Affect labeling in fear extinction: can use of word labels enhance extinction?

A Master's Thesis submitted by

By: Meghan Whalen

in partial fulfillment of the requirements for the degree of

Master of Science

in

Psychology

Tufts University

February 2025

Advisor: M. Alexandra Kredlow

Abstract

Affect labeling is a strategy for decreasing negative emotions, such as fear and anxiety, and is considered a form of implicit emotion regulation. Initial studies have examined affect labeling in individuals with anxiety disorders as a strategy to enhance exposure therapy, a common treatment for these disorders. There is some evidence to suggest that physiological responses are attenuated in response to viewing labeled emotional stimuli, specifically negative images. However, to our knowledge there have been no studies to date that have examined affect labeling in healthy participants in a laboratory setting using the laboratory analog for exposure – fear extinction, nor have these studies considered potential individual differences in affect labeling ability. The present study aims to 1) investigate how individual differences in affect labeling ability relate to fear extinction outcomes in participants and 2) examine whether viewing labeled emotional stimuli during fear extinction results in reduced threat responses during extinction and during a test for return of fear (the laboratory analog for clinical relapse). Healthy participants (n_1 = 54 and n_2 = 37) underwent a classical fear acquisition and extinction procedure with a test of spontaneous recovery and performed an affect labeling test. Some participants were assigned to modified extinction procedures that included either negative emotion word labels or neutral content word labels. Skin conductance responses and shock expectancy ratings were recorded as indicators of fear response. Results investigating fear extinction success and scores on the affect labeling test revealed no significant relationship of individual ability to label emotions and extinguish and prevent return of fear. Preliminary observations comparing the groups that underwent a modified negative word extinction and that underwent a modified neutral word extinction indicate that the design of the study is mostly effective and might point to patterns of altered extinction and return of fear when a negative emotion word is present during extinction. A larger sample is needed to confirm these findings. This type of research provides an important foundation for future investigations of affect labeling as a potential strategy to enhance extinction learning. Continued work should further attempt to understand how affect labeling functions and its impact on extinction outcomes, so as to enhance extinction outcomes and improve clinical treatments.

Acknowledgements

I first want to thank the most supportive, encouraging, and understanding advisor, Dr. Xandra Kredlow. I have learned so much from her mentorship the past few years and this project was made possible by her guidance and support.

I thank my committee members Dr. Heather Urry and Dr. Joseph Dunsmoor for their expertise and collaboration.

I am grateful to the Human Extinction Workgroup for the use of the paradigm and to Dr. Samuel E. Cooper for analysis consultation.

I am lucky to have a whole team of past and current colleagues that have aided in all phases of this project, especially my undergraduate research assistants and Uku-Kaspar Uustalu in the Data Lab.

Finally, the perseverance in completing this project is owed in large part to my family and friends, especially Brittany Blasetti, M.S.

Table of Contents

I. Introduction
   a. Study Aims
II. Methods
   a. Participants
   b. Procedures
   c. Analysis Plan
III. Results
IV. Discussion
V. Supplementary Materials
VI. References

Introduction

Anxiety disorders, obsessive compulsive disorder (OCD), and other fear-related disorders have a high lifetime prevalence (e.g., anxiety disorders lifetime prevalence: 33.7%; Kessler et al., 2012). These disorders are not only damaging for individuals, but also have detrimental costs for the healthcare system and broader society (Bandelow & Michaelis, 2015). Although current cognitive and exposure-based treatment methods have been shown to be effective, a range of 10-50% of individuals receiving these treatments remain symptomatic or experience a relapse of clinical symptoms following treatment (Arch & Craske, 2009; Bystritsky, 2006; Craske & Mystkowski, 2006; Schottenbauer et al., 2008). A central component of many of these disorders is negative fear memories. It is important to study negative fear memories in the lab so that we can seek to understand how fears develop and importantly, how negative memories can be targeted as a potential avenue for treatment (Craske et al., 2018).

Fear conditioning and extinction paradigms[1] are used as a laboratory analog to study how fear memories develop and diminish over time, respectively, in healthy populations (Milad & Quirk, 2012). Extinction learning is also an analog for exposure therapy, a common treatment for fear-based disorders (Craske et al., 2018). Return of fear responses, by which the initial fear reappears some period of time after undergoing extinction, is a parallel for relapse following exposure (Rachman, 1989). Extinction is a relatively weak process, and for decades, researchers have been using fear conditioning and extinction paradigms to examine potential strategies to enhance extinction learning in the lab, with the ultimate goal to be able to translate these findings back to the clinic (Craske et al., 2018). A key part of this initial research is understanding first how individual difference factors in the population impact fear extinction outcomes. This may provide insight into how different people would respond to exposure-based treatments and factors that could be targeted to influence exposure therapy success.

The existence of individual variability in how people acquire and recover from fear has been extensively documented (Bush et al., 2007; Holmes & Singewald, 2013; Lonsdorf & Merz, 2017; Shumake et al., 2014). A number of previous neuroimaging studies have also confirmed individual

[1] Note, the choice to use "fear conditioning" is based on commonly used terminology in the literature. We acknowledge that there is some debate as to whether the term "fear" correctly portrays the emotions experienced by participants (Barrett, 2012; LeDoux, 2017)

differences in how people neurobiologically respond to fear conditioning and extinction (Dunsmoor et al., 2011; Hartley et al., 2011; Indovina et al., 2011). Because there is an inherent role of emotional processing that occurs during fear reduction, emotional processing is an important direction of future research on individual differences (Foa & Kozak, 1986; Foa et al., 2006; Olatunji et al., 2007). Fear memories are highly emotional, and the method of eradicating those memories also involves processing the associated emotions. Indeed, some researchers conceptualize fear extinction as a form of emotion regulation (Hartley & Phelps, 2010). Extinction of a fear memory requires inhibition of fear responses to the associated stimulus, and attention to threat cues and self-awareness (Craske et al., 2022). Therefore, emotion processing and regulation are likely important to the process of fear reduction, and fear extinction in the lab.

Across countries, cultures, and generations, emotion is experienced differently from person to person (Winter & Kuiper, 1997). And further, how people recognize, identify, and regulate those emotions also varies across individuals (Bonanno & Burton, 2013; Gohm, 2003; Lieberman et al., 2011). Understanding how individual differences in the ability to process and regulate emotion therefore may be important to the study of fear learning and especially, fear extinction (Hartley & Phelps, 2010). Other emotion regulation strategies (e.g., cognitive reappraisal) are already known to improve fear extinction outcomes (Hermann et al., 2014; Kitamura et al., 2022), but the emotion regulation strategy of affect labeling has not yet been examined.

Affect labeling is a strategy for managing negative emotions and is considered a form of implicit emotion regulation (Torre & Lieberman, 2018). Affect labeling is consistently defined and studied in laboratory settings as describing an emotion you observe in another person or feel in yourself (Hariri et al., 2000; Lieberman et al., 2007). It is both an internal and external skill that is reflexive according to established literature (Salovey & Mayer, 1990). Neuroimaging research suggests that affect labeling dampens the fear memory brain circuitry. Specifically, affect labeling results in reduced activity in the amygdala, and increased activity in the right ventrolateral prefrontal cortex and medial prefrontal cortex – a top-down inhibitory pathway (Lieberman et al., 2007). These same neural circuits were also observed to

function in a top-down manner in a large systematic review of functional connectivity during emotion processing in healthy participants (Underwood et al., 2021). In further support of this dampening theory, there is physiological evidence of an attenuation of skin conductance response (a physiological measure of emotional arousal assessed by sweat levels often on the hand or fingertips) in people who view labeled emotional stimuli compared to non-labeled emotional stimuli (Tabibnia et al., 2008). Additionally, there is evidence to support that labeling may also decrease self-reported distress while viewing negative stimuli (Burklund et al., 2014; Lieberman et al., 2011; Constantinou et al., 2014).

Given these findings from cognitive neuroscience research on how affect labeling dampens fear memory circuitry, there have been some initial studies examining affect labeling in individuals with anxiety disorders and OCD as a strategy to enhance exposure therapy (Kircanski et al., 2012; Kreiser et al., 2019; Niles et al., 2015). Even though it may seem like a simple addition, prior research suggests that instructing people to label emotions during exposures may improve therapy outcomes in populations with OCD (Kreiser et al., 2019), public speaking anxiety (Niles et al., 2015), and specific phobia (Kircanski et al., 2012). For example, in the Kreiser et al. (2019) study, when participants chose the emotion they felt from a word list as they were undergoing an exposure event, there was an observed reduction in skin conductance response over time during the exposure and at a 1-week follow-up test compared to participants undergoing exposure alone.

However, all three previous studies showed that reduction in physiological responses were not always accompanied by a reduction in self-reported fear. Kircanski et al. (2012), Kreiser et al. (2019), and Niles et al. (2015) found that SCR and/or heart rate decreased with affect labeling in exposure groups but self-reported fear levels did not show the same decrease. In addition, the participants who chose more anxiety words during the affect labeling activity showed lower SCR responses at a one week re-test (Kircanski et al., 2012; Niles et al., 2015). Finally, Kreiser et al. (2019) interestingly found that choosing an emotion word from a list, but not spontaneously coming up with an original label, resulted in a decreased SCR during exposure. Given these somewhat mixed results, still not enough is known about affect labeling to establish and maximize it as an enhancement to exposure therapy treatment.

Additionally, previous studies were conducted in clinical populations without taking into account any potential individual differences in affect labeling ability. In people with anxiety and other fear-related disorders, there are often deficits in these inhibitory learning and top-down regulatory pathways (Craske et al., 2022), especially during the inhibitory learning that is necessary during extinction (Sehlmeyer et al., 2011). As such, it is likely that in these populations there may also be deficits in affect labeling ability. It is also reasonable to expect that individual variability in affect labeling ability exists within the general population. Differences in emotional experience and emotion regulation can be observed in clinical populations and healthy populations alike (Burklund et al., 2014; Sheppes et al., 2015). Further, in healthy populations past studies have observed differences specifically in how people label emotions and rate affect (Constantinou et al., 2014). It is useful to study affect labeling as it relates to fear reduction in the general population prior to examination in clinical populations in order to gain understanding of the healthy functioning and pre-existing differences of affect labeling ability.

Although "affect labeling" has been specifically defined as describing an emotion you observe in another person or feel in yourself, studies have operationalized it in various ways (Torre & Lieberman, 2018). Previous cognitive neuroscience studies that explored affect labeling most often used affect matching designs, e.g., matching pairs of emotional faces or choosing between two listed emotion words corresponding to an emotional face presented (Lieberman et al., 2007). This design emphasizes direct recognition and simple discrimination of emotions in others and does not require spontaneous verbal labelling. In contrast, in clinical studies, participants will often choose an emotion word from a given list or will choose a subjective self-label for what they are feeling during an exposure event (Burklund et al., 2014; Constantinou et al., 2014; Kircanski et al., 2012; Kreiser et al., 2019; Niles et al., 2015). Yet, there are other implicit aspects to putting feelings into words that may be important, for example, affect naming of emotional faces or verbal expressions of emotion. The studies conducted here aim to address these limitations and examine: 1) whether affect labeling as an individual difference factor relates to extinction learning and return of fear, and 2) whether an affect labeling manipulation can serve to enhance extinction and prevent the return of fear.

Study 1 Aims

The primary goal of Study 1 was to investigate whether individual differences in affect labeling ability relate to fear extinction outcomes. To our knowledge, no study has examined the relationship between affect labeling ability and fear extinction. The Comprehensive Affect Testing System (CATS; Froming et al., 2006) was used as a measure of affect labeling ability. The CATS, developed to be a more comprehensive assessment of more nuanced affect identification and labeling abilities, has been effective in prior research in detecting how certain individual characteristics, like age, are related to affect labeling performance in healthy individuals (Schaffer et al., 2009). Given the prior research of affect labeling in clinical populations and studies observing changes in neural, physiological, and self-reported fear responses when healthy individuals are instructed to label affect, we expected that affect labeling performance would be worse in participants who also had poorer fear extinction outcomes. Study 1 completed independent data collection (n = 54) and analyses also include the control group collected during Study 2 (n = 13), for a total sample analyzed n = 67.

