RUNNING HEAD: TRAIT ANXIETY AND GENERALIZATION

Trait anxiety and fear generalization:

Overgeneralization of fear or undergeneralization of safety learning?

Jessica C. Lee[1,2]

Tracy Dang[1]

Jack H. Collins[1]

1 University of Sydney

2 University of New South Wales

Please address correspondence to:
jessica.c.lee@sydney.edu.au
School of Psychology, Griffith Taylor (A19)
University of Sydney, Camperdown, 2006, NSW

Acknowledgements
This research was supported by a Discovery Early Career Researcher Award (DE210100292) awarded from the Australian Research Council to JL.

Abstract

The tendency to overgeneralize fear learning has been identified as a potential risk factor for the maintenance of anxiety disorders. In this study, we used a morphed shape dimension to separately measure generalization of fear (learning when an aversive stimulus is present) and generalization of safety (learning when an aversive stimulus is absent) in participants classified as high or low in trait anxiety. In two experiments, we found that high trait anxious participants undergeneralized safety learning relative to low trait anxious participants, but there were no differences in overgeneralization of fear. A control group who learned about a neutral outcome did not show this pattern of results, and the aversiveness of the outcome predicted undergeneralization of safety learning, but did not predict overgeneralization of fear learning. Our findings suggest that overgeneralization of fear in clinically or high trait-anxious participants may be partly driven by undergeneralization of safety learning. Therefore, promoting generalization of safety learning may be a useful strategy to reduce the spread of fear.

Keywords: trait anxiety, generalization, fear conditioning, fear learning, safety learning, individual differences

Trait anxiety and fear generalization:
Overgeneralization of fear or undergeneralization of safety learning?

To understand pathological anxiety, it is important to study the origins of fear, as well as how that fear spreads. Cues that are associated with a traumatic event (e.g., a car crash, getting robbed) often acquire the ability to elicit fear. In the laboratory, we can study this phenomenon using a fear conditioning procedure, which involves pairing neutral conditioned stimuli (CS) with an aversive outcome (+, e.g., an electric shock or loud scream). Over repeated pairings, these CS+ acquire the ability to elicit the fear response that was previously elicited to the aversive outcome. Decades of research has shown that fear learning is not confined to stimuli associated with the original aversive outcome. Instead, fear often spreads or *generalizes* to related or similar stimuli (Beckers et al., 2023; Dunsmoor & Paz, 2015; Dymond et al., 2015; Lonsdorf et al., 2017).

The acquisition and generalization of fear is adaptive, since it allows us to successfully avoid harm and maximize survival chances in the future. When fear generalization is overly broad however, it may extend to innocuous stimuli, leading to excessive avoidance and anxiety. In fact, a tendency to overgeneralize fear has been cited as a potential risk factor for the development and maintenance of anxiety disorders (Beckers et al., 2022, Dunsmoor & Paz, 2015; Lissek, 2012; Cooper et al., 2022; Fraunfelter et al., 2022). Understanding how and why people differ in how they generalize fear is therefore of clinical importance (Lonsdorf & Merz, 2017) and has driven increased research efforts in recent years to uncover the predictors of overgeneralization of fear.

In this study, we focus on trait anxiety. Trait anxiety has been identified as one such predictor of fear overgeneralization (Sep et al., 2019) and has been proposed as a risk factor for anxiety disorders (Chambers et al., 2004; Gershuny & Sher, 1998).

Natural variation in trait anxiety in the population may predispose individuals to overgeneralize fear learning, and may therefore provide information about the susceptibility of individuals to developing anxiety disorders (Lonsdorf & Merz, 2017; Sep et al., 2019; Vervliet & Boddez, 2020). Further, studying preclinical levels of anxiety has the advantage of avoiding the issue of comorbidity when studying patients diagnosed with anxiety disorders. Understanding how identical experiences can lead to different learning outcomes can inform personalized interventions to reduce excessive spread of fear.

Such individual differences however, are not always easy to observe. Lissek and colleagues (2006) have proposed that "strong" situations involving unambiguous fear may not be conducive to uncovering variability in responding between individuals. Instead, they propose that "weak" situations involving uncertainty may be optimal. This distinction may explain why fear extinction (where there are two competing memories) is more susceptible to revealing anxiety differences than acquisition (Duits et al., 2015, Kausche et al., 2025, but see Morriss et al., 2021). Likewise, generalization appears to be a prime example of a weak situation since it is inherently open-ended and uncertain. Combined with the fact that all situations involve some degree of novelty, this makes generalization a powerful tool for identifying risk factors associated with anxiety disorders.

Much of the fear generalization literature uses a procedure developed by Lissek et al. (2008). This procedure uses a continuous stimulus dimension (circle size) ranging from small to large, and participants undergo training where a circle of a particular size signals an aversive outcome (i.e., CS+, the fear cue), and a circle of another (e.g., larger, counterbalanced) size signals nothing (i.e., CS-, the safety cue). Afterwards, participants are tested on novel circles of varying intermediate sizes between the CS+

and CS-, with responses along the dimension plotted to form a generalization gradient. Studies using this procedure have found that participants with GAD (Lissek et al., 2014), panic disorder, (e.g., Lissek et al., 2010), and post-traumatic stress disorder (e.g., Kaczkurkin et al., 2017), show heightened fear responses to novel generalization stimuli (GS) relative to non-anxious control participants. This overgeneralization of fear is also seen in participants who are high (relative to low) in trait/state anxiety (Baumann et al., 2017; Dunsmoor, White, & LaBar, 2011; Huggins et al., 2021; Haddad et al., 2012, but see Torrents-Rodas et al., 2013), and is observed on a variety of dependent measures including self-report measures (e.g., predictions of the outcome, risk ratings), as well as psychophysiological measures (e.g., startle responses, skin conductance, heart rate). The robustness of this finding is supported by recent meta-analyses which show a small-medium positive effect size for overgeneralization of fear in participants with anxiety disorders (Cooper et al., 2022; Fraunfelter et al., 2022), and a small positive effect size for the relationship between fear generalization and anxious personality traits (e.g., trait anxiety, neuroticism, Sep et al., 2019). In summary, although there are some failures to observe the effect (Ahrens et al., 2016; Tinoco-Gonzalez, 2015; Greenberg et al., 2013), overgeneralization of fear in anxious participants is seen as a robust phenomenon (Beckers et al., 2023; Cooper et al., 2022; Fraunfelter et al., 2022).

The circle size dimension has proved to be sensitive in uncovering predictors of overgeneralization of fear. There are, however, features of the circle size dimension that do not seem optimal for investigating generalization. Generalization is assessed by presenting circles of varying size between the CS+ and CS-, but no stimuli outside of this range. This type of procedure therefore assesses *interpolation* (generalization between), but not *extrapolation* (generalization beyond what is known). This

distinction from the function learning literature is relevant since the latter is considered a more diagnostic test of the content of learning (e.g., DeLosh, Busemeyer, & McDaniel, 1997). Consistent with this idea, work identifying individual differences in rule learning involves measurement beyond the training stimuli to demonstrate different gradient shapes (e.g., Lee et al., 2018, 2021; Lovibond et al., 2020). While it may be argued that any novel stimulus presented in the generalization test is ambiguous and therefore constitutes a weak situation (Lissek et al., 2006), extrapolation is arguably more ambiguous than interpolation, as this type of test goes beyond the training stimuli where there is less information.

Of relevance to the current study, the most important limitation of the circle size dimension is that it potentially confounds two underlying processes. Early theories of stimulus generalization (e.g., Spence, 1937) assumed that the post-discrimination generalization gradient could be derived from generalized excitation arising from the CS+, combined with generalized inhibition arising from the CS-. This duality is also implemented in later elemental associative models of generalization (e.g., Blough, 1975; Ghirlanda & Enquist, 1998; McLaren & Mackintosh, 2002). In these models, there are two generalization processes that can occur – generalization of excitatory (i.e., fear) learning originating from the CS+, and generalization of inhibitory (i.e., safety) learning originating from the CS-. On the circle size dimension, since the CS+ and CS- are located at the two extreme ends of the dimension, it is not possible to distinguish these two processes because they extend along the same size dimension and will overlap due to the CS+ and CS- sharing common features. Indeed, associative models employing elemental representation of stimuli rely on this exact process (interaction between overlapping excitatory and inhibitory gradients) to explain generalization phenomena such as peak and area shift (Hanson, 1959). In summary, while the circle

size dimension has been immensely useful to date, it only captures a limited way in which generalization might differ between individuals. Greater understanding of individual differences may be obtained if we could distinguish between generalization of fear and generalization of safety learning, and use stimuli that better capture extrapolation, which is arguably the essence of generalization.

