Novelty-facilitated extinction and the reinstatement of conditional human fear

Katherine Lucas[1], Camilla C. Luck[1,2] & Ottmar V. Lipp[1,2]

[1] School of Psychology, Curtin University, Australia

[2] ARC-SRI: Science of Learning Research Centre, University of Queensland, Australia

Running Head:

Novelty-facilitated extinction

Addresses for correspondence:

Ottmar V. Lipp, PhD, School of Psychology, Curtin University, GPO Box U1987, Perth, WA, 6845, Australia; phone: +61 8 9266 5112; fax: +61 8 9266 2464; Email: ottmar.lipp@curtin.edu.au

Word count: 5655

Acknowledgments:

This work was supported by grants number DP120100750 and SR120300015 from the Australian Research Council

Abstract

Although contemporary treatments for anxiety disorders are very efficient in reducing anxiety, return of fear after successful treatment is common which signifies a need for interventions that have a more enduring outcome. A recent laboratory study suggested that novelty-facilitated extinction, a simple modification of standard extinction which involves presenting a novel non-aversive stimulus during extinction, prevents spontaneous recovery, one laboratory analogue of return of fear. The current study assessed whether novelty-facilitated extinction can also prevent reinstatement, a second laboratory analogue of return of fear. Following differential fear conditioning, one group of participants underwent standard extinction training whereas the second was presented with a novel tone after the conditional stimulus that previously predicted the aversive unconditional stimulus (US). Three presentations of the USs alone reinstated differential electrodermal fear responses after standard extinction, but not after novelty-facilitated extinction. Moreover, replicating previous findings, the extent of return of fear was correlated with self-reported intolerance of uncertainty after standard extinction, but not after novelty-facilitated extinction. These results support the proposal that novelty-facilitated extinction training can reduce the extent of return of fear.

Keywords: Fear conditioning, return of fear, reinstatement, electrodermal response, extinction learning, Intolerance of Uncertainty

Anxiety disorders are highly prevalent in developed countries, with a lifetime prevalence of 25% (Graham & Milad, 2011). This high prevalence rate is not surprising given our current understanding of the mechanisms that are instrumental in the acquisition, maintenance, and reduction of fear and anxiety (Mineka & Zinbarg, 2006). Human fear learning is highly efficient and enables us to acquire fear responses to signals of danger in a manner that is quick and enduring. It is thus fortuitous that past research has developed interventions such as exposure therapy and cognitive based treatments which can treat anxiety disorders with significant success. However, many patients experience relapse after successful treatment due to treatment effects failing to persist long-term (Craske Treanor, Conway, Zbozinek, & Vervliet, 2014; Graham & Milad, 2011). Again, given our current understanding of the basic mechanisms that are reflected in evidence based treatments, this is not surprising. A prominent theory holds that extinction learning, which is thought to underlie most behaviour based interventions, renders the stimuli that previously signalled danger ambiguous by adding an inhibitory association without removing the original fear learning (Bouton, 2002). This makes it likely that the fear response will return after successful extinction following encounters with highly arousing events (reinstatement) or changes in context away from the extinction context (renewal; Craske et al., 2014). To counteract return of fear, two strategies seem feasible, either to target the original fear learning and eradicate the fear memory (see Schiller, Monfils, Raio, Johnson, LeDoux, & Phelps, 2010; Thompson & Lipp, 2017) or to strengthen the inhibitory learning during extinction (see Dunsmoor, Campese, Ceceli, LeDoux, & Phelps, 2015; Zbozinek, Holmes, & Craske, 2015; Culver, Stevens, Fanselow, & Craske, 2018; Thompson, McEvoy, & Lipp, 2018).

Dunsmoor et al. (2015) proposed a novelty-facilitated extinction procedure to strengthen extinction. After training in a differential fear conditioning paradigm in which one conditional stimulus (CS^+) was paired with an aversive unconditional stimulus (US) whereas

a second was presented alone (CS^-), one group of participants was presented with standard extinction training whereas the second was presented with the CS^+ not alone, but paired with a novel, non-aversive stimulus, a tone. Across two experiments, one involving rodents and the second humans, spontaneous recovery of conditional fear responses was absent after novelty-facilitated extinction such that twenty-four hours after acquisition and extinction training differential responses to CS^+ and CS^- were larger after standard extinction than after novelty-facilitated extinction with no significant difference observed in the latter condition. In the human experiment, the spontaneous recovery test was followed by a test of reinstatement which involved the presentation of three unconditional stimuli alone followed by further presentations of the CSs. After the reinstatement treatment, differential responding was significantly different from zero in participants trained with standard extinction, but not after novelty-facilitated extinction. However, when compared between groups, the extent of differential responding did not differ leaving it unclear whether novelty-facilitated extinction indeed protected against reinstatement. This failure to find a clear pattern of results for reinstatement may occur because reinstatement testing was preceded by a test for spontaneous recovery (Lonsdorf et al., 2017).

The promising results reported by Dunsmoor et al. (2015) are somewhat pulled into question by a recent failure to find an effect of novelty-facilitated extinction on conditioned avoidance. Krypotos and Engelhard (2018) presented two groups of participants with a differential fear conditioning procedure involving two CS^+-US pairings and two CS^- alone presentations followed by an avoidance conditioning phase during which participants could prevent the occurrence of the US by pressing the space bar. This was followed by extinction training (12 trials per CS), which was novelty-facilitated for one group and standard for another, the presentation of three USs to induce reinstatement, and a reinstatement test. The dependent measures, US expectancy ratings, self-reported fear of the CSs, and avoidance

behaviour, did not show any difference between the groups. Reinstatement was evident in both groups in all measures and enhanced in the fear ratings in the novelty-facilitated extinction group. It should be noted, however, that the design used by Krypotos and Engelhard (2018) differed from that employed by Dunsmoor et al. (2015) in that pictures of spiders were used as CSs instead of angry faces, electrodermal activity was not measured, the actual fear conditioning phase was brief, and the different experimental phases were separated by the measurement of self-reported fear. Nevertheless, the study highlights the need for replication of the benefits of novelty-facilitated extinction.