Study 2 Aims

The goal of Study 2 was to examine whether viewing affectively labeled conditioned stimuli during fear extinction resulted in reduced fear responses during extinction and during a test for return of fear. Study 2 provides pilot data to analyze this question and serves as the foundation for the future fully powered study examining this manipulation. The sample size was n = 37, with n = 13 control participants receiving standard extinction and two additional experimental groups (n=12 each).

This study adds to the current literature in that no study has examined affect labeling within the context of a fear conditioning and extinction paradigm. This is an important step because extinction is a well-established model for exposure therapy. Additionally, neutral cues that become threatening after repeated pairings within an unconditioned stimulus during conditioning, may better represent how fear and anxiety responses develop to neutral cues in the natural world in the context of fear-related disorders. The impact

of affectively labeling these stimuli may differ from labeling stimuli that are inherently negative in valence (e.g., IAPS pictures) as has been done in past research. The fear conditioning model also allows for examination of the impact of affect labeling at the time of labeling and at a future test of return of fear responses.

Methods

Participants

Participants were sampled from the community surrounding Tufts University. Recruitment efforts involved flyers and online postings for participation credit of academic credit for university students and cash payment for community members, or a combination of the two. The study inclusion criteria were: (1) Healthy adults aged 18-50; (2) Normal or corrected-to-normal vision and hearing; (3) Fluent in English.

The study exclusion criteria were: (1) Participation in another experiment involving electric stimulation within the past 6 months; (2) Current neurological, endocrine or psychiatric disorder, or treatment for one of these disorders within the past year; (3) Medical issues (e.g., seizures, heart conditions, pregnancy) or implantable medical devices (e.g., pacemakers, defibrillators) that contraindicate fear conditioning; (4) Color blindness; (5) History of unusual adverse reactions to pain or uncomfortable physical stimuli; (6) Current regular medication use, with the exception of oral contraceptives and acetaminophen; (7) Cigarette smoking or any drug use (i.e., marijuana, illegal drug use) weekly or more; (8) Consuming more than 3 alcoholic drinks/day on average; (9) Consumption of caffeine within 2 hours of the study appointments; (10) Consumption of alcohol, marijuana, or nicotine on the day of study appointments; (11) Use of illegal substances within 24 hours of a study appointment. These exclusion criteria were chosen in order to recruit a relatively healthy sample, reduce confounding factors that could impact physiological responses, and/or reduce risks to participants.

We predicted, based on our experience in this field and previous research studies (Schaffer et al., 2009), that a sample size of 60 was adequately powered to detect a moderate to large effect hypothesized for Study 1 and that a sample size of 36 was sufficient to ensure feasibility of the design and detect a trend in the associations we hypothesized for Study 2.

Procedures

Procedures common to both studies 1 and 2 are described first, followed by an explanation of unique components of studies 1 and 2.

I. Overview

This study was approved by the Tufts Institutional Review Board and preregistered with Open Science Framework prior to the start of data collection (January 25, 2023; osf.io/xby6m). The fear conditioning protocol used was developed by the Human Extinction Workgroup of the Exposure Therapy Consortium (osf.io/r42tu). This protocol has been tested by multiple research groups and results have demonstrated adequate acquisition, extinction, and return of threat responses to assess extinction enhancement strategies.

The study procedures took place across two subsequent days at the same time of day. On Day 1, all participants read and sign the informed consent form. Consented individuals completed an eligibility screening questionnaire assessing the criteria outlined above. If eligible, participants were set up with the equipment and underwent a simple cued fear conditioning and extinction procedure that consisted of the following phases: habituation, acquisition, extinction. On Day 2, participants underwent a test of spontaneous recovery. An overview of the procedure can be found in supplementary materials Figure 1.

The conditioned stimuli (CS+ and CS-) were colored shapes on the computer screen and the unconditioned stimulus (US) is a 200 ms shock to the wrist. The CS duration was 8 seconds and the inter-trial interval (ITI) was 10 +/- 2 seconds for all phases. After the test of spontaneous recovery on Day 2, participants completed the affect labeling task to assess affect labeling ability and other questionnaires to assess related constructs.

II. Fear Conditioning Procedures

Equipment, Shock Setting, and Baseline Period. Non-invasive disposable electrodes containing electrode gel were attached to the participant's right wrist for shock and connected to the Biopac STMISOC equipment. The shock level (0.4 – 4.0 mA) was calibrated before the start of the experiment using an ascending staircase procedure, during which participants chose their level of shock to be "highly annoying/uncomfortable but not painful." They also rated their perceived intensity for the shock on an Intensity Measurement Scale from 0 (no sensation) - 9 (very high intensity). Participants were also affixed with two noninvasive electrodes on the fingers of their left hand to measure skin conductance per established guidelines (Fowles et al., 1981; Lonsdorf & Merz, 2017). Participants went through calibration procedures to ensure they were set up with the equipment and skin conductance was being recorded correctly. Once the equipment was set up, participants then underwent a five-minute baseline resting period during which they sat in the dark experiment room with no stimuli presented to ensure that skin conductance levels returned to baseline following the shock selection procedure.

Habituation and Acquisition. Participants were next given instructions for the task. They were instructed to pay attention in order to learn the association between the colored shapes and the shock. During the habituation phase, participants saw 2 CS+ and 2 CS- stimuli that were not associated with shock for the purpose of initial familiarization to the stimuli. During the acquisition phase, 10 CS+ and 10 CS- were presented, with 70% of the CS+ trials co-terminating with the shock. The order of stimuli was pseudorandomized, such that each participant had a different stimuli presentation order (with no more than 3 of the same shape in a row). The CS+ conditioned stimuli shape (orange square or blue circle) was counterbalanced across participants, using block randomization.

Extinction. Immediately after acquisition, the extinction phase began. During the extinction phase, 20 CS+ and 20 CS- stimuli were presented without shock. The order was pseudorandomized as above.

Spontaneous Recovery. Twenty-four hours later, participants returned for a test of spontaneous recovery, a measure of return of fear response following extinction. They were set up with the equipment in the same manner as day 1, except the shock setting procedure was not repeated. Participants were, however, attached to the shock electrodes, the shock was set to the same level as day 1, and participants were told that if they were to be shocked it would be at the same level that they selected on day 1. During the test of spontaneous recovery, 12 CS+ and 12 CS- stimuli were once again presented without shock. The first stimulus presented (CS+ or CS-) was counterbalanced across participants to account for the fact that participants tend to have larger responses to the first stimulus presented on the test day.

III. Assessments

Skin Conductance

Skin conductance levels were recorded for all phases of conditioning using Biopac AcqKnowledge software.

Skin conductance responses (i.e., SCRs) were calculated for each CS+ and CS- using the Autonomate software (Green et al., 2014) method, by taking the trough to peak amplitude for the largest deflection during the period of time from stimulus onset (0 seconds) to just prior to the end (7.8 seconds). Because the US shock was 200 msec and co-terminated with the end of the stimulus at the 8 second mark, 7.8 seconds was used as the endpoint for the stimulus period, to separate it from capturing any skin conductance response to the shock itself. Non-measurable or no responses were scored as zeros. Per standard practice (Lonsdorf & Merz, 2017), skin conductance responses were square-root transformed prior to analysis. Differential SCRs were calculated by subtracting the average SCR to the CS- from the average SCR to the CS+. Analyses were conducted with all participants and then repeated after removing participants who had no differential SCR (differential SCR < 0) during the last half of acquisition given debate over this practice in the literature (Lonsdorf et al., 2019).

US Expectancy Ratings

Expectancy ratings are the most commonly used self-reported rating during fear conditioning procedures and given that it is unclear whether participants are experiencing fear or another feeling, expectancy ratings are the best representation of assessing anticipation of a negative contingency (Boddez et al., 2013; Constantinou et al., 2021). Participants rated US expectancy for each CS+ and CS- trial during all fear conditioning phases using a keyboard. Instructions were to indicate whether they expected to receive a shock to each CS presentation every time from the following options: "yes" (1), "unsure" (2), "no" (3). “Unsure” was used as a middle value between “yes” and “no” to indicate more of the transition of contingency learning that happens during each phase. For analyses, ratings were reverse-coded, such that higher ratings indicated higher expectancy.

Self-Report Questionnaires

Participants completed various self-report questionnaires throughout the study, including the DASS-21 and other questionnaires assessing demographics.

IV. Study 1. Assessment of Individual Differences in Affect Labeling

Following the spontaneous recovery session on Day 2, participants were disconnected from the fear conditioning equipment then completed the CATS-R abbreviated version (Froming et al., 2006) via Millisecond online testing software on a laptop. The testing system consisted of 13 subtests that measured different aspects of external affect awareness and emotional processing via pictures of faces with emotional expressions, sentences with emotional language, and audio of phrases that have emotional valence. The subtests followed this order: (1) Identity discrimination; (2) Facial affect discrimination; (3) Non emotional prosody discrimination; (4) Emotional prosody discrimination; (5) Name affect; (6) Identify emotional prosody; (7) Match affect; (8) Select affect; (9) Conflict – attend prosody; (10) Conflict – attend meaning; (11) Match prosody to face; (12) Match face to prosody; (13) Three faces. The faces were sourced from the Ekman and Friesen database of universally recognized and exhibit the six

major emotional expressions: happy, sad, angry, surprised, fearful, and disgusted (Ekman, 1994; Ekman & Friesen, 1976). The testing software was self-paced and instructions were automatically delivered via audio prior to the start of each subtest. There was no time limit and participants were allowed to click the "repeat" button to hear sentence audio for each question repeated within the subtests that involved audio. Research suggests that the CATS-R is a reliable and valid assessment of affect labeling differences in psychological or neurological populations compared to healthy controls and of individual differences in a large healthy population (Schaffer et al., 2009).

V. Study 2. Extinction Manipulation

Participants were randomized in a 1:1:1 ratio to one of three groups: 1) standard extinction (Study 1); 2) modified-negative extinction procedure; 3) modified-neutral extinction procedure. In the modified-negative extinction condition, emotion labels such as "scary" (negative) were placed on the CS+ stimuli such that each CS+ was labeled 50% of the time with a negative word and unlabeled 50% of the time. The CS- was not labeled at all. The following negative words were used based on previous studies and emotion word labels related to fear generated by ChatGPT: scary, negative, unpleasant, distressing, uneasy. In the modified-neutral group, neutral content labels such as "spatial" were placed on the CS+ stimuli such that each CS+ was labeled 50% of the time with a neutral word and unlabeled 50% of the time. The CS- was never labeled. The following neutral words were used: spatial, symmetrical, geometric, outlined, structured. Presentation order and labels were pseudorandomized across all participants, such that there were no more than 2 labeled or unlabeled stimuli in a row and no more than 2 of the same label presented in a row.

The rationale for the third group is an additional level of across group comparison that could not be achieved with just a 2 group design that would attempt to analyze responses to negative versus neutral labels. With this modified-neutral extinction group, responses to a negatively labeled CS+ (group 2) and

to a neutrally labeled CS+ (group 3) can be compared to analyze whether just mere presence of a label has an effect, or rather whether any effects are specific to negative emotion labels.

We considered the implications of conditioned inhibition (Hermans et al., 2006) by just altering extinction stimuli, not spontaneous recovery or acquisition, and decided that leaving spontaneous recovery and acquisition stimuli unchanged would be the best model for the real world relapse in treatment. To decrease the likelihood of conditioned inhibition, we presented the label on the same screen just below the CS+ shape 2-3 seconds following CS+ onset and also only presented the labels on 50% of the CS+ stimuli. One drawback of this is timing is that participants could input a US expectancy rating prior to seeing the label. This issue is likely only relevant to the first labeled CS+ where the participant may respond within the three seconds prior to the appearance of the word label. However, it is expected that overall US expectancy ratings will be influenced by general perceptions of the CS+ not just a single CS+ over the course of extinction. We will look at ratings across all CS+ regardless of label presentations. Additionally, the first CS+ during extinction was fixed such that it was always unlabeled in order to facilitate continuity from acquisition into extinction.

