of the decreased safety behaviors on anxiety and negative thoughts in social phobics. *Journal of Anxiety Disorders, 19*(1), 69–86. https://doi.org/10.1016/j.janxdis.2003.11.002

Krypotos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neurosicence, 9*, 189. https://doi.org/10.3389/fnbeh .2015.00189

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018). Peak shift and rules in human generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(22), 1955–1970. https://doi.org/10.1037/xlm0000558

Le, T. M., Wang, W., Zhornitsky, S., Dhingra, I., Zhang, S., & Li, C. R. (2019). Reward sensitivity and electrodermal responses to actions and outcomes in a go/no-go task. *PLoS One, 14*(7), Article e0219147. https://doi.org/10.1371/journal.pone.0219147

Lommen, M. J. J., Engelhard, I. M., & van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? *Personality and Individual Differences, 49*(8), 1001–1006. https://doi.org/10.1016/j.paid.2010.08.012

Lovibond, P. F. (2006). Fear and avoidance: An integrated expectancy model. In *Fear and learning: From basic processes to clinical implications* (pp. 117–132). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/11474 -006.

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

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

McManus, F., Sacadura, C., & Clark, D. M. (2008). Why social anxiety persists: An experimental investigation of the role of safety behaviors as a maintaining factor. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(2), 147–161. https://doi.org/10.1016/j.jbtep.2006.12.002

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

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

Nilges, P., & Essau, C. (2015). Die depressions-angst-stress-skalen: Der DASS–ein screeningverfahren nicht nur Fur schmerzpatienten [depression, anxiety and stress scales: DASS–A screening procedure not only for pain patients]. *Schmerz, Der, 29*(6), 649–657. https://doi.org/10.1007/s00482-015-0019-z

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

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

Pittig, A., Boschet, J. M., Glück, V. M., & Schneider, K. (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

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

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

Salkovskis, P. M. (1991). The importance of behavior in the maintenance of anxiety and panic: A cognitive account. *Behavioural and Cognitive Psychotherapy, 19*, 6–19.

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

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

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

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: Application to models of anxiety and depression. *Psychological Medicine, 43*(11), 2241–2244. https://doi.org/10.1017/S0033291712002267

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

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

Wong, A. H. K., & Lovibond, P. F. (2017). Rule-based generalisation in single-cue and differential fear conditioning in humans. *Biological Psychology, 129*, 111–120. https://doi.org/10.1016/j.biopsycho.2017.08.056

Wong, A. H. K., & Lovibond, P. F. (2018). Excessive generalization of conditioned fear in trait anxious individuals under ambiguity. *Behaviour Research and Therapy, 107*, 53–63. https://doi.org/10.1016/j.brat.2018.05.012

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

Wong, A. H. K., & Pittig, A. (2022). Avoiding a feared stimulus: Modelling costly avoidance of learnt fear in a sensory preconditioning paradigm. *Biological Psychology, 168*, Article 108249. https://doi.org/10.1016/j.biopsycho.2021.108249