Starting fear is a stronger predictor of long-term fear than rate of change in fear in human fear conditioning

Tomislav D. Zbozinek
California Institute of Technology, Pasadena, CA, USA

Alexandra S. Tanner
University of California Los Angeles, Los Angeles, CA, USA

Michelle G. Craske
University of California Los Angeles, Los Angeles, CA, USA

Abstract
In rodent studies, faster extinction rate has been shown to predict less long-term fear. However, this has scarcely been studied in humans. The present report investigated the association between extinction rate and long-term fear in humans. We secondarily evaluated specificity of extinction rate by including other fear conditioning values as predictors, including acquisition intercept, acquisition rate, and extinction intercept. Lastly, we investigated whether trait measures of behavioral approach, behavioral inhibition, anxiety, and depression predicted long-term fear. Results show that slower extinction rate predicted less long-term fear when tested alone in the model. However, when including other fear conditioning variables, extinction rate no longer predicted long-term fear. Instead, greater fear at the beginning of acquisition was the most robust predictor of greater long-term fear (all three measures of fear), followed by greater fear at the beginning of extinction (unconditional stimulus expectancy only). These effects occurred for both the danger signal (i.e., conditional stimulus; CS+) and safety signal (i.e., CS−). The results suggest that fear at the start of acquisition and, secondarily, extinction are predictors of long-term fear. Lastly, there were no effects of trait behavioral approach, behavioral inhibition, anxiety, or depression. This report has relevance for improving our understanding and treatment of anxiety disorders.

Keywords
Fear conditioning, individual differences, acquisition, extinction, anxiety, extinction rate

Date received: 21 June 2021; accepted: 16 June 2022

Extinction of an acquired fear response is essential to adaptive emotional responding and resilience across species. According to inhibitory learning models of Pavlovian extinction, memory of the conditional stimulus (CS)/unconditional stimulus (US) association is not erased during extinction, but rather a new inhibitory CS/no US association is learned that competes with the original CS/US association (Bouton, 1993). Individuals with anxiety disorders demonstrate deficits in extinction learning (Duits et al., 2015; Lissek et al., 2005), possibly rendering them more...
susceptible to the return of fear and persistence of anxiety (Craske et al., 2018). Exposure therapy is a major component of cognitive-behavioral therapy—the gold standard treatment for anxiety disorders—and is thought to be largely based on extinction learning (Craske et al., 2014, 2022). However, cognitive-behavioral therapy is effective in only approximately 50% of individuals (Loerinc et al., 2015). Because anxiety disorders affect approximately 41.7% of individuals in the United States during their lifetime (Kessler et al., 2012), this leaves millions of people underserved. Thus, identifying factors that contribute to impaired extinction learning may identify at-risk individuals and inform treatment.

Rapid treatment responsiveness is commonly viewed as a positive prognostic sign for overall treatment effectiveness; in the context of exposure therapy, faster rates of fear reduction have been assumed to predict superior outcomes. Yet, evidence regarding the relationship between rate of extinction—the mechanism underlying exposure therapy—and long-term fear outcomes is relatively sparse.

Several animal studies suggest that faster reduction of conditional fear during extinction is predictive of lower fear at follow-up test (Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015). First, Reznikov et al. (2015) compared “weak” extinguishers and “strong” extinguishers in rodents. Weak extinguishers demonstrated more than 70% freezing behaviors to the CS+ during the last four trials of extinction (12 trials total), whereas strong extinguishers demonstrated less than 30% freezing. Weak extinguishers displayed significantly more freezing than strong extinguishers one day and 15 days after extinction. A second rodent study (Bush et al., 2007) bred “fast” extinguishers and “slow” extinguishers based on the top versus bottom 20% of CS+ freezing on trials 5–7 of extinction (out of 20 trials total). Results showed that fast extinguishers had less fear than slow extinguishers at test 1 day later. Furthermore, King et al. (2017) separated rodents based on the number of trials needed to reach <35% freezing to the CS+ during extinction; “fast” extinguishers did so in ≤13 blocks of three trials, whereas “slow” extinguishers did so in ≥16 blocks. The results showed that fast extinguishers had less fear 1 day after extinction than slow extinguishers. In a subsequent series of three experiments, King et al. (2018) calculated extinction rate using the same CS+ freezing criteria above, finding that faster extinction rate predicted less ABC renewal, less reinstatement, and less spontaneous recovery 8 days after extinction. Thus, the animal literature consistently suggests that faster extinction rate predicts less long-term fear (Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015). In contrast to rodent studies, only one human study has reported rate of extinction in relation to long-term fear using self-report fear and startle reflex to the CS+ and CS− (Brown et al., 2017). Rate of extinction of the startle reflex did not predict startle reflex 1 week later; similarly, changes in self-report fear from before to after extinction did not predict self-report fear 1 week later.

Lastly, there is some evidence to suggest that individual differences in trait emotionality could affect long-term fear. First, based on meta-analyses, clinically anxious individuals have been shown to have greater fear of the CS+ and CS− during acquisition and greater fear of the CS+ during extinction compared to healthy controls (Duits et al., 2015; Lissek et al., 2005). Such effects may extend to long-term fear, as assessed by an extinction test phase conducted one or more days after extinction. Similar effects could occur when using measures of behavioral inhibition as an indicator of anxious tendencies, as evidenced by poorer discrimination of CS+ versus CS− fear during both spontaneous recovery and renewal (Staples-Bradley et al., 2018). Second, several studies have shown that non-delivery of an aversive US produces reward signals in the brain and that dopaminergic neurons are involved in extinction learning (Abraham et al., 2014; Felsenberg et al., 2018; Gerlicher et al., 2018; Hikind & Maroun, 2008; Luo et al., 2018; Mueller et al., 2010). Relatedly, individual differences in anhedonia—a construct tied to reward hypersensitivity (Wang et al., 2020)—have been associated with neural indices during extinction and extinction test (Rosenberg et al., 2021; Young et al., 2021). Furthermore, behavioral studies show an association between low state positive affect (a central feature of anhedonia) prior to extinction and less stable long-term fear extinction, as measured by stronger reacquisition (Zbozinek & Craske, 2017b) and reinstatement (Zbozinek et al., 2015) of conditional fear. Thus, measures related to reward processing (as measured by reward sensitivity or symptoms of depression) may predict long-term fear after extinction.

The present report extends the existing literature (Brown et al., 2017; Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015) examining whether individual differences in extinction rate predict long-term fear. We do so by investigating whether extinction rate predicts long-term fear reduction from four human fear conditioning studies conducted with similar procedures in the same laboratory. We conducted a series of a priori and secondary/exploratory tests to evaluate extinction rate, other fear conditioning variables, and individual difference variables (i.e., trait anxiety, depression, behavioral approach, and behavioral inhibition) as predictors of long-term fear outcomes (measured during extinction test 1 week after extinction). First, with our a priori analyses, and consistent with the animal literature, we hypothesized that faster extinction rate would predict less long-term fear. Second, in secondary/exploratory analyses, we evaluated specificity of the effects of extinction rate by including other potentially relevant fear conditioning variables from acquisition and extinction (i.e., acquisition intercept, acquisition rate, and extinction intercept). Lastly, we evaluated which individual difference variables (i.e., trait anxiety,
depression, behavioral approach, and behavioral inhibition) predicted long-term fear. Our exploratory analyses build upon previous studies (which investigated extinction rate only) by a) assessing the specificity of extinction rate as a predictor of long-term fear reduction in comparison to other fear conditioning variables and b) identifying the individuals with less long-term fear.

**Methods**

**Participants**

Participants ($N = 264$) were students from the University of California, Los Angeles, who participated for course credit or cash: Study 1 ($n = 100$) (Zbozinek et al., 2015), Study 2 ($n = 62$) (Zbozinek & Craske, 2017b), Study 3 ($n = 58$; Experiment 2) (Zbozinek, 2017), and Study 4 ($n = 44$) (Zbozinek & Craske, 2018). Twenty-two participants dropped out, leaving 242 completers. Participants were 61.1% female (38.9% male); mean age 21.83 years (SD = 3.93); and 3.6% Black or African-American, 0.4% American-Indian or Alaska Native, 41.1% Asian or Asian-American, 20.6% Hispanic/Latinx, 0.4% Native Hawaiian or Other Pacific Islander, 26.6% White, and 7.3% Multiracial or Other. These studies were approved by the University of California, Los Angeles Institutional Review Board, and all participants provided informed consent prior to commencing the study.