Study 1 Analysis Plan

Normality of data was assessed using Shapiro-Wilkes tests and confirmed by inspection of histograms. The SCR and expectancy rating data were non-normal for all phases of fear conditioning, as were the CATS-R scores. We proceeded with nonparametric statistical analyses. P values are presented in the results section and full statistical statements for all analyses are presented in Supplementary Table 1.

Preliminary Analyses:

First, robust linear mixed effects regression models were used to ensure that, on average, the sample demonstrated the expected pattern of responses during habituation, acquisition, extinction, and the test of

spontaneous recovery. These linear mixed models were used in place of Wilcoxon signed rank tests for their effectiveness in handling missing data, non-normal data, and individual differences. All linear mixed effects models contained a random effect of participant and trial number. The models were used to examine differences between the CS+ and CS- at various timepoints (first half, last half) using both SCR and expectancy ratings.

Outcome Analyses:

Second, to assess the relationship between fear extinction outcomes and performance on the affect labeling test, robust linear mixed effects regression models were performed with affect labeling score as a predictor. The two fear extinction outcomes utilized were speed of extinction and spontaneous recovery of fear.

Speed of Extinction was operationalized in the mixed models as interaction of trial and CS type from trial 1 to trial 3, thereby yielding a differential SCR/expectancy change. Although typically the greatest reduction in SCR will be observed in the 4-5 initial trials of extinction (Orr et al., 2000), in our data we observed that the greatest change in SCR occurred within the first 3-4 trials of extinction. As such, we chose to examine the change from Trial 1 to 3, rather than 1 to 4 as done by Lommen et al. (2013). If extinction is rapid, a large change in differential SCR would be demonstrated between trial 1 (high differential SCR) and trial 3 (low SCR). If extinction is slow, a smaller change in differential SCR would be demonstrated between trial 1 (high differential SCR) and trial 3 (medium-high differential SCR).

Spontaneous Recovery of Fear was operationalized as the differential of CS type during the first 2 trials of spontaneous recovery, with a large differential indicative or greater return of fear than a small differential for both SCR and expectancy ratings.

The main analysis of affect labeling used the CATS-R overall scores (out of 126), calculated from the sum of each subtests' number of correct answers out of total questions. Five of the subtests' scores (2, 5, 7, 10, 13) were corrected for male gender and age per the test manual guidelines. Subtests 1 and 3 were excluded from the total score because they were non-emotional tests, per the test manual guidelines. Additionally as secondary analyses, individual subtests 5 (name affect) and 7 (match affect) were examined as predictors in linear mixed models of the same two outcomes of extinction success, speed of extinction and return of fear, to compare any differences in modality of affect labeling (i.e., labeling vs. matching) which may elicit different responses (Lieberman et al., 2007; Torrisi et al., 2013).

Study 2 Analysis Plan

Normality of data was assessed using Shapiro-Wilkes tests and confirmed by inspection of histograms. The SCR and expectancy rating data were non-normal for all phases of fear conditioning, as were the CATS-R scores. We proceeded with nonparametric statistical analyses. Due to small sample size, preliminary statistics are presented for aspects of conditioning which should show expected patterns, but not main outcomes for which there may be too small of effect sizes to observe any meaningful patterns. P values are presented in the results section and full statistical statements for all analyses are presented in Supplementary Table 2.

Preliminary Analyses:

For Study 2, linear mixed effects regression models were not used due to small sample size. First, in order to examine whether there was a difference in SCR and expectancy rating fear responses between the three groups to the CS- stimuli, a Kruskal-Wallis rank sum test was used for each phase of conditioning. Next, Friedman rank sum tests were conducted separately for each group on SCR and ratings with the factor of CS type (CS+, CS-) across each phase or half phase to confirm adequate acquisition and extinction. Finally, Kruskal-Wallis tests were performed to ensure that all groups displayed similar differential SCR and ratings and in habituation and acquisition. Another Kruskal-Wallis test was performed on differential

SCR and ratings in the last quarter of extinction to check that there were no differences between the groups at the end of extinction, and that all participants successfully extinguished fear.

Outcome Analyses:

The means and standard deviations for SCR and expectancy ratings of each CS in each phase (first half, etc.) were calculated for each group and are reported in the Supplementary information.

For the main hypotheses, we examined group differences at two outcomes, extinction of fear and spontaneous recovery of fear. The first was operationalized as the differential SCR and ratings averaged across the first half of extinction. A differential that was smaller on average would indicate more rapid extinction. Next, after confirming that the differential SCR and ratings are typically small or near zero at the end of extinction, return of fear response will be operationalized as the average differential during the first 2 trials of spontaneous recovery. A large differential would be indicative of greater return of fear than a small differential.

Additionally, the means and standard deviations calculated for the average differential SCR and ratings in the first half of extinction and during the first two trials of spontaneous recovery were used to generate effect sizes for differences between pairs of the groups (i.e., negative vs. control and neutral vs. control). Hedges g was used as the effect size metric as it is more appropriate than Cohen's d for small sample sizes. Recommendations for effect size interpretation are: small ($g = 0.2$), medium ($g = 0.5$), and large ($g = 0.8$).

Results

Timeline

Study 1 data collection occurred between Spring 2023 to Spring 2024. Study 2 data collection occurred between Summer 2024 to Fall 2024.

Study 1 Results

Sample Characteristics

Of the 75 participants that consented to participate in the study, 9 failed to meet the inclusion criteria, 8 were withdrawn for not completing Day 2 procedures, and 4 participants' data were excluded from analyses for potential confounds (e.g., asked for shock level to be lowered on day 2). Thirteen additional participants consented, met study inclusion criteria, and had usable data as a part of Study 2 that were included in the final sample for Study 1 (n = 67). The demographic information and average CATS total scores, and subtest scores are presented in Table 1.

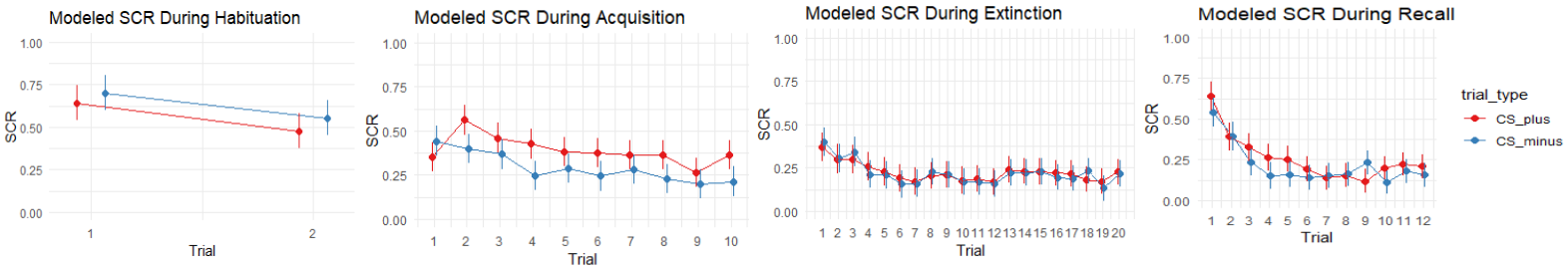
Table 1. Demographic Information of Study 1 Participants

	Overall (N=67)
Age	
Mean (SD)	22.0 (6.16)
Median [Min, Max]	19.0 [18.0, 49.0]
Sex	
Female	39 (58.2%)
Male	28 (41.8%)
Race	
Asian	30 (44.8%)
Black	5 (7.5%)
Multiracial	7 (10.4%)
Other	1 (1.5%)
White	24 (35.8%)
Ethnicity	
Hispanic or Latino	6 (9.0%)
Not Hispanic or Latino	61 (91.0%)
CATS_score	
Mean (SD)	94.3 (8.78)
Median [Min, Max]	96.0 [57.0, 112]
T5_score	
Mean (SD)	3.79 (1.38)
Median [Min, Max]	4.00 [1.00, 6.00]
T7_score	
Mean (SD)	8.93 (1.79)
Median [Min, Max]	9.00 [3.00, 12.0]

Preliminary Analyses

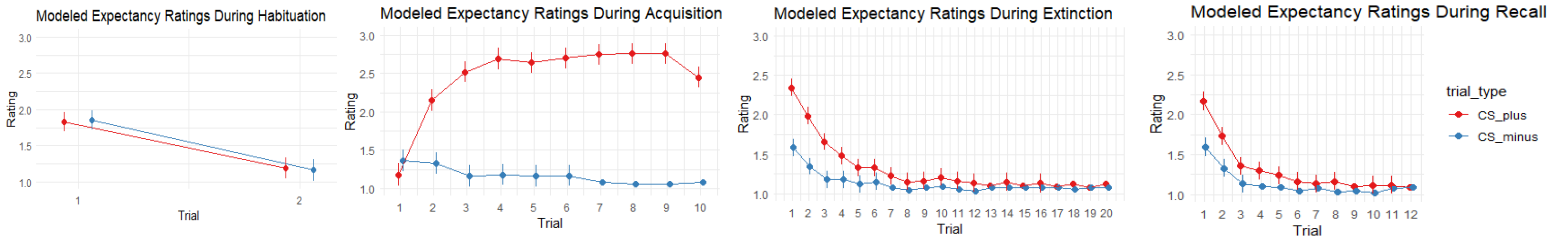
Linear mixed model analyses (n=68) showed no main effect of CS during habituation ($p = 0.068$), suggesting that participants did not come into the study with any pre-existing differences in response to the CS+ versus CS- stimulus. A significant main effect of CS during the second half of acquisition demonstrating successful conditioning ($p < 0.001$). To note, there were twenty identified participants that did not have a differential SCR by the second half of acquisition, and of those only one participant additionally did not display a differential rating response by the second half of acquisition. There was no effect of CS during the last half of extinction, indicating successful extinction ($p = 0.144$). There was a main effect of CS during the first half of spontaneous recovery evidencing robust return of fear ($p < 0.001$). Additionally, there was a main effect of CS type specifically within the first two trials of spontaneous recovery ($p = 0.048$). See Figure 1 for observations of these patterns.

Figure 1. Modeled SCR During all Phases of Conditioning and Extinction



The US expectancy followed this same pattern of model results. There was no main effect of CS during habituation ($p = 0.39$). In the last half of acquisition, there was an observed main effect of CS type ($p < 0.001$). There was a main effect of CS during the last half of extinction ($p < 0.001$). There was a main effect of CS observed during both the first half of spontaneous recovery ($p < 0.001$) and the first two trials ($p < 0.001$). Of note, the US expectancy rating results followed the same pattern as the SCR results except for extinction which resulted in a main effect of CS type during the last half and also during the last quarter.

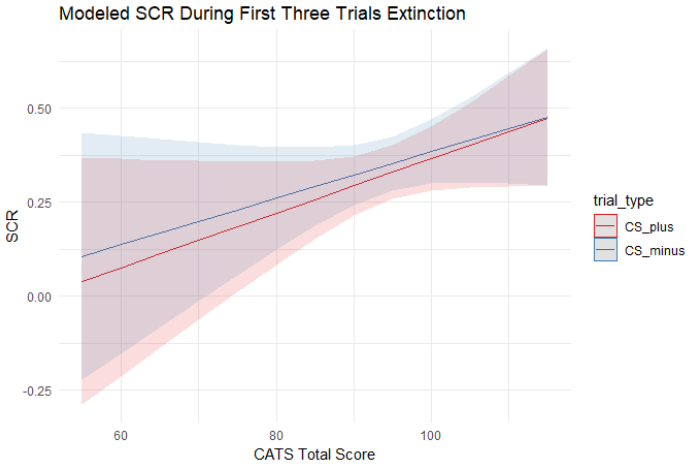
Figure 2. Modeled Expectancy Ratings During all Phases of Conditioning and Extinction



Outcome Analyses

Over the course of trial 1-3, there was no significant interaction of CS type and CATS total score for SCR (p=0.15), indicating that there was no observable relationship of CATS total score predicting differential speed of extinction. During the first two trials of spontaneous recovery, CATS total score did not significantly predict differential SCR because there was no CATS x CS interaction (p=0.33). For the expectancy ratings, CATS total score did not significantly predict responses to the CS either during the course of the first three trials of extinction (p=0.21), nor at the first two trials of recall (p=0.100). Figures 3 and 4 show graphs of these outcomes.