Although the empirical findings reported by Dunsmoor et al. (2015) are very encouraging, the question remains as to what mediates the protection against spontaneous recovery afforded by novelty-facilitated extinction training. Dunsmoor et al. (2015) suggest that the presentation of a novel, surprising stimulus after the CS^+ supports the formation of a stronger extinction memory than does the mere omission of the US. It may do so by creating a bigger prediction error than standard extinction in particular after acquisition training that utilized a partial reinforcement schedule. The notion that novelty-facilitated extinction training strengthened extinction learning is supported by an accessory observation reported by Dunsmoor et al. (2015). Participants were asked to complete the Intolerance of Uncertainty Scale (IUS; Buhr & Dugas, 2002) before commencement of the experiment, a measure that has been shown to capture individual differences in human fear conditioning (Lonsdorf & Merz, 2017). The level of self-reported Intolerance of Uncertainty correlated with the extent of spontaneous recovery after standard extinction ($R^2 = .24$), but not after novelty-facilitated extinction. A similar pattern of results was reported for reinstatement with a significant correlation between reinstatement and IUS after standard extinction, but not after novelty-facilitated extinction training. This may suggest that novelty-facilitated extinction training reduces the uncertainty about the potential recurrence of the US after the

CS^+ relative to standard extinction training.

One factor that has been discussed as a potential mediator of the return of fear is residual negative CS valence after successful extinction (Hermans, Dirikx, Vansteenwegen, Baeyens, Van den Bergh, & Eelen, 2005; Luck & Lipp, 2015). During acquisition, a CS^+ that is paired with an aversive event will not only come to elicit fear responses, but will also acquire negative valence, such that it becomes unpleasant or disliked. Like conditional fear responses, this negative valence is subject to extinction (Lipp, Oughton, & LeLievre, 2003), however, this extinction seems to progress at a slower rate and some residual negative valence for the CS^+ may remain at the end of extinction. Residual negative valence is said to be a predictor of the extent of fear recovery after reinstatement (Hermans et al., 2005). The results reported by Krypotos and Engelhard (2018) question whether novelty-facilitated extinction training will affect CS valence acquired during differential conditioning, but this study assessed self-reports of fear, rather than CS valence, and after each conditioning phase, not continuously. Past research has shown differences between online and offline measures of CS valence (Lipp et al., 2003) and thus, an online measure of CS valence was included in the current study.

The current study was designed to conceptually replicate and extend the finding reported by Dunsmoor et al. (2015) that novelty-facilitated extinction training reduced the return of extinguished fear as indexed by electrodermal responses. Rather than spontaneous recovery, the current study assessed whether novelty-facilitated extinction would also reduce fear reinstatement induced by the presentation of three unpaired USs after successful extinction. In order to replicate the relationship between Intolerance of Uncertainty and fear recovery shown by Dunsmoor et al. (2015) participants completed the IUS-12, an abbreviated version of the IUS (Carleton, Norton, & Asmundson, 2007). Finally, we wanted to assess whether novelty-facilitated extinction training only affects recovery of fear as indexed by

electrodermal responses or extends to other indices of fear learning such as self-reported stimulus valence which was assessed online in parallel to electrodermal responses.

Method

Participants

Forty-eight university students and community members (mean age: *M* = 25.60, *SD* = 10.53, range: 18 – 62; 29 female) volunteered participation in exchange for course credit or AU$15 and provided informed consent. Participant numbers were based on the sample size used by Dunsmoor et al. (2015) for statistical analyses. Upon arrival at the laboratory participants were allocated to one of two groups, Novelty-Facilitated Extinction (NFE) or standard Extinction (EXT), alternatingly with the proviso of keeping the sex ratios balanced between the groups. Participants completed the experimental protocol relevant to their group, a post-experimental questionnaire and the IUS-12 (Carleton, Norton, & Asmundson, 2007). The study protocol was approved by the Curtin University Human Research Ethics Committee.

Apparatus and materials

The conditional stimuli were four angry, male Caucasian faces (poses An_O, of posers 20, 23, 32, 34; Tottenham et al., 2009), with each participant presented with two of the faces. The faces were presented for eight seconds, centred on a light grey background on a 17-inch LCD screen. The two faces used, whether the first trial was a CS^+ or CS^-, and which face served as the CS^+/CS^- was counterbalanced across participants. The 200 ms electro-tactile US was generated by a Grass SD9 stimulator and presented through a concentric electrode secured to the participant's dominant forearm. US intensity was set individually to a level they experienced as 'unpleasant but not painful'. During extinction, an 80 dBA 800 Hz pure tone was presented for 1.5 seconds through headphones (Sennheiser HD 25-1) in group NFE. DMDX software (Forster & Forster, 2003) was used to control stimulus presentation

and timing.

Physiological responses were recorded with a Biopac MP150 system at 1000 Hz. Respiration was monitored with a respiration belt (TSD201) attached around the participants' lower torso, and SCR was recorded with two 8-mm Ag/AgCl pre-gelled electrodes (EL507) attached to the thenar and hypothenar eminences of the participants' non-dominant hand and connected to a EDA100C amplifier (gain: 2μS/V). Participants provided continuous CS valence ratings using a TSD115 variable assessment transducer with a scale anchored from 'very unpleasant' (0) to 'very pleasant' (9). After completion of the experimental protocol, participants completed a post experimental questionnaire comprising a) a check of contingency knowledge requiring participants to select out of the four possible faces the two presented during the experimental protocol and the one followed by the US, b) pleasantness ratings of the four CS faces and the electro-tactile US on a 7 point Likert scale anchored 'Pleasant' and 'Unpleasant', c) the IUS-12 (Carleton, Norton, & Asmundson, 2007), and d) a request for demographic information, including age, gender, and ethnicity. The IUS-12 is a 12 item self-report measure that assesses intolerance of uncertainty on a 5 point Likert scale (Anchors: Not at all characteristic of me – Entirely characteristic of me). It is reported to have excellent internal consistency (α=.91).