One solution is a stimulus set used by Lemmens et al. (2021). Participants in Lemmens et al. (2021) were trained with a black triangle and a black square as the fear and safety cues (counterbalanced). Rather than morphing between the CS+ and CS- to create the stimulus dimension, they included a discrete morphing stimulus (MS, a circle) in the middle of the dimension and morphed the CS+ and CS- with the MS (see Figure 1). Unlike the circle size dimension, Lemmens et al.'s (2021) morphed shape dimension allows for assessment of extrapolation of learning. Since the morphing stimulus (the circle) is novel, all GSs can be considered to extend "beyond" the training stimuli (interpolation would involve morphing between the triangle and square without the circle). An additional but more minor advantage is that these stimuli do not constitute an intensity dimension. Intensity dimensions are known to lead to monotonic gradients even in animals (Razran, 1949), and can lead to asymmetries in learning (see Inman & Pearce, 2018, for a review), creating unnecessary noise when the dimension is counterbalanced, and unreliable results if the dimension is not counterbalanced.

Figure 1

Morphed Stimulus Dimension Separating Generalization from CS+ from CS-

Note. CS+ = conditioned stimulus predicting the outcome, CS- = conditioned stimulus predicting nothing, GS = Generalization Stimulus, MS = morphing stimulus.

More importantly for our current purposes, Lemmens et al.'s (2021) use of the morphing stimulus capitalizes on our tendency to discretize continuous dimension using verbal labels (e.g., "left" and "right" in Wong & Lovibond, 2017), and effectively separates the dimension into two categories of Generalization Stimuli (GSs). The GS+ are clearly more similar to the CS+, while the GS- are clearly more similar to the CS-. This morphed dimension therefore allows for separate measurement of generalization of fear learning from the CS+ (GS+) from generalization of safety learning from the CS- (GS-), allowing us to disentangle these two processes in studying individual differences. Previous studies have attempted to separate fear and safety learning using a conditional discrimination (AX+ BX-, Jovanovic et al., 2005), but these studies tested transfer by presenting the safety cue (B) with the fear cue (A) or a novel fear cue in a summation test (Jovanovic et al., 2012), rather than changing features of the safety cue. Thus to date, there has been no study that has attempted to characterize trait anxiety differences in generalization by separately measuring generalization of fear learning, and generalization of safety learning, along a continuous stimulus dimension.

The aim of this study was to test whether individuals high in trait anxiety overgeneralize from fear cues and/or undergeneralize from safety cues, relative to those who are low in trait anxiety using the Depression and Anxiety Stress Scales (DASS,

Lovibond & Lovibond, 1995). The DASS was chosen over other measures of trait anxiety such as the State Trait Anxiety Inventory (STAI, Spielberger et al., 1983) to minimize construct overlap with negative affect (Kennedy et al., 2001; Knowles & Olatunji, 2020)[1]. The DASS has good psychometric properties in that the stress, depression, and anxiety scales load onto separate factors (Antony et al., 1998; Brown et al., 1997) and show stability over long periods of time (3-8 years later, Lovibond, 1998). In line with previous studies (e.g., Grillon & Ameli, 2003; Wong & Lovibond, 2020b, 2021), we classified participants into high and low trait anxious groups. We ran two experiments using a fear conditioning procedure where participants learned about one conditioned stimulus (e.g., a black triangle) that signaled an outcome (CS+, i.e., the fear cue) consisting of an aversive loud female scream presented with a scary image, and another conditioned stimulus (e.g., a black square) that signaled nothing (CS-, i.e., the safety cue). Following Lemmens et al. (2021), we morphed the CS+ and CS- with a morphing stimulus (a black circle) situated in the center of the dimension to create a dimension ranging from the fear to the safe cue (see Figure 1).

We report the results from two experiments. The aim of Experiment 2 was to replicate and extend the results of Experiment 1 by testing whether any trait anxiety differences were dependent on the use of an aversive outcome and therefore specific to safety learning. Experiment 2 included an additional control group where participants learned about a neutral outcome (a tone) instead of the aversive scream. The results from this control group can tell us whether any anxiety differences in generalization are dependent on outcome valence (and are therefore inherent to fear and safety) or represent a domain-general bias to generalize positive and negative associations. Due

[1] In Experiment 1, correlations between trait anxiety and depression measures were higher for the STAI than for the DASS (see Supplemental Materials).

to the high degree of similarity between experiments, we report the experiments together.

Experiments

Method

Participants

The participants were recruited from the undergraduate Psychology pool at the University of Sydney. There were 128 participants (107 female, 20 male, 1 other/no answer, mean age = 19.8, SD = 2.1) in Experiment 1 and 169 participants (111 female, 53 male, 5 other/no answer, mean age = 20.2, SD = 5.0) in Experiment 2. For both experiments, we prescreened participants using the anxiety subscale of the Depression and Anxiety Stress Scales (DASS, Lovibond & Lovibond, 1995) and also discarded participants who scored in the Mild range (8) on the DASS administered within the experiment (27 participants in Experiment 1, 30 participants in Experiment 2). Participants who scored 1-6 (classified as Normal) were allocated to the Low Anxious group, and participants who scored 10 or above (classified as Moderate, Severe, or Extremely Severe) were allocated to the High Anxious group. These cutoffs were chosen to maximize usage of the undergraduate convenience sample and to avoid reliability issues associated with median splits (Lonsdorf & Merz, 2017).

Materials

We assessed trait anxiety using the 21-item version of the DASS (Lovibond & Lovibond, 1995). The stimuli were based on Lemmens et al. (2021) and consisted of solid black shapes on a white square background. A triangle and square served as the training stimuli (CS+ and CS-, randomized), while the morphing stimulus (MS) used to morph the dimension was a circle (see Figure 1). All stimuli were created and

morphed in R using the "transformr" package (Pedersen, 2024) and presented on screen as a 400x400 pixel square image. For the aversive groups, the outcome consisted of a 1s audio of a female scream (used in Purves et al., 2019; McGregor et al., 2021) and an image of a scary face ("ScaryFace1") from the Open Affective Standardized Image Set (OASIS) database (Kurdi et al., 2017) that was presented simultaneously on screen. For the neutral group, the outcome was a 1s 440Hz tone presented simultaneously with an image of a clipart speaker (see https://osf.io/avpkr/?view_only=68436efc39b0427eb9426078b040e7ee). The audio files were normalized to the same volume using software (Audacity).

Procedure

The experiment was approved by the University of Sydney Human Research Ethics Committee. The experiment was programmed using the jsPsych library (de Leeuw, 2015), hosted using JATOS (Lange, Kühn, & Filevich, 2015) and run on a web browser in the lab. Participants signed up for the experiment using an online platform and attended a face-to-face testing session in the lab in groups of up to 4. They were asked to wear headphones during the task, and told that the volume would be set at a predetermined level. The experimenter informed participants that they could ask to turn down the volume at any time, but that it was preferable to keep the volume at the preset level as the experiment involved sound. All ensuing tasks, questionnaires, and instructions were administered through the program, and the total task took about 20 minutes to complete.

After providing informed consent and answering basic demographic questions (gender, age, language/s), participants were given instructions as to how to respond in the main task. They were told that their task was to predict the occurrence of an outcome

by pressing “L” if they predicted the outcome, or “A” if they predicted no outcome. They were presented with a brief instructions quiz and had to answer all questions correctly to proceed. If they answered any questions incorrectly, they were taken back to the beginning of the instructions and presented with the same instruction quiz again.

Baseline Affective Ratings. Prior to the training phase, participants were asked to complete some initial judgements of the CS+ and CS-. Similar to Purves et al. (2019), the CS+ and CS- were presented on separate screens and participants were asked to rate how each stimulus made them feel by moving sliders on three visual analogue scales. The happiness scale ranged from “happy/pleased/content” to “unhappy/annoyed/despairing”, the anxious scale ranged from “calm/sleepy/dull” to “anxious/aroused/jittery”, and the fearful scale ranged from “unafraid/safe/unconcerned” to “fearful/afraid”. The arrangement of the three scales was fixed (happiness on the top, anxious in the middle, fearful on the bottom). The midpoint of all three scales was marked with a tick and participants had to respond on all 3 scales before they could click on a button to continue. The CS+ and CS- were presented on different screens, with the order of presentation randomized.

Training. The allocation of the triangle and square to CS+ and CS- was randomized. CS+ always led to the outcome, and CS- always led to nothing. The training phase consisted of 12 presentations of the CS+ and CS-, randomized in blocks of 2 presentations of each stimulus (4 trials in each block). On each trial, participants were presented with either the CS+ or CS- at the top of the screen, along with the text “What do you think will happen?” underneath. Participants had to press “A” or “L” to make their prediction. After 4s, the stimulus disappeared and the outcome was either presented or not presented. The outcome was a 1s audio of a female scream (Purves et al., 2019; McGregor et al., 2021) presented simultaneously with an image of a scary

face (Kurdi et al., 2017) for all participants in Experiment 1, and for the Aversive group in Experiment 2. The outcome was a 1s 440Hz tone presented simultaneously with an icon of a speaker for the Neutral group in Experiment 2 (see https://osf.io/avpkr/?view_only=68436efc39b0427eb9426078b040e7ee). The outcome feedback duration was 1s, where participants either heard the outcome audio and saw the outcome image if it was a CS+ trial, or heard nothing and saw a blank screen if it was a CS- trial. There was a blank inter-trial interval (ITI) of 4s between training trials.