**Study selection and design**

We combined the data from four fear conditioning studies to maximize sample size and power. The four studies shared the following characteristics: experimental phases of habituation, acquisition, extinction, and test (i.e., CS+ and CS−); similar intertrial interval duration; electric shock US; no changes in context; number of CS trials (i.e., 7–8 for acquisition and 8–9 for extinction); time between acquisition and extinction (i.e., <1 hour) and between extinction and test (i.e., 1 week); and measures of fear (i.e., US expectancy, skin conductance response (SCR), and startle reflex). Participants underwent habituation, acquisition, and extinction on Day 1. 1 week later (i.e., Day 8), participants underwent extinction test. All studies were conducted within the same laboratory. Throughout the manuscript, we use the term “fear” to refer to the broad latent construct of fear, which was measured during the CS+ and CS− using US expectancy, SCR, and startle reflex.

**Materials and apparatus**

The Pavlovian conditioning procedures were programmed using E-Prime 2. CSs were 8 sec images on a computer screen, including a CS+ (paired with shock during acquisition) and CS− (never paired with shock). Intertrial intervals (ITIs) varied between 25, 30, and 35 sec. The CS+ and CS− were either a picture of a neutral human face (Study 1, (Zbozinek et al., 2015); Study 2, (Zbozinek & Craske, 2017b); Study 3, Experiment 2, (Zbozinek, 2017)) or a picture of a green triangle or a fork (Study 4) (Zbozinek & Craske, 2018). The CSs were pseudo-randomized with no more than two consecutive presentations of the CS+ or CS− in a given phase for every phase. The first trial of acquisition and extinction had a 50% chance of being the CS+ and a 50% chance of being the CS− (all studies). This same notion applied during extinction test for Studies 1–3; Study 4 tested the CS− and three generalization stimuli before the CS+, in which the CS− had a 50% chance of being the first stimulus presented. The US was 0.5 sec electric shock to the dominant arm bicep delivered 7.5 sec after CS+ onset, using the STMEPM, two LEAD110A (BIOPAC, Inc), and two Tele electrode T716 Ag/AgCl electrodes. The shock consisted of 10 consecutive pulses .05 seconds in duration, totaling to .5 sec. The CS+/US reinforcement rate during acquisition was 100%. BIOPAC MP150 hardware unit and AcqKnowledge version 4.2 software (BIOPAC Systems, Inc) were used to acquire US expectancy, skin conductance, and electromyography (i.e., startle reflex) data.

**US expectancy.** For US expectancy, participants used a sliding scale (BIOPAC model TSD115) and were asked to “Please rate how certain you are that you will receive electric shock in the next few moments.” The values ranged from 0 = “Certain no electric shock,” 4.5 = “Uncertain,” and 9 = “Certain electric shock.” US expectancy was calculated as the mean rating 4.5–5 sec (Studies 1–3) or 6.5–7 sec (Study 4) after CS onset. Studies 1–3 used the term “muscle stimulation” instead of “electric shock.”

**Skin conductance response.** SCR was recorded as a measure of arousal from two EL507 11 mm diameter Ag/AgCl electrodes placed on the distal phalanx of the index and middle fingers of the non-dominant hand (Bradley et al., 1990). Using a GSR100C amplifier and two LEAD110A, SCR data were sampled at a rate of 31.25 Hz and filtered using an FIR low pass filter with a frequency cutoff fixed at 2 Hz. SCR was calculated as a difference score between the maximum skin conductance value 1–6 sec after CS or US onset minus the mean skin conductance value −2 to 0 sec prior to CS or US onset. This difference score was then square-root-transformed to normalize the data, as well as range-corrected by dividing the difference score by the largest observed skin conductance response. SCRs less than zero were coded as zero.

**Startle reflex.** Startle reflex was measured with electromyography (EMG) orbicularis oculi activity under the left eye.
using two EL254S 4 mm Ag/AgCl electrodes filled with SigmaGel electrode gel and measured with the EMG100C amplifier. Electrode placement was 8–10 mm beneath the outer corner of the eye. The second electrode was placed 1 cm medial to the first electrode and 8–10 mm beneath the bottom eyelid (Fridlund & Cacioppo, 1986). The startle probes consisted of 50 milliseconds, 65 dB (Study 1) or 102 dB (Studies 2–4) bursts of white noise. Startle probes occurred 5 or 6.5 sec after CS onset (Study 1) or 7 sec after CS onset (Studies 2–4) and at 10, 15, or 25 sec after onset of the intertrial interval (i.e., ITI), averaging at 16.25 sec. Startle habituation was conducted for Studies 2–4 prior to acquisition (Day 1) and extinction test (Day 8), consisting of 10 startle probes across a 2.5-minute period. EMG data was sampled at a rate of 2 kHz and filtered using an FIR band pass filter with a low frequency cutoff of 30 Hz (Study 1) or 60 Hz (Studies 2–4) and a high frequency cutoff of 1,000 Hz (Study 1) or 500 Hz (Studies 2–4). Data were then converted to an absolute value and smoothed using .005 samples mean value smoothing. Startle reflex was calculated as the difference between the absolute maximum EMG level in volts during the 20 ms–150 ms immediately after the startle probe and the mean EMG level in volts during the 200 ms immediately preceding the startle probe. EMG data were then transformed into a within-subjects t-score using responses to all CSs and ITIs across both days (startle habituation trials were excluded from t-score calculations).

**Trait individual difference variables.** Individual difference variables included the Behavioral Inhibition Scale/Behavioral Activation Scale (i.e., BIS/BAS) (Carver & White, 1994) and the State-Trait Anxiety Inventory (STAI) Trait version (Bieling et al., 1998; Spielberger & Gorsuch, 1983). BIS/BAS was measured in all four studies, but STAI was measured in Studies 1, 3, and 4.

The BIS/BAS includes four subscales: BIS relates to anticipation of punishment, Drive (BAS-D) relates to the persistent pursuit of desired goals, Reward Responsiveness (BAS-RR) relates to positive responses to the occurrence or anticipation of reward, and Fun Seeking (BAS-FS) relates to a desire for new rewards and a willingness to spontaneously approach a potentially rewarding event. Scores were calculated as the average score for each subscale on a 1–4 scale, where 1 = “very false for me,” 2 = “somewhat false for me,” 3 = “somewhat true for me,” and 4 = “very true for me.” Higher scores indicate higher BIS and BAS.

Although initially viewed as a measure of anxiety (Spielberger & Gorsuch, 1983), the STAI was reevaluated and determined to have two subscales: one measuring anxiety and the other measuring depression (Bieling et al., 1998). The anxiety subscale (STAI-A) and depression subscale (STAI-D) have 7 and 13 items, respectively. For example, the anxiety items include, “I feel nervous and restless,” “I worry too much over something that really doesn’t matter,” and “I get in a state of tension and turmoil as I think about my recent concerns and interests.” Conversely, the depression items include “I wish I could be as happy as others seem to be,” “I feel like a failure,” “I am happy,” and “I feel inadequate.” Scores were calculated as the average score for each subscale on a 1–4 scale, where 1 = “Almost never,” 2 = “Sometimes,” 3 = “Often,” and 4 = “Almost always.” Higher scores indicate higher trait anxiety (STAI-A) and depression (STAI-D).

**Procedure**

On Day 1, participants provided informed consent, physiological equipment was attached, and a shock workup procedure was conducted. In the shock workup procedure, shocks started at a low intensity and increased to the level the participant considered “uncomfortable but not painful” using a 0–10 discomfort scale (0 = “Not at all,” 5 = “Moderately,” and 10 = “Very”; M = 6.13 SD = 1.72). Across all studies, shock voltage ranged from 5 to 80 with a mean of 45.68 (SD = 11.97). Each study had a similar shock voltage: (F (3,266) = 2.56, p = .055; Study 1 (M = 43.57, SE = 1.166); Study 2 (M = 46.23, SE = 1.484); Study 3 (M = 48.89, SE = 1.534); and Study 4 (M = 45.49, SE = 1.737)). On Day 1, the primary experimental phases included startle habituation (10 startle probes during an ITI; Studies 2–4), habituation (2 CS+ and 2 CS−), acquisition (CS+/US and CS−; Studies 1–3: 8 trials each; Study 4: 7 trials each), and extinction (CS+/US and CS−; Studies 1–3: 8 trials each; Study 4: 9 trials each). Between acquisition and extinction, Studies 1–3 included mood inductions (Zbozinek, 2017; Zbozinek & Craske, 2017b; Zbozinek et al., 2015). Mood induction was not a significant predictor of Extinction Test fear in any of the individual studies; it was not of theoretical interest to this paper, and including it as a predictor reduced our sample size (since Study 4 had no mood induction). However, to ensure no confounding effects of mood induction, we conducted secondary analyses including mood induction as a predictor. On Day 8, after physiological equipment was attached, procedures included startle habituation (10 startle probes during an ITI; Studies 2–4) and extinction test (1 CS+ and 1 CS−).