Figure 3. Modeled Fear Responses During the Speed of Extinction



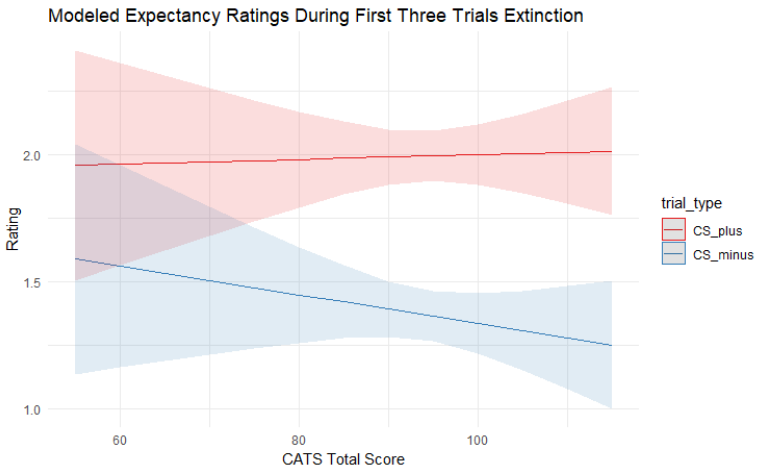
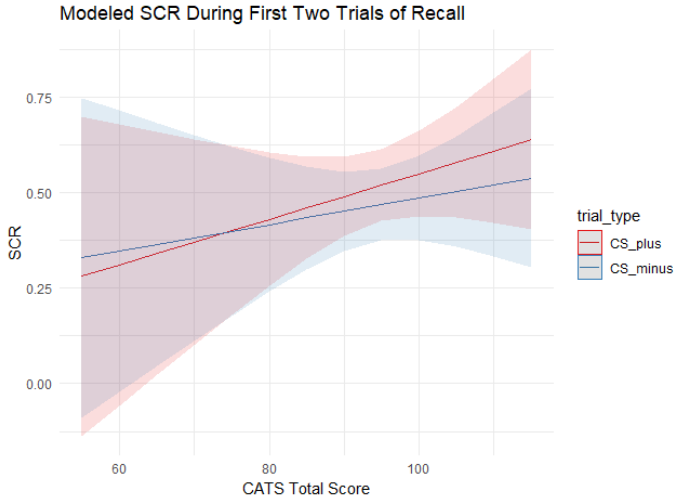
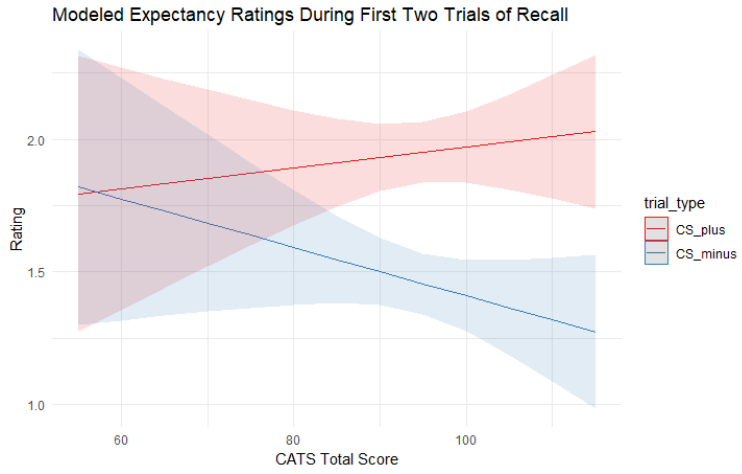


Figure 4. Modeled Fear Responses at the Beginning of Spontaneous Recovery

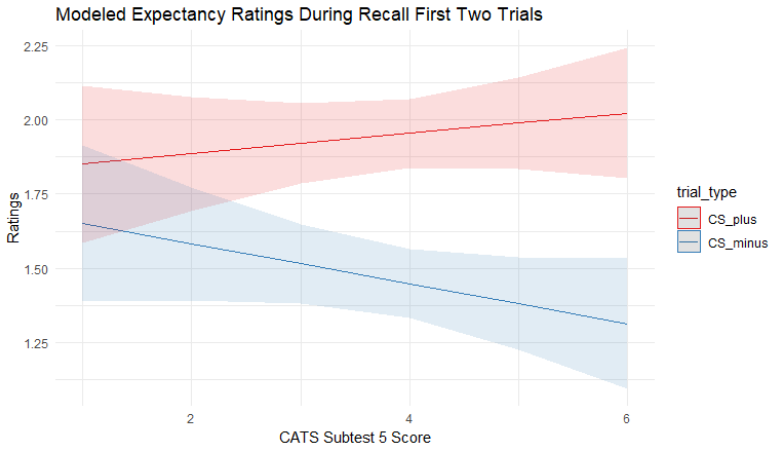




Additionally, the relationship of each of the CATS subtests 5 (name affect) and 7 (match affect) and the fear conditioning outcomes were assessed. There were no significant interaction of CATS subtest 5*CS type*trial for the first three trials of extinction for either SCR (p=0.09) or expectancy ratings (p=0.18). However, scores on subtest 5 did have a significant interaction with CS type during the first two trials of spontaneous recovery only for expectancy ratings (p=0.048), but not for SCR (p=0.10). The interaction was such that higher scores on subtest 5 were associated with higher ratings to the CS+ relative to the CS-. See Figure 5 for visualization.

Scored on the CATS subtest 7 did not significantly interact with CS and trial during the first three trials of extinction for SCR (p=0.39) or for expectancy ratings (p=0.34). Further, there was no significant interaction of subtest 7 and CS type during the first two trials of spontaneous recovery either for SCR (p=0.51) or expectancy ratings (p=0.14).

Figure 5. Modeled Expectancy Ratings for CATS Subtest 5 Interaction in Spontaneous Recovery



Study 2 Results

Sample Characteristics

Of the 44 participants that consented to participate in the study, 3 failed to meet the inclusion criteria and 4 were withdrawn for not completing Day 2 procedures (n = 37). The demographic information of the total sample, broken down by group is presented in Table 2.

Table 2. Demographic Information of Study 2 Participants

	Control (N=13)	Negative (N=12)	Neutral (N=12)	Overall (N=37)
Age				
Mean (SD)	22.2 (5.16)	23.3 (7.02)	25.8 (9.85)	23.7 (7.48)
Median [Min, Max]	19.0 [18.0, 33.0]	19.0 [18.0, 39.0]	22.5 [18.0, 50.0]	20.0 [18.0, 50.0]
Sex				
Female	10 (76.9%)	9 (75.0%)	7 (58.3%)	26 (70.3%)
Male	3 (23.1%)	3 (25.0%)	5 (41.7%)	11 (29.7%)
Race				
Asian	6 (46.2%)	7 (58.3%)	5 (41.7%)	18 (48.6%)
Multiracial	1 (7.7%)	1 (8.3%)	1 (8.3%)	3 (8.1%)
Other	1 (7.7%)	0 (0%)	2 (16.7%)	3 (8.1%)
White	5 (38.5%)	4 (33.3%)	4 (33.3%)	13 (35.1%)
Ethnicity				
Hispanic or Latino	1 (7.7%)	0 (0%)	1 (8.3%)	2 (5.4%)
Not Hispanic or Latino	12 (92.3%)	12 (100%)	11 (91.7%)	35 (94.6%)

Preliminary Analyses:

There were no significant differences in fear responses to the CS- between the three groups for SCR during habituation ($p=0.23$), acquisition ($p=0.57$), extinction ($p=0.37$), or spontaneous recovery ($p=0.46$). Likewise, there were also no differences in expectancy ratings to the CS- between the groups during habituation ($p=0.72$), acquisition ($p=0.81$), extinction ($p=0.09$), or spontaneous recovery ($p=0.91$). This was an important indicator that all three groups had similar responses to the safety stimuli, which was not modified, in order to be confident in subsequent comparisons between the group responses for the CS+ and differential.

Next, conditioning and extinction were assessed across all groups. For the Control group, as expected, there was no significant effect of CS type during habituation ($p = 0.052$), although this is closer to the 0.05 level of significance than is typical. The differential pattern can be observed visually in Figure 6. During the second half of acquisition, there was an observed significant effect of CS type, indicating successful acquiring of fear ($p = 0.01$). By the last half of extinction, there was no significant effect of CS type, meaning participants extinguished fear ($p = 0.56$). Expectancy ratings followed the same pattern for habituation ($p=0.71$), second half of acquisition ($p<0.001$), and second half of extinction ($p=0.16$).

For the Neutral label group, there was no significant effect of SCR to CS type during habituation ($p = 0.37$), no significant difference during the second half of acquisition ($p = 0.32$), and there was a significant difference during the last half of extinction ($p = 0.03$). It was surprising that the Neutral group did not condition by the end of acquisition even prior to any manipulation, however, this may be due to the small sample size. The Neutral group did, however, extinguish by the last quarter of extinction ($p = 0.06$). The expectancy ratings followed the same pattern for habituation ($p= 1$) and extinction (no p value, see note in Table). Acquisition showed expected results of significant differences during the last half ($p = 0.004$).

For the Negative label group, as expected, SCR to CS type was not significantly different during habituation ($p = 0.25$), significantly different during the second half of acquisition ($p = 0.002$), not significantly different during the last half of extinction ($p = 0.13$). Expectancy ratings followed the same pattern for habituation ($p=0.16$) and acquisition ($p<0.001$). The CS type was still significant in the second half of extinction ($p=0.045$), however, it diminished by the last quarter of extinction ($p=0.08$).

Last, differential SCR in habituation and the end of acquisition was assessed across groups for similarity. Differential SCR during habituation was not significantly different across groups ($p = 0.12$). Same for the second half of acquisition ($p = 0.07$). To note, there were eleven identified participants that did not have a

differential SCR by the second half of acquisition, and of those only one participant additionally did not display a differential rating response by the second half of acquisition. Additionally, A Kruskal-Wallis test was performed on Differential SCR at the end of extinction to check that there were no differences between the groups at the end of extinction. In the last quarter of extinction, there were no significant differences of group in Differential SCR ($p= 0.08$). The expectancy ratings followed the same pattern of significance for habituation ($p=0.59$), acquisition ($p=0.45$), and extinction ($p=0.19$).

Outcome Analyses:

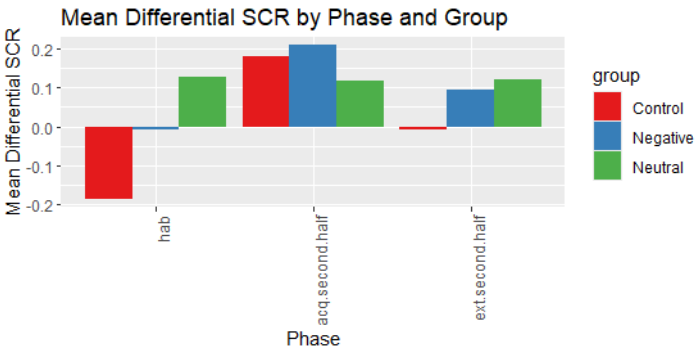
The means and standard deviations of each CS+ and CS- in each phase for each of the three groups are shown in Supplementary Table 3. The expectancy ratings are presented in Supplementary Table 4.