Post-Training Affective Ratings. After the training phase, participants were again asked to provide affective ratings to the CS+ and CS- in the same way (and order) as before.

Generalization Test. For the generalization test, participants were told that they would be presented with more stimuli and should continue to make predictions, but that they would no longer receive feedback about the outcome. This no-feedback testing procedure has been shown to effectively avoid extinction allowing for accurate measurement of the whole generalization gradient (Lee et al., 2022). Participants were first presented with all 9 stimuli on the morphed shape dimension in randomized order (see Figure 1). On each trial, a stimulus was presented along with the text "What is the likelihood of this stimulus leading to a [scream/tone]?" Participants made their judgement by clicking on a visual analogue scale ranging from "Certain NO [scream/tone] (0% chance of scream/tone]" on the left, to "Certain [scream/tone] (100% chance of scream/tone]" on the right. The midpoint of the scale was marked with a tick. Participants had unlimited time to respond and could change their rating as many times as they wanted. Once they were happy with their rating, they clicked a button at the bottom of the screen to continue to the next trial. There was a blank ITI of 2s between trials.

We also presented additional test stimuli after the morphed shape dimension for exploratory purposes. These stimuli were a set of shape outlines (triangle, circle, square), as well as "compound" stimuli consisting of the CS+ and CS- presented together (with left to right order of presentation counterbalanced across 2 presentations). These exploratory stimuli did not produce interesting results and will not be reported here (see Supplementary Materials).

Questionnaire Phase. Participants completed the 21item version of the DASS (Lovibond & Lovibond, 1995), and then the short 12 item version of the Intolerance to Uncertainty Scale (Carleton et al., 2007). In Experiment 1, participants completed additional exploratory questions (see Supplemental Materials) as well as the State-Trait Anxiety Inventory (STAI, Spielberger et al., 1983). In Experiment 2 only, after completing the questionnaires participants were presented with the outcome image (scary face or tone) and asked to rate how aversive they found the scream/tone on a visual analogue scale ranging from "Not aversive at all" to "Extremely aversive". Participants were then asked whether they kept their volume at the same level for the whole experiment by choosing amongst the options: I kept my volume at the same level for the whole experiment, I reduced or muted the volume for some, or all of the experiment, I increased the volume for some, or all of the experiment. This question was included to measure compliance but we did not exclude participants on the basis of this response since there were very few people who reported turning down the volume.[2]

Results

Exclusion Criteria

[2] 1 participant in the Neutral group, and 6 participants in the Aversive group reported reducing the volume.

Participants were excluded if they failed the training criterion, which was calculated using the last 4 CS+ and CS- trials in training. Participants had to be correct on 3 or 4 trials (i.e. accuracy >=75%) for both trial types in order to pass (consistent with Lee et al., 2018). This resulted in 15 participants being excluded in Experiment 1 (11.7%), and 17 participants being excluded in Experiment 2 (10.1%). We chose to exclude participants on the basis of training performance since any differences in acquisition may mean that generalization gradients originate from different starting points, complicating interpretation. When we analyzed the data without training exclusions, the pattern of results was largely the same (see Supplemental Materials).

As stated above, we additionally excluded participants who did not score within our predetermined ranges for low and high anxiety groups. This resulted in an additional 12 (Experiment 1) and 18 (Experiment 2) exclusions. After applying these exclusions there were 101 participants in Experiment 1 (59 high anxiety, 42 low anxiety) and 135 participants in Experiment 2 (35 high anxiety/neutral outcome, 35 low anxiety/neutral outcome, 35 high anxiety/aversive outcome, 30 low anxiety/aversive outcome).[3]

Training

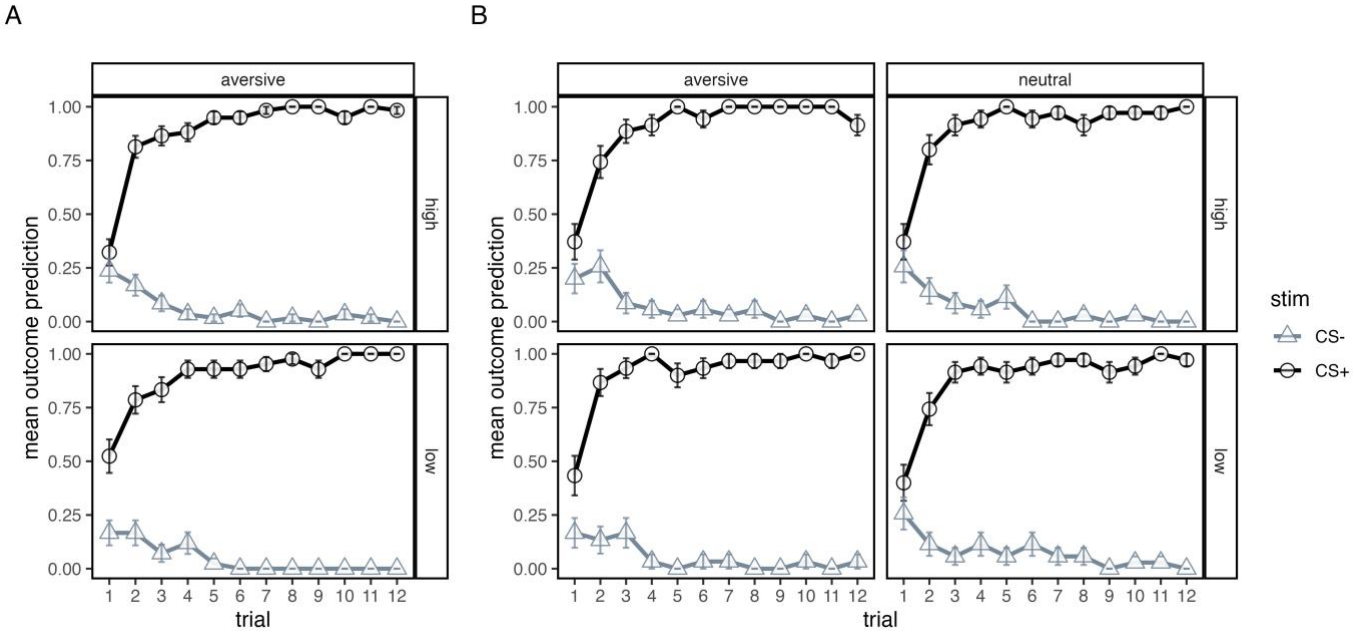
Figure 1 shows the training data for experiments 1 (panel A) and 2 (panel B). Overall, it appears that predictions of the outcome increased for CS+ and decreased for CS- over trials in a similar fashion for low and high anxiety groups, and for aversive and neutral outcomes in Experiment 2. To analyze the data, we computed mean predictions for the CS+ and CS- in the final block of training, and tested whether this

[3] Note that the pattern of results was very similar when we removed the participants scoring in the Moderate range from the High Anxiety group (see Supplemental Materials).

terminal performance differed as a function of the grouping variables in an ANOVA. In Experiment 1, there was a large main effect of stimulus (comparing CS+ to CS-), $F(1,99) = 26057.8$, $p < .001$, $\eta_p^2 = .996$, but no main effect of anxiety group, $F(1,99) = 2.73$, $p = .102$, $\eta_p^2 = .027$, and no interaction, $F < 1$. In Experiment 2, there was a main effect of stimulus, $F(1,131) = 8647.1$, $p < .001$, $\eta_p^2 = .985$, no main effect of anxiety group, $F < 1$, no main effect of outcome group (comparing neutral to aversive), $F(1,131) = 2.20$, $p = .140$, $\eta_p^2 = .017$, and no interactions, $Fs < 1.37$. Thus, there were no differences in terminal training performance between high and low trait anxiety groups in either experiment, and no differences between participants learning about a neutral or aversive outcome in Experiment 2.

Figure 2

Mean Outcome Predictions during Training for Experiments 1 (A) and 2 (B).