**Data analysis**

We used Stata 15.1 for all analyses. For model-free acquisition and extinction (analyzed separately), level 1 predictors included CS Type (CS+ and CS−), Linear Slope (Acquisition: Trials 1–7; Extinction: Trials 1–8), and Quadratic Slope (Acquisition: Trials 1–7; Extinction: Trials 1–8). If the Quadratic Slope was not significant, it was removed from the model. For Extinction Test, level
1 predictors included CS Type (CS+ and CS−) and Linear Slope (Extinction Trial 8, Extinction Test).

Furthermore, we used structural equation modeling to calculate latent intercept, linear slope, and quadratic slope separately for CS+ acquisition, CS+ extinction, CS− acquisition, and CS− extinction for each measure of fear. The latent intercept, linear slope, and quadratic slope were calculated from acquisition trials 1–7 and extinction trials 1–8, where the intercept trials maintained its value across all trials, linear slope increased its value linearly by increments of 1 across trials, and quadratic slope increased its value exponentially to the power of 2 across trials. We first estimated the quadratic model in which intercept, linear slope, and quadratic slope were estimated separately. If this model did not converge, we dropped the quadratic slope and estimated the intercept and linear slope. The quadratic slope did not converge for CS+ acquisition (startle), CS− acquisition (US expectancy and SCR, startle), CS+ extinction (SCR and startle), and CS− extinction (US expectancy and SCR, startle)—most likely because there seemed to be no true quadratic slope in the data (i.e., quadratic slope ≈ 0). In all models, the intercept and linear slope converged. In the dataset, we saved participants’ estimated intercept, linear slope, and quadratic slope for the CS+ and CS− for acquisition and extinction and for each measure of fear. We then used these saved values in subsequent analyses to predict long-term fear (i.e., extinction test), in which CS+ variables predicted CS+ extinction test fear and CS− variables predicted CS− extinction test fear (details in next paragraph). Although the startle reflex CS+ acquisition linear slope converged, the estimates were highly invariant, and our subsequent structural equation modeling analyses predicting extinction test fear showed that their standard errors were generally over 100x that of the other standard errors; thus, we removed CS+ acquisition linear slope from analyses.

For predictors of long-term fear, we used our model-based measures of latent intercepts and linear slopes to predict extinction test fear for the CS+ and CS− separately. We excluded the model-based quadratic slopes as predictors in these analyses since they were theoretically of little interest and offered low interpretability, though we kept them in the models (if they converged) in order to model the data more accurately. In our analyses, we focused on model-based linear slope as our measure of acquisition or extinction rate. We first conceptually replicated the literature by conducting linear regressions using only extinction rate (i.e., extinction linear slope) to predict extinction test fear (Brown et al., 2017; Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015). Then, we built upon the previous literature by examining the specificity of effects of extinction rate versus other variables that could plausibly predict long-term fear. Specifically, we added extinction intercept to the regression model (with extinction slope) to see which of the two would be a stronger predictor of extinction test fear. Next, we added acquisition intercept and acquisition slope to the model to determine the specific effects of acquisition/extinction intercept/slope in predicting long-term fear. This was conducted using structural equation modeling (SEM) in which each predictor was entered in the model chronologically to predict all subsequent variables in the model, ultimately ending with extinction test (i.e., acquisition intercept → acquisition linear slope → extinction intercept → extinction linear slope → extinction test). Lastly, we built upon these SEM models by including individual differences variables (i.e., trait anxiety, depression, behavioral inhibition, and behavioral approach; BIS/BAS, STAI) together as chronologically the first variables in the models. Thus, our final model was individual differences variables → acquisition intercept → acquisition linear slope → extinction intercept → extinction linear slope → extinction test. We standardized all variables in the regression and SEM analyses to produce β values. Additionally, our SEM models were just-identified, which did not allow us to quantitatively assess model fit. However, model comparison was not the goal of our analyses, and our models qualitatively showed very good fit with the observed data (see Figure 1). We chose a SEM approach because it allowed us to estimate direct, indirect, and total effects (a regression model would only assess direct effects in predicting extinction test fear). Also, our model-based estimates showed moderate-to-strong correlations with the model-free data (see Results).

Results

Model-free acquisition

See Figures 1(a)–(c) for fear across experimental phases. Overall, there was significant differential acquisition for all three measures of fear.

For US expectancy and SCR, there was a significant CS Type × Linear Slope × Quadratic Slope interaction ($\chi^2 (1) = 97.77, f = .174$ (US expectancy) and $\chi^2 (1) = 30.02, f = .096$ (SCR); ps < .001) with greater deceleration for the CS+ than the CS−. There was also a significant CS Type × Linear Slope interaction ($\chi^2 (1) = 219.19, f = .260$ (US expectancy) and $\chi^2 (1) = 65.12, f = .142$ (SCR); ps < .001) with a steeper incline for the CS+ than the CS−.

For startle reflex, the CS Type × Linear Slope × Quadratic Slope interaction was not significant ($\chi^2 (1) = 2.11, f = .026, p = .146$). However, there was a significant CS Type × Linear Slope interaction ($\chi^2 (1) = 40.83, f = .113, p < .001$) with a steeper decline for the CS− than the CS+. 
Model-free extinction

Overall, there was significant differential extinction for US expectancy and SCR in which the CS+ extinguished to a greater extent than the CS−; startle reflex also showed differential extinction, though the CS− extinguished more quickly than the CS+.

For US expectancy, there was a significant CS Type × Linear Slope × Quadratic Slope interaction ($\chi^2 (1) = 25.44$, $f = .083$, $p < .001$) with greater acceleration for the CS+ than the CS−. There was also a significant CS Type × Linear Slope interaction ($\chi^2 (1) = 52.23$, $f = .119$, $p < .001$) with a steeper decline for the CS+ than the CS−.

For SCR, the CS Type × Linear Slope × Quadratic Slope interaction was not significant ($\chi^2 (1) = 1.62$, $f = .021$, $p = .203$). However, there was a significant CS Type × Linear Slope interaction ($\chi^2 (1) = 12.92$, $f = .059$, $p = .000$).

Figure 1. Fear Conditioning Results. Panels a–c present model-free (observed) fear conditioning data across habituation, acquisition, extinction, and extinction test. Measures include unconditional stimulus (US) expectancy, skin conductance response (SCR), and startle reflex. US Expectancy was measured on a 0–9 scale, where 0 = “Certain no electric shock,” 4.5 = “Uncertain,” and 9 = “Certain electric shock.” Error bars represent standard error. Panels d–f present model-based results in which latent intercepts, linear slopes, and quadratic slopes were estimated for acquisition and extinction. If quadratic slopes did not converge during model-based estimation, they were dropped and re-run with just intercepts and linear slopes. All models contain intercepts and linear slopes. Results show that our model-based results closely match the model-free results, suggesting the validity of our estimated latent intercepts, linear slopes, and quadratic slopes.
p < .001) with a steeper decline for the CS+ than the CS–.

For startle reflex, there was a significant CS Type × Linear Slope × Quadratic Slope interaction ($\chi^2 (1) = 4.98, f = .037, p = .026$) with greater acceleration for the CS– than the CS+ (i.e., the CS– approached a minimum asymptote more quickly). There was also a significant CS Type × Linear Slope interaction ($\chi^2 (1) = 4.18, f = .034, p = .041$) with a steeper decline for the CS– than the CS+.

**Model-free extinction test**

Overall, there was a significantly greater increase in US expectancy and startle reflex for the CS+ than the CS– from the end of extinction to extinction test. For SCR, there was a non-differential increase for the CS+ and CS–.

For US expectancy and startle reflex, there was a significant CS Type (CS+, CS–) × Linear Slope (Extinction Trial 8, Extinction Test) interaction ($\chi^2 (1) = 40.50, f = .227, p < .001$) (US expectancy) and $\chi^2 (1) = 4.99, f = .081, p = .026$ (startle reflex)) with a steeper incline for the CS+ than the CS–.

For SCR, the CS Type (CS+, CS–) × Linear Slope (Extinction Trial 8, Extinction Test) interaction was not significant ($\chi^2 (1) = 0.28, f = .019, p = .594$), nor was the main effect of CS Type ($\chi^2 (1) = 1.08, f = .013, p = .299$). However, there was a significant main effect increase in Linear Slope for both CSs ($\chi^2 (1) = 79.40, f = .498, p < .001$).