The average SCR and expectancy ratings to each CS in each half of each phase across all groups is shown in Supplementary Figure 2. The average differential SCR and expectancy ratings in the first half of extinction specifically are shown in Figure 7.

Figure 8 shows the SCR and expectancy ratings to the labeled CS plus and unlabeled CS plus for each of the three groups in both the first and second halves of extinction (note: there are no labeled CS+ for the control group, only unlabeled). Figures 9 and 10 show the differential SCR and differential expectancy ratings, respectively, in the last two trials of extinction compared to the first two trials of spontaneous recovery.

Figure 6. Average Differential Response by Phase and Group

A.



B.

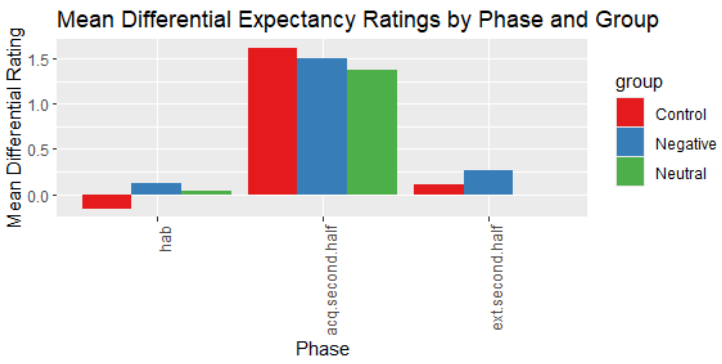
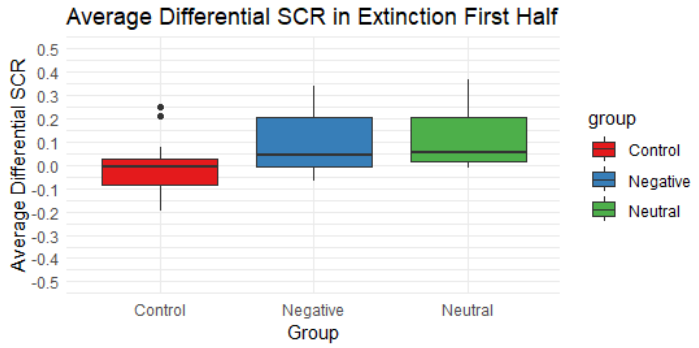


Figure 7. Average Differential Fear Response in the First Half of Extinction by Group

A.



B.

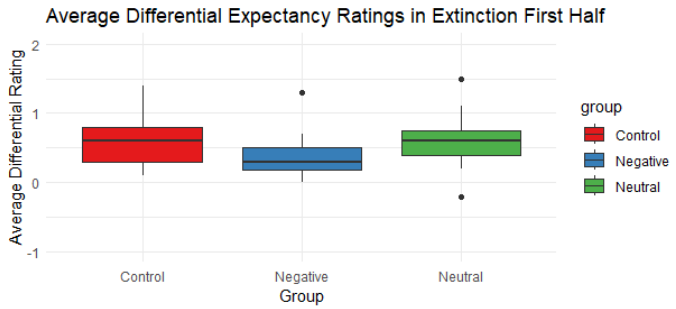
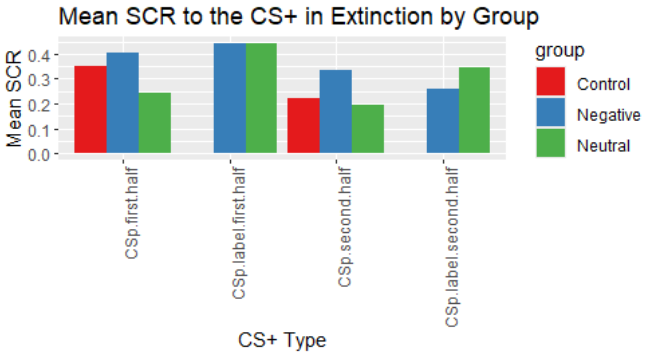


Figure 8. Average Fear Response to CS+ in Extinction

A.



B.

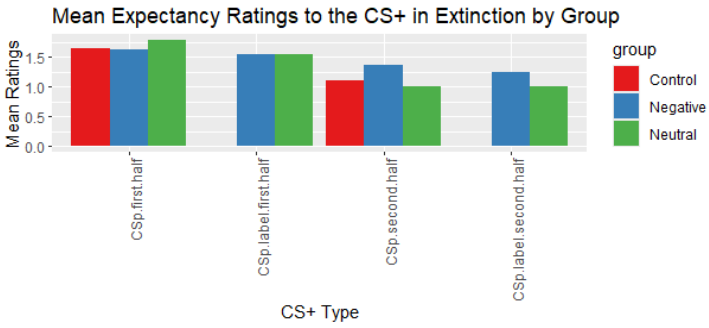
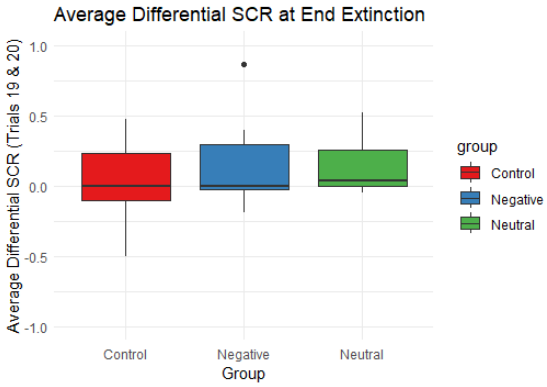


Figure 9. Average Differential SCR Between End of Extinction and Recall Start



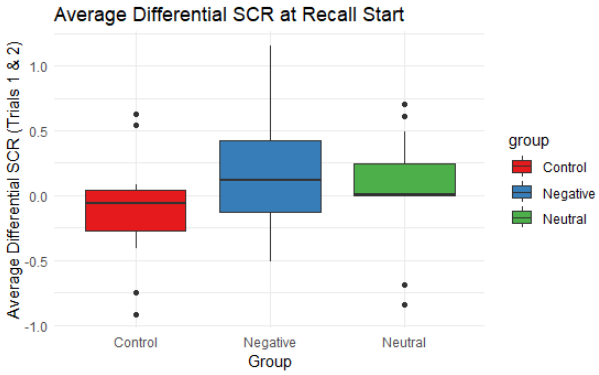
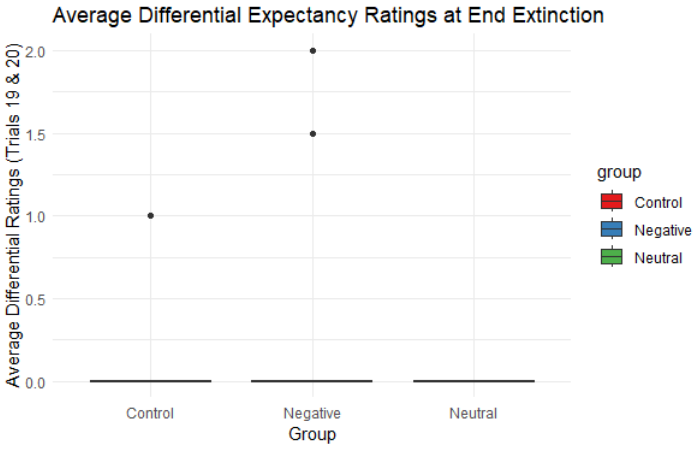
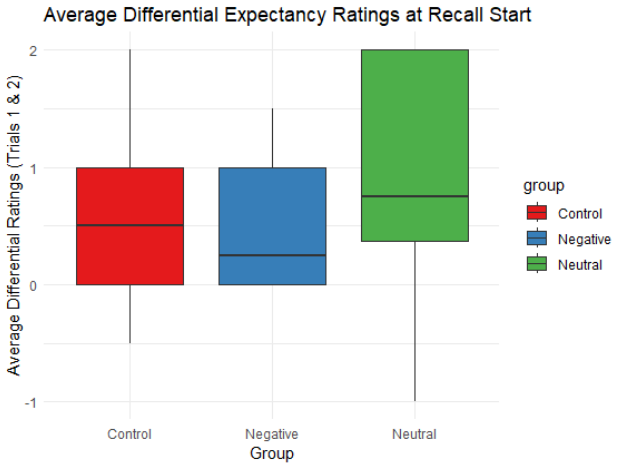


Figure 10. Average Differential Expectancy Ratings Between End of Extinction and Recall Start





The effect size of differential response in the first two trials of spontaneous recovery between negative vs. control groups was small to medium for SCR (Hedge's g = 0.49, 95% CI[-0.28, 1.26]) and in the opposite direction than expected, and negligible for expectancy ratings (Hedge's g = -0.09, 95% CI[-0.85, 0.67]) but in the expected direction. The effect size of differential between neutral vs. control groups was small for SCR (Hedge's g = 0.29, 95% CI[-0.47, 1.05],) and small for expectancy ratings (Hedge's g = 0.30, 95% CI[-0.46, 1.07]).

The effect size of differential response in the first half of extinction between negative vs. control groups was small for SCR (Hedge's g = 0.23, 95% CI[-0.53, 0.99]) and once again in the opposite direction than expected, and small for expectancy ratings (Hedge's g = -0.26, 95% CI[-1.02, 0.50]) in the expected direction. The effect size of differential between neutral vs. control groups was negligible for SCR (Hedge's g = 0.14, 95% CI[-0.62, 0.90]) and near zero for expectancy ratings (Hedge's g = 0.02, 95% CI[-0.74, 0.77]).

Discussion

Study 1

This study examined 1) whether individual differences in affect labeling ability were related to fear extinction outcomes and 2) if the presence of negative affect labels would improve extinction success and decrease return of fear. It was hypothesized that scores on the Comprehensive Affect Testing System would be associated with improved speed of extinction and diminished return of fear after twenty-four hours.

The fear conditioning and extinction procedure in Study 1 was successful. Participants demonstrated differential SCR by the end of acquisition and diminished that difference by the end of extinction. Additionally, differences in return of fear response at spontaneous recovery were significant and variable, which is in alignment with previous research demonstrating the individual variability in how people respond to fear conditioning and extinction (Bush et al., 2007; Holmes & Singewald, 2013; Shumake et al., 2014). This finding thereby lends support for the larger importance in conducting this type of research – that individual differences in the return of fear exist and should be studied further to identify factors that may predict return of fear.

In a sample of 68 participants, there were noticeable differences in the ability to identify and label emotions, as exhibited by the overall raw scores on the CATS. Identifying that individual differences exist for this ability and were able to be demonstrated using this task was an important outcome of the study. The average score we observed in 67 participants was 75.2%, which aligns closely with the average score of 78.3% determined from a study of 60 healthy participants in Schaffer et al. (2009). These preliminary findings suggest that broader differences may likely exist within the wider population, although this should be studied further. Given the paucity of research on the CATS, it is difficult to assess whether our participant scores on the CATS are “normal” and representative of the broader population. There was no ceiling effect observed, however, the range was 45.2% to 88.9%. Only a handful of participants scored below 60%. As such, we may be missing individuals with very “poor” affect labeling

ability. Future comparisons of a clinical group's performance or recruitment of participants with particularly low scores may be necessary in the future.

Regarding our primary hypotheses for Study 1, participants that were slower to extinguish fear in early extinction and had higher return of fear at spontaneous recovery were not shown to also have poorer performance on the affect labeling task. This result did not support our hypothesis which was based on prior research demonstrating a decrease in skin conductance responses to labelled emotional stimuli (Tabibnia et al., 2008) and as well as a decrease in skin conductance responses during exposure therapy when choosing the emotion label that was felt (Kreiser et al., 2019). There are a few possibilities. The first is that there is in no relationship between affect labeling and extinction outcomes (a true null result) or our study did not provide a good test of the hypothesized association. Other studies examining the relationship between psychological traits (e.g., intolerance of uncertainty, trait anxiety) and fear extinction outcomes have typically used samples of 40-60 or more, so it is possible that our study was powered sufficiently to detect individual differences if present. Alternatively, it could be that the effect size for CATS predicting extinction learning success is smaller than that of other psychological traits, necessitating a larger sample to detect. Another possibility is that there was not sufficient variability in the sample for affect labeling or fear extinction success to test the hypothesis. It is also possible that affect labeling is an inhibitory emotion regulation strategy that simply does not overlap with the processes of regulating emotions during fear extinction, and therefore does not predict how successful extinction will be in individuals (Craske et al., 2022).

