Enhancing extinction learning: Occasional presentations of the unconditioned stimulus during extinction eliminate spontaneous recovery, but not necessarily reacquisition of fear

Alina Thompson[a,*], Peter M. McEvoy[a,b], Ottmar V. Lipp[a,c]

[a] School of Psychology, Curtin University, Kent Street, Bentley, WA 6102, Australia
[b] Centre for Clinical Interventions, 223 James Street, Northbridge, WA 6003, Australia
[c] ARC-SRI: Science of Learning Research Centre, University of Queensland, Brisbane, QLD 4072, Australia

ABSTRACT

Background: Fears underlying anxiety disorders are commonly treated with exposure-based therapies, which are based on the principles of extinction learning. While these treatments are efficacious, fears may return after successful treatment. Past research suggested that post-extinction recovery of fear could be reduced through extinction training that involves occasional presentations of the aversive unconditioned stimulus (US), paired with the conditioned stimulus (CS). Here, we examined whether extinction training with occasionally paired or unpaired US presentations is superior in the reduction of fear recovery to non-reinforced extinction.
Method: Following differential fear conditioning to neutral cues, participants ($N = 72$; *M age* = 21.61 years, $SD = 3.95$) underwent either non-reinforced, partially reinforced, or unpaired extinction training.
Results: Extinction involving paired or unpaired US presentations, but not non-reinforced extinction, eliminated spontaneous recovery of differential skin conductance responses (SCRs). Results further suggested that unpaired, but not paired, US presentations may guard against rapid reacquisition of differential SCRs. No benefits of US presentations during extinction were found on the reinstatement of SCRs or recovery of differential negative CS + valence.
Conclusion: Presenting USs during extinction training was more effective than non-reinforced extinction in the reduction of fear recovery, as indexed by SCRs, with unpaired extinction being more effective than partially reinforced extinction.

1. Introduction

Past research has provided us with a good understanding of mechanisms underlying the development and reduction of fears, phobias, and anxiety disorders. Fears are acquired through association of neutral cues (conditioned stimuli, CSs), such as animals, with aversive outcomes (unconditioned stimuli [USs]; Davey, 1992), such as an animal bite. Through CS-US pairings we learn to predict which cues signal the arrival of aversive and potentially threatening events. While learning to fear cues which may pose a threat to our survival is an important adaptive mechanism that can protect us from harm and facilitate survival (Öhman & Mineka, 2001), fears may also become maladaptive and contribute to the development of anxiety and stress disorders, which can interfere with daily functioning (Foa & McLean, 2016). The current global prevalence rate of anxiety disorders is estimated at 7.3%, with approximately 11.6% of the population experiencing an anxiety

disorder in a given year (Baxter, Scott, Vos, & Whiteford, 2013; Craske & Stein, 2016). Anxiety disorders are commonly treated with exposure-based therapies, which are based on the principles of extinction learning (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). In its basic form, extinction training involves the repeated presentation of the CS, in the absence of the US, until a reduction of fear is achieved (Bouton, 2000).

While the efficacy of exposure therapies is well established, not all individuals respond to these treatments, while others experience a return of fear after successful treatment (Craske & Mystkowski, 2006; Weisman & Rodebaugh, 2018). Research suggests that extinguished fear may return, because extinction training does not result in the unlearning or persistent elimination of the original fear learning (i.e. the CS-US association), but creates a new, inhibitory association (CS-no US) that co-exists with the fear association (Bouton, 1993). As such, future CS presentations may activate the CS-no US or CS-US association,

---

whereby the latter would allow for a return of fear (Bouton, 1993). Recovery from extinction phenomena are well-documented in the conditioning literature and include recovery of extinguished responding in a new context (renewal), after the passage of time (spontaneous recovery), after the unsignaled presentation of the US (reinstatement), or after additional post-extinction CS-US pairings (reacquisition; Bouton, 2002). Findings from animal research indicate that reacquisition after extinction may occur at a faster rate than de novo conditioning (Napier, Macrae, & Kehoe, 1992; but see; Ricker & Bouton, 1996), suggesting that the original fear learning is preserved during extinction and, thereby, may be retrieved through future cue encounters. Taken together, research has identified several pathways that may result in the return of fear following successful extinction training or the successful completion of exposure therapy.

When applied to an example of relapse in the clinical setting, for instance the return of social anxiety, which is characterized by fear of social situations in which individuals may be exposed to rejection, embarrassment, or negative evaluations by others (American Psychiatric Association, 2013), fear may recover when individuals are re-exposed to previously feared social situations (CS) or through exposure to additional CS-US pairings (reacquisition), such as receiving negative feedback (US) during a meeting at work (CS). Given the frequency with which feared cues and outcomes may be encountered in daily life, whether in a paired (CS-US) or unpaired manner (CS or US), the likelihood of fear recovery appears high - this may seem discouraging from a clinical point of view. However, recent evidence suggests that exposure therapy may be optimized in a way that would minimize recovery of extinguished fear, even in light of occasional post-extinction CS-US pairings.

A method of exposure therapy for reducing the return of fear proposed