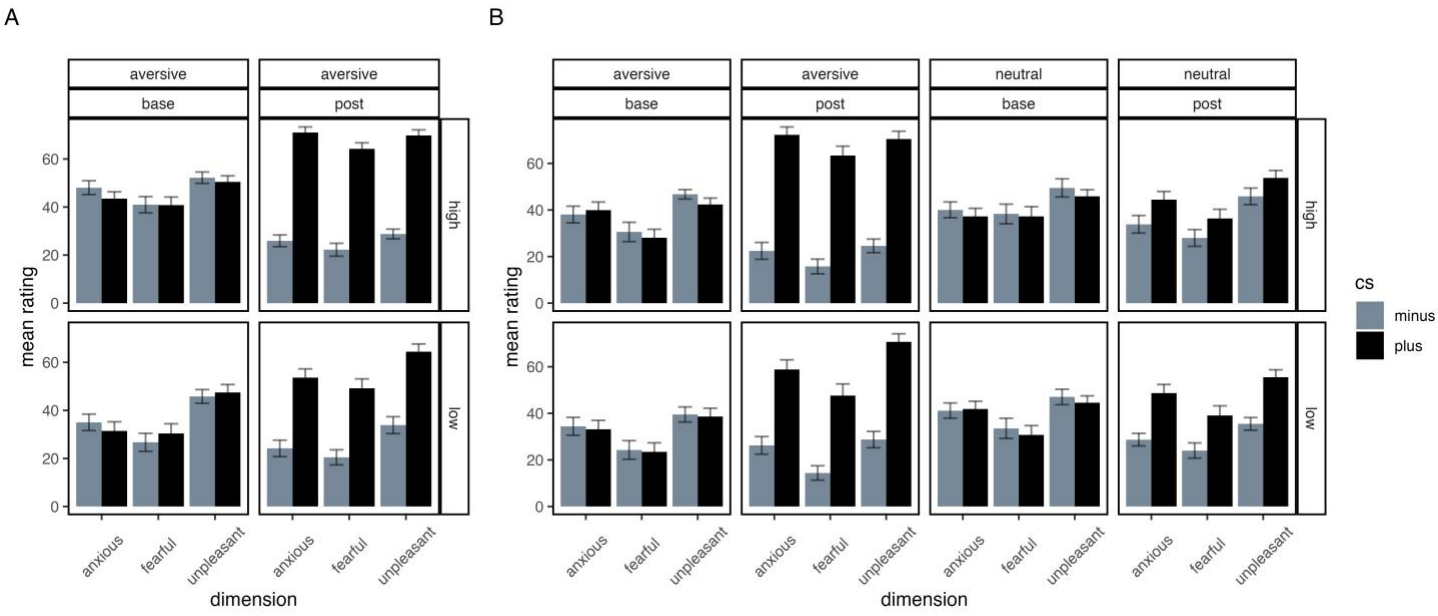
Note. Error bars represent standard error of the mean.

Affective Ratings

Figure 3 shows the mean ratings for each of the affective ratings at baseline (prior to training) and post-training in Experiments 1 (panel A) and 2 (panel B). From the figure, it is clear that for both experiments, baseline ratings do not differ between CS+ and CS-, but they diverge as expected after training with an aversive outcome, with the CS+ rated as more unpleasant, anxious, and fearful than the CS-. Interestingly, the pattern of results appears similar but weaker for the neutral outcome in Experiment 2. Although the tone was intended to be a neutral outcome, playing the tone at a similarly loud volume to the scream during training may have resulted in small changes in affective ratings to the CS+. Still, the affective ratings show that the scream experienced by the aversive group was much more aversive than the tone by the end of training.

Figure 3

Mean Affective Ratings (Anxious/Fearful/Unpleasant) Prior to (Baseline) and After (Post) training in Experiments 1 (A) and 2 (B)



Note. Error bars represent standard error of the mean.

The affective ratings were analyzed using ANOVAs run separately for each dimension (anxious, fearful, unpleasant). In Experiment 1, an ANOVA with anxiety group (low vs. high), CS (CS+ vs. CS-), and time (baseline vs. post) was run. In Experiment 2, the ANOVA contained outcome group (neutral vs. aversive) as an additional factor. We also ran ANOVAs on the CS+ and CS- separately which showed changes on all affective measures indicating successful fear and safety learning. The full set of results are reported in Supplemental Materials and were broadly consistent for the three affective dimensions. Therefore, we will report the results that were common across conditions and of most relevance for our research aims.

For the three affective dimensions (anxious, fearful, unpleasant) and in both experiments, there was a significant interaction between CS and time, smallest $F(1,131) = 106.8$, $p < .001$, $\eta_p^2 = .449$, reflecting the larger increase in affective ratings for CS+ than for CS- after training. In Experiment 1, this interaction further interacted with anxiety group for the anxious, fear, and unpleasantness ratings, smallest $F(1,99) = 5.41$, $p = .022$, $\eta_p^2 = .052$, reflecting the tendency for a greater divergence in ratings for the CS+ and CS- following training for high anxious participants. This three-way interaction between anxiety group, time, and CS did not occur in any of the affective ratings in Experiment 2, Fs < 1[4], perhaps because this interaction averages across outcome groups (aversive and neutral). Looking at the aversive group in Experiment 2 in Figure 3b, there does appear to be greater differential change in affect in the high anxious participants, similar to Experiment 1 (Figure 3a).

Unsurprisingly in Experiment 2, there was a three-way interaction between outcome group, CS, and time, across all three affective dimensions, smallest $F(1,131) = 19.7$, $p < .001$, $\eta_p^2 = .131$. This interaction reflects the greater increase in affect to

[4] The interactions were also not significant when we analyzed the aversive group alone.

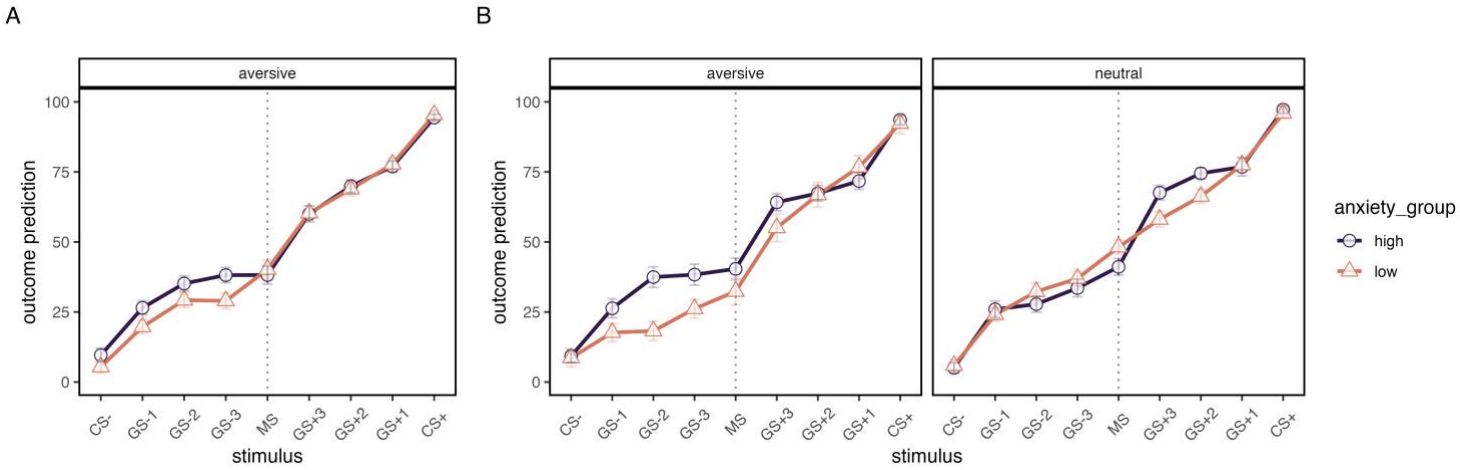
CS+ for participants trained with an aversive outcome, relative to those trained with a neutral outcome. Although the pattern of results appeared similar for all three dimensions, there was a further four-way interaction with anxiety group for the fearfulness ratings only, $F(1,131) = 5.73$, $p = .018$, $\eta_p^2 = .042$. This interaction can be explained by the observation that high trait anxious participants show greater differences in affective ratings between CS+ and CS- after training when they are trained with an aversive outcome, but for some reason, this pattern is reversed when trained with a neutral outcome (see Figure 3B). In summary, the aversive scream outcome produced successful fear conditioning in the aversive groups, and the changes in affect were stronger when an aversive rather than a neutral outcome was used. There was evidence that the changes in anxiety and fear were more pronounced for high trait anxious participants trained with the aversive outcome in both experiments.

Generalization Test

Figure 4 shows the generalization gradients for both experiments. Surprisingly, the figure shows that for participants trained with an aversive outcome in Experiment 1, there were no trait anxiety group differences in generalization from the CS+ (the right-hand side of the gradient). Instead, the group differences appear to be confined to generalization from the CS- (the left-hand side of the gradient) in Experiment 1 (see Figure 4A) and are much more pronounced in Experiment 2 (aversive group in Figure 4B). Interestingly, participants trained with a neutral outcome seem to show the opposite pattern of results, with greater generalization from the CS+ in high anxious participants, but no differences in generalization from the CS-. For these analyses, we report Greenhouse-Geisser corrections (Greenhouse & Geisser, 1959) for violations of sphericity.