Procedure

Prior to arrival participants were assigned to one of the two groups; Novelty-Facilitated Extinction (NFE) or Extinction (EXT). On arrival at the laboratory participants were greeted, presented with information about the experiment and asked to provide informed consent. Participants were seated in front of the computer screen and the respiratory belt, electrodermal, and US electrodes were attached. Participants were then instructed how to use the variable assessment transducer to rate CS valence. A shock work-up was performed to set the US to an intensity that each participant indicated was 'unpleasant, but not painful',

which was used for the remainder of the experiment. Participants were then asked to wear a set of headphones to block out background noise and to allow them to focus on the task and instructed to relax while a 3-minute electrodermal baseline was recorded.

The experimenter initiated the experimental protocol comprising habituation, acquisition, extinction, reinstatement, and reinstatement test. During habituation, participants viewed four presentations of each of the two CSs for eight seconds. During the 24 acquisition trials (12 per CS), the electro-tactile US was presented during the last 200 ms of half of the presentations of one CS (CS^+) whereas the other CS (CS^-) was presented alone. The US presentations were distributed at random with the restrictions that the first CS^+ of acquisition was followed by a US and that no more than two consecutive CS^+ were presented without a US. Extinction training comprised 16 presentations of each CS. No electro-tactile USs were presented, however the 1.5 s, 80 dBA, 800 Hz tone was presented during all CS^+ trials in group NFE such that tone and CS^+ co-terminated. The reinstatement manipulation comprised three presentations of the electro-tactile US alone 14, 26, and 38 s after the last extinction trial. This was followed by a reinstatement test comprising four presentations of each CS without the US. The first CS during the reinstatement test was presented 12 s after the last US. In all phases CS onsets were separated by a random intertrial interval of 22, 24 or 26s. CS sequence was random with the restriction that no more than two consecutive CSs could be the same and two counterbalanced CS sequences were used. After completion of the experimental protocol, participants were asked to complete the post-experimental questionnaire, the IUS-12, and to provide demographic information.

Response definition and data analysis

The number of spontaneous electrodermal responses during the 3-minute baseline was recorded to provide a measure of overall responsiveness (Dawson, Schell, & Filion, 2007). SCRs were scored in three latency windows (Prokasy & Kumpfer, 1973; Luck & Lipp,

2016). First interval responses (FIR) as the largest responses starting between 1 to 4 s after CS onset, second interval responses (SIR) as the largest responses starting between 4 to 8.8 seconds after CS onset, and third interval responses (TIR) as the largest responses starting between 8.8 to 11.8 seconds after CS onset (Prokasy & Kumpfer,1973). SCRs were square root transformed and range corrected prior to data analysis to reduce skewness and the impact of individual differences in electrodermal responding (Dawson, Schell, & Filion, 2007; Lykken, 1972). Range correction was performed by dividing each response by the largest response produced by the participant, usually the response to the first US.

Valence ratings were scored by subtracting the largest voltage deviation occurring during the 8 s CS presentation from the 1 s pre-CS baseline voltage which represented a 'neutral' setting. The reinstatement index for the correlational analysis was derived by subtracting the response to the first CS^- from that to the first CS^+ during reinstatement test. The IUS scores were determined as the mean of the participants' responses across the available items rather than as a total score as one participant had missed one item.

Habituation data for one participant were lost due to equipment failure, however this participant provided valid data in all other phases. SCRs and valence ratings from habituation, acquisition, and extinction were averaged into blocks of two consecutive trials and subjected to 2 × 2 × n (Group [NFE vs. EXT) × CS [CS^+ vs. CS^-] × Block [2, 6 or 8 respectively]) factorial mixed model ANOVAs using SPSS 22. For each significant main effect and interaction, Pillai's Trace F values and partial n^2 are reported (Vasey & Thayer, 1987) adopting a significance level of .05. Only the results for electrodermal FIRs are reported as the analysis of SIRs did not add additional information.

Results

Preliminary analyses

The two groups did not differ in gender ratio (female:male; NFE: 14:10; EXT: 15:9),

number of contingency non-verbalisers (NFE: 5; EXT: 6), age (NFE: *M* = 26.17 years, *SD* = 11.87; EXT: *M* = 25.04 years, *SD* = 9.22), perceived unpleasantness of the electro-tactile stimulus (NFE: *M* = 5.83, *SD* = 1.09; EXT: *M* = 5.50, *SD* = 0.98), number of spontaneous SCRs during 3-min baseline (NFE: *M* = 34.29, *SD* = 18.60; EXT: *M* = 30.96, *SD* = 17.79), and IUS-12 scores (NFE: *M* = 2.67, *SD* = 0.87, range: 1.25 – 4.50; EXT: *M* = 2.74, *SD* = 0.76, range: 1.50 – 4.58), all $t(46) < 1.13$, $p > .270$. Participants rated the CS^+ as more unpleasant than the CS^- post-experimentally (CS^+: *M* = 6.19, *SD* = 1.23; CS^-: *M* = 5.57, *SD* = 1.10; $F(1, 45) = 9.68$, $p = .002$, $\eta p^2 = .18$), with no differences between the groups, all $F < 1.0$, $p > .960$, $\eta p^2 < .01$. A 2 × 6 (Group × Block) factorial ANOVA revealed that responses to the electro-tactile unconditional stimulus during acquisition declined across blocks of trials, $F(5, 42) = 4.17$, $p = .004$, $\eta p^2 = .33$, but did not differ between groups, all $F < 1.0$, $p > .960$, $\eta p^2 < .01$. The tone stimulus presented in group NFE during extinction elicited larger electrodermal responses than seen in the same latency window in group EXT as indicated by a main effect for block, $F(7, 40) = 3.64$, $p = .004$, $\eta p^2 = .39$, and a Group × Block interaction, $F(7, 40) = 4.35$, $p = .001$, $\eta p^2 = .43$.

Habituation, Acquisition, and Extinction

Analyses of data from all participants and from participants who were able to verbalise the contingencies in the post-experimental questionnaire only yielded the same pattern of results. Hence the current report is based on the data from the entire sample. As shown in the left panel of Figure 1, electrodermal FIRs declined across blocks of habituation, $F(1, 45) = 27.16$, $p < .001$, $\eta p^2 = .38$. During acquisition (see middle panel of Figure 1), electrodermal FIRs to CS^+ exceeded those to CS^-, $F(1, 46) = 19.74$, $p < .001$, $\eta p^2 = .30$, and declined across blocks, $F(5, 42) = 2.55$, $p = .042$, $\eta p^2 = .23$. The CS × Block, $F(5, 42) = 1.61$, $p = .179$, $\eta p^2 = .16$, and Group × CS × Block interactions, $F(5, 42) = 0.83$, $p = .536$, $\eta p^2 = .09$, were not significant.