**Latent variable modeling produces accurate estimates of acquisition and extinction intercepts, linear slopes, and quadratic slopes**

See Figure 1(d)–(f) for aggregate results of our model-based estimations of latent intercepts, linear slopes, and quadratic slopes for acquisition and extinction across trials. Also, see Figure 2 for means, standard errors, and individual values of model-based intercepts and slopes. In all cases, the intercepts and linear slopes converged in the models. Additionally, quadratic slopes converged for 1) US expectancy CS+ acquisition, CS+ extinction, and CS– extinction and 2) SCR CS+ acquisition. We saved all participants’ model-estimated intercepts and slopes in the dataset to use in subsequent regressions and structural equation models to predict long-term fear (i.e., extinction test).

Visual inspection of Figure 1 shows that the model-free and model-based results are very similar, suggesting the validity of our model. Additionally, we correlated participants’ latent acquisition and extinction intercepts with observed fear averaged across trials 1 and 2 in each phase. We found that the correlations were very high, ranging .732–.940 (mean = .847): US expectancy (CS+ acquisition: $r = .732$; CS+ extinction: $r = .940$; CS– acquisition: $r = .779$; CS– extinction: $r = .904$), SCR (CS+ acquisition: $r = .879$; CS+ extinction: $r = .844$; CS– acquisition: $r = .799$; CS– extinction: $r = .755$), and startle reflex (CS+ acquisition: $r = .848$; CS+ extinction: $r = .892$; CS– acquisition: $r = .911$; CS– extinction: $r = .882$). Similarly, we correlated participants’ latent acquisition and extinction linear slopes with their observed fear averaged across trials 3 and 4 minus the average of trials 1 and 2 in each phase. We found the correlations were moderately positively correlated, ranging .158 to .719 (mean = .413): US expectancy (CS+ acquisition: $r = .469$; CS+ extinction: $r = .676$; CS– acquisition: $r = .173$; CS– extinction: $r = .719$), SCR (CS+ acquisition: $r = .650$; CS+ extinction: $r = .276$; CS– acquisition: $r = .305$; CS– extinction: $r = .158$), and startle reflex (CS+ acquisition: $r = .301$; CS+ extinction: $r = .387$; CS– acquisition: $r = .388$; CS– extinction: $r = .458$).

Moreover, for all measures of fear, we found significant covariance between the model-based intercept and model-based linear slope for a given CS within a conditioning phase (e.g., covariance of CS+ acquisition intercept and CS+ acquisition linear slope was significant for US expectancy). The correlations between acquisition intercept and linear slope were high in two of three cases for the CS+ (US expectancy: $r = .967$; SCR: $r = .756$, startle reflex: $r = 0.941$) and low for the CS– (US expectancy: $r = .568$; SCR: $r = .763$, startle reflex: $r = 0.715$). The low intercept/linear slope correlation for acquisition CS+ startle reflex should be interpreted with caution, though, since the slope was overall very close to 0, had a very restricted range (range $-0.039$ to $-0.007$), and was highly invariant across participants. This led us to drop CS+ acquisition slope for startle reflex in subsequent models in order to maintain validity of our results. The correlations between extinction intercept and linear slope were low for the CS+ (US expectancy: $r = -0.680$; SCR: $r = -0.952$, startle reflex: $r = -0.742$) and CS– (US expectancy: $r = -0.471$; SCR: $r = -0.981$, startle reflex: $r = -0.674$). Thus, a higher acquisition intercept predicted a more positive acquisition slope for the CS+ as measured by US expectancy and SCR (i.e., participants with higher initial fear had a faster acquisition rate), and a higher acquisition intercept predicted a more negative acquisition slope for the CS– across all measures. Additionally, a lower extinction intercept predicted a less negative extinction slope across both CSs and all measures (i.e., participants with lower fear at the start of extinction had a less negative slope, presumably because there was less room to decrease fear).
When alone in regression, slower extinction rate predicts lower long-term fear

First, to conceptually replicate previous experiments examining just CS+ extinction rate in its prediction of long-term CS+ fear (Brown et al., 2017; Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015), we entered extinction slope as the only predictor of extinction test fear in a regression. The results for all three measures of fear indicate that a less negative (i.e., slower) extinction linear slope predicted less extinction test fear: US expectancy ($\beta = -.316$, $SE = .192$, $t (228) = -5.02$, $p < .001$), SCR ($\beta = -.230$, $SE = 2.050$, $t (213) = -3.45$, $p = .001$), and startle reflex ($\beta = -.242$, $SE = 1.446$, $t (223) = -3.72$, $p < .001$).
This result is the opposite of those found in animal studies (Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015), which found that a more negative (i.e., faster) extinction rate predicted less extinction test fear.

For specificity, we assessed this with the CS+ and found similar but less robust results. A less negative extinction linear slope (i.e., slower) predicted lower extinction test fear as measured by SCR ($\beta = -.210$, SE = 2.953, $t (215) = -3.14, p = .002$) and startle reflex ($\beta = -.176$, SE = 1.538, $t (223) = -2.66, p = .008$) but not US expectancy ($\beta = -.070$, SE = .294, $t (222) = -1.04, p = .300$).

**When entered in model with extinction intercept, extinction rate no longer predicts lower long-term fear, but higher extinction intercept does**

The findings in the previous paragraph could be interpreted to suggest that slower extinction rate predicts lower long-term fear for the CS+ and CS−. However, as stated previously, both i) acquisition intercept and linear slope and ii) extinction intercept and linear slope were significantly covaried. The data shows that individuals with a lower extinction intercept also have a less negative extinction slope. Thus, when entering just extinction slope into the model to predict long-term fear, this effect could be significant due to the intercept rather than the slope.

To investigate this, we included both extinction linear slope and extinction intercept into the regressions predicting extinction test fear. The results strongly changed with both the CS+ and CS−. For the CS+, with all three measures of fear, extinction linear slope no longer predicted long-term fear: US expectancy ($\beta = -.045$, SE = .201, $t (228) = -.68, p = .497$), SCR ($\beta = -.217$, SE = 5.615, $t (213) = -1.19, p = .237$), and startle reflex ($\beta = -.070$, SE = 1.827, $t (223) = -.85, p = .395$). Instead, higher extinction intercept predicted higher extinction test fear with US expectancy ($\beta = .512$, SE = .070, $t (228) = 7.74, p < .001$) and startle reflex ($\beta = .272$, SE = .132, $t (223) = 3.30, p = .001$) but not SCR ($\beta = .014$, SE = .412, $t (213) = .08, p = .937$). Results were very similar for the CS−. Extinction linear slope did not predict extinction test fear: US expectancy ($\beta = .062$, SE = .305, $t (222) = .89, p = .373$), SCR ($\beta = -.177$, SE = 13.578, $t (215) = -.58, p = .565$), and startle reflex ($\beta = -.042$, SE = 1.717, $t (223) = -.57, p = .569$). Instead, higher extinction intercept predicted higher extinction test fear with US expectancy ($\beta = .335$, SE = .081, $t (222) = 4.81, p < .001$) and startle reflex ($\beta = .273$, SE = .126, $t (223) = 3.70, p < .014$) but not SCR ($\beta = .033$, SE = .737, $t (215) = .11, p = .914$). Thus, it appears that extinction intercept is a stronger predictor of CS+ and CS− long-term fear than extinction slope.

**Structural equation models with acquisition and extinction intercepts and linear slopes show that acquisition intercept is the strongest predictor of long-term fear**

To further investigate specificity of effects, we included acquisition intercept, acquisition linear slope, extinction intercept, and extinction linear slope in the models in most cases. The only exception was the exclusion of acquisition linear slope for CS+ startle reflex due to very little variability in the estimation of this measure, which compromised its ability to be a valid predictor (e.g., standard errors were extremely high when predicting CS+ extinction test startle reflex). Additionally, to achieve greater validity in our analyses, we transitioned from regressions to structural equation modeling (SEM) in which variables were arranged in chronological order to predict each subsequent variable, ultimately ending with all variables predicting extinction test fear (e.g., acquisition intercept $\rightarrow$ acquisition linear slope $\rightarrow$ extinction intercept $\rightarrow$ extinction linear slope $\rightarrow$ extinction test fear). This was done separately for the CS+ and CS− for each measure of fear. In our SEM models, all chronologically “earlier” variables were given direct pathways to predict all “later” variables (e.g., acquisition intercept directly predicted all subsequent variables). Previous experiments had not included or reported the effects of acquisition in predicting extinction test fear (Brown et al., 2017; Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015), which provides our analysis with a more specific evaluation of predictive variables than previous work.