Figure 4

Generalization Gradients in Experiments 1 (panel A) and 2 (panel B)



Note. CS+ = conditioned stimulus predicting the outcome, CS- = conditioned stimulus predicting nothing, GS = Generalization Stimulus, MS = morphing stimulus. The CS+ and CS- was randomly selected to be a black triangle or a black square, the MS was always a black circle. Error bars represent standard error of the mean.

Experiment 1. To test these observations in Experiment 1, we analyzed the generalization stimuli (GS) from the CS+ (i.e., GS+1, GS+2, GS+3) and generalization from the CS- (i.e., GS-1, GS-2, GS-3), in separate ANOVAs[5] with anxiety group (high vs. low), and stimulus (1-3) as factors. For the GS+ in Experiment 1, there was a significant main effect of stimulus, $F(1.95,193.4) = 44.4$, $p < .001$, $\eta_p^2 = .187$, but no effect of anxiety group, $F < 1$, and no interaction, $F < 1$. In contrast for the GS- in Experiment 1, there was a significant main effect of anxiety group, $F(1,99) = 4.57$, $p = .035$, $\eta_p^2 = .044$[6], and stimulus, $F(1.8,180.1) = 29.0$, $p < .001$, $\eta_p^2 = .139$, but a non-significant interaction, $F < 1$. There were no significant anxiety group differences for the CS+, CS-, or MS, highest $t(196.6) = 1.85$, $p = .066$. Thus, there were differences between high and low anxiety groups for the GS-, but not for the GS+.

[5] In addition to ANOVAs, we also conducted hierarchical regressions treating anxiety as a continuous predictor. These results are reported in Supplemental Materials and we have noted the minor differences with our main analysis.
[6] This effect was not significant was anxiety was treated as a continuous predictor.

Experiment 2. The data in Experiment 2 were analyzed in an ANOVA with outcome group (aversive vs. neutral), anxiety group (high vs. low) and stimulus (1-3) as factors. As in Experiment 1, the GS+ data were analyzed separately from GS-. For the GS+, there was a significant main effect of stimulus, $F(2.0,261.0) = 38.9$, $p < .001$, $\eta_p^2 = .130$, and significant interaction between stimulus and anxiety group, $F(2.0,261.0) = 6.92$, $p = .001$, $\eta_p^2 = .026$. This interaction reflects the higher generalization in the high anxiety group as the GSs become more dissimilar to the CS+. Critically, like Experiment 1, the main effect of anxiety group was not significant $F(1,131) = 1.96$, $p = .164$, $\eta_p^2 = .015$. All other main effects and interactions were not significant, Fs < 1.49.

In contrast for the GS-, there was a significant main effect of anxiety group, $F(1,131) = 4.77$, $p = .031$, $\eta_p^2 = .035$[7], and significant interaction between anxiety group and outcome group, $F(1,131) = 8.74$, $p = .004$, $\eta_p^2 = .063$, which further interacted with stimulus, $F(1,9,253.8) = 3.56$, $p = .031$, $\eta_p^2 = .014$[8]. These results support our observation that the high anxiety groups showed undergeneralization of safety learning with an aversive scream outcome, while there was no such effect for the neutral tone outcome (Figure 4B). There was also a significant effect of stimulus, $F(1,9,253.8) = 20.9$, $p < .001$, $\eta_p^2 = .076$. All other main effects and interactions were not significant, Fs < 1.2.

We also conducted analyses for the aversive group only in Experiment 2 to test whether the pattern of results in Experiment 1 replicated. For the GS+ in the aversive group, there was no main effect of anxiety group, $F < 1$, a significant main effect of stimulus, $F(1.9,118.6) = 15.7$ $p < .001$, $\eta_p^2 = .117$, and a significant interaction between

[7] This result was not significant when anxiety was treated as a continuous predictor.
[8] This result was not significant when anxiety was treated as a continuous predictor.

anxiety group and stimulus, $F(1.9,118.6) = 3.70$, $p = .030$, $\eta_p^2 = .030$. One difference between Experiment 1 is that in Experiment 2, the trait anxiety differences for the GS+ varied with similarity to the CS+ (i.e., interacted with stimulus). However, the lack of a significant main effect of anxiety group was consistent with Experiment 1. For the GS- in the aversive group, there was a significant main effect of anxiety group, $F(1,63) = 11.3$, $p = .001$, $\eta_p^2 = .152$, and stimulus, $F(1.9,120.9) = 8.15$, $p < .001$, $\eta_p^2 = .063$, but no interaction between anxiety group and stimulus, $F(1.9,120.9) = 2.23$, $p = .114$, $\eta_p^2 = .018$. Thus in the aversive group, we replicated the finding of greater expectancy of the scream outcome for the GS- (i.e., undergeneralization of safety learning) in the high anxious group.

For the GS+ in the neutral group, although there appears to be higher responding in the high anxious group (see Figure 4B), there was no significant main effect of anxiety group, $F(1,68) = 3.71$, $p = .058$, $\eta_p^2 = .052$. There was a significant main effect of stimulus, $F(1.9,128.3) = 24.8$, $p < .001$, $\eta_p^2 = .162$, and a significant interaction between anxiety group and stimulus, $F(1.9,128.3) = 3.73$, $p = .029$, $\eta_p^2 = .028$. For the GS-, there was a significant main effect of stimulus, $F(1.9,127.2) = 14.0$, $p < .001$, $\eta_p^2 = .099$, but no main effect of anxiety group, $F < 1$, and no interaction, $F(1.9,127.2) = 1.55$ $p = .218$, $\eta_p^2 = .012$.

For both neutral and aversive groups in Experiment 2, there were no significant anxiety group differences for the familiar training stimuli (CS+ or CS-), ts < 0.874. However, the high anxiety group showed higher outcome predictions for the MS in the aversive, $t(121.8) = 1.99$, $p = .049$, and lower outcome predictions in the neutral group, $t(130.4) = 2.30$, $p = .023$.

Predictors of Generalization

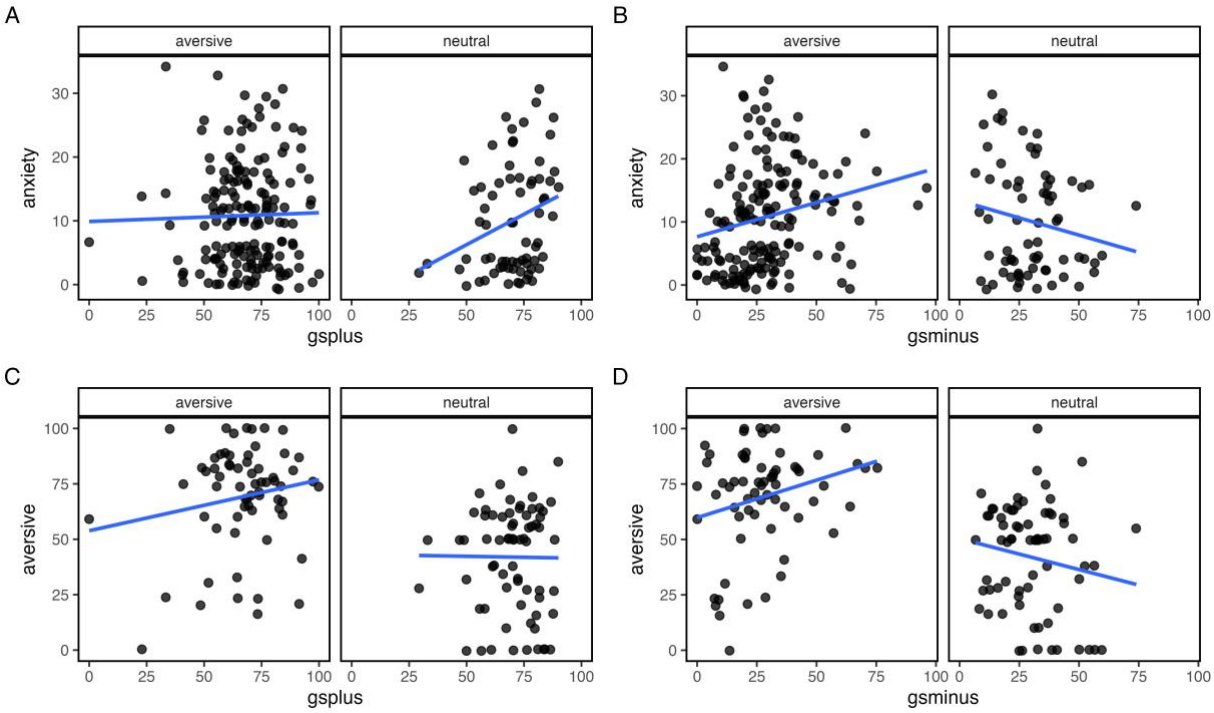
We conducted exploratory analyses to test whether participants' subjective ratings of the aversiveness of the outcome predicted generalization from the CS+ and CS- in Experiment 2 (Experiment 1 did not include aversiveness ratings). In these analyses, we computed 2 dependent variables (GS+, GS-) which were the average ratings for the GS+1, GS+2, and GS+3, and the GS-1, GS-2, and GS-3, respectively. We ran regressions with group as a categorical predictor, anxiety as a continuous predictor (mean-centered), and aversive rating as a continuous predictor (mean-centered). For the GS+, there were no significant main effects or interactions, highest $F(1,127) = 1.86$, $p = .175$. For the GS-, there was a significant main effect of anxiety, $F(1,127) = 7.00$, $p = .009$, and outcome group interacted with anxiety score, $F(1,127) = 5.48$, $p = .021$, as well as the aversiveness rating, $F(1,127) = 4.18$, $p = .043$, while the remaining effects were not significant, highest $F(1,127) = 2.71$, $p = .102$.

