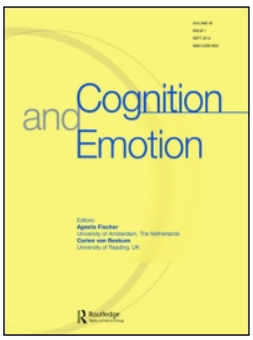

Routledge
Taylor & Francis Group

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

Associative learning versus fear habituation as predictors of long-term extinction retention

Lily A. Brown, Richard T. LeBeau, Ka Yi Chat & Michelle G. Craske

To cite this article: Lily A. Brown, Richard T. LeBeau, Ka Yi Chat & Michelle G. Craske (2017) Associative learning versus fear habituation as predictors of long-term extinction retention, Cognition and Emotion, 31:4, 687-698, DOI: 10.1080/02699931.2016.1158695

To link to this article: https://doi.org/10.1080/02699931.2016.1158695

COGNITION AND EMOTION, 2017
VOL. 31, NO. 4, 687–698
http://dx.doi.org/10.1080/02699931.2016.1158695

Routledge
Taylor & Francis Group

Associative learning versus fear habituation as predictors of long-term extinction retention

Lily A. Brown[a], Richard T. LeBeau[a,b], Ka Yi Chat[c] and Michelle G. Craske[a]

[a]Department of Psychology, University of California, Los Angeles, CA, USA; [b]VA West Los Angeles Medical Center, CA, USA; [c]Department of Psychology, Boston University, Boston, MA, USA

ABSTRACT

Violation of unconditioned stimulus (US) expectancy during extinction training may enhance associative learning and result in improved long-term extinction retention compared to within-session habituation. This experiment examines variation in US expectancy (i.e., expectancy violation) as a predictor of long-term extinction retention. It also examines within-session habituation of fear-potentiated startle (electromyography, EMG) and fear of conditioned stimuli (CS) throughout extinction training as predictors of extinction retention. Participants ($n = 63$) underwent fear conditioning, extinction and retention and provided continuous ratings of US expectancy and EMG, as well as CS fear ratings before and after each phase. Variation in US expectancy throughout extinction and habituation of EMG and fear was entered into a regression as predictors of retention and reinstatement of levels of expectancy and fear. Greater variation in US expectancy throughout extinction training was significantly predictive of enhanced extinction performance measured at retention test, although not after reinstatement test. Slope of EMG and CS fear during extinction did not predict retention of extinction. Within-session habituation of EMG and self-reported fear is not sufficient for long-term retention of extinction learning, and models emphasizing expectation violation may result in enhanced outcomes.

ARTICLE HISTORY
Received 15 May 2015
Revised 19 February 2016
Accepted 22 February 2016

KEYWORDS
Expectation violation; inhibitory learning; fear conditioning; fear extinction; anxiety

Research has long noted the importance of assessing both anxiety, a future-oriented mood state associated with preparation for possible upcoming negative events, and fear, an alarm response to present or imminent danger regardless of whether it is real or perceived (e.g., Barlow, 1988; Craske et al., 2009). Anxiety and fear in turn are comprised three key response components: characterised as subjective-verbal, physiological and behavioural (Lang, 1971). As such, comprehensive measurement should include the subjective-verbal symptoms of fear [i.e., arousal ratings to a conditional stimulus (CS)] and anxiety [i.e., worry about future encounters with an unconditional stimulus (US)], physiological symptoms of fear (i.e., startle response) and anxiety (i.e., muscle tension), and the behavioural symptoms of fear (i.e., escape) and anxiety (i.e., avoidance and safety behaviours). A further indication of associative learning during fear conditioning is US expectancy, or a measure of expectancy about the likelihood of US occurrence in the presence of a CS. Recent reviews have highlighted the importance and relevance of US expectancy as one component of conditioned fear (Boddez et al., 2013), but no studies to our knowledge have examined US expectancy during extinction training as a predictor of subsequent extinction retention.

Dissociation between performance during extinction training and extinction learning is well established in animal studies (Akirav, Raizel, & Maroun,

CONTACT Michelle G. Craske craske@psych.ucla.edu
This article was originally published with errors. This version has been corrected. Please see Erratum (http://dx.doi.org/10.1080/02699931.2016.1170464)

2006; Plendl & Wotjak, 2010) and has been replicated in human studies using psychophysiology and functional magnetic resonance imaging (Phelps, Delgado, Nearing, & LeDoux, 2004). Such dissociation is demonstrated by a test of extinction retention following completion of extinction training. Further, dissociation across fear response systems has been observed across measures of conditioned fear, including US expectancy, or the expectation that an aversive event will occur, self-reported fear, skin conductance response and fear-potentiated startle (Boddez et al., 2013). The goal of this experiment is to extend upon extant research in humans to determine whether performance during extinction training is predictive of retention of extinction learning using measures of perceived contingency between the conditional stimulus and the US as well as subjective and objective measures of conditional fear. The results may inform clinical practice by clarifying the specific changes during extinction that predict long-term retention of extinction learning, which is the ultimate goal of exposure therapy.

Evidence for the dissociation between extinction training and extinction learning is abundant in the rodent literature. For example, reduction in the level of conditional fear responding across repeated extinction trials is not a reliable predictor of conditional fear measured at extinction retention in mice (Plendl & Wotjak, 2010). Also, a number of studies demonstrate that some rats successfully recall extinction at retention whereas others do not, despite the fact that the two groups perform equivalently during fear acquisition and fear extinction (e.g., Peters, Dieppa-Perea, Melendez, & Quirk, 2010). Dissociation is also observed in neural activation. Neuronal activation in the ventromedial prefrontal cortex (vmPFC), a brain region implicated in extinction learning, is apparent during delayed test of extinction, but not during immediate extinction training in rodents (Milad & Quirk, 2002). Furthermore, whereas immediate fear reduction requires intact cannabinoid receptor type 1 signalling, fear reduction from one extinction training to the next does not (Plendl & Wotjak, 2010).