See Table 1 for statistical details. Results showed that a higher total effect of acquisition intercept was the most robust predictor of higher extinction test fear for both the CS+ and CS− across all three measures of fear. This was complemented by direct effects of CS+ acquisition intercept predicting extinction test fear for US expectancy and startle reflex. There was also one significant negative indirect effect of acquisition slope predicting less extinction test fear US expectancy. Additionally, higher direct and total effects of extinction intercept predicted higher extinction test fear for the CS+ and CS− as measured by US expectancy. Thus, the results suggest that both the CS+ and CS−, acquisition intercept and extinction intercept are concurrent predictors of long-term fear with US expectancy, acquisition intercept becomes a predictor with SCR, and acquisition intercept replaces extinction intercept with startle reflex. Extinction slope was not a significant predictor of extinction test fear with any measure, and acquisition slope had one indirect effect with US expectancy. Thus, the data emphasize initial fear levels at the beginning of a phase (especially acquisition) in predicting extinction...
### Table 1. CS+/CS− acquisition/extinction intercepts and slopes predicting extinction test fear.

| CS   | Predictor | Effect | US Expectancy | SCR | Startle Reflex |
|------|-----------|--------|---------------|-----|----------------|
|      |           | β      | SE            | Z   | p              | β     | SE  | Z   | p  |
|      |           |        |               |     |                |       |     |     |    |
| CS+  | Acq Intercept | Direct | 0.391 | 0.379 | 2.38 | <.001 | 0.17 | 0.083 | 0.270 | 0.85 | 0.396 | 0.050 | 0.247 | 0.180 | 2.55 | 0.011 | 0.017 |
|      |           | Indirect | 0.063 | 0.362 | 0.40 | 0.689 | — | 0.116 | 0.194 | 1.67 | 0.096 | — | 0.100 | 0.138 | 1.35 | 0.177 | — |
|      | Total     | Direct | 0.454 | 0.136 | 7.68 | <.001 | 0.17 | 0.199 | 0.192 | 2.87 | 0.004 | 0.050 | 0.346 | 0.117 | 5.50 | <.001 | 0.025 |
|      |           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
|      | Ext Intercept | Direct | 0.385 | 0.076 | 5.41 | <.001 | 0.17 | — | 0.042 | 0.432 | 0.22 | 0.827 | 0.050 | 0.104 | 0.167 | 1.00 | 0.317 | 0.025 |
|      |           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
|      | Total     | Direct | 0.402 | 0.068 | 6.34 | <.001 | 0.17 | — | 0.119 | 0.188 | 1.43 | 0.152 | 0.025 | 0.123 | 0.154 | 1.29 | 0.199 | 0.050 |
|      |           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
|      | Ext Slope | Direct | 0.259 | 0.090 | 3.33 | 0.001 | 0.17 | — | 0.014 | 0.764 | 0.04 | 0.966 | 0.050 | 0.068 | 0.176 | 0.66 | 0.510 | 0.025 |
|      |           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
|      | Total     | Direct | 0.238 | 0.082 | 3.36 | 0.001 | 0.17 | — | 0.136 | 0.209 | 1.56 | 0.119 | 0.025 | 0.083 | 0.163 | 0.87 | 0.384 | 0.050 |
|      |           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |

Note. Significant results passing Holm–Bonferroni cutoff in bold. “HB” = Holm–Bonferroni p-value cutoff. We corrected analyses for the same analysis across the three measures of fear, leading to cutoffs of .017, .025, and .050 in ascending order. “Acq” = acquisition; “Ext” = extinction. Unconditional stimulus (US) expectancy rated on 0–9 scale (0 = “Certain no electric shock,” 4.5 = “Uncertain,” and 9 = “Certain electric shock”) and “SCR” = skin conductance response.
test fear rather than changes in fear levels within acquisition or extinction.  

Sample distributions of trait behavioral inhibition, behavioral approach, anxiety, and depression

Figure 3 shows the distribution of values for the BIS/BAS (trait behavioral inhibition and behavioral activation) and STAI (trait anxiety and depression). Values were widely distributed, providing us with a strong foundation from which to predict long-term fear in subsequent analyses: BAS-RR (M = 3.504, SE = .026), BAS-D (M = 2.829, SE = .038), BAS-FS (M = 2.982, SE = .035), STAI-A (M = 1.975, SE = .043), and STAI-D (M = 2.133, SE = .034).

Trait behavioral inhibition, behavioral approach, anxiety, and depression do not predict long-term fear

First, the individual difference variables (i.e., BAS-RR, BAS-D, BAS-FS, BIS, STAI-A, and STAI-D) had low variance inflation factor (VIF; Thompson et al., 2017) scores (VIFs ≤1.98; VIFs ≥10 indicate concerns of multicollinearity), suggesting no concerns of multicollinearity. This supports the concurrent use of these measures within the same analysis as predictors of long-term fear. Second, SEM results showed that none of the individual difference variables (BIS/BAS and STAI) significantly predicted extinction test fear (US expectancy, SCR, and startle reflex) and higher extinction intercept (US expectancy) predicted long-term fear largely remained significant even when including our trait measures in the analyses. See Tables 2 and 3 for details.

Discussion

The present report investigated prediction of long-term fear from four human fear conditioning studies. We estimated participants’ latent acquisition intercept, acquisition slope, extinction intercept, and extinction slope for the CS+ and CS− for each subjective and physiological measure of fear. Our model showed strong convergence with the real data and a wide distribution of estimated values. To conceptually replicate previous studies, we first evaluated the isolated effects of extinction slope on long-term conditional fear (measured 1 week after extinction), hypothesizing that individuals with faster extinction (i.e., a more negative slope) would show less long-term fear. Subsequently, we built upon the previous literature by assessing the specificity of extinction slope by including other fear conditioning variables into the model: we included extinction intercept, followed by acquisition intercept and acquisition slope. These analyses were conducted using structural equation modeling in which variables predicted each other in chronological fashion. We then used trait measures of behavioral approach, behavioral inhibition, anxiety, and depression to assess “who” had greater or lower long-term fear in fear conditioning. We used a multimodal approach to the measurement of fear during every trial, including unconditional stimulus (US) expectancy, skin conductance response (SCR), and startle reflex.

When assessed in isolation, our results showed that slower extinction rate predicted less long-term fear across
Table 2. Individual difference and fear conditioning variables predicting CS+ extinction test fear.