Additionally, only CATS subtest 5 (name affect), and not Subtest 7 (match affect), was a significant predictor of SCR during the first two trials of spontaneous recovery. The direction of this relationship was such that higher score on the naming affect subtest predicted higher response to the CS+ stimuli, which is the opposite of what was hypothesized. However, I am reluctant to interpret this since we have conducted many statistical tests, thereby increasing our chances of acquiring an error. This leads into the first inherent limitation of "affect labeling" research more broadly. There is a lack of consistency in terminology used, as well as determining the exact construct that the CATS is measuring when it

comes to labeling emotions. Due to the relationship and overlap of concepts and terms, such as emotion regulation, emotional intelligence, emotion granularity, affect discrimination, and affect labeling, there is additional research needed to characterize the functional relationship of these concepts. A next step for this study is to analyze our results in comparison with the questionnaires that were collected which assessed emotional intelligence, emotion regulation, and alexithymia.

Further, an additional limitation comes with the use of the CATS as an overall measure of affect labeling ability, when there may in fact be differences in the modality of affect "labeling" (i.e., affect matching, affect labeling, matching, discrimination). Previous neuroimaging research has demonstrated some potential differences between the action of labeling versus matching (Lieberman et al., 2007; Torrisi et al., 2013). Our secondary comparisons between the applicable CATS subtests 5 (name affect) and 7 (match affect) attempted to shed light on this, but further research is needed to more confidently disentangle the relationship between modalities of affect labeling. In the Schaffer et al. (2019) study, the researchers also observed differences in "simple" versus "complex" facial affect recognition which included these two subtests and a few others. This would support the hypothesis that different modalities of affect labeling have varying effects (Torrisi et al., 2013). Additionally, the CATS assesses affect labeling of others. We believe this is related to ability to label one's own affect but there may be other factors that influence labeling one's own affect as well. A future direction aimed to parse this apart would be to collect a large enough sample to compare extinction outcomes with responses to the Affect Labeling Questionnaire, which gets at more internal labeling.

Study 2

Given that study 2 was a pilot study, it was hypothesized that we would observe preliminary trends in the data of the negative-label group showing lower overall return of fear following extinction compared to the neutral-label and control group. Study 2 was a somewhat successful implementation of an extinction-enhancement research design, despite only preliminary trends available for analysis due to small sample size. First, all three groups showed expected patterns of conditioning and extinction. Only

the neutral group showed slight variability in the pattern – SCR was not significantly different by the last half of acquisition, however the expectancy ratings did show a differential acquired. Further, the neutral group also did not extinguish the differential until the last quarter of extinction, compared to the last half which is not surprising given that they may have had some variability in conditioning to begin with.

In support of a check on our experimental design, we observed no significant differences in either SCR or expectancy ratings to the CS- during extinction between the groups. A concern was potential disruption of the natural safety learning of extinction with a label on the CS- driven by a change in extinction procedures from acquisition. The considerations of labeling the CS- made us decide to label only the CS+ in the modified-negative extinction group. Given that there were no significant differences in responses to the CS- between groups in extinction, we consider this in preliminary support of the current design which will then enable confident analyses in differences in responses to the CS+ across groups in future full samples. This is based on small group sizes in the present analysis and a future step is to compute Bayes Factors or equivalence testing within the larger sample.

An additional check on design outcomes was the differential responses between groups during habituation and acquisition prior to any manipulation in extinction. We found no significant differences across the three groups in either habituation or acquisition for SCR or expectancy ratings, indicating that there were likely no group differences prior to the manipulation. This was slightly at odds with the finding that the Neutral group did not acquire SCR differential by the second half of acquisition. But as we observed, the expectancy ratings showed acquired fear, thereby supporting this group-wide finding of no significant differences. Furthermore, we also confirmed that the differential responses by the last quarter of extinction were not significantly different across groups. Both SCR and expectancy ratings confirmed that all three groups were able to successfully extinguish fear by the end of extinction indicated by near-zero differential responses.

In initial observation of general patterns in the data concerning differences in extinction and return of fear across the three groups, we calculated means and standard deviations. Visually, it appears that the Neutral group had lower overall SCR responses compared to the Negative and Control groups

when contrasting the overall CS+ and CS- responses. The expectancy ratings appear similar across groups. When observing the CS+ labeled alongside the CS+ unlabeled and CS- stimuli, there are some minor differences emerging with the labeled CS+ stimuli in the Negative and Neutral groups being higher than the CS+ (always unlabeled) in the Control group throughout extinction for SCR, but not expectancy ratings.

To examine how the speed of extinction might be impacted by the manipulation, we looked at patterns of the differential SCR and expectancy ratings (CS+ - CS-) during the first half of extinction. The SCR differential shows more variability in the Negative and Neutral groups compared to the Control group. Both the Negative and Neutral groups show positive but small differentials into the first half of extinction compared to the Control group, which has already almost diminished its differential SCR. It is important to note that skin conductance responses are somewhat more variable, especially in a small sample, which is another important reason to observe the self-reported expectancy ratings alongside these results.

In fact, the patterns of the differential expectancy ratings in the first half of extinction are in the direction of expected results. The average differential in the negative group is lower than both the control and neutral groups, indicating the negative group participants might be quicker to extinguish their subjective fear in response to the presence of a negative emotion word label. Further, the average differential responses of the neutral group and the control group are relatively similar, which is also in line with the original expectation that there would be no difference of response due to the presence of a non-emotional label in extinction, and that it would only be due to a negative word that reductions in fear would be observed.

Comparing the differential SCR during the last two trials of extinction to the first two trials of spontaneous recovery shows some apparent differences in the return of fear across groups. At the end of extinction, all groups look very similar with low SCR differential responses. However, the Negative group shows an increased differential SCR during recall compared to either group. The neutral group also has comparably elevated return of fear relative to the control group. It might be indicative of a more

general impact of a modification to extinction, which may be having an impact on return of fear. This finding is present despite labeling the CS+ only 50% of the time in extinction, so that participants would not see unlabeled CS+ only in acquisition and recall, thereby skipping over extinction.

Moreover, the control group actually has negative differential responses in the first two trials of spontaneous recovery. Making comparisons to the SCR in this control group difficult because we should be seeing return of fear occurring. With a larger sample size, separate ANOVAs or non-parametric equivalents would be used to compare differences in groups between extinction and the beginning of recall. Overall, this pattern of findings is inconsistent with our hypothesis that the Negative group would show the lowest return of fear relative to the Control and Neutral groups.

The expectancy ratings, again, are in the direction of expected patterns across the groups. The negative group has the lowest return of fear compared to the control and neutral groups. Importantly, the control group is showing a differential return of hear, but the neutral group curiously has the highest and most upwardly variable return of fear. This was key to observe in support of our hypotheses that a negative word modification to extinction would both improve immediate fear extinction and reduce future return of fear. However, the neutral group finding does bring up the potential for alternative hypotheses to the original hypothesis that the neutral group would have similar responses to the control group. One such alternative hypothesis is that the addition of neutral labels might be considered an intervention in itself, similar to cognitive reappraisal, and would thereby potentially result in the neutral group having altered responding compared to the control group. Future research with larger samples to flesh out any potential impacts on the neutral group is prudent.

Given that there were relatively small, or small to medium, effect sizes between the groups, this lends support to the idea that a much larger sample size is needed to observe any effects if present. This is also not to mention that the effect size for SCR was in the opposite direction of expected. This is a good example of prior evidence of discrepancies between physiological and self-reported fear outcomes (Kircanski et al., 2012; Kreiser et al., 2019; and Niles et al., 2015). Future analyses might also consider using Bayesian statistics which would be more well-suited for smaller samples.

Limitations specific to Study 2 are mainly related to small sample size. We were only able to preliminarily observe trends and not quantify statistical differences between groups based on the intervention. Additionally, given the small sample size, individual differences are more likely to impact a group like the Neutral group for example that looks to have lower overall mean SCR to both stimuli during acquisition. Although the SCR was square-root transformed, there are additional procedures to alleviate some differences if present, like range-correcting for responses to the unconditioned stimulus. Another limitation stems from differences to extinction. For future analyses, we are considering a scoring window that only captures 3-8 seconds of stimuli presentation and disregards the first 3 seconds when the stimuli were unlabeled. We might also choose to discard the first labeled CS+ as an orienting stimulus that may cause higher responses in the beginning of extinction. Confirmatory analyses should also consider comparing results with all CS+ stimuli (both labeled and unlabeled together) with results of just the unlabeled CS+ for any differences. Lastly, the full sample collected was used for analysis. For future analyses, a consideration is to exclude participants that are non-responsive to the aversive stimulus during the shock setting procedure. An immediate next step for this sample is to re-run the main analyses while excluding for participants that do not show a conditioned fear to the CS+ by the end of acquisition. Some research maintains that if participants do not acquire a conditioned fear, that they may also not exhibit typical patterns of extinction and return of fear. However, consistent with the view of others in the field (Lonsdorf et al., 2017), I do think there is some merit to capturing the entire spectrum of individual differences regarding the process of fear conditioning. Future studies should also utilize other assessment methods in addition to SCR, like pupillometry and self-reported fear, for example.

One potential limitation that may impact the results is the words used for the labels. These were identified by the experimenter prior to the start of the study and may not have necessarily been representative of the subjective experience of each individual participant. For example, one participant may not necessarily find the CS+ “distressing,” but rather more along the lines of bothersome. We did include a post-participation questionnaire assessing how relevant the words felt to each participant during the experiment. A future direction is to analyze whether these survey attitudes related to fear responses

during extinction in the labeled groups. Additionally, these words were externally applicable to the stimulus and not "self" descriptive words (i.e., scary vs. scared). Although prior affect labeling research has suggested that this is a reflexive ability (Salovey & Mayer, 1990), there may in fact be differences in labeling one's own emotions while undergoing fear extinction not explicitly captured by this external descriptive label. A future iteration of this study should use self-descriptive words or even cue the participant to generate their own descriptive word to compare results.

Although we did not observe the expected relationship between affect labeling ability and fear extinction success or have yet to determine the statistical relationship of the presence of negative emotion word labels on extinction success, this project was novel and important for advancing the field. There may or may not be a relationship of affect labeling to extinction and/or return of fear, but regardless the ultimate goal of this research remains to improve exposure therapy for anxiety and other fear-related disorders. If established that physiological responses to conditioned negative stimuli can be diminished through implementation of an emotion label, the goal would then be to implement an affect labeling intervention during extinction. However, if there is no observable impact on extinction success in the lab, then more research is needed into affect labeling as a skill, enhancements to extinction, and translatability to clinical populations.