Few human studies have evaluated the dissociation between extinction performance and extinction learning. The studies that are available are correlational, reflective of their preliminary nature. In humans, the central nucleus of the amygdala, involved in fear expression through projections to other brain regions, is inhibited by direct input from the vmPFC (see Phelps et al., 2004 for a review). Connectivity analyses have implicated that functional activation of both the hippocampus, involved in contextual modulation of extinction (Corcoran, Desmond, Frey, & Maren, 2005), and vmPFC are associated with recall of the extinction memory in humans (Milad et al., 2007). Although the amygdala is 'downregulated' during both extinction training and tests of extinction retention, skin conductance responding to the CS correlates with amygdala activation only during extinction training trials and not at extinction retention (Phelps et al., 2004). Furthermore, vmPFC activation was negatively correlated with amygdala activation, and positively correlated with attenuation of psychophysiological responses to the CS at extinction retention but was not correlated with psychophysiological responses to the CS during extinction training (Phelps et al., 2004). These data suggest that the neuronal underpinnings of extinction training and retention of extinction learning and performance are dissociable.

The dissociation between extinction training and extinction retention has inspired a growing line of research aimed at the prediction of long-term extinction learning from immediate extinction performance. In the context of the clinical proxy of extinction, or exposure therapy, sustained fear (Culver, Stoyanova, & Craske, 2012) and variability in fear (Kircanski et al., 2012) throughout exposure were found to be more robust indicators of long-term outcome (as measured by fear during a behavioural avoidance test at follow-up) than fear habituation (see also Baker et al., 2010). These findings have been interpreted to be consistent with an inhibitory learning model of extinction (Bouton, 2004), in which fear habituation during extinction training is not an essential element (see Craske, Liao, Brown, & Vervliet, 2012). The emphasis on variability in particular is also consistent with decades of findings on the importance of surprise and expectation violation as a robust predictor of long-term extinction learning (Rescorla & Wagner, 1972). In other words, variations in subjective fear may reflect repeated instances of elevations in expectancy for aversive outcomes followed by violation of such expectancies and subsequent decreases in expectancy. Translation of the exposure variability data to extinction training would suggest that greater absolute differences in expectancy for the US across all trials of extinction training may serve as an index of expectation violation.

US expectancy is an understudied but important dimension of conditioned in both clinical and experimental contexts. US expectancy has demonstrable

face validity, clinical utility, predictive validity and construct validity, particularly in fear conditioning designs that include some degree of ambiguity, such as studies using multiple CS and reinforcement rates <100% (Boddez et al., 2013). As reviewed in prior research (Boddez et al., 2013), US expectancy disturbances are present in individuals with elevations in trait anxiety, panic disorder and posttraumatic stress disorder. Further, other researchers have demonstrated that when other response systems habituate, US expectancies remain elevated (Soeter & Kindt, 2010). It is possible that the maintenance of danger (US) expectancies may leave individuals susceptible to return of fear even in the absence of elevations in physiological responding. The primary goal of the current study was to directly assess expectancy violation during extinction training as a predictor of extinction retention using measures of explicit expectancy of an US (see Mitchell & Lovibond, 2002; Purkis & Lipp, 2001). US expectancy is highly relevant to fear reactivity as it reflects real-world awareness of the relationship between stimuli and aversive outcomes (Boddez et al., 2013). This relationship was assessed across two indices of extinction retention: retention of extinction after 7 days of consolidation; and after reinstatement via unpaired US presentations. We hypothesised that the more instances of expectancy violation during extinction training (i.e., the greater the absolute difference in expectancy across trials), the greater the retention of the extinction memory after consolidation and after reinstatement. In addition, we hypothesised that extinction retention would be more strongly related to expectancy violation during extinction training than to indices of fear habituation during extinction training, as measured using subjective ratings of fear to the CS and defensive eye blink startle reflex to the CS.

Methods

Participants

Undergraduate students in psychology ($n = 63$) were recruited to participate in the experiment for course credit. They were mostly female (61.9%) and their average age was 20.4 ($SD = 2.0$). Participants were racially and ethnically diverse, with Asian-Americans (39.7%), Caucasians (20.6%) and Hispanics/Latinos (19.1%) being the most represented groups. Exclusion criteria were as follows: under 18 years of age; lack of English language proficiency; a score of greater than 18 on the Beck Depression Inventory (BDI; Beck, Ward, & Mendelson, 1961); recommendation by a physician to stay away from stressful situations; serious respiratory, cardiovascular and pulmonary conditions or diseases; and current pregnancy.

Measures

Expectancy of shock (US)

Expectancy for the US is a common measure of associative learning (Chan & Lovibund, 1996; Lipp, 2006). Expectancy of the US, an electric shock, was measured using a sliding dial on a small portable unit, on which 0 indicates "certain no shock," 5 indicates "uncertain" and 9 indicates "certain shock" on the BIOPAC MP150 equipment (BIOPAC, 2010). Participants could rate any value between 0 and 9 by sliding the dial, and they were prompted to rate US expectancy 3–4 s after the CS onset.

Startle electromyography

Eye blink responses to startle probes were collected using BIOPAC's MP150 (BIOPAC, 2010) using two Ag–AgCl electrodes (11 mm outer diameter, 4 mm inner diameter) affixed to the orbicularis oculi muscle (one directly beneath and one approximately 1 cm to the right below the right eye) using electrode gel and adhesive collars. Startle stimuli included 50 ms, 100 decibel bursts of white noise delivered binaurally via headphones. Eight habituation startle probes were presented prior to the beginning of acquisition and one probe was presented during every CS (7 s post-onset) on half of intertrial intervals (ITI; in the middle of the ITI). Raw data were filtered using Acknowledge (Version 4.1, BIOPAC, 2010) using a 30–1000 Hz bandpass filter. Data were sampled at 2000 Hz and rectified over 11 ms. Electromyography (EMG) responses were analysed as the difference between mean response during the 200 ms prior to startle probe and the maximum response in the 20–150 ms post probe. Responses were square-root transformed, and outliers (those greater than 3 *SD* above the mean for a trial of the same Stimulus time and trial number, 1.5% of all responses) were Winsorized (i.e., replaced with the closest non-outlier value; Tukey, 1977).