| Predictor | Effect  | US Expectancy | SCR | Startle Reflex |
|-----------|---------|---------------|-----|---------------|
|           | $\beta$ | SE  | Z   | $P$ | HB | $\beta$ | SE  | Z   | $P$ | HB | $\beta$ | SE  | Z   | $P$ | HB |
| BAS-RR    | Direct  | -.172 | .638 | -.220 | .028 | .017 | -.019 | .081 | -.19 | .850 | .050 | -.089 | 3.109 | -.95 | .344 | .025 |
|           | Indirect| -.001 | .491 | -.001 | .989 | .050 | -.085 | .034 | -.19 | 1.97 | .048 | .017 | -.011 | 1.056 | -.34 | .733 | .025 |
|           | Total   | -.173 | .794 | -.177 | .076 | --- | -.104 | .083 | -.10 | 1.00 | .319 | --- | -.100 | 3.247 | -.102 | .309 | --- |
| BAS-D     | Direct  | -.058 | .421 | -.074 | .457 | --- | .153 | .050 | 1.62 | .105 | --- | -.088 | 1.976 | -.97 | .334 | --- |
|           | Indirect| .043  | .325 | .072  | .474 | --- | .012 | .019 | 0.34 | .734 | --- | .026 | 6.959 | .82  | .413 | --- |
|           | Total   | -.015 | .520 | -.015 | .878 | --- | .165 | .052 | 1.67 | .095 | --- | -.062 | 2.053 | -.65  | .514 | --- |
| BAS-FS    | Direct  | .009  | .472 | .013  | .900 | --- | .180 | .056 | 1.89 | .058 | --- | -.079 | 2.257 | -.87  | .387 | --- |
|           | Indirect| .032  | .362 | .054  | .586 | --- | .064 | .023 | 1.67 | .095 | --- | .019 | 8.070 | .56  | .572 | --- |
|           | Total   | .041  | .589 | .043  | .664 | --- | .116 | .058 | 1.17 | .241 | --- | -.061 | 2.340 | -.64  | .522 | --- |
| BIS       | Direct  | .142  | .546 | 1.73  | .083 | --- | -.051 | .068 | 0.51 | .611 | --- | -.030 | 2.712 | -.30  | .764 | --- |
|           | Indirect| .005  | .417 | .009  | .931 | .050 | .026 | .025 | 0.69 | .487 | .025 | .083 | 1.071 | 2.12 | .034 | .017 |
|           | Total   | .147  | .684 | 1.44  | .151 | --- | -.026 | .072 | 0.24 | .809 | --- | .053 | 2.767 | .53  | .598 | --- |
| STAI-A    | Direct  | .025  | .482 | 0.29  | .773 | --- | .020 | .059 | 0.18 | .858 | --- | .125 | 2.357 | 1.20  | .232 | --- |
|           | Indirect| .038  | .372 | -.57  | .570 | --- | .038 | .024 | 0.85 | .396 | --- | .043 | 8.242 | 1.18  | .239 | --- |
|           | Total   | -.013 | .596 | -.012 | .903 | --- | .058 | .061 | 0.51 | .607 | --- | .082 | 2.452 | .75  | .451 | --- |
| STAI-D    | Direct  | -.149 | .600 | -.170 | .088 | --- | .025 | .074 | 0.23 | .820 | --- | -.142 | 2.941 | 1.35  | .176 | --- |
|           | Indirect| .110  | .462 | 1.64  | .102 | --- | .010 | .027 | 0.25 | .805 | --- | .029 | 1.055 | .77  | .439 | --- |
|           | Total   | .039  | .744 | -.36  | .720 | --- | .034 | .078 | 0.30 | .765 | --- | -.113 | 3.048 | 1.04 | .300 | --- |
| Acq Intercept | Direct   | .395  | .449 | 2.08  | .038 | .017 | .075 | .321 | 0.67 | .501 | .050 | .182 | 2.122 | 1.67  | .095 | .025 |
|           | Indirect| .108  | .429 | 0.59  | .553 | .050 | .158 | .227 | 2.01 | .045 | .017 | .103 | 1.545 | 1.30  | .192 | .025 |
|           | Total   | .503  | .515 | 7.52  | .001 | .017 | .233 | .242 | 2.78 | .005 | .050 | .285 | 1.492 | 3.73  | <.001 | .025 |
| Acq Slope | Direct  | -.095 | .963 | -.52  | .603 | --- | .054 | .576 | 0.51 | .609 | --- | --- | --- | --- | --- | --- |
|           | Indirect| .133  | .426 | 1.65  | .100 | --- | .063 | .204 | 1.67 | .095 | --- | --- | --- | --- | --- | --- |
|           | Total   | -.228 | 1.022 | 1.18  | .239 | --- | .117 | .571 | 1.11 | .266 | --- | --- | --- | --- | --- | --- |
| Ext Intercept | Direct   | .369  | .087 | 4.50  | <.001 | .017 | -.129 | .535 | -0.57 | .568 | .050 | .070 | .196 | 0.58  | .562 | .025 |
|           | Indirect| .014  | .038 | .39   | .695 | --- | .353 | .481 | 1.74 | .081 | --- | .056 | .082 | 1.11  | .265 | --- |
|           | Total   | .383  | .078 | 5.21  | <.001 | .017 | .225 | .239 | 2.23 | .026 | .025 | .126 | .182 | 1.14  | .254 | .050 |
| Ext Slope | Direct  | -.029 | .244 | 0.39  | .694 | --- | .380 | 6.990 | 1.75 | .081 | --- | -.110 | 2.061 | 1.14  | .256 | --- |
|           | Indirect| ---   | ---  | ---   | ---  | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

Note: Significant results passing Holm–Bonferroni cutoff in bold. "HB" = Holm–Bonferroni p-value cutoff. We corrected analyses for the same analysis across the three measures of fear, leading to cutoffs of .017, .025, and .050 in ascending order. "Acq" = acquisition; "Ext" = extinction; "Cov" = covariance of intercept and linear slope. Unconditional stimulus (US) expectancy rated on 0-9 scale (0 = Certain no electric shock, 4 = Uncertain, and 9 = Certain electric shock); and "SCR" = skin conductance response. "BAS-RR" = Behavioral Activation Scale Reward Responsiveness, "BAS-D" = BAS Drive, "BAS-FS" = BAS Fun Seeking, "BIS" = Behavioral Inhibition Scale, "STAI-A" = State-Trait Anxiety Inventory Anxiety subscale, "STAI-D" = STAI Depression subscale.
Table 3. Individual difference and fear conditioning variables predicting CS− extinction test fear.

| Predictor | Effect | US Expectancy | SCR | Startle Reflex |
|-----------|--------|---------------|-----|----------------|
|           |        | β | SE | Z | P | HB | β | SE | Z | P | HB | β | SE | Z | P | HB |
| BAS-RR    | Direct | .059 | 0.627 | 0.65 | 0.515 | — | — | — | — | — | — | 0.066 | 2.854 | 0.70 | 0.482 | — |
|           | Indirect | —0.028 | .236 | —0.83 | .407 | — | — | — | — | — | — | 0.008 | .998 | —0.26 | 0.796 | — |
|           | Total | .031 | 0.656 | 0.32 | 0.746 | — | — | — | — | — | — | 0.057 | 3.004 | 0.58 | 0.561 | — |
| BAS-D     | Direct | .128 | 0.411 | 1.42 | 0.155 | — | — | — | — | — | — | —1.28 | 1.823 | —1.41 | 0.159 | — |
|           | Indirect | .030 | 0.155 | 0.87 | 0.382 | — | — | — | — | — | — | 0.039 | 0.685 | 1.13 | 0.259 | — |
|           | Total | .158 | 0.430 | 1.67 | 0.094 | — | — | — | — | — | — | —0.089 | 1.899 | —0.94 | 0.345 | — |
| BAS-FS    | Direct | .130 | 0.468 | 1.45 | 0.147 | — | — | — | — | — | — | —1.39 | 2.047 | —1.55 | 0.121 | — |
|           | Indirect | —0.022 | 0.183 | —0.62 | 0.532 | — | — | — | — | — | — | 0.005 | 0.700 | 0.17 | 0.867 | — |
|           | Total | .108 | 0.487 | 1.16 | 0.245 | — | — | — | — | — | — | —1.13 | 2.161 | —1.41 | 0.158 | — |
| BIS       | Direct | —0.035 | 0.538 | —0.37 | 0.709 | — | — | — | — | — | — | 0.000 | 2.505 | <0.001 | 0.997 | — |
|           | Indirect | —0.041 | 0.197 | —1.17 | 0.241 | — | — | — | — | — | — | 0.068 | 0.982 | 1.78 | 0.076 | — |
|           | Total | —0.076 | 0.565 | —0.76 | 0.446 | — | — | — | — | — | — | 0.068 | 2.590 | 0.67 | 0.503 | — |
| STAI-A    | Direct | .112 | 0.469 | 1.13 | 0.257 | — | — | — | — | — | — | 0.058 | 2.160 | 0.56 | 0.577 | — |
|           | Indirect | .021 | 0.179 | 0.56 | 0.576 | — | — | — | — | — | — | 0.039 | 0.767 | —1.06 | 0.287 | — |
|           | Total | .134 | 0.489 | 1.56 | 0.146 | — | — | — | — | — | — | 0.019 | 2.269 | 0.17 | 0.864 | — |
| STAI-D    | Direct | .055 | 0.593 | 0.53 | 0.595 | — | — | — | — | — | — | —0.077 | 2.669 | —0.75 | 0.451 | — |
|           | Indirect | .087 | 0.243 | 2.06 | 0.039 | .017 | — | — | — | — | — | 0.016 | 0.930 | 0.45 | 0.654 | 0.025 |
|           | Total | .142 | 0.610 | 1.34 | 0.181 | — | — | — | — | — | — | —0.061 | 2.815 | —0.57 | 0.571 | — |
| Acq Intercept | Direct | .107 | 0.104 | 1.11 | 0.268 | — | — | — | — | — | — | 0.177 | 2.222 | 1.36 | 0.175 | — |
|           | Indirect | .094 | 0.068 | 1.49 | 0.137 | .025 | — | — | — | — | — | 0.124 | 1.821 | 1.16 | 0.248 | 0.050 |
|           | Total | .201 | 0.085 | 2.54 | 0.11 | .025 | — | — | — | — | — | 0.302 | 1.288 | 3.99 | <0.001 | 0.017 |
| Acq Slope | Direct | —0.021 | 0.578 | —0.24 | 0.813 | — | — | — | — | — | — | —0.079 | 2.204 | —0.82 | 0.415 | — |
|           | Indirect | .041 | 0.204 | 1.32 | 0.186 | — | — | — | — | — | — | —0.001 | 0.704 | —0.02 | 0.983 | — |
|           | Total | .020 | 0.584 | 0.23 | 0.820 | — | — | — | — | — | — | —0.080 | 2.112 | —0.86 | 0.391 | — |
| Ext Intercept | Direct | .230 | 0.101 | 2.65 | 0.008 | .017 | — | — | — | — | — | 0.066 | 1.822 | 0.57 | 0.569 | 0.025 |
|           | Indirect | .037 | 0.041 | 1.04 | 0.297 | — | — | — | — | — | — | 0.038 | 0.604 | 0.93 | 0.354 | — |
|           | Total | .267 | 0.093 | 3.34 | 0.001 | .017 | — | — | — | — | — | 0.105 | 1.711 | 0.95 | 0.342 | 0.050 |
| Ext Slope | Direct | .084 | 0.350 | —1.06 | 0.288 | — | — | — | — | — | — | —0.080 | 1.843 | —0.95 | 0.344 | — |
|           | Indirect | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |

Note. Significant results passing Holm–Bonferroni cutoff in bold. "HB" = Holm–Bonferroni p-value cutoff. We corrected analyses for the same analysis across the three measures of fear, leading to cutoffs of .017, .025, and .050 in ascending order. "Acq" = acquisition; "Ext" = extinction; "Cov" = covariance of intercept and linear slope. Uncovntional stimulus (US) expectancy rated on 0–9 scale (0=Certain no electric shock,"4.5=Uncertain," and 9=Certain electric shock") and "SCR" = skin conductance response. "BAS-RR" = Behavioral Activation Scale Reward Responsiveness, "BAS-D" = BAS Drive, "BAS-FS" = BAS Fun Seeking, "BIS" = Behavioral Inhibition Scale, "STAI-A" = State-Trait Anxiety Inventory Anxiety subscale, "STAI-D" = STAI Depression subscale.
all three measures of fear for the CS+ and across two measures for the CS− (SCR and startle reflex). These results counter previous animal studies (Bush et al., 2007; King et al., 2017; Reznikov et al., 2015), which demonstrated that faster extinction rate predicted less long-term fear. However, a less negative extinction slope was significantly covaried with a lower extinction intercept. When extinction intercept was entered into the model along with extinction slope, extinction slope no longer predicted extinction test fear for the CS+ or CS−; instead, higher extinction intercept predicted higher extinction test fear across US expectancy and startle reflex for the CS+ and CS− (but not SCR). Thus, the extinction slope effect seems to be better explained by extinction intercept, ultimately leading to no effect of extinction slope.

We further expanded our extinction analyses to include acquisition intercept and acquisition slope. Results showed that higher acquisition intercept predicted higher extinction test fear across all three measures of fear, and higher extinction intercept predicted higher extinction test using US expectancy. Thus, in our most stringent analyses that included acquisition intercept, acquisition slope, extinction intercept, and extinction slope to predict extinction test fear, the results suggest that higher acquisition intercept is the strongest predictor of long-term fear with some unique predictive ability from extinction intercept; this pattern occurred for both the CS+ and CS−. In this model, there were no significant effects of extinction slope. Our results thus emphasize the importance of starting fear (during acquisition and secondarily during extinction) in predicting long-term fear and de-emphasize rate of acquisition or extinction. Given that test–retest reliability has been generally shown to be strong when using startle reflex, SCR, and US expectancy as measures of fear across periods of 3 weeks to 8 months with 2–3 rounds of fear conditioning (Fredrikson et al., 1993; Torrents-Rodas et al., 2014; Zeidan et al., 2012), fear conditioning is considered to be reliable within individuals across multiple sessions. Hence, the individual difference assessments we conducted in our study likely reflect stable effects.

Our final structural equation model simultaneously included trait measures of behavioral approach, behavioral inhibition, anxiety, and depression as predictors. There were no significant effects of these trait variables, and their inclusion did not change the primary findings.

The finding that extinction slope had no effects above and beyond acquisition intercept, acquisition rate, or extinction intercept could suggest that the process of extinction is not critical to long-term fear reduction. On the other hand, it may reflect discord between level of fear expressed during extinction and associative mechanisms of inhibitory learning. It has long been recognized that the level of fear expressed during extinction is not an index of long-term learning and that fear at the end of extinction is not related to fear at test (Plendl & Wotjak, 2010).

The effects of acquisition intercept, acquisition slope, extinction intercept, and extinction slope on long-term fear were consistent between the CS+ and CS−. There are at least two ways to interpret this finding. First, it may be indicative of fear generalization, in that individuals who show high initial fear to a novel stimulus also show elevated fear to that stimulus 1 week later, regardless of whether the stimulus becomes a danger or safety signal. Fear generalization from the start of acquisition (or, secondarily, from the start of extinction) may be a metric of heightened anticipatory anxiety that is often observed in anxiety disorders.

For example, meta-analyses show that clinically anxious individuals have elevated fear of the CS− (acquisition) as well as the CS+ (acquisition and extinction) compared to healthy controls (Duits et al., 2015; Lissek et al., 2005). It is also notable that acquisition intercept is the point during which uncertainty in the experiment is maximal—participants know they will get shocked at some point, but they do not know what will predict shock. Anticipatory anxiety may be greatest at this point of maximal uncertainty. Second, non-associative processes of higher initial reactivity (measured by US expectancy, SCR, and startle reflex) predicted higher responding when tested 1 week later. Whether this level of reactivity is specific to fear or broader to emotions in general is not discernible in the current experiment where we tested fear only: fear versus emotionality could be distinguished, for example, using a within-subjects design that included both fear and reward conditioning. Thus, it is unclear whether our findings indicate that i) poorer Pavlovian discrimination and higher fear at the start of acquisition and start of extinction predict greater long-term fear or ii) whether greater fear/emotional reactivity at the start of acquisition and start of extinction predict greater long-term fear/emotional responding. Regardless, our results highlight that initial responding in a fear conditioning task is a stronger predictor of long-term fear than acquisition rate, extinction intercept, or extinction rate.

Clinically, our null extinction rate findings are consistent with data from exposure therapy—the clinical proxy of extinction—which shows that the amount of within-session fear reduction is not a reliable predictor of long-term fear (Brown et al., 2017) nor necessary for positive outcomes from exposure therapy (Baker et al., 2010; Culver et al., 2012). Our finding that extinction rate is relatively non-predictive of long-term fear is neither supportive nor unsupportive of two major models of exposure therapy: the inhibitory learning model (Craske et al., 2014, 2022) and the habituation model (Foa & Kozak, 1986). The inhibitory learning model argues that elevated US expectancy during exposure will result in greater prediction error that generates competing CS–noUS associations, albeit contextually...
specific (Bouton, 1993; Rescorla & Wagner, 1972), whereas the habituation model argues that fear reduction during exposure will result in greater long-term fear reduction. Thus, the inhibitory model favors elevated US expectancy during exposure (which often slows fear reduction, thus creating slower extinction rate but is not theoretically reliant upon fear levels), whereas the habituation model favors fear reduction during exposure (i.e., faster extinction rate). Our results suggest that extinction rate altogether is unimportant when assessed alongside initial acquisition fear, acquisition rate, and initial extinction fear, which poses a challenge for the habituation model. Our results indicate that relatively greater focus in clinical settings could be placed on preventing the acquisition of fear.

Lastly, we found no effects of trait anxiety, depression, behavioral approach, or behavioral inhibition in predicting long-term fear. This is somewhat surprising given meta-analytic evidence that clinically anxious individuals show greater fear of the CS+ during acquisition and extinction and the CS− during acquisition compared to healthy controls (Duits et al., 2015; Lissek et al., 2005). One possible explanation for the discrepancy is that the moderating effect of trait anxiety upon extinction is only observable in clinically severe samples. On the other hand, the meta-analytic results did not evaluate long-term extinction retest and considered trait anxiety in isolation rather than as part of a comprehensive set of predictors of fear responding. Additionally, there is some evidence suggesting that reward processing is involved in extinction learning and that anhedonia or deficits in reward processing are predictive of poorer extinction learning (Abraham et al., 2014; Admon & Pizzagalli, 2015; Felsenberg et al., 2018; Gerlicher et al., 2018; Hikind & Maroun, 2008; Luo et al., 2018; Mueller et al., 2010; Rosenberg et al., 2021; Staples-Bradley et al., 2018; Zbozinek et al., 2015; Zbozinek & Craske, 2017a, 2017b); however, our reward processing and depressive predictors (behavioral approach and depressive) were not predictive of long-term fear. One possibility is that our measures of reward processing were not optimal for assessing this (e.g., we measured depression more broadly rather than anhedonia specifically).