These results are broadly consistent with those reported above. In addition, they show that that both anxiety scores and subjective ratings of the aversiveness of the outcome predict generalization from the CS-, but not the CS+, with these relationships being stronger for aversive over neutral outcomes. These findings were supported by individual correlations conducted within each group, and stimulus set (GS+/GS-) separately (see Figure 5). For the aversive group, the aversive rating was positively correlated with the GS-, $r = .248$, $t(63) = 2.04$, $p = .046$, but not with the GS+, $r = .169$, $t(63) = 1.36$, $p = .178$. In the neutral group, the aversiveness rating was not correlated with the GS-, $r = -.160$, $t(68) = -1.34$, $p = .185$, nor the GS+, $r = -.009$, $t(68) = -0.076$, $p = .940$. Anxiety was positively correlated with the GS- in the aversive group, $r = .372$, $t(63) = 3.18$, $p = .002$, but not in the neutral group, $r = -.179$, $t(68) = -1.50$, $p = .138$,

while anxiety was correlated with the GS+ in the neutral group, $r = .279$, $t(68) = 2.39$, $p = .020$, but not the aversive group, $r = .073$, $t(63) = 0.580$, $p = .564$.

Figure 5

Scatterplots of Predictors (Ratings of Outcome Aversiveness [C and D] and Anxiety Score [A and B]) of Generalization from the CS+ (A and C) and CS- (B and D)



General Discussion

In two experiments, we used a morphed shape dimension created by Lemmens et al. (2021) to separately measure generalization from a fear cue predicting an aversive outcome, and generalization from a safety cue predicting nothing. This design allowed us to test whether differences in generalization between high and low trait anxious participants were driven by overgeneralization of fear learning, or undergeneralization of safety learning. Our main finding was that high trait anxious participants undergeneralized safety learning, but did not overgeneralize their fear learning. These results were consistent between experiments, and Experiment 2 showed that this result

was dependent on the use of an aversive outcome (the scream). In a control group of participants trained with a neutral tone outcome, there were no significant differences between high and low anxiety groups, the subjective aversiveness of the scream predicted generalization from the CS- but not the CS+, and the aversiveness of the tone predicted did not predict generalization from the CS- nor CS+. Since there were no trait anxiety group differences for the safety cue in training nor test, these deficits appear to be specific to *generalization* of safety learning rather than safety learning itself.

To our knowledge, this was the first study to investigate trait anxiety effects by separating overgeneralization of fear from undergeneralization of safety learning on a continuous stimulus dimension. Surprisingly, generalization of safety learning varied more as a function of trait anxiety than fear learning. In other words, stimuli that were clearly dissimilar from known fearful stimuli differed more as a function of trait anxiety than more similar stimuli. This result is significant given that the typical characterization of differences in generalization between anxiety groups is in terms of overgeneralization of fear learning (e.g., Dunsmoor & Paz, 2015; Lissek, 2012; Beckers et al., 2022; Cooper et al., 2022; Fraunfelter et al., 2022). While heightened fear responses to intermediate stimuli varying between the CS+ and CS- is consistent with overgeneralization of fear, the stimulus dimensions that are typically used do not allow us to determine whether undergeneralization of safety learning contributes to these results. Our work builds on previous studies by separating fear and safety generalization to determine which process underlies these findings.

While there are very few studies that have investigated generalization of safety learning, there is at least one other study that has shown consistent results. Greenberg et al. (2013) compared patients with GAD with healthy controls in their neural responses to fear generalization. They found that patients with GAD showed impaired

reactivity in the ventral medial prefrontal cortex in response to novel generalization stimuli (but see Kaczkurkin et al., 2017), suggesting ineffective recruitment of fear *regulation* (similar to inhibition of fear). Interestingly, there were no differences in regions coding for fear *responding* even though there was no safety stimulus in their study. It should be noted however, that the authors found no group differences on the generalization test using likelihood ratings of the shock (similar to what we used here). Nevertheless, our results are compatible with Greenberg et al.'s results in suggesting that anxious individuals have a specific deficit in inhibition of fear (see also Jovanovic et al., 2010, 2012).

There are some peculiar aspects of our results that are worth commenting on. First, although the results were non-significant, there seemed to be a trend towards greater generalization for the GS+ (stimuli resembling the CS+) in the neutral group in Experiment 2. At present, we do not have a good explanation for this trend. Whatever the explanation, we note that this effect was absent when we re-analyzed the data without training exclusions (see Supplemental Materials). There was also an unexpected three-way interaction obtained in the affective ratings in Experiment 2. In the aversive group, the high anxiety group showed larger affective changes to CS+ (anxious/fearful/unpleasantness ratings) than the low anxiety group, which makes sense. The reverse pattern however, was found for the neutral group, with greater affective changes for the low anxiety group. As discussed, the change in affective ratings to the CS+ suggest that the neutral tone outcome may not have been completely neutral since it was played at a loud volume to match the aversive scream. As such, it may have acquired some aversive properties. This makes the interaction puzzling since if high trait anxious participants show greater affective changes to conditioning with aversive outcomes, the same result, albeit weaker, should have resulted in the neutral

group. Further experiments are needed to understand how the valence of the outcome interacts with generalization from the CS+ and CS-, and how this differs as a function of trait anxiety.

Safety Learning

Our results suggest that greater attention should be paid towards safety learning (cues that predict the absence of aversive outcomes) when it comes to understanding and treating maladaptive generalization of fear. At a practical level, for any traumatic event, there will be a single set of CS+ associated with that episode, but countless CS- that are not associated with the aversive event that may become implicit safety signals. If minor variations in these safety cues create uncertainty about their meaning (i.e., constitute a "weak" situation, Lissek et al., 2006), then the ubiquity of these cues may pose a greater problem than the possibility of encountering stimuli resembling the original threat. Our results therefore have important clinical implications for the types of interventions we should use to target maladaptive generalization.

Many gold-standard treatments for anxiety disorders focus on addressing responses to fear-related cues and attempt to maximize expectancy violation during exposure therapy (Craske et al., 2014) or promote generalization of extinction by using multiple fear-relevant stimuli or extinction in multiple contexts (Bustamante et al., 2024). Other approaches aim to directly target excessive fear generalization by increasing discrimination between GSs resembling the CS+. Our results suggest that these approaches may be complemented by directly targeting the subjective ambiguity and potential threat value of safety cues, and by encouraging generalization of safety in novel situations. One way to achieve this would be encouraging encoding of safety as a discrete, concrete outcome in and of itself which may potentially reduce ambiguity

(see Laing et al., 2024). Framing the absence of a fearful event as a distinct, concrete, and predictable event may reduce the ambiguity involved in generalization, and produce a second memory that competes more effectively with the fear memory (Laing et al., 2024). These conceptual considerations are important for future research as they may help explain why safety learning is encoded and generalized differently from fear learning.