CS fear

Fear ratings of the CSs (two facial images described below) were collected using the prompt "On a scale from 0 to 10, how fearful does this image make you?" The following anchors were provided: 0 indicates

"not at all", 3 indicates "a little", 5 indicates "moderately", 8 indicates "quite a lot" and 10 indicates "very". These ratings were collected after habituation, after acquisition/before extinction, after extinction and prior to test of extinction retention. We included two indices of fear habituation (EMG and CS fear) because subjective and objective ratings of fear habituation are often discordant and thus allows for a more stringent test of the role of fear habituation.

Materials

Conditional stimuli

Two facial images from the NimStim (Tottenham et al., 2009) data set were counterbalanced as the CS+ (paired with shock) and CS− (not paired with shock) using E-Prime (Psychology Software Tools, 2012). Both images depicted neutral facial expressions due to prior findings that non-neutral facial expressions are associated with increased physiological reactivity (Lang, Greenwald, Bradley, & Hamm, 1993). Kappa agreements of the neutrality of the facial expressions were rated at .79 and .68 (Tottenham et al., 2009). Facial images were matched to the modal membership of the University of California, Los Angeles Psychology subject pool: young adult, female and Asian.

Unconditional stimulus

A STM200 (BIOPAC, 2010) administered a 0.5-s shock. Two electrodes were affixed to either side of the participant's left bicep muscle. The level of the shock was tailored to each participant and was determined by a work-up procedure in which participants rated their level of discomfort as the shock was gradually increased. Participants were encouraged to raise the level of shock to a level that was uncomfortable, but not painful.[1]

Procedure

All experiment procedures were approved by the Institutional Review Board of the University of California, Los Angeles. Participants were recruited from a subject pool comprising students enrolled in undergraduate psychology courses and received course credit in exchange for their participation. Eligible and interested participants attended the laboratory for two sessions scheduled 1 week apart.

Day 1 procedures

After providing informed consent, participants completed self-report questionnaires, had electrodes attached and underwent the US work-up procedure described above. Next, a 5-min baseline period was initiated during which participants sat quietly. After baseline, a habituation phase involved eight presentations of unpaired startle probes (15 s ITI), and two 8 s presentations of each of two facial images (including the CS+, CS−; 15 s ITI) in randomised order. Participants rated CS fear for each image.

Prior to all phases (with the exception of the 5-min baselines), participants were informed that some of the images may be followed by the muscle stimulation; however, they were not provided with any additional details about the contingency between stimuli. During acquisition, each CS (8 s) was presented eight times, with a 25–35 s ITI. The CS+ had a 75% reinforcement rate (i.e., six pairings with a 0.5 s shock that co-terminated with the CS). The CS− was never paired with shock. The trial order during acquisition was semi-randomised. Expectancy of the US was rated for each CS trial in the time interval between CS onset and US onset. At completion of acquisition, participants rated CS fear for each image.

Following acquisition, participants viewed a 7-min neutral movie of nature scenes. This relatively brief window of consolidation time was chosen in order to reduce participant burden, and because it is consistent with other published research in this area demonstrating no difference in rates of extinction responding in individuals with 10 min versus 6 h of a consolidation window following acquisition training (Schiller et al., 2010). Human fear conditioning research has found effects with consolidation windows as brief as 1 min (Milad et al., 2007). Next, extinction training consisted of eight presentations of each CS (CS+ and CS−; 8 s), with 25–35 s ITI, in the absence of the US. At completion of extinction training, participants rated CS fear for each image.[2]

Day 8 procedures

Prior literature has demonstrated tests of extinction retention at 24 (Milad et al., 2007) or 72 h post-extinction training (Kalisch et al., 2009). We chose a 1-week delay in extinction retention to provide a test that has greater relevance for clinical applicability. Upon arriving to the laboratory, electrodes were affixed and participants completed CS fear ratings for each image, followed by a 5-min baseline phase and

8 presentations of each CS. Extinction retention, hereafter referred to as Test 1, included the first presentations of each CS (8 s) in the absence of the US. One unpaired, reinstating shock was presented in a counterbalanced order, after either 4 or 5 CS+ presentations, to test the effects of reinstatement. Test of reinstatement, hereafter referred to as Test 2, included the CS+ and CS− trial immediately following the reinstatement probe, regardless of the timing of the probe (i.e., on either the fifth or sixth CS+ on Day 8).

Data analysis

All data analyses were completed using Stata version 13 (StataCorp, 2013). Analyses of variance (ANOVAs) were used to determine differences in the variables of interest by experimental condition. Multilevel modelling, with trials nested within participants, was used to confirm differential responding between the CS+ and CS− across all experimental phases. The strongest measure of contingency awareness, the crux of this study, is US expectancy. We analysed initial fear responding to capture a pure measure of spontaneous recovery. Responding on the first trial of an extinction retention test is an important clinical observation point, as reactivity at the first test following extinction might influence a patient's willingness to continue with extinction training, or exposure therapy in a clinical context. To ensure that the findings of this paper were not specific to only a single trial, analyses were replicated for the initial test of extinction retention using an outcome of average responding immediately prior to the reinstatement test (i.e., on the first four trials of Day 8). No differences were found based on single trial or averaged-trial outcomes, and thus single trial outcomes are reported.

A series of analyses were conducted to examine US expectancy variation, EMG habituation and CS fear habituation as predictors of CS fear, US expectancy and EMG at test of extinction retention (Test 1) and reinstatement (Test 2). All analyses controlled for the pre-acquisition/Trial 1 response on the dependent variable. US expectancy variation throughout extinction training was calculated as the sum of absolute value of differences in US expectancy on each consecutive trial across all trials of CS + . Greater variation in US expectancy is an indicator that participants were alternating between expecting and not expecting the US on consecutive trials. Those with greater variation therefore had opportunities to expect the US in its absence, resulting in expectation violation. In contrast, those with little variation in US expectancy throughout extinction training would not have experienced major discrepancies in expectancy on consecutive trials, and would therefore not have had the opportunity to violate their US expectancy. CS fear habituation was calculated as a change score from the first trial to the last trial of extinction training. Finally, EMG habituation was calculated as the extracted slope during extinction training via multilevel modelling (see Table 1). The multilevel models used continuous time as a predictor, nesting observations within participants for each phase of acquisition and extinction separately for the CS+ and CS−. The residuals for each model were inspected and confirmed that assumptions regarding of normality, nonlinearity and heteroscedasticity were met. Random intercepts and linear slopes were estimated for these models, as were correlations between the slope and intercept.