Electrodermal responses during extinction are shown in the right panel of Figure 1. FIRs to CS^+ were larger than responses to CS^-, F (1, 46) = 16.23, $p < .001$, $\eta p^2 = .26$, and declined across blocks of trials, F (7, 40) = 2.34, $p = .042$, $\eta p^2 = .29$. This decline differed between groups, Group × Block interaction, F (7, 40) = 2.48, $p = .033$, $\eta p^2 = .30$, with responses in group NFE larger on Block 1 than on Block 8 whereas there was no such difference in group EXT. The CS × Block, F (7, 40) = 0.86, $p = .545$, $\eta p^2 = .13$, and Group × CS × Block interactions, F (7, 40) = 1.30, $p = .274$, $\eta p^2 = .19$, were not significant. To confirm that differential electrodermal responses extinguished in both groups, a supplementary analysis compared responses elicited early (trials 2 and 3 – trial 1 was excluded to ensure that all participants had experienced at least two consecutive CS^+ presented without the US) and late during extinction (trials 15 and 16). This analysis revealed main effects for CS, F (1, 46) = 11.66, $p = .001$, $\eta p^2 = .202$, and block, F (1, 46) = 8.15, $p = .006$, $\eta p^2 = .15$, and a CS × Block interaction, F (1, 46) = 4.29, $p = .044$, $\eta p^2 = .085$. The Group × CS and Group × CS × Block interactions were not significant, both F (1, 46) < 1.0, $p > .840$, $\eta p^2 < .002$. Follow up analyses confirmed that differential responding was significant in both groups early, both F (1, 46) > 5.96, $p < .020$, $\eta p^2 > .114$, but not late during extinction, both F (1, 46) < 2.05, $p > .160$, $\eta p^2 < .042$.

CS evaluations did not differ across stimuli, groups or blocks during habituation (see left panel of Figure 2), all $F < 2.53$, $p > .119$, $\eta p^2 < .05$. The analysis of the CS evaluations during acquisition yielded a CS × Block interaction, F (5, 42) = 4.60, $p = .002$, $\eta p^2 = .36$, however, follow up analyses failed to yield any significant results (largest difference between CS^+ and CS^- on block 6: F (1, 46) = 3.99, $p = .052$, $\eta p^2 = .080$). The Group × CS × Block interaction, F (5, 42) = 1.38, $p = .251$, $\eta p^2 = .14$, was not significant. During extinction, CS^+ was evaluated as more unpleasant than CS^-, F (1, 46) = 7.48, $p = .009$, $\eta p^2 = .14$. The CS × Block, F (7, 40) = 0.35, $p = .926$, $\eta p^2 = .06$, and Group × CS × Block interactions, F (7, 40) =

0.91, $p = .509$, $\eta p^2 = .13$, were not significant. A supplementary analysis based on evaluations from early and late during extinction confirmed this pattern of results, revealing a main effect for CS, F (1, 46) = 7.06, $p = .011$, $\eta p^2 = .133$.

Reinstatement

To assess the effect of the reinstatement manipulation electrodermal responses and evaluations from the last CS^+ and CS^- trials of extinction and the first CS^+ and CS^- trials of the reinstatement test were subjected to 2 × 2 × 2 (Group × CS × Trial) factorial ANOVAs. As can be seen in the left panel of Figure 3, electrodermal responses to CS^+ seemed to exceed those to CS^- on the first trial of the reinstatement test after standard extinction, but not after novelty-facilitated extinction. The analysis confirmed this impression yielding main effects for CS, F (1, 46) = 5.54, $p = .023$, $\eta p^2 = .11$, and trial, F (1, 46) = 18.11, $p < .001$, $\eta p^2 = .28$, as well as a marginal Group × CS × Trial interaction, F (1, 46) = 3.58, $p = .065$, $\eta p^2 = .07$. Responses to CS^+ were larger than responses to CS^- on the first reinstatement trial in group EXT, F (1, 46) = 9.57, $p = .003$, $\eta p^2 = .17$, but not in group NFE, F (1, 46) = 0.001, $p = .976$, $\eta p^2 < .001$. Responses to CS^+ and CS^- did not differ on the last trial of extinction in either group, both $F < 1.10$, $p > .315$, $\eta p^2 < .03$. Responding to CS^+ increased from the last trial of extinction to the first trial of the reinstatement test in group EXT, F (1, 46) = 13.23, $p = .001$, $\eta p^2 = .22$, but only marginally so in group NFE, F (1, 46) = 3.22, $p = .079$, $\eta p^2 = .065$. The increase in responding to CS^- from the last trial of extinction to the first trial of the reinstatement test was significant in group NFE, F (1, 46) = 6.04, $p = .018$, $\eta p^2 = .116$, but not in group EXT, F (1, 46) = 0.16, $p = .691$, $\eta p^2 = .003$. The increase in responding to CS^+ or to CS^- from the last trial of extinction to the first trial of reinstatement test did not differ between groups, both $t(46) < 1.50$, $p > .150$.

The corresponding analysis of the CS evaluations (see Figure 3, right panel) yielded a main effect for CS, F (1, 46) = 8.57, $p = .005$, $\eta p^2 = .16$, suggesting more unpleasant

evaluations of the CS^+ and a marginal CS × Trial interaction, F (1, 46) = 3.03, p = .089, ηp^2 = .06. The latter reflects more negative evaluations of CS^+ after the reinstatement manipulation, (M = -1.66, SD = 0.88 vs. M = -1.77, SD = 0.91; F (1, 46) = 5.26, p = .027, ηp^2 = .10), whereas there was no difference for CS^- (M = -1.32, SD = 0.84 vs. M = -1.32, SD = 0.90; F (1, 46) = 0.01, p = .972, $\eta p^2 < .01$). All other F (1, 46) < 1.80, $p > .190$, $\eta p^2 < .04$.