Supplementary Materials

Supplementary Table 1. Study 1 Statistical Statements

Preliminary Analyses	Skin Conductance Responses (SCR)	US Expectancy Ratings
CS type during Habituation	β = -0.12, twald(264) = -1.83, p = 0.068, 95% CI [-0.25, 0.01]	β = -1.54e-9, twald(258) = -0.86, p = 0.392, 95% CI [0.00, 0.00]
CS type during second half Acquisition	β = 0.32, twald(666) = 6.38, p < 0.001, 95% CI [0.22, 0.42]	β = 2.12, twald(658) = 8.94e+8, p < 0.001, 95% CI [2.12, 2.12]
CS type during second half Extinction	β = 0.05, twald(1336) = 1.46, p = 0.144, 95% CI [-0.02, 0.12]	β = 4.99e-11, twald(1316) = 5.29, p < 0.001, 95% CI [0.00, 0.00]
CS type during last quarter Extinction	N/A	β = 2.93e-14, twald(658) = 1.97, p = 0.050, 95% CI [0.00, 0.00]
CS type during first half Spontaneous Recovery	β = 0.18, twald(800) = 4.46, p < 0.001, 95% CI [0.10, 0.26]	β = 0.96, twald(2790) = 12.79, p < 0.001, 95% CI [0.81, 1.11]
CS type during first two trials Spontaneous Recovery	β = 0.14, twald(264) = 1.99, p = 0.048, 95% CI [0.00, 0.28]	β = 0.88, twald(262) = 5.81, p < 0.001, 95% CI [0.58, 1.18]
Outcome Analyses		
CATS Total Score*CS*trial during first three trials Extinction	β = -0.08, twald(393)=-1.44, p=0.151, CI[-0.19, 0.03]	β = -0.11, twald(393)= -1.26, p=0.207, CI[-0.28, 0.06]
CATS Total Score*CS during first two trials Spontaneous Recovery	β = 0.07, twald(262)=0.97, p=0.333, CI[-0.07, 0.20]	β = 0.18, twald(261)= 1.65, p=0.100, CI[-0.03, 0.39]
CATS Subest 5 Score*CS*trial during first three trials Extinction	β = -0.10, twald(393)= -1.70, p=0.090, CI[-0.21, 0.02]	β = -0.11, twald(393)= -1.35, p=0.179, CI[-0.28, 0.05]
CATS Subtest 5 Score*CS during first two trials Spontaneous Recovery	β = 0.11, twald(262)=1.65, p=0.100, CI[-0.02,0.25]	β = 0.21, twald(261)=1.98, p=0.048, CI[0.00,0.42]
CATS Subest 7 Score*CS*trial during first three trials Extinction	β = -0.05, twald(393)=-0.86, p=0.388, CI[-0.16,0.06]	β = -0.08, twald(393)=-0.96, p=0.338, CI[-0.25,0.09]
CATS Subtest 7 Score*CS during first two trials Spontaneous Recovery	β = 0.05, twald(262)=0.65, p=0.513, CI[-0.09,0.18]	β = 0.16, twald(261)=1.48, p=0.141, CI[-0.05,0.37]

Supplementary Table 2. Study 2 Statistical Statements

Preliminary Analyses	Skin Conductance Responses (SCR)	US Expectancy Ratings
CS- during Habituation between groups	Kruskal-Wallis Chi-squared = 2.98, df = 2, p = 0.225	Kruskal-Wallis Chi-squared = 0.66401, df = 2, p = 0.7175
CS- during Acquisition between groups	Kruskal-Wallis Chi-squared = 1.1241, df = 2, p = 0.57	Kruskal-Wallis Chi-squared = 0.42812, df = 2, p = 0.8073
CS- during Extinction between groups	Kruskal-Wallis Chi-squared = 1.9652, df = 2, p = 0.3743	Kruskal-Wallis Chi-squared = 4.8884, df = 2, p = 0.0868
CS- during Spontaneous Recovery between groups	Kruskal-Wallis Chi-squared = 1.5533, df = 2, p = 0.4599	Kruskal-Wallis Chi-squared = 0.1867, df = 2, p = 0.9109
Control Group: CS type during Habituation	Friedman Chi-squared = 3.7692, df = 1, p = 0.0522	Friedman Chi-squared = 0.14286, df = 1, p = 0.7055
Control Group: CS type during second half Acquistion	Friedman Chi-squared = 7.3636, df = 1, p = 0.0067	Friedman Chi-squared = 13, df = 1, p < 0.001
Control Group: CS type during second half Extinction	Friedman Chi-squared = 0.3333, df = 1, p = 0.5637	Friedman Chi-squared = 2, df = 1, p = 0.1573
Neutral Label Group: CS type during Habituation	Friedman Chi-squared = 0.8182, df = 1, p = 0.3657	Friedman Chi-squared = 0, df = 1, p = 1
Neutral Label Group: CS type during second half Acquistion	Friedman Chi-squared = 1, df = 1, p = 0.3173	Friedman Chi-squared = 8.3333, df = 1, p = 0.0039
Neutral Label Group: CS type during second half Extinction	Friedman Chi-squared = 4.4545, df = 1, p = 0.03481	N/A*
Neutral Label Group: CS type during last quarter Extinction	Friedman Chi-squared = 3.6, df = 1, p = 0.05778	N/A
Negative Label Group: CS type during Habituation	Friedman Chi-squared = 1.3333, df = 1, p = 0.2482	Friedman Chi-squared = 2, df = 1, p = 0.1573
Negative Label Group: CS type during second half Acquistion	Friedman Chi-squared = 10, df = 1, p = 0.0016	Friedman Chi-squared = 12, df = 1, p < 0.001
Negative Label Group: CS type during second half Extinction	Friedman Chi-squared = 2.2727, df = 1, p = 0.1317	Friedman Chi-squared = 4, df = 1, p = 0.0455
Negative Label Group: CS type during last quarter Extinction	NA	Friedman Chi-squared = 3, df = 1, p = 0.08326
Differential during Habituation between groups	Kruskal-Wallis Chi-squared = 4.2395, df = 2, p = 0.1201	Kruskal-Wallis Chi-squared = 1.0484, df = 2, p = 0.592
Differential during second half Acquisition between groups	Kruskal-Wallis Chi-squared = 5.2821, df = 2, p = 0.07129	Kruskal-Wallis Chi-squared = 1.6025, df = 2, p = 0.4488
Differential during last quarter Extinction between groups	Kruskal-Wallis Chi-squared = 5.0482, df = 2, p = 0.08013	Kruskal-Wallis Chi-squared = 3.3463, df = 2, p-value = 0.1877

***Note, this value could not be computed because all rating values were 1, therefore a CS difference of exactly 0**

Supplementary Table 3. Means and standard deviations of SCR to CS type in each phase by each group

1A. Habituation

group	CSplus.mean	CSplus.sd	CSminus.mean	CSminus.sd
Control	0.6759683	0.5577146	0.8627174	0.4988301
Negative	0.8544809	0.6016317	0.8616479	0.6496391
Neutral	0.6895218	0.6235639	0.5632085	0.5091728

1B. Acquisition

group	CSplus.first.half.mean	CSplus.first.half.sd	CSminus.first.half.mean	CSminus.first.half.sd	CSplus.second.half.mean	CSplus.second.half.sd	CSminus.second.half.mean	CSminus.second.half.sd
Control	0.4983510	0.4948897	0.3820619	0.4477001	0.4494848	0.4739515	0.2687390	0.3552301
Negative	0.6828145	0.5462129	0.4291472	0.5161799	0.4537991	0.4952364	0.2463735	0.3963201
Neutral	0.3091691	0.3880125	0.2789821	0.4202388	0.3187018	0.4660972	0.2028398	0.3505023

1C. Extinction

group	CSplus.first.half.mean	CSplus.first.half.sd	CSminus.first.half.mean	CSminus.first.half.sd	CSplus.second.half.mean	CSplus.second.half.sd	CSminus.second.half.mean	CSminus.second.half.sd
Control	0.3526193	0.5061509	0.2899424	0.4247529	0.2217644	0.3381604	0.2305962	0.3957948
Negative	0.4250830	0.4997128	0.2378531	0.4158075	0.2993482	0.5063196	0.2065736	0.4089727
Neutral	0.3378605	0.5125364	0.1954721	0.3423213	0.2736696	0.3840715	0.1531941	0.3104089

1D. Spontaneous Recovery

group	CSplus.first.two.mean	CSplus.first.two.sd	CSminus.first.two.mean	CSminus.first.two.sd
Control	0.5337903	0.6110700	0.6555097	0.7081125
Negative	0.8005837	0.6538906	0.6034778	0.6360424
Neutral	0.4566819	0.5354599	0.4109244	0.5681238

Supplementary Table 4. Means and standard deviations of expectancy ratings to CS type in each phase in each group

2A. Habituation

group	CSplus.mean	CSplus.sd	CSminus.mean	CSminus.sd
Control	1.384615	0.5710988	1.520000	0.7141428
Negative	1.695652	0.6349504	1.625000	0.7109394
Neutral	1.583333	0.7172815	1.565217	0.6623709

2B. Acquisition

group	CSplus.first.half.mean	CSplus.first.half.sd	CSminus.first.half.mean	CSminus.first.half.sd	CSplus.second.half.mean	CSplus.second.half.sd	CSminus.second.half.mean	CSminus.second.half.sd
Control	1.723077	0.8928132	1.384615	0.7844645	2.676923	0.6639769	1.061538	0.2998397
Negative	1.916667	0.8692811	1.333333	0.6806444	2.533333	0.7471225	1.033333	0.2581989
Neutral	1.916667	0.9259291	1.150000	0.4809947	2.583333	0.7431419	1.216667	0.6131792

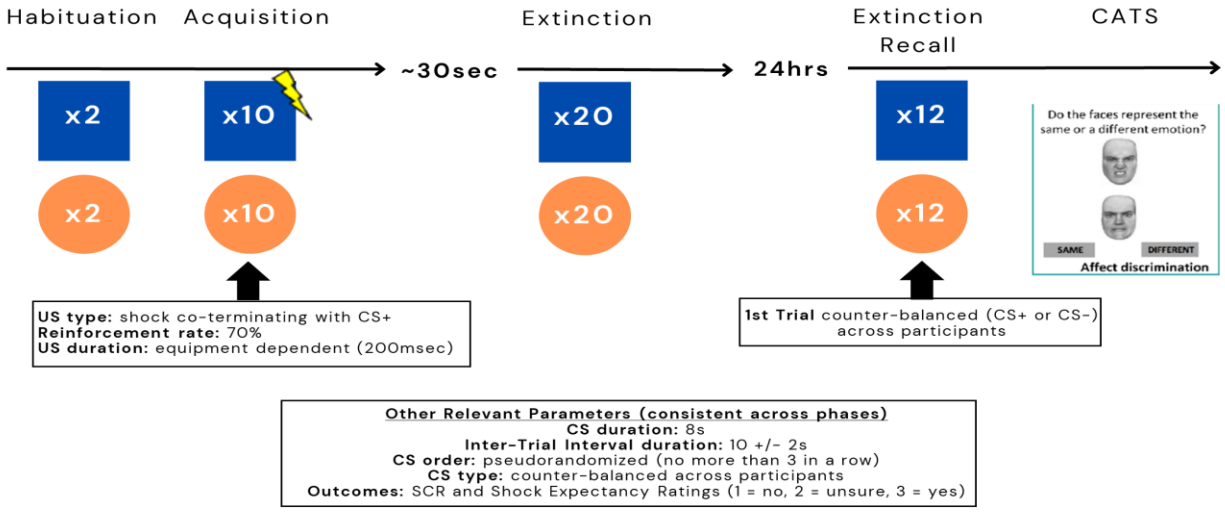
2C. Extinction

group	CSplus.first.half.mean	CSplus.first.half.sd	CSminus.first.half.mean	CSminus.first.half.sd	CSplus.second.half.mean	CSplus.second.half.sd	CSminus.second.half.mean	CSminus.second.half.sd
Control	1.638462	0.8445341	1.030769	0.1733599	1.107692	0.3985663	1.000000	0.0000000
Negative	1.583333	0.8156385	1.200000	0.4953566	1.300000	0.7053219	1.033613	0.2581805
Neutral	1.675000	0.8417918	1.050420	0.2866741	1.000000	0.0000000	1.000000	0.0000000

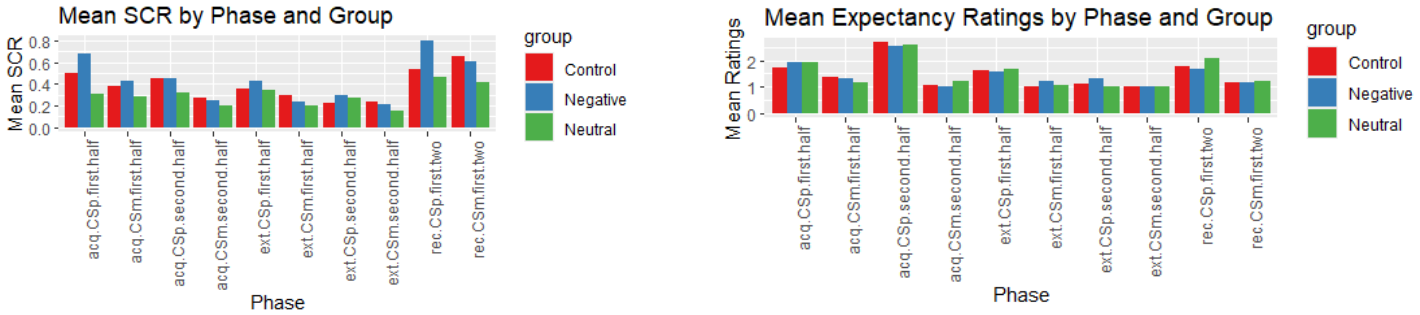
2D. Spontaneous Recovery

group	CSplus.first.two.mean	CSplus.first.two.sd	CSminus.first.two.mean	CSminus.first.two.sd
Control	1.769231	0.7103629	1.192308	0.4019185
Negative	1.666667	0.8164966	1.166667	0.4815434
Neutral	2.083333	0.8297022	1.208333	0.5089774

Supplementary Figure 1. Overview of Study 1 procedures, including the fear conditioning paradigm and the affect labeling task.