Finally, models were run that combined all predictors to consider the predictive power of each independent variable controlling for the other possible predictor and the pre-acquisition/trial 1 value of the outcome variable of interest.

Results

Differences by reinstatement timing

There were no significant differences by reinstatement timing (either after 4 or 5 CS+ presentations) in responding to the first CS+ trial after reinstatement in US expectancy ($F(1, 61) = .43$, $p = .512$, partial $\eta^2 = .007$) or EMG ($F(1, 61) = .08$, $p = .774$, partial $\eta^2 = .001$). Reinstatement timing was also not

Table 1. Parameters from multilevel models used to extract slopes and intercepts.

EMG CS+	Time (slope) estimate (*SE*)	σ slope	Intercept estimate (*SE*)	σ intercept	ICC
Acquisition	−.008 (.003)	0.00005	0.680 (.02905)	0.047	0.758
Extinction	−.018 (.003)	0.0002	0.642 (.0307695)	0.054	0.810
EMG CS−	Time (slope) estimate (*SE*)	σ slope	Intercept estimate (*SE*)	σ intercept	ICC
Acquisition	−.019 (.003)	0.0002	0.731 (.033)	0.061	0.789
Extinction	−.020 (.003)	0.0002	0.647 (.029)	0.048	0.773

Note: Quadratic effects of time were only included in the multilevel model for US expectancy during acquisition for the CS+ and CS−.

significantly associated with degree of variation in US expectancy during extinction training ($F(1, 61) = .03$, $p = .873$, partial η^2=.000).

Responses to the CS+/CS− during acquisition, extinction, retention and reinstatement

US expectancy ratings to the CS+ were significantly higher than the CS− on the final trial of acquisition ($t(62) = 6.92$, $p < .0001$, Cohen's $D = 1.31$), and multilevel modelling confirmed that there was a Trial × Stimulus interaction on US expectancy during acquisition (see Table 2). Tests of simple slope were significant for each Stimulus (p's < .05), though the slope for the CS+ and CS− were positive and negative, respectively. Similarly, Time × Stimulus interactions were significant on US expectancy during extinction, though during extinction the simple slopes for each Stimulus were significant and negative (p's < .001). There was not a Time × Stimulus interaction for US expectancy during the first four trials of Day 8 (i.e., before any reinstating probes were presented); however, there was a main effect of Stimulus (with responding at the beginning of Day 8 higher for the CS+) and Time, with significant and negative simple slopes for each Stimulus ($p < .05$). A paired sample t-test revealed that responding was higher to the CS+ than the CS− at Test 2, the reinstatement test.

Consistent with non-associative effects of habituation on psychophysiological measures (Vervliet, Baeyens, Van den Bergh, & Hermans, 2013), EMG decreased throughout acquisition, and responses to the CS+ were not significantly higher than the CS− on the final trial of acquisition ($t(62) = 1.444$, $p = .154$, Cohen's $D = .125$). However, mixed level modelling resulted in a significant Time × Stimulus effect, and while the simple slopes for each Stimulus were significant and negative (p's < .01), the slope for the CS− was steeper than the CS +. This appears to be largely driven by enhanced responding to the CS− in the beginning of acquisition ($t(62) = -2.954$, $p < .01$). Therefore, there is no evidence of greater responding to the CS + relative to the CS− for EMG during acquisition. Of note, EMG during the ITI also decreased throughout acquisition. For extinction, there was a significant reduction over Time in EMG, with significant and negative slopes for each Stimulus (p's < .001), but there was not a Time × Stimulus interaction or a main effect of Stimulus. Similarly, EMG responding to the ITI significantly decreased throughout extinction. For test of extinction retention, there was a significant

Table 2. Differences in responding to CS+ and CS−.

	Acquisition	Extinction	Retention	Reinstatement
US expectancy ICC	0.13	0.244	0.333	–
Time × Stimulus	Coefficient = .395 (*SE*: .064), $z = 6.20$, $p < .001$	Coefficient = −.230 (*SE*: .053), $z = -4.31$, $p < .001$	Coefficient = −.100 (*SE*: .151), $z = -.66$, $p = .508$	–
Time	Coefficient = −.132 (*SE*: .049), $z = -2.68$, $p < .01$	Coefficient = −.430 (*SE*: .051), $z = -8.40$, $p < .001$	Coefficient = −.360 (*SE*: .102), $z = -3.54$, $p < .001$	–
Stimulus	Coefficient= 1.048 (*SE*: .266), $z = 3.93$, $p < .001$	Coefficient = 2.483 (*SE*: .223), $z = 11.11$, $p < .001$	Coefficient = 1.676 (*SE*: .169), $z = 9.89$, $p < .001$	$t(62) = 2.481$, $p < .05$
EMG ICC	0.751	0.757	0.674	
Time × Stimulus	Coefficient = .010 (*SE*: .003), $z = 2.90$, $p < .01$	Coefficient = .001 (*SE*: .003), $z = .45$, $p = .653$	Coefficient = −.004 (*SE*: .011), $z = -.35$, $p = .723$	–
Time	Coefficient = −.019 (*SE*: .003), $z = -6.82$, $p < .001$	Coefficient = −.019 (*SE*: .002), $z = -8.09$, $p < .001$	Coefficient = −.039 (*SE*: .007), $z = -6.03$, $p < .001$	–
Stimulus	Coefficient = −.051 (*SE*: .015), $z = -3.51$, $p < .001$	Coefficient = .000 (*SE*: .007), $z = .09$, $p = .932$	Coefficient = .029 (*SE*: .012), $z = 2.42$, $p < .05$	$t(62) = 1.014$, $p = .315$
EMG ITI	Coefficient = −.027 (*SE*: .004), $z = -7.68$, $p < .001$	Coefficient = −.015 (*SE*: .003), $z = -5.54$, $p < .001$	Coefficient = −.011 (*SE*: .003), $z = -3.50$, $p < .001$)	–
CS Fear ICC	0.359	0.363	–	–
Time X Stimulus	Coefficient: 2.540 (*SE*: .389), $z = 6.53$, $p < .001$	Coefficient = −1.698 (*SE*:.393), $z = -4.32$, $p < .001$	–	–
Time	Coefficient: −.159 (*SE*: .330), $z = -.48$, $p = .630$	Coefficient: −.476 (*SE*: .278), $z = -1.71$, $p = .087$	$t(62) = 2.91$, $p < .01$	–
Stimulus	Coefficient: −.159 (*SE*: .273), $z = -.58$, $p = .564$	Coefficient: 4.079 (*SE*: .621), $z = 6.56$, $p < .001$	–	–