Relation to IUS-12

Figure 4 shows the relationship between reinstatement of conditional electrodermal responding, defined as the difference in electrodermal response to CS^+ and CS^- on the first trial of the reinstatement test, and the IUS-12 score in groups NFE and EXT. As can be seen, this relationship was significant in group EXT, r_{xy} = .41, p = .049, but not in group NFE, r_{xy} = -.03, p = .874. A similar analysis for CS evaluations yielded no significant results (EXT: r_{xy} = -.16, p = .457; NFE: r_{xy} = .11, p = .604).

Supplementary analyses

The approach to conceptualize reinstatement used in the current study differs from that employed by Dunsmoor et al. (2015) who subjected responses to CS^+ and CS^- in the early phase of the reinstatement test (the first three trials of the reinstatement test respectively) to a Group × CS ANOVA and calculated a reinstatement index as a ratio of the mean SCR to the first three CS^+ presentations during the reinstatement test divided by the largest SCR to a CS^+ during acquisition. Like Dunsmoor et al. we did not find a Group × CS interaction, F (1, 46) = 0.99, p = .326, ηp^2 = .021, although the main effect for CS was significant in our study, F (1, 46) = 4.52, p = .039, ηp^2 = .089. Like Dunsmoor et al. we find larger responses to CS^+ than to CS^- during early reinstatement test in group EXT, F (1, 46) = 4.87, p = .032, ηp^2 = .096, but not in group NFT, F (1, 46) = 0.64, p = .427, ηp^2 = .014. Like in Dunsmoor et al., the reinstatement index did not differ between groups NFE (M = 0.365, SD = .30) and EXT (M = 0.497, SD = 0.390), t(46) = 1.316, p = .195.

Discussion

The current study aimed to conceptually replicate and extend the findings of Dunsmoor et al. (2015) who reported that return of fear is reduced after extinction training in which the CS^+ is paired with a novel, non-aversive tone stimulus, novelty-facilitated extinction. Rather than using spontaneous recovery as an index of return of fear, the current study assessed the effects of novelty-facilitated extinction on reinstatement. Following three unpaired presentations of the unconditional stimulus, differential electrodermal responding was reinstated after standard extinction, but not after training with the novelty-facilitated extinction procedure. It should be noted, however, that the critical three way interaction was only marginal in the omnibus analysis (p = .065) and not significant at the pre-set level. Also consistent with Dunsmoor et al. (2015), self-reported Intolerance of Uncertainty predicted the extent of return of fear after standard extinction training, but not after novelty-facilitated extinction. Novelty-facilitated extinction did not affect differential conditional stimulus evaluations, however, which remained stable across extinction training in both groups. This may be due to using angry faces as CSs which were evaluated as negative prior to acquisition training or the use of a partial reinforcement schedule during acquisition which may have enhanced uncertainty and delayed extinction. The a-priori negative valence of the CSs may have limited the extent to which differential evaluations were acquired during acquisition and may have slowed or prevented extinction which renders the observation of relapse difficult. On the other hand, past research on evaluative conditioning has documented failures to extinguish acquired CS valence (e.g., Baeyens, Crombez, van den Bergh, & Eelen, 1988) although a recent meta-analysis has supported the notion that evaluative learning is subject to extinction (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Future research will have to clarify the conditions under which extinction of acquired stimulus valence can be observed.

The current study provides some support for the notion that extinction training can be strengthened to the extent that return of fear can be avoided. However, currently it remains unclear how the addition of novel stimuli that had not been encountered before can achieve this. Dunsmoor et al. (2015) liken the novelty-facilitated extinction procedure to counterconditioning in an attempt to explain the finding, but concede that given the novel stimulus is neutral and does not elicit a behavioural response it is difficult to accommodate the current findings within traditional theories of counter conditioning that assume a competition between opposing motivational tendencies. Moreover, the failure to see any change in stimulus evaluations during novelty-facilitated extinction training does not support a motivational explanation.

Alternatively, one might argue that pairing the CS^+ with a non-aversive stimulus during extinction may enhance the prediction error which drives extinction learning. The CS^+-novel tone pairing may promote a stronger learning of the CS^+-noUS association than does the mere omission of the US. This discrepancy between groups may be even stronger due to the intermittent reinforcement schedule used during acquisition which may have resulted in the simultaneous acquisition of CS^+-US and CS^+-noUS associations. Thus, little additional learning may have occurred during standard extinction training whereas a new association was added in the novelty-facilitated extinction training.

The observation that self-reported Intolerance of Uncertainty covaried with reinstatement after standard extinction training, but not after novelty-facilitated extinction training gives rise to a different interpretation of what mediates the effect of novelty-facilitated extinction training observed by Dunsmoor et al. (2015) and in the present study. Both studies employed an intermittent reinforcement schedule during acquisition, but a continuous reinforcement schedule during novelty-facilitated extinction training. One might argue that the switch from the CS^+ being an unreliable predictor of the US to the CS^+ being a

reliable predictor of a neutral stimulus during novelty-facilitated extinction removed the ambiguity that the CS^+ had acquired during acquisition (and that was maintained or even enhanced in standard extinction training). As a reliable predictor of a low intensity tone, the CS^+ was no longer a potential signal of an aversive outcome even after three presentations of this outcome alone. Thus, rather than mediated by the pairing of the CS^+ with a novel stimulus, the reduction of return of fear after novelty-facilitated extinction training may occur because the CS^+ transitions from an unreliable to a reliable predictor. This interpretation can be readily tested by varying the reinforcement schedules during acquisition or novelty-facilitated extinction training. It seems worth noting that finding that the effect of novelty-facilitated extinction reflects on a change in outcome certainty does not render the phenomenon uninteresting in the context of the return of fear. Rather, it would provide novel information as to which conditions enable it and which impair it.

Another factor which may have affected the outcome of the current study is the use of angry faces as CSs. Fear conditioned to angry faces has been shown to resist extinction (Öhman, & Dimberg, 1978) which may have reduced the efficacy of standard extinction training. Using this CS material may also have limited the extent to which self-reported evaluations could reflect changes in stimulus valence during acquisition and extinction. As indicated by CS evaluations from habituation, these faces were disliked prior to pairing with the aversive electro-tactile US. Thus replication of the effects of novelty-facilitated extinction training with non fear-relevant stimuli seems required.