Safety learning is often understood as an example of *conditioned inhibition*, where a safety cue has a negative contingency with an outcome and acquires negative associative strength (Laing et al., 2025; Laing and Harrison, 2021; Christianson et al., 2012)[9]. A conditioned inhibitor (X) is obtained from a feature negative procedure (A+ AX-) and inhibition is inferred if X suppresses responding to another trained excitor in a summation test (Rescorla, 1969, but see Chow et al., 2022 and Lee et al., 2025 for an alternative view). Although a CS- in a differential design like the one employed here does generate some degree of inhibition according to associative models (e.g., Rescorla & Wagner, 1972), it is not the same as a feature negative procedure. We can therefore question whether true safety learning occurred. While we did not include a formal summation test in our experiment, there are two features of our data that indicate that participants learned that the CS- signaled safety. The first is that test trials where the CS+ and CS- were presented together in compound generated much lower responses (M=61.6 in Experiment 1 and M = 62.1 in Experiment 2) compared to the CS+ alone (M=94.8 in Experiment 1 and M=94.8 in Experiment 2, see Supplemental Materials). Thus, the CS- showed successful transfer, suppressing responding to an excitatory fear cue. The other source of evidence comes from the affective ratings. In both

[9] Note that conditioned inhibition is interpreted in different ways by different theories (e.g., Rescorla & Wagner, 1972; Konorski, 1967; Fraser & Holland, 2019; Pearce & Hall, 1980) and there seem to be individual differences in how humans learn about negative contingencies (Lee & Lovibond, 2021; Lovibond & Lee, 2021; Chow et al., 2022; Lee et al., 2025) which align with different theories.

experiments, there were significant reductions in fear, anxiety, and unpleasantness ratings from pre- to post-training for the CS-. Together, these findings suggest that participants did not learn that the CS- was redundant or meaningless, rather, it signaled safety from the aversive outcome.

Fear vs Safety Generalization

One of the aims of our study was to separate fear and safety generalization. It is possible that our use of Lemmens et al.'s (2021) morphed shape dimension was not fully effective in this regard. Since the assessment of inhibitory (i.e., safety) learning necessarily requires excitatory learning to counteract, an alternative explanation of our results is that high anxious participants show higher levels of fear generalization, but only at low levels. While valid, there are a number of reasons to dispute this interpretation. As discussed above, this idea goes against our natural tendency to discretize continuous dimensions using verbal labels (e.g., "left" and "right" in Wong & Lovibond, 2017). This tendency is evident in the unique gradient shapes obtained from these segmented dimensions, and is so pervasive that it has driven researchers to create complex, non-verbalizable stimuli to isolate associative generalization in humans (e.g., Lee & Livesey, 2018; Wills & Mackintosh, 1998; Livesey & McLaren, 2007).

Nevertheless, it is still possible that some fear generalized from the CS+ to the GS-, or that the generalized fear in the GS- originated from some other source. For example, some degree of fear may have conditioned to the common features of the fear and safety cues (black shapes, location on screen). While it is difficult to rule out this possibility, we note that theories (e.g., Mackintosh, 1975, see Le Pelley et al., 2016 for a review) and empirical findings (relative signal validity, Wagner et al., 1968) from the

associative learning literature suggest that this learning should be minimal since diagnostic features (i.e., shape) should gain attention over non-diagnostic features. This means that associative strength (i.e., fear) should accrue primarily to the *unique* features of the CS+ (e.g., square-ness), which should be absent in the GS- since these stimuli resemble the other stimulus. In other words, if we conceptualize the CS+ as AX and the CS- as BX, A should acquire excitatory associative strength, B should acquire inhibitory associative strength, and X should be ignored because it is irrelevant. Even if we assume that the common features (X) acquired some associative strength, we still need to explain why we only found group differences for GS- when these common features are present in GS+ and GS- (and the morphing stimulus, which did not show clear differences). At the very least, our results demonstrate how trait anxiety differences in fear generalization can be parameter-dependent, and provide information about boundary conditions for what types of situations can be considered "weak" (Lissek et al., 2006).

We hope that our results motivate researchers to consider generalization of safety learning as a distinct process when explaining differences in fear generalization. As stated above, this idea aligns with mechanisms in formal models of stimulus generalization (Spence, 1937, Blough, 1975; Ghirlanda & Enquist, 1998; McLaren & Mackintosh, 2002). Associative models of generalization posit the existence of two processes: excitation originating from the CS+ and inhibition originating from the CS-. These usually take the form of a Gaussian activation function, which will overlap if the CS+ and CS- are similar and lie on the same dimension. This overlap leads to interaction via error-correction mechanisms (Blough, 1975; Ghirlanda & Enquist, 1998; McLaren & Mackintosh, 2002) or simple subtraction (Spence, 1937). Thus, the

idea that generalization of fear learning occurs alongside generalization of safety learning is consistent with established theoretical approaches to generalization.

The differences we observed in generalization are somewhat consistent with asymmetries in trait anxiety differences found in *acquisition* of fear learning. Duits et al. (2015) found in a meta-analysis that clinically anxious participants show elevated fear responses to the CS-, but not to the CS+. Interestingly, the idea that anxious individuals may overgeneralize fear was borne from observations of impaired safety learning (e.g., Huggins et al., 2021; Gazendam et al., 2013, Lissek et al., 2008, 2010, 2014; McGregor et al., 2021, Grillon & Ameli, 2003, for reviews see Duits et al., 2015, Kauche et al., 2025; Beckers et al., 2023). Lissek et al. (2008) suggested that this result might be explained by generalization from the fear cue to the safety cue, since the two cues often share common features. Consistent with this interpretation, greater impairment in discrimination learning is seen in anxiety participants when the CS- is highly similar to the CS+ (Haddad et al., 2012). Therefore, impairment in safety over fear learning in anxious participants may be a feature of acquisition *and* generalization.

Limitations and Future Directions

There are some limitations worth noting in our study. First, we used a scream outcome (audio and image) in our fear conditioning procedure. Although this type of procedure has been shown to produce successful fear conditioning (McGregor et al., 2021; Purves et al., 2019), motivationally significant outcomes like electric shock are arguably more potent and ecologically valid. Our use of the scream also meant that our measure of fear was restricted to self-report (the affective ratings). Our affective ratings were taken at the beginning and end of training, rather than online in trial-by-trial recordings. Further, we capitalized on natural variations in trait anxiety in a

convenience sample. It is therefore unknown whether similar results would be found in participants with diagnosed anxiety disorders relative to non-anxious controls, but we note that the same pattern of results was found when we restricted the high anxiety group to the most severe DASS classifications (see Supplemental Materials).

A feature of the DASS is that it asks participants to rate how much each statement applied over the last week, which may indicate that it is capturing state rather than trait anxiety. However, the DASS has been shown to be stable over time (Lovibond, 1998), and the anxiety subscale was more highly correlated with trait rather than state anxiety in the STAI (see Supplemental Materials). A final critique is that we chose to dichotomize participants into high and low anxious groups which meant that we failed to use the full range of the trait anxiety continuum. Although we used criterion-based cutoffs instead of a median split, dichotomization can lead to inconsistent findings between studies employing different classifications making it difficult to compare results across studies (Londsdorf & Merz, 2017). Although the majority of our results were unchanged when we re-analyzed the data treating trait anxiety as a continuous variable, and using the STAI to divide participants into low and high anxiety groups, it would be useful to test whether our results replicate across stimulus sets, measures, and samples.

Conclusion

In conclusion, we believe that researchers should pay greater attention to generalization of safety learning as a potential transdiagnostic mechanism underlying anxiety disorders. We have argued that a carefully selected stimulus dimension can allow for separate measurement of generalization of fear learning and generalization of safety learning, allowing for assessment of individual differences in each. The results

in our study suggest that generalization of safety learning is more variable between individuals than generalization of fear learning. Thus, finding ways to promote the acquisition and generalization of safety learning may be a more useful way to treat disorders characterized by excessive spread of fear.

References

Ahrens, L. M., Pauli, P., Reif, A., Mühlberger, A., Langs, G., Aalderink, T., & Wieser, M. J. (2016). Fear Conditioning and Stimulus Generalization in Patients with Social Anxiety Disorder. *Journal of Anxiety Disorders, 44*, 36–46.

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.

Baumann, C., Schiele, M. A., Herrmann, M. J., Lonsdorf, T. B., Zwanzger, P., Domschke, K., Reif, A., Deckert, J., & Pauli, P. (2017). Effects of an Anxiety-Specific Psychometric Factor on Fear Conditioning and Fear Generalization. *Zeitschrift Für Psychologie, 225(3),* 200–213.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy, 35(1*), 79-89.

Bustamante, J., Soto, M., Miguez, G., Quezada-Scholz, V. E., Angulo, R., & Laborda, M. A. (2024). Extinction in multiple contexts reduces the return of extinguished responses: A multilevel meta-analysis. *Learning & Behavior*, *52*(3), 209-223.

Carleton, R. N., Norton, M. J. P., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders, 21(1)*, 105-117.

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*, 587-607.

Chow, J. Y.-L., Lee, J. C., & Lovibond, P. F. (2022). Inhibitory summation as a form of generalisation. *Journal of Experimental Psychology: Animal Learning and Cognition, 48(2),* 86-104.

Christianson, J. P., Fernando, A. B., Kazama, A. M., Jovanovic, T., Ostroff, L. E., & Sangha, S. (2012). Inhibition of fear by learned safety signals: a mini-symposium review. *Journal of Neuroscience*, *32*(41), 14118-14124.