Note: Retention included the first four trials on Day 8 prior to any reinstating US presentations, random slopes were not estimated for CS fear during extinction as the *SE* could not be estimated.

decrease in EMG over Time as well as an effect of Stimulus with responding higher to the CS+ relative to the CS−, but not a Time × Stimulus interaction. Similarly, EMG significantly decreased during the ITI for the retention test. EMG did not differ by Stimulus at test of reinstatement.

There was a significant Time × Stimulus interaction for CS Fear during acquisition, with tests of simple slopes revealing a significant positive slope for the CS+ ($p < .001$), whereas the slope for the CS− was not significant ($p = .630$). Similarly, there was a Time × Stimulus interaction for CS Fear during extinction, and while both slopes were negative during this phase, only the slope for the CS+ was significant ($p < .001$). CS fear ratings were higher to the CS+ relative to the CS− at the beginning of Day 8. All subsequent results reported are relevant to responses to the CS+ (Figures 1–3).

Violation of US expectancy

Controlling for US expectancy on the first trial of acquisition, larger sum differences in trial-to-trial US expectancy during extinction training was a significant predictor of US expectancy on Test 1, the first trial of extinction retention ($\beta = -.262$, $t(60) = -2.12$, $p < .05$, partial $\eta^2 = .07$, see Figure 4) such that higher variation during extinction was predictive of lower US expectancy at retention. Variation in US expectancy during extinction training was not a significant predictor of US expectancy at Test 2, following reinstatement ($\beta = -.094$, $t(60) = -.74$, $p = .462$, partial $\eta^2 = .009$).

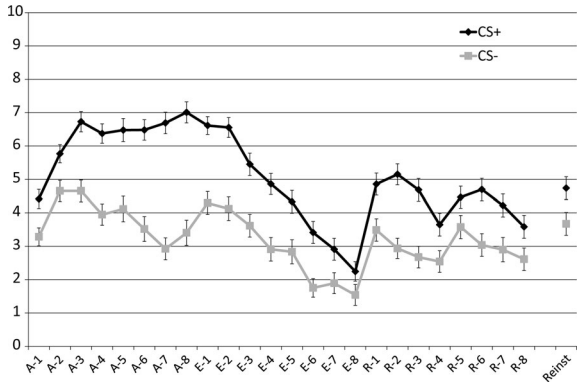

Figure 1. Mean Expectancy. Note: Expectancy with standard errors throughout acquisition ("A"), extinction ("E") and retention ("R"). A reinstating unpaired US was presented in counterbalanced order after either trial 4 or 5 during an ITI and the results of this trial are presented under "Reinst" to clearly demonstrate responding to the first trial after the reinstating US.

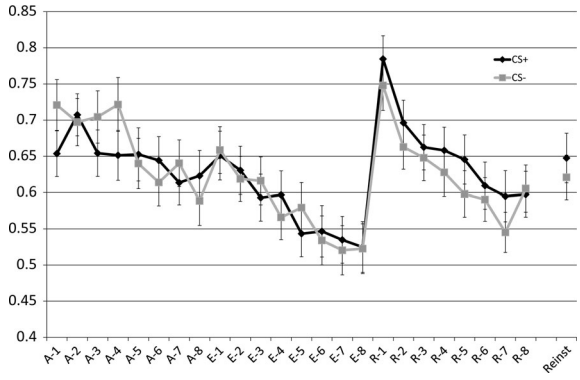

Figure 2. Mean EMG responses. Note: EMG with standard error bars throughout trials 1–8 of acquisition ("A"), extinction ("E") and retention ("R"). A reinstating unpaired US was presented in a counterbalanced order after either trial 4 or 5 during an ITI and the results of this trial are reported under "Reinst" to clearly demonstrate responding to the first trial after the reinstating US.

Controlling for responses on the outcome variable at pre-acquisition/trial 1, variation in US expectancy during extinction training was not a significant predictor of CS fear ($\beta = -.205$, $t(60) = -1.66$, $p = .102$, partial $\eta 2 = .044$) or EMG at Test 1 ($\beta = -.033$, $t(60) = -.28$, $p = .777$, partial $\eta^2 = .001$) or Test 2 ($\beta = -.048$, $t(60) = -.41$, $p = .683$, partial $\eta^2 = .003$).

Fear habituation

EMG

Visual inspection of the EMG data suggested that a quadratic effect of time may be necessary, but model estimation revealed a non-significant parameter estimate for quadratic time during both acquisition and extinction. Therefore, a linear effect of time (trials 1–8) was included as the sole predictor in these

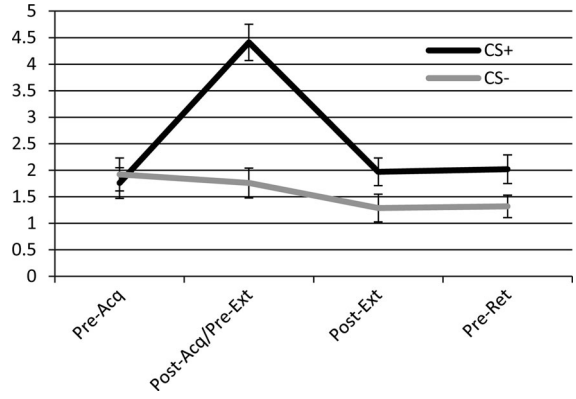

Figure 3. CS fear ratings. Note: Fear ratings were collected after habituation, acquisition and extinction, and before and after retention.