There were a few limitations of the present study. First, while the four studies in this report were largely similar, there were minor differences in methods. Second, three of the studies contained mood inductions between acquisition and extinction, which were not of theoretical interest in the present manuscript. However, analyzing the data with and without the mood induction produced very similar results. Third, acquisition and extinction were conducted on the same day, which may have reduced excitatory learning via “unlearning” mechanisms (Myers et al., 2006) and thus reduced the fear observed 1 week later at test. Fourth, we chose a baseline correction approach to SCR calculations (Sjouwerman et al., 2021) in which negative SCRs were coded as zeros. This is commonly used and necessitated when using square root or log transformation. As is consistent with other reports (Sjouwerman et al., 2021), most zero or negative SCRs using baseline correction occurred to the CS− (59.71% of trials), followed by CS+ (46.83%) and US (3.48%). Negative values in the baseline correction could reflect i) a true non-response to the CS with a habituation drift or ii) a relatively small response to the CS that is overshadowed by the relatively greater habituation drift. A comparison of zero-valued SCRs from the baseline-corrected approach (the approach we used) versus trough-to-peak SCRs showed that both approaches had values of 0 for 69.1% of trials (indicating a true habituation drift), whereas ~30.9% of zero-valued baseline-corrected SCRs had positive values in the trough-to-peak approach (indicating a relatively small SCR that is overshadowed by a relatively greater habituation drift) (Sjouwerman et al., 2021). Thus, the baseline-corrected approach we used may have missed some small positive SCR values that the trough-to-peak approach would have detected. Despite the potential limitation of this SCR calculation, our SCR results for a) acquisition, extinction, and extinction test and b) the association of acquisition/extinction intercept/slope with extinction test are cross-validated with the US expectancy and startle reflex measures. This consistency is encouraging and suggests that the SCR results are valid and informative. Lastly, our experiments model a 1-week interval between extinction and test. It is unclear how these results would extend over longer periods of time or multiple sessions of extinction, which would be relevant for exposure treatment of anxious individuals.

In conclusion, the present report suggests that fear at the start of acquisition of fear conditioning is the strongest predictor of fear 1 week after extinction, followed by fear at the start of extinction. Slower extinction rate initially predicted lower long-term fear when entered in the model alone, which countered previous animal studies (Bush et al., 2007; King et al., 2017, 2018; Reznikov et al., 2015). However, this effect was no longer significant when including extinction intercept, acquisition intercept, and acquisition slope. These results were consistent across both the CS+ and CS−, suggesting that individuals who are initially fearful of a stimulus will likely be afraid of it later—regardless of whether the stimulus becomes a danger signal or safety signal. Lastly, there were no effects of trait behavioral approach, behavioral inhibition, anxiety, or depression in predicting long-term fear.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of California, Los Angeles Dissertation Year Fellowship award.

ORCID iD
Tomislav D. Zbozinek https://orcid.org/0000-0003-0187-671X

Notes
1. Because of this methodological difference in Study 4 versus Studies 1–3, we ran the main analyses with and without Study 4 data. The results were consistent with each other, suggesting this testing procedure had little-to-no impact on our results.
2. We estimated the intercepts/slopes of the CS+ and CS– separately and analyzed the CS+ and CS– separately because using the CS+ and CS– as a within-subjects factor would greatly complicate the SEM model, making it more difficult to estimate the latent intercepts and slopes. Nonetheless, for completeness, we estimated the intercepts and slopes using the CS+ minus CS– difference scores. Overall, the estimations converged with US expectancy and the results were consistent with the results when analyzing the CS+ and CS– separately. However, the models had more difficulty converging with SCR and startle reflex, ultimately preventing us from estimating extinction rate (SCR and startle reflex) and acquisition rate (startle reflex).
3. We evaluated potential multicollinearity within each model by running regressions including all predictors within a model and assessing the variance inflation factor (VIF; Thompson et al., 2017), where VIFs > 10 may indicate multicollinearity. All VIFs ranged from 1.23 to 9.72, except for SCR CS– extinction intercept and extinction slope, which had VIFs of 21.61 and 21.37, respectively (total mean = 4.762). We thus re-ran these analyses excluding either SCR CS– extinction intercept or extinction slope and found that the results did not change (i.e., CS– acquisition intercept continued to significantly predict CS– extinction test SCR, but CS– extinction intercept and extinction linear slope remained non-significant). Thus, for congruency with other analyses, we present the results that include both SCR CS– extinction intercept and extinction slope in the model with the knowledge that results remain the same regardless of whether we include both or exclude one.
4. Studies 1–3 had a between-subjects mood induction condition (positive vs. negative) before extinction. We conducted the same SEM analyses but inserted mood induction as a predictor before extinction intercept. Across the CS+ and CS– and all three measures of fear, mood induction did not have any significant direct, indirect, or total effects in predicting extinction test fear (ps > .110).
5. For acquisition, results were significant when using simple conditioning procedures (CS+ only), not differential conditioning (CS+ and CS–).

References
Abraham, A. D., Neve, K. A., & Lattal, K. M. (2014). Dopamine and extinction: A convergence of theory with fear and reward circuitry. Neurobiology of Learning and Memory, 108, 65–77. https://doi.org/10.1016/j.nlm.2013.11.007
Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. Current Opinion in Psychology, 4, 114–118. https://doi.org/10.1016/j.copsyc.2014.12.011
Baker, A., Mystkowski, J., Culver, N., Yi, R., Mortazavi, A., & Craske, M. G. (2010). Does habituation matter? Emotional processing theory and exposure therapy for acrophobia. Behaviour Research and Therapy, 48(11), 1139–1143. https://doi.org/10.1016/j.brat.2010.07.009
Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The state–trait anxiety inventory, trait version: Structure and content re-examined. Behaviour Research and Therapy, 36(7–8), 777–788. https://doi.org/10.1016/s0005-7967(98)00023-0
Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. Psychological Bulletin, 114(1), 80–99. https://doi.org/10.1037/0033-2909.114.1.80
Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle reflex modification: Emotion or attention? Psychophysiology, 27(5), 513–522. https://doi.org/10.1111/j.1469-8868.1990.tb01966.x
Brown, L. A., LeBeau, R. T., Chat, K. Y., & Craske, M. G. (2017). Associative learning versus fear habituation as predictors of long-term extinction retention. Cognition & Emotion, 31(4), 687–698. https://doi.org/10.1080/02699931.2016.1158695
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
Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. Journal of Personality and Social Psychology, 67(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319
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., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. Behaviour Research and Therapy, 58, 10–23. https://doi.org/10.1016/j.brat.2014.04.006
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
Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology, 39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Andión, O., Molinuevo, B., Caseras, X., & Torrubia, R. (2014). Testing the temporal stability of individual differences in the acquisition and generalization of fear. *Psychophysiology, 51*(7), 697–705. https://doi.org/10.1111/psyp.12213

Wang, Z., Li, Q., Nie, L., & Zheng, Y. (2020). Neural dynamics of monetary and social reward processing in social anhedonia. *Social Cognitive and Affective Neuroscience, 15*(9), 991–1003. https://doi.org/10.1093/scan/nsaa128

Young, K. S., Bookheimer, S. Y., Nusslock, R., Zinbarg, R. E., Damme, K. S. F., Chat, I. K.-Y., Kelley, N. J., Vinograd, M., Perez, M., Chen, K., Cohen, A. E., & Craske, M. G. (2021). Dysregulation of threat neurocircuitry during fear extinction: The role of anhedonia. *Neuropsychopharmacology, 46*(9), 1650–1657. https://doi.org/10.1038/s41386-021-01003-8

Zbozinek, T. D. (2017). The effect of positive affect on extinction learning and return of fear. [UCLA]. https://escholarship.org/uc/item/33m547fv

Zbozinek, T. D., & Craske, M. G. (2017a). The role of positive affect in enhancing extinction learning and exposure therapy for anxiety disorders. *Journal of Experimental Psychopathology, 8*(1), 13–39. https://doi.org/10.5127/jep.052615

Zbozinek, T. D., & Craske, M. G. (2017b). Positive affect predicts less reacquisition of fear: Relevance for long-term outcomes of exposure therapy. *Cognition & Emotion, 31*(4), 712–725. https://doi.org/10.1080/02699931.2016.1142428

Zbozinek, T. D., & Craske, M. G. (2018). Pavlovian extinction of fear with the original conditional stimulus, a generalization stimulus, or multiple generalization stimuli. *Behaviour Research and Therapy, 107*, 64–75. https://doi.org/10.1016/j.brat.2018.05.009

Zbozinek, T. D., Holmes, E. A., & Craske, M. G. (2015). The effect of positive mood induction on reducing reinstatement fear: Relevance for long term outcomes of exposure therapy. *Behaviour Research and Therapy, 71*, 65–75, https://doi.org/10.1016/j.brat.2015.05.016

Zeidan, M. A., Lebron-Milad, K., Thompson-Hollands, J., Im, J. J. Y., Dougherty, D. D., Holt, D. J., Orr, S. P., & Milad, M. R. (2012). Test–retest reliability during fear acquisition and fear extinction in humans. *CNS Neuroscience & Therapeutics, 18*(4), 313–317. https://doi.org/10.1111/j.1755-5949.2011.00238.x