The current study adds some support to the notion that novelty-facilitated extinction training can reduce the return of fear. However, more basic work is needed to reach a better understanding as to how this training enhances the effectiveness of extinction training to make it more lasting. Such an understanding has the potential to modify the manner in which exposure training is designed in a clinical setting (Scheveneels, Boddez, Vervliet, &

Hermans, 2016). One might, for instance, consider training patients to imagine neutral situations whenever confronted with signals that were previously associated with negative outcomes. Such an intervention could resemble the approach used in ‘association splitting’, a technique used to reduce unwanted intrusive thoughts in obsessive compulsive disorder (Moritz, Jelinek, Klinge, & Naber, 2007). However, more basic research work will be required to enhance our conceptual understanding of what mediates the effects of novelty-facilitated extinction training before a translation into applied settings can be considered.

References

Baeyens, F., Crombez, G., van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*, 179-199.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986. doi:10.1016/S0006-3223(02)01546-9

Bouton, M. E., Woods, A. M., & Pineño, O. (2004). Occasional reinforced trials during extinction can slow the rate of rapid reacquisition. *Learning and Motivation, 35*(4), 371-390. doi:10.1016/j.lmot.2004.05.001

Buhr, K., & Dugas, M. J. (2002). The intolerance of uncertainty scale: psychometric properties of the English version. *Behaviour Research and Therapy, 40*(8), 931-945. doi:10.1016/S0005-7967(01)00092-4

Carleton, R. N., Norton, M. A., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders, 21*, 105-117. doi:10.1016/j.janxdis.2006.03.014

Culver, N. C., Stevens, S., Fanselow, M. S., & Craske, M. G. (2018). Building physiological toughness: Some aversive events during extinction may attenuate return of fear. *Journal of Behavior Therapy and Experimental Psychiatry, 58*, 18-28. doi:10.1016/j.jbtep.2017.07.003

Craske, M., Treanor, M., Conway, C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10-23. doi:10.1016/j.brat.2014.04.006

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.). *Handbook of psychophysiology* (pp.

159–181). Cambridge: Cambridge University Press.

Dunsmoor, J. E., Campese, V. D., Ceceli, A. O., LeDoux, J. E., & Phelps, E. A. (2015). Novelty-Facilitated Extinction: Providing a Novel Outcome in Place of an Expected Threat Diminishes Recovery of Defensive Responses. *Biological Psychiatry, 78*(3), 203-209. doi:10.1016/j.biopsych.2014.12.008

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*, 116-124. doi:10.3758/BF03195503

Graham, B., & Milad, M. (2011). The study of fear extinction: Implications for anxiety disorders. *The American Journal of Psychiatry, 168*, 1255-1265. doi:10.1176/appi.ajp.2011.11040557

Hermans, D., Dirikx, T., Vansteenwegenin, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*(4), 533-551. doi:10.1016/j.brat.2004.03.013

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative Conditioning in Humans: A Meta-Analysis. *Psychological Bulletin, 136*, 390-421. doi:10.1037/a0018916

Kerkhof, I., Vansteenwegen, D., Beckers, T., Dirikx, T., Baeyens, F., D'Hooge, R., et al. (2012). The role of negative affective valence in return of fear. In A. D. Gervaise (Ed.), Psychology of fear: New research (pp. 153e170). New York: Nova Science Publishers.

Krypotos, A.-M., & Engelhard, I. M. (2018). Testing a novelty-based extinction procedure for the reduction of conditioned avoidance. *Journal of Behavior Therapy and Experimental Psychiatry, 60*, 22-28. doi:https://doi.org/10.1016/j.jbtep.2018.02.006

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

doi:10.1016/S0023-9690(03)00011-0

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience & Biobehavioral Reviews, 80,* 703-728. doi:10.1016/j.neubiorev.2017.07.007

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., . . . Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews, 77,* 247-285. doi:j.neubiorev.2017.02.026

Luck, C. C., & Lipp, O. V. (2015). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy, 66,* 18-31. doi:10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2016). When orienting and anticipation dissociate - a case for scoring electrodermal responses in multiple latency windows in studies of human fear conditioning. *International Journal of Psychophysiology, 100,* 36-43. doi:10.1016/j.ijpsycho.2015.12.003

Lykken, D. T. (1972). Range correction applied to heart rate and to GSR data. *Psychophysiology, 9,* 373-379. doi:10.1111/j.1469-8986.1972.tb03222.x

Mineka, S., & Zinbarg, R. (2006). A Contemporary Learning Theory Perspective on the Etiology of Anxiety Disorders: It's Not What You Thought It Was. *American Psychologist, 61,* 10-26. doi:10.1037/0003-066X.61.1.10

Moritz, S., Jelinek, L., Klinge, R., & Naber, D. (2007). Fight Fire with Fireflies! Association Splitting: A Novel Cognitive Technique to Reduce Obsessive Thoughts. *Behavioural and Cognitive Psychotherapy, 35,* 631-635. doi:10.1017/S1352465807003931

O'Brien, F., & Cousineau, D. (2014). Representing Error bars in within-subject designs in typical software packages. *The Quantitative Methods for Psychology, 10*(1), 56-67. doi:10.20982/tqmp.10.1.p056

Öhman, A., & Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses: A case of "Preparedness"? *Journal of Personality and Social Psychology, 36,* 1251-1258. doi:10.1037/0022-3514.36.11.1251

Prokasy, W. F., & Kumpfer, K. L. (1973). Classical conditioning. In W. F. Prokasy, &D. C. Raskin (Eds.), *Electrodermal activity in psychological research* (pp. 157-202). San Diego, CA: Academic Press.