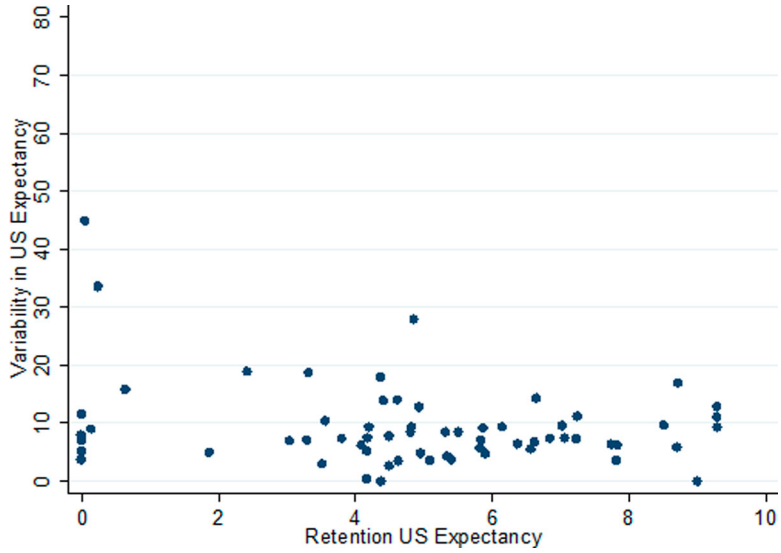

Figure 4. US expectancy extinction retention by variability. Note: Greater variability (as indicated by higher scores on the X axis) reflects greater discrimination in expectancy ratings by trial.

analyses. Controlling for EMG during the first trial of acquisition, the slope of EMG during extinction was not predictive of EMG at Test 1 (β=.076, $t(60)$= .62, p = .538, partial η^2=.006) or Test 2 (β=.132, $t(60)$ = 1.069, p = .293, partial η^2=.018). Controlling for the pre/trial 1 acquisition response on the outcome variable, slope of EMG during extinction was not predictive of CS fear at Test 1 (β=–.037, $t(60)$=–.29, p = .770, partial η^2 = .001) or US Expectancy at Test 1 (β=–.158, $t(60)$=–1.24, p = .219, partial η^2 = .025) or Test 2 (β=–.082, $t(60)$=–.64, p = .525, partial η^2 = .007).

CS fear

Controlling for CS fear before acquisition, change in CS fear from before to after extinction training was not a significant predictor of CS fear at Test 1 (β=–.038 $t(60)$ =–.29, p = .773, partial η^2=.001). Similarly, controlling for the pre-acquisition/trial 1 value of the outcome variable, change of CS fear from before to after extinction training was not a significant predictor of EMG at Test 1 (β=–.071, $t(60)$=–.62, p = .537, partial η^2 = .006) or Test 2 (β=–.140, $t(60)$=–1.20, p = .234, partial η^2 = .024), or US expectancy at Test 1 (β=–.159, $t(60)$ = –1.25, p = .215, partial η^2 = .026) or Test 2 (β=–.106, $t(60)$=–.83, p = .410, partial η^2 = .011).

Combined models

Finally, we tested a series of models in which all three predictors (i.e., US expectancy variation, EMG habituation and CS Fear habituation) were included in one model to predict each outcome variable of interest, controlling for the pre-acquisition/trial 1 response on the outcome variable. Over and above habituation in EMG during extinction, habituation in CS fear during extinction and US expectancy on the first trial of acquisition, variation in US expectancy during extinction training was a significant predictor of US expectancy at Test 1 (β=–.275, $t(58)$=–2.26, p < .05, partial η^2 = .081). No other significant predictors of any outcome variables emerged from these models (see Table 3).

Discussion

Habituation of fear is often considered an important goal during exposure treatment (Foa & Kozak, 1986, 1991). Decades of research have found that violation of the expectancy of an aversive event, or enhancing surprise throughout extinction training, may increase retention of the CS–no US association during extinction training (Rescorla & Wagner, 1972). No studies to our knowledge have directly compared within-session habituation of fear-potentiated startle and self-reported fear during extinction training to expectation violation during extinction training as predictors of retention of extinction learning. In the current study, habituation of objective and subjective fear responses (as measured by EMG and ratings of CS fear) was not predictive of extinction retention or reinstatement.

Table 3. Combined model analyses

Outcome	US expectancy variation	EMG habituation	CS fear habituation
US expectancy test 1	**$\beta = -.275$, $t(58) = -2.26$, $p \leq .05$, partial $\eta^2 = .081$**	$\beta = -.208$, η^2 (58) $= -1.68$, $p = .10$, partial $\eta^2 = .047$	$\beta = -.190$, $t(58) = -1.54$, $p = .13$, partial $\eta^2 = .039$
US expectancy test 2	$\beta = -.101$, $t(58) = -.79$, $p = .436$, partial $\eta^2 = .011$	$\beta = -.109$, $t(58) = -.83$, $p = .409$, partial $\eta^2 = .012$	$\beta = -.122$, $t(58) = -.94$, $p = .351$, partial $\eta^2 = .015$
EMG test 1	$\beta = -.029$, $t(58) = -.25$, $p = .805$, partial $\eta^2 = .001$	$\beta = .065$, $t(58) = .52$, $p = .608$, partial $\eta^2 = .005$	$\beta = -.063$, $t(58) = -.53$, $p = .596$, partial $\eta^2 = .005$
EMG test 2	$\beta = -.042$, $t(58) = -.35$, $p = .724$, partial $\eta^2 = .002$	$\beta = .111$, $t(58) = .88$, $p = .385$, partial $\eta^2 = .013$	$\beta = -.125$, $t(58) = -1.05$, $p = .296$, partial $\eta^2 = .019$
CS fear test 1	$\beta = -.208$, $t(58) = -1.66$, $p = .101$, partial $\eta^2 = .046$	$\beta = -.057$, $t(58) = -.45$, $p = .654$, partial $\eta^2 = .003$	$\beta = -.044$, $t(58) = -.34$ $p = .737$, partial $\eta^2 = .002$

Note: Bold values represent the only significant test.

However, variation in US expectancy, as measured by the sum of absolute discrepancies in scores across all trials, was a significant predictor of extinction retention. This finding is consistent with other recent studies demonstrating that habituation of fear is neither necessary nor sufficient as a goal of exposure or extinction (Baker et al., 2010; Culver et al., 2012; Plendl & Wotjak, 2010), and that expectation violation may be a more robust predictor of long-term extinction retention (Rescorla & Wagner, 1972).

Although greater variation in US expectancy predicted greater extinction retention, it did not predict reinstatement. Perhaps, experiencing an elevation in US expectancy following an unpredictable US presentation may be an adaptive and flexible defensive response, provided that the expectancy can be lowered with subsequent extinction trials. It is possible that more trials following reinstatement would have allowed for the benefit of variation in US expectancy during extinction training to re-emerge, as the current study only included three or four post-reinstatement CS presentations. Future research should explore the effect of variation in US expectancy on reinstatement both immediately following the reinstating probe, several trials thereafter, and at a later test. In addition, variation in US expectancy was not predictive of non-US expectancy outcomes in terms of CS fear and EMG. One potential reason may be reduced error when predicting within measures relative to across measures, allowing for greater power. Further, fear response systems are known to diverge, potentially explaining these discrepancies (Boddez, et al., 2013).

The finding that habituation during extinction training across subjective fear and objective physiology was not predictive of either extinction retention or reinstatement suggests that exposure therapies (the clinical analogue of extinction training) should not emphasize within-session habituation as an important indicator of improvement. Instead, based on the finding that increased variability in associative learning through US expectancy is associated with enhanced retention of extinction learning, expectation violation might be considered a more potent target of exposure therapy. Examples include heightening emotional and stimulus variability throughout extinction (Culver, Stoyanova, & Craske, 2011), exposure to multiple conditioned stimuli (i.e., deepened extinction; Culver, Vervliet, & Craske, 2014; Rescorla, 2006) and the use of supplementary strategies such as affect labelling (see Kircanski et al., 2012; Tabibnia, Lieberman, & Craske,

2008 for examples). For more specific clinical recommendations for enhancing expectation violation, see Craske et al. (2012, 2014). However, the lack of a differential response between the CS+ and CS– during acquisition in terms of EMG responding, but not CS fear or US expectancy, warrants additional caution when interpreting these results in particular.

One limitation of the present research is that explicit US expectancy ratings may not be an ideal method for measuring the associative mechanism of extinction learning, as they rely on that which is accessible to declarative knowledge. Associative mechanisms may involve implicit processes, which are important to consider in future designs. Measures of implicit approach and avoidance, for example, might provide useful opportunities to replicate these findings (for examples of such measures, see Krypotos, Effting, Arnaudova, Kindt, & Beckers, 2013; Marsh, Ambady, & Kleck, 2005; Rinck & Becker, 2007). However, other researchers have noted the importance of US expectancy as one component of fear and anxiety as this measure of contingency awareness reflects prediction of aversive outcomes in anxiety disorders (Boddez et al., 2013). Another limitation is that CS fear was not collected after each trial during extinction training to reduce participant burden, preventing more in-depth analysis of the subtle changes in CS fear throughout extinction training. Future studies should therefore collect more frequent CS fear ratings. Further, the lack of greater responding to the CS+ relative to the CS– during acquisition in EMG responding suggests that the EMG results should be interpreted with caution. The generalisability of these findings were limited by a relatively small, non-clinical sample of young and mostly female individuals that underwent fear conditioning. Finally, extinction training commenced only after a short (7 min) break following acquisition training, and retention test occurred after only 7 days. Future research should include a longer window of time for both acquisition consolidation and retention test.

The results described herein have important implications for the understanding of long-term fear reduction. Habituation of EMG and CS fear during extinction training did not predict retention of extinction learning. In contrast, expectation violation was associated with long-term retention in extinction learning. If replicated, these findings suggest that clinicians should format their exposures to optimise inhibitory learning rather than relying solely on habituation. In addition to enhancing expectation violation in exposures as mentioned above, additional strategies for optimising inhibitory learning during exposure are warranted. In conclusion, these findings warrant replication, but suggest that habituation of fear-potentiated startle and self-reported fear throughout extinction training is not sufficient for retention of extinction learning.

Notes

1. The mean US level chosen was 42.6 (SD = 11.6; range 20–70; maximum possible = 85).
2. Participants were randomized to one of three experimental conditions that were collapsed for all analyses in the current study due to lack of significant findings on all outcome variables. As described in Appendix 1 and 2, these conditions only differed in the writing task to which participants were assigned following extinction; a no writing condition, writing about the extinction training, and writing about topics unrelated to extinction.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akirav, I., Raizel, H., & Maroun, M. (2006). Enhancement of conditioned fear extinction by infusion of the GABA(A)agonist muscimol into the rat prefrontal cortex and amygdala. *European Journal of Neuroscience*, *23*(3), 758–764.

Baker, A. S., Mystkowski, J. L., Culver, N. C., 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.

Barlow, D. H. (1988). *Anxiety and its disorders*. New York: Guilford Press.

Beck, A. T., Ward, C., & Mendelson, M. (1961). Beck Depression Inventory (BDI). *Archives of General Psychiatry*, *4*(6), 561–571.

BIOPAC. (2010). *BIOPAC MP150 psychophysiological data acquisition system*. Goleta, CA: BIOPAC Systems, Inc.

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*, 201–206.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory*, *11*(5), 485–494.

Chan, C. K. Y., & Lovibund, P. F. (1996). Expectancy bias in trait anxiety. *Journal of Abnormal Psychology*, *105*, 637–647.

Corcoran, K. A., Desmond, T. J., Frey, K. A., & Maren, S. (2005). Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *The Journal of Neuroscience*, *25*(39), 8978–8987.

Craske, M. G., Liao, B., Brown, L. A., & Vervliet, B. (2012). The role of inhibition in exposure therapy. *Journal of Experimental Psychopathology*, *3*, 322–345.

Craske, M. G., Rauch, S. L., Ursano, R., Prenoveau, J., Pine, D. S., & Zinbarg, R. E. (2009). What is an anxiety disorder?. *Depression and Anxiety*, *26*(12), 1066–1085.