Supplementary Figure 2. Average fear response to CS type by phase and group



References

Arch, J. J., & Craske, M. G. (2009). First-line Treatment: A Critical Appraisal of Cognitive Behavioral Therapy Developments and Alternatives. *Psychiatric Clinics of North America*, *32*(3), 525–547. https://doi.org/10.1016/j.psc.2009.05.001

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335. https://doi.org/10.31887/DCNS.2015.17.3/bbandelow

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. https://doi.org/10.1037/a0027555

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 201–206. https://doi.org/10.1016/j.jbtep.2012.08.003

Bonanno, G. A., & Burton, C. L. (2013). Regulatory Flexibility: An Individual Differences Perspective on Coping and Emotion Regulation. *Perspectives on Psychological Science*, *8*(6), 591–612. https://doi.org/10.1177/1745691613504116

Burklund, L., Creswell, J., Irwin, M., & Lieberman, M. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00221

Bush, D. E. A., Sotres-Bayon, F., & LeDoux, J. E. (2007). Individual differences in fear: Isolating fear reactivity and fear recovery phenotypes. *Journal of Traumatic Stress*, *20*(4), 413–422. https://doi.org/10.1002/jts.20261

Bystritsky, A. (2006). Treatment-resistant anxiety disorders. *Molecular Psychiatry*, *11*(9), Article 9. https://doi.org/10.1038/sj.mp.4001852

Constantinou, E., Purves, K. L., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M., Craske, M. G., & Eley, T. C. (2021). Measuring fear: Association among different measures of fear learning. *Journal of Behavior Therapy and Experimental Psychiatry*, *70*, 101618. https://doi.org/10.1016/j.jbtep.2020.101618

Constantinou, E., Van Den Houte, M., Bogaerts, K., Van Diest, I., & Van den Bergh, O. (2014). Can words heal? Using affect labeling to reduce the effects of unpleasant cues on symptom reporting. *Frontiers in Psychology*, *5*, 807. https://doi.org/10.3389/fpsyg.2014.00807

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. https://doi.org/10.1098/rstb.2017.0025

Craske, M. G., & Mystkowski, J. L. (2006). Exposure Therapy and Extinction: Clinical Studies. In *Fear and learning: From basic processes to clinical implications* (pp. 217–233). American Psychological Association. https://doi.org/10.1037/11474-011

Craske, M. G., Treanor, M., Zbozinek, T. D., & Vervliet, B. (2022). Optimizing exposure therapy with an inhibitory retrieval approach and the OptEx Nexus. *Behaviour Research and Therapy*, *152*, 104069. https://doi.org/10.1016/j.brat.2022.104069

Dunsmoor, J. E., Prince, S. E., Murty, V. P., Kragel, P. A., & LaBar, K. S. (2011). Neurobehavioral mechanisms of human fear generalization. *NeuroImage*, *55*(4), 1878–1888. https://doi.org/10.1016/j.neuroimage.2011.01.041

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*(2), 268–287. https://doi.org/10.1037/0033-2909.115.2.268

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75. https://doi.org/10.1007/BF01115465

Foa, E. B., Huppert, J. D., & Cahill, S. P. (2006). Emotional Processing Theory: An Update. In *Pathological anxiety: Emotional processing in etiology and treatment* (pp. 3–24). The Guilford Press.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication Recommendations for Electrodermal Measurements. *Psychophysiology*, *18*(3), 232–239. https://doi.org/10.1111/j.1469-8986.1981.tb03024.x

Froming, K. B., Ekman, P., & Levy, M. (2006). *Comprehensive Affect Testing System* [Dataset]. https://doi.org/10.1037/t06823-000

Gohm, C. L. (2003). Mood regulation and emotional intelligence: Individual differences. *Journal of Personality and Social Psychology*, *84*(3), 594–607. https://doi.org/10.1037/0022-3514.84.3.594

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *NeuroReport*, *11*(1), 43.

Hartley, C. A., Fischl, B., & Phelps, E. A. (2011). Brain Structure Correlates of Individual Differences in the Acquisition and Inhibition of Conditioned Fear. *Cerebral Cortex (New York, NY)*, *21*(9), 1954–1962. https://doi.org/10.1093/cercor/bhq253

Hartley, C. A., & Phelps, E. A. (2010). Changing Fear: The Neurocircuitry of Emotion Regulation. *Neuropsychopharmacology*, *35*(1), Article 1. https://doi.org/10.1038/npp.2009.121

Hermann, A., Keck, T., & Stark, R. (2014). Dispositional cognitive reappraisal modulates the neural correlates of fear acquisition and extinction. *Neurobiology of Learning and Memory*, *113*, 115–124. https://doi.org/10.1016/j.nlm.2014.03.008

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. *Biological Psychiatry*, *60*(4), 361–368. https://doi.org/10.1016/j.biopsych.2005.10.006

Holmes, A., & Singewald, N. (2013). Individual differences in recovery from traumatic fear. *Trends in Neurosciences*, *36*(1), 23–31. https://doi.org/10.1016/j.tins.2012.11.003

Indovina, I., Robbins, T. W., Núñez-Elizalde, A. O., Dunn, B. D., & Bishop, S. J. (2011). Fear-Conditioning Mechanisms Associated with Trait Vulnerability to Anxiety in Humans. *Neuron*, *69*(3), 563–571. https://doi.org/10.1016/j.neuron.2010.12.034

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, *21*(3), 169–184. https://doi.org/10.1002/mpr.1359

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings Into Words: Contributions of Language to Exposure Therapy. *Psychological Science*, *23*(10), 1086–1091. https://doi.org/10.1177/0956797612443830

Kitamura, H., Strodl, E., Johnston, P., & Johnson, L. R. (2022). The influence of dispositional cognitive reappraisal and expressive suppression on post-retrieval and standard extinction. *Psychophysiology*, *59*(9), e14048. https://doi.org/10.1111/psyp.14048

Kreiser, I., Moyal, N., & Anholt, G. E. (2019). Regulating Obsessive-Like Thoughts: Comparison of Two Forms of Affective Labeling with Exposure Only in Participants with High Obsessive-Compulsive Symptoms. *Clinical Neuropsychiatry*, *16*(1), 25–32.

LeDoux, J. E. (2017). Semantics, Surplus Meaning, and the Science of Fear. *Trends in Cognitive Sciences*, *21*(5), 303–306. https://doi.org/10.1016/j.tics.2017.02.004

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting Feelings into Words: Affect Labeling Disrupts Amygdala Activity in Response to Affective Stimuli. *Psychological Science*, *18*(5), 421–428.

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion*, *11*(3), 468–480. https://doi.org/10.1037/a0023503

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 & Biobehavioral Reviews*, *80*, 703–728. https://doi.org/10.1016/j.neubiorev.2017.07.007

Milad, M. R., & Quirk, G. J. (2012). Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annual Review of Psychology*, *63*(1), 129–151. https://doi.org/10.1146/annurev.psych.121208.131631

Niles, A. N., Craske, M. G., Lieberman, M. D., & Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy*, *68*, 27–36. https://doi.org/10.1016/j.brat.2015.03.004

Olatunji, B. O., Forsyth, J. P., & Feldner, M. T. (2007). Implications of emotion regulation for the shift from normative fear-relevant learning to anxiety-related psychopathology. *American Psychologist*, *62*(3), 257–259. https://doi.org/10.1037/0003-066X.62.3.257

Orr, S. P., Metzger, L. J., Lasko, N. B., Macklin, M. L., Peri, T., & Pitman, R. K. (2000). De novo conditioning in trauma-exposed individuals with and without posttraumatic stress disorder. *Journal of Abnormal Psychology*, *109*(2), 290–298. https://doi.org/10.1037/0021-843X.109.2.290

Rachman, S. (1989). The return of fear: Review and prospect. *Clinical Psychology Review*, *9*(2), 147–168. https://doi.org/10.1016/0272-7358(89)90025-1

Salovey, P., & Mayer, J. D. (1990). *Emotional Intelligence*. https://journals-sagepub-com.ezproxy.library.tufts.edu/doi/abs/10.2190/DUGG-P24E-52WK-6CDG

Schaffer, S. G., Wisniewski, A., Dahdah, M., & Froming, K. B. (2009). The Comprehensive Affect Testing System–Abbreviated: Effects of Age on Performance. *Archives of Clinical Neuropsychology*, *24*(1), 89–104. https://doi.org/10.1093/arclin/acp012

Schottenbauer, M. A., Glass, C. R., Arnkoff, D. B., Tendick, V., & Hafter Gray, S. (2008). Nonresponse and Dropout Rates in Outcome Studies on PTSD: Review and Methodological Considerations. *Psychiatry: Interpersonal & Biological Processes*, *71*(2), 134–168. https://doi.org/10.1521/psyc.2008.71.2.134

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., Zwitserlood, P., Schiffbauer, H., Heindel, W., Arolt, V., & Konrad, C. (2011). Neural correlates of trait anxiety in

fear extinction. *Psychological Medicine*, *41*(4), 789–798. https://doi.org/10.1017/S0033291710001248

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion Regulation and Psychopathology. *Annual Review of Clinical Psychology*, *11*(1), 379–405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Shumake, J., Furgeson-Moreira, S., & Monfils, M. H. (2014). Predictability and heritability of individual differences in fear learning. *Animal Cognition*, *17*(5), 1207–1221. https://doi.org/10.1007/s10071-014-0752-1

Tabibnia, G., Lieberman, M. D., & Craske, M. G. (2008). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion*, *8*(3), 307–317. https://doi.org/10.1037/1528-3542.8.3.307

Torre, J. B., & Lieberman, M. D. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, *10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Torrisi, S. J., Lieberman, M. D., Bookheimer, S. Y., & Altshuler, L. L. (2013). Advancing understanding of affect labeling with dynamic causal modeling. *NeuroImage*, *82*, 481–488. https://doi.org/10.1016/j.neuroimage.2013.06.025

Underwood, R., Tolmeijer, E., Wibroe, J., Peters, E., & Mason, L. (2021). Networks underpinning emotion: A systematic review and synthesis of functional and effective connectivity. *NeuroImage*, *243*, 118486. https://doi.org/10.1016/j.neuroimage.2021.118486

Winter, K. A., & Kuiper, N. A. (1997). Individual differences in the experience of emotions. *Clinical Psychology Review*, *17*(7), 791–821. https://doi.org/10.1016/S0272-7358(97)00057-3