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (2016). The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment. *Behaviour Research and Therapy, 86,* 87-94. doi:10.1016/j.brat.2016.08.015

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. doi:10.1038/nature08637

Thompson, A., & Lipp, O. V. (2017). Extinction during reconsolidation eliminates recovery of fear conditioned to fear-irrelevant and fear-relevant stimuli. *Behaviour Research and Therapy, 92,* 1-10. doi:10.1016/j.brat.2017.01.017

Thompson, A., McEvoy, P. M., & Lipp, O. V. (2018). Enhancing extinction learning: Occasional presentations of the unconditioned stimulus during extinction eliminate spontaneous recovery, but not necessarily reacquisition of fear. *Behaviour Research and Therapy, 108,* 29-39. doi: 10.1016/j.brat.2018.07.001

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. doi:10.1016/j.psychres.2008.05.006

Vasey, M. W., & Thayer, J. F. (1987). The conditioning problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology, 24*, 479-486. doi:10.1111/j.1469-8986.1987.tb00324.x

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. doi:10.1016/j.brat.2015.05.016

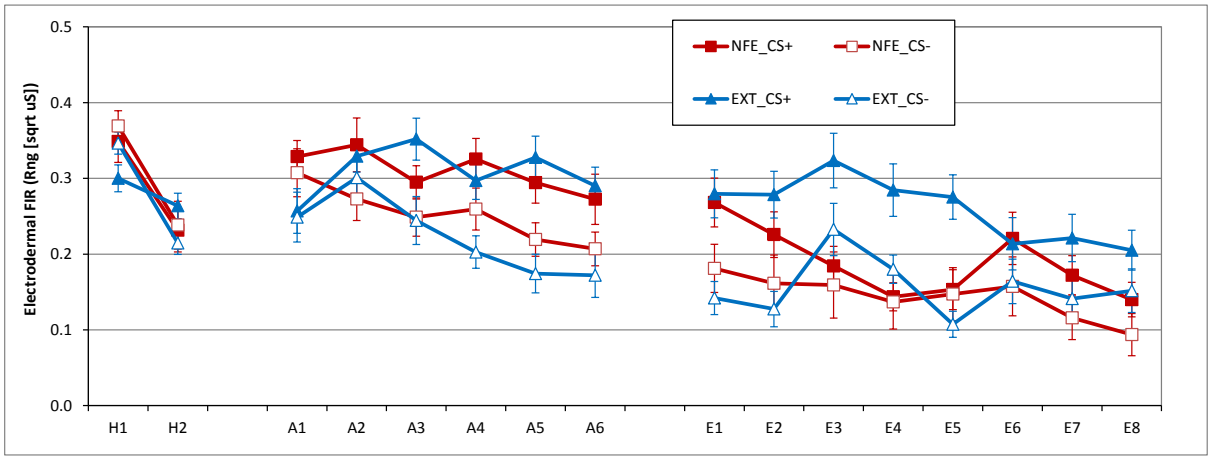
Figure Captions

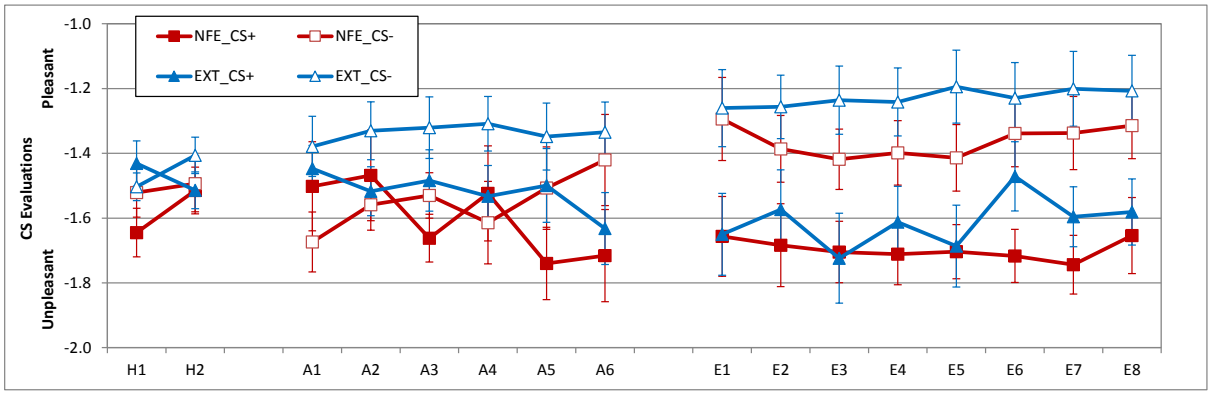
Figure 1: Electrodermal first interval responses during habituation (H1 – H2), acquisition (A1-A6), and extinction (E1-E8) as a function of group and CS condition. Error bars represent SEMs for within subject designs based on O'Brien and Cousineau (2014).

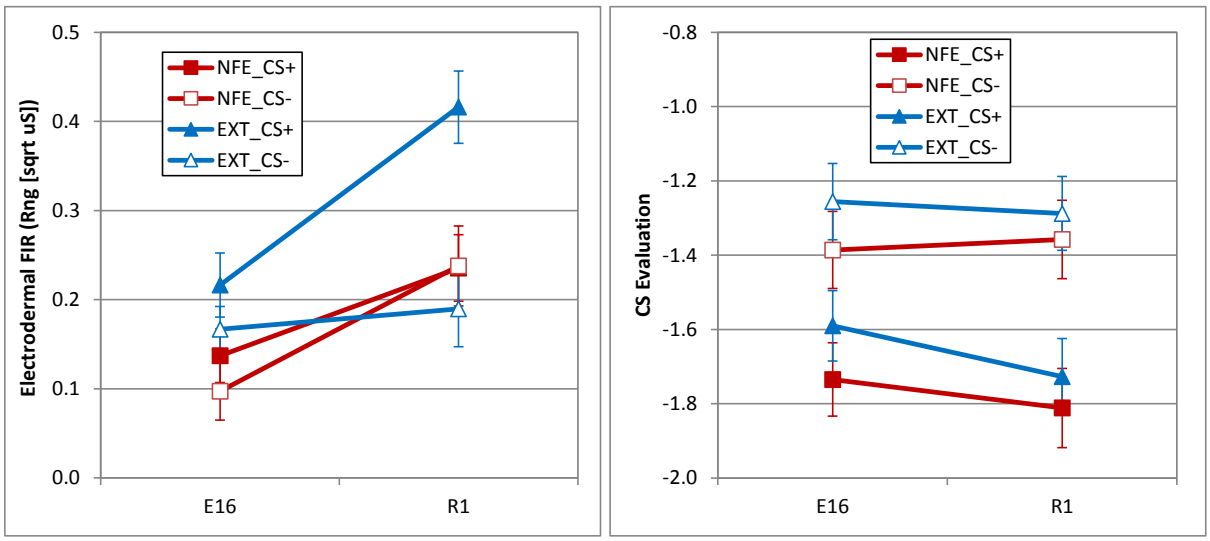
Figure 2: Stimulus evaluations during habituation (H1 – H2), acquisition (A1-A6), and extinction (E1-E8) as a function of group and CS condition (possible range: -2.5 – 2.5). Error bars represent SEMs for within subject designs based on O'Brien and Cousineau (2014).