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.

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1742), 20170025.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10-23.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*, 1-12.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The Sine Qua Non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(4),* 968-986.

Dibbets, P., van den Broek, A., & Evers, E. A. T. (2015). Fear conditioning and extinction in anxiety- and depression-prone persons. *Memory (Hove), 23(3),* 350–364.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety, 32*, 239-253.

Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: Behavioral and neural mechanisms. *Biological Psychiatry*, *78*(5), 336-343.

Dunsmoor, J. E., White, A. J., & LaBar, K. S. (2011). Conceptual similarity promotes generalization of higher order fear learning. *Learning & Memory, 18*, 156-160.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. *Behavior Therapy, 46*, 561-582.

Fraser, K. M., & Holland, P. C. (2019) Occasion setting. *Behavioral Neuroscience, 133(2),* 145-175.

Fraunfelter, F., 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 and Biobehavioral Reviews, 139,* 104707.

Greenberg T., Carlson, J. M., Cha, J., Hajcak, G., Mujica-Parodi, L. R. (2013). Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. *Depression and Anxiety, 30*, 242-250.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24(2),* 95–112.

Grillon, C., & Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology, 38*, 807-815.

Haaker, J., Lonsdorf, T. B., Schümann, D., Menz, M., Brassen, S., Bunzeck, N., Gamer, M., & Kalisch, R. (2015). Deficient inhibitory processing in trait

anxiety: Evidence from context-dependent fear learning, extinction recall and renewal. *Biological Psychology, 111*, 65–72.

Haddad, A. D. M., Pritchett, D., Lissek, S., & Lau, J. Y. F. (2012). Trait Anxiety and Fear Responses to Safety Cues: Stimulus Generalization or Sensitization? *Journal of Psychopathology and Behavioral Assessment, 34(3)*, 323–331.

Huggins, A. A., Weis, C. N., Parisi, E. A., Bennett, K. P., Miskovic, V., & Larson, C. L. (2021). Neural substrates of human fear generalization: A 7T-fMRI investigation. *NeuroImage, 239,* 118308–118308.

Inman, R. A., & Pearce, J. M. (2018). The discrimination of magnitude: A review and theoretical analysis. *Neurobiology of Learning and Memory, 153,* 118-130.

Jovanovic, T., Kazama, A., Bachevalier, J., & Davis, M. (2012). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology*, *62*(2), 695-704.

Jovanovic, T., Keyes, M., Fiallos, A., Myers, K. M., Davis, M., & Duncan, E. J. (2005). Fear potentiation and fear inhibition in a human fear-potentiated startle paradigm. *Biological psychiatry*, *57*(12), 1559-1564.

Jovanovic, T., Norrholm, S. D., Blanding, N. Q., Davis, M., Duncan, E., Bradley, B., & Ressler, K. J. (2010). Impaired fear inhibition is a biomarker of PTSD but not depression. *Depression and anxiety*, *27*(3), 244-251.

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.

Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the state-trait anxiety inventory. *Clinical Psychology Review, 82*, 101928.

Konorski, J. (1967). *Integrative activity of the brain*. Chicago, IL: University of Chicago Press.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). *Behavior Research Methods, 49,* 457-470.

Laing, P. A. F., & Harrison, B. J. (2021). Safety learning and the Pavlovian conditioned inhibition of fear in humans: Current state and future directions. *Neuroscience and Biobehavioral Reviews, 127*, 659-674.

Laing, P. A. F., Vervliet, B., Dunsmoor, J. E., & Harrison, B. J. (2025). Pavlovian safety learning: An integrative theoretical review. *Psychonomic Bulletin & Review, 32*, 176-202.

Lange, K., Kühn, S., & Filevich, E. (2015). " Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS One*, *10*, e0130834.

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin, 142(10),* 1111-1140.

Lee, J. C., Cahyadi, T., Lovibond, P. F., & Schlegelmilch, R. (2024). Effects of discrimination difficulty on peak shift and generalization. In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* (pp 541-547).

Lee, J. C., Chow, J. Y.-L., & Lovibond, P. F. (2025). Occasion setting in humans: Norm or exception? *Comparative Cognition and Behavior Reviews, 20*, 45-49.

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(12)*, 1955-1970.

Lee, J. C., Le Pelley, M. E., & Lovibond, P. F. (2022). Nonreactive testing: Evaluating the effect of withholding feedback in predictive learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 48(1*), 17-28.

Lee, J. C., & Livesey, E. J. (2018). Rule-based generalization and peak shif tin the presence of simple relational rules. *PLoS ONE, 13(9),* e0203805.

Lee, J. C., & Lovibond, P. F. (2021). Individual differences in causal structures inferred during feature negative learning. *Quarterly Journal of Experimental Psychology, 74(1),* 150-165.

Lemmens, A., Beckers, T., Dibbets, P., Kang, S., & Smeets, T. (2021). Overgeneralization of fear, but not avoidance, following acute stress. *Biological Psychology, 164*, 108151.

Lissek, S. (2012). Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms of Pavlovian fear-learning: The case for conditioned overgeneralization. *Depression and Anxiety, 29,* 258-263.

Lissek, S., Pine, D. S., & Grillon, C. (2006). The *strong situation*: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biological Psychology, 72*, 265-270.

Lissek, S., Power, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy, 43*, 1391-1424.

Livesey, E. J., & McLaren, I. P. L. (2007). Elemental associability changes in human discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes, 33(2)*, 148-159.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Drexler, S. M., Meulders, A., Nees, F., Pittig, A., Richter, J., Romer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas, J. M. P., & Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews, 77*, 247-285.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans – Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience and Biobehavioral Reviews, 80*, 703-728.

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology, 107(3),* 520-526.

Lovibond, P. F., & Lee, J. C. (2021). Inhibitory causal structures in serial and simultaneous feature negative training. *Quarterly Journal of Experimental Psychology, 74(12),* 2165-2181.

Lovibond, P. F., Lee, J. C., & Hayes, B. K. (2020). Stimulus discriminability and induction as independent components of generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(6*), 1106-1120.

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82(4),* 276-298.

Morriss, J., Wake, S., Elizabeth, C., & van Reekum, C. M. (2021). I doubt it is safe: A meta-analysis of self-reported Intolerance of Uncertainty and threat extinction training. *Biological Psychiatry Global Open Science, 1(3),* 171-179.

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

McGregor, T., Purves, K. L., Constantinou, E., Baas, J. M., Barry, T. J., Carr, E., Craske, M. G., Lester, K. J., Palaiologou, E., Breen, G., Young, K. S., & Eley, T. C. (2021). Large-scale remote fear conditioning: Demonstration of associations with anxiety using the FLARe smartphone app. *Depression and Anxiety*, *38*(7), 719-730.

Pedersen, T. (2024). transformr: Polygon and Path Transformations. R package version 0.1.5, https://CRAN.R-project.org/package=transformr.

Purves, K. L., Constantinou, E., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M., Sun, M., Margraf, J., Craske, M. G., Breen, G., & Eley, T. C. (2019). Validating the use of a smartphone app for remote administration of a fear conditioning paradigm. *Behaviour Research and Therapy*, *123*, 103475.

Razran, G. (1949). Stimulus generalization of conditioned responses. *Psychological Bulletin, 46(5*), 337-365.

Rescorla, R.A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*. NY: Appleton-Century-Crofts.

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 and Biobehavioral Reviews, 107*, 320-328.

Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P., Jacobs, G.A. (1983). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologist Press.

Tinoco-González, D., Fullana, M. A., Torrents-Rodas, D., Bonillo, A., Vervliet, B., Blasco, M. J., Farré, M., & Torrubia, R. (2015). Conditioned Fear Acquisition and Generalization in Generalized Anxiety Disorder. *Behavior Therapy, 46(5),* 627–639.

Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Caseras, X., Andión, O., & Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biological Psychology, 92(2*), 185–190.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171-180.

Wake, S., Hedger, N., van Reekum, C. M., & Dodd, H. (2024). The effect of social anxiety on threat acquisition and extinction: A systematic review and meta-analysis. *PeerJ, 12*:e17262.

Wong, A. H. K. & Lovibond, P. F. (2018). Excessive generalisation of conditioned fear in trait anxious individuals under ambiguity. *Behaviour Research and Therapy, 107*, 53-63.

Wong, A. H. K. & Lovibond, P. F. (2020a). Generalization of extinction of a generalization stimulus in fear learning. *Behaviour Research and Therapy, 125*, 103535.

Wong, A. H. K. & Lovibond, P. F. (2020b). Breakfast or bakery? The role of categorical ambiguity in overgeneralization of learned fear in trait anxiety. *Emotion, 21(4*), 856-870.