Craske, M. G., Treanor, M., Conway, C., Zbozinek, T. & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10–23.
Culver, N. C., Stoyanova, M., & Craske, M. G. (2011). Clinical relevance of retrieval cues for attenuating context renewal of fear. *Journal of Anxiety Disorders*, *25*(2), 284–292.
Culver, N. C., Stoyanova, M., & Craske, M. G. (2012). Emotional variability and sustained arousal during exposure. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 745–751.
Culver, N. C., Vervliet, B., & Craske, M. G. (2014). Compound extinction using the Rescorla–Wagner model to maximize exposure therapy effects for Anxiety disorders. *Clinical Psychological Science*, *3*(3), 1–14. doi:10.1177/2167702614542103
Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: exposure to corrective information. *Psychological Bulletin*, *99* (1), 20–35.
Foa, E. B., & Kozak, M. J. (1991). Emotional processing: Theory, research, and clinical implications for anxiety disorders. In J. D. Safran & L. S. Greenberg (Eds.), *Emotion, psychotherapy, and change* (pp. 21–49). New York: Guilford Press.
Kalisch, R., Holt, B., Petrovic, P., De Martino, B., Kloppel, S., Buchel, C., & Dolan, R.J. (2009). The NMDA agonist D-cycloserine facilitates fear memory consolidation in humans. *Cerebral Cortex*, *19*, 187–196.
Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., & Craske, M. G. (2012). Challenges to the traditional exposure paradigm: Variability in exposure therapy for contamination fears. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 745–751.
Krypotos, A. M., Effting, E., Arnaudova, I., Kindt, M., & Beckers, T. (2013). Avoided by association: Acquisition, extinction, and renewal of avoidance tendencies toward conditioned fear stimuli. *Clinical Psychological Science*, *2*, 336–343.
Lang, P.J. (1971). The application of psychophysiological methods to the study of psychotherapy and behavior modification. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 75–125). New York: Wiley.
Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. doi:10.1111/j.1469-8986.1993.tb03352.x
Lipp, O. V. (2006). Of snakes and flowers: Does preferential detection of pictures of fear-relevant animals in visual search reflect on fear relevance? *Emotion*, *6*, 296–308.
Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach and avoidance-related behaviors. *Emotion*, *5*(1), 119–124.
Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Neuron*, *420*(6911), 70–74.
Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry*, 62(5), 446–454.
Mitchell, C. J., & Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *The Quarterly Journal of Experimental Psychology: Section B*, *55*(4), 311–329.
Peters, J., Dieppa-Perea, L. M., Melendez, L. M., & Quirk, G. J. (2010). Induction of fear extinction with hippocampal-infralimbic BDNF. *Science*, *328*(5983), 1288–1290.
Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron*, *43*(6), 897–905.
Plendl, W., & Wotjak, C. T. (2010). Dissociation of within- and between-session extinction of conditioned fear. *The Journal of Neuroscience*, *30*(14), 4990–4998.
Psychology Software Tools, I. (2012). E-Prime 2.0. Retrieved from http://www.pstnet.com.
Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation*, *32*(1), 84–99.
Rescorla, R. A. (2006). Deepened extinction from compound stimulus presentation. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 135–144.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.
Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(2), 105–120.
Schiller, D., Monfils, M.H., Raio, C.M., Johnson, D.C., LeDoux, J.E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463, 49–53.
Soeter, M. & Kindt, M. (2010). Dissociating response systems: erasing fear from memory. *Neurobiology of Learning and Memory*, *94*(1), 30–41.
StataCorp. (2013). *Stata statistical software: Release 13*. College Station, TX: StataCorp LP.
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.
Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.
Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalisation, and return of fear: A critical review of renewal research in humans. *Biological Psychiatry*, *92*, 51–58.

Appendix 1. Details of experimental conditions in experiment I

As described in note 2, participants were randomised to one of three experimental writing conditions for a hypothesis unrelated to the analyses in the current study. There were no differences in expectancy of the US or fear by experimental group, but the details of the conditions are described herein. Participants were randomised to one of three experimental conditions that differed only in the writing task to which they were assigned following extinction. Participants in the no writing condition (NW) sat

quietly and watched a 10-min nature video with no sound. The extinction-relevant writing condition (RW) wrote continuously for 10 min about the following topics: (1) During which image the muscle stimulation was received?; (2) What did each facial image look like?; (3) What was your emotional reaction to each of the facial images?; (4) What was your emotional reaction to receiving the muscle stimulation? This task was intended to retrieve participants' fear and extinction memory. The extinction-irrelevant writing condition (IW) wrote continuously for 10 min about the following topics: (1) What did you eat yesterday and when?; (2) What work did you do yesterday and why?; (3) What interactions did you have with family or friends yesterday and what was it like?; (4) What was your emotional reaction to the things that happened yesterday? This task was designed to control for the task of writing. In both conditions, the instructions were read to the participants, and then they were provided with the instructions in writing, along with the images of both CS labelled "A" and "B" for ease of explaining their answers as appropriate. Participants in both writing conditions were instructed to repeat sections that they had already written if they ran out of new material in order to encourage their continual linguistic processing.

Appendix 2. Mean scores by experimental condition

		First trial/ beginning of extinction mean (*SD*)	Last trial/ end of extinction mean (*SD*)	Reinstatement mean (*SD*)
No writing	*Expectancy*	6.62 (1.94)	2.00 (2.53)	4.98 (2.54)
	EMG	.57 (.22)	.48 (.28)	.63 (.23)
	CS fear	3.42 (2.60)	1.25 (1.59)	1.29 (1.83)
Irrelevant writing	*Expectancy*	6.55 (2.13)	2.35 (2.41)	4.69 (2.58)
	EMG	.64 (.27)	.47 (.26)	.56 (.24)
	CS fear	4.58 (2.85)	2.16 (2.12)	2.37 (2.43)
Relevant writing	*Expectancy*	6.67 (2.41)	2.45 (2.03)	4.50 (3.04)
	EMG	.75 (.30)	.63 (.28)	.75 (.32)
	CS fear	4.60 (2.66)	2.65 (2.28)	2.55 (1.93)

Note: There were no significant differences in any of the outcome measures by experimental condition.