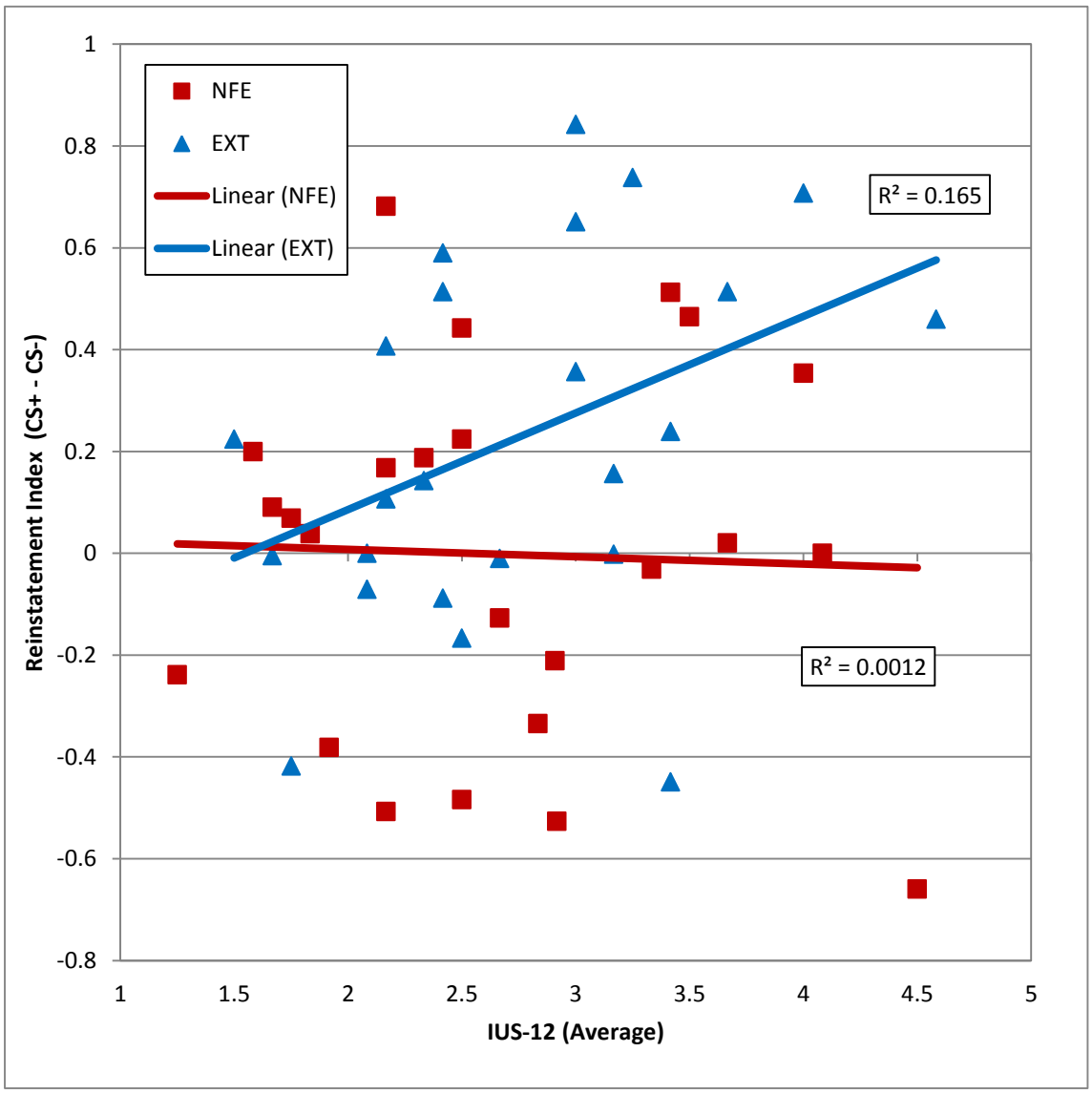
Figure 3: Electrodermal first interval responses (left panel) and stimulus evaluations (right panel) on the last trial of extinction (E16) and the first trial of reinstatement test (R1) as a function of group and CS condition. Error bars represent SEMs for within subject designs based on O'Brien and Cousineau (2014).

Figure 4: Scatterplot of average scores on the IUS-12 and Reinstatement index as a function of group.


Electrodermal FIR (Rng [sqrt uS])
0.5
0.4
0.3
0.2
0.1
0.0
NFE_CS+
NFE_CS-
EXT_CS+
EXT_CS-
H1
H2
A1
A2
A3
A4
A5
A6
E1
E2
E3
E4
E5
E6
E7
E8


NFE_CS+
NFE_CS-
EXT_CS+
EXT_CS-
CS Evaluations
Pleasant
Unpleasant
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0
H1
H2
A1
A2
A3
A4
A5
A6
E1
E2
E3
E4
E5
E6
E7
E8


Electrodermal FIR (Rng [sqrt uS])
0.5
0.4
0.3
0.2
0.1
0.0
NFE_CS+
NFE_CS-
EXT_CS+
EXT_CS-
E16
R1
CS Evaluation
-0.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0
NFE_CS+
NFE_CS-
EXT_CS+
EXT_CS-
E16
R1


NFE
EXT
Linear (NFE)
Linear (EXT)
R² = 0.165
R² = 0.0012
Reinstatement Index (CS+ - CS-)
IUS-12 (Average)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
1
1.5
2
2.5
3
3.5
4
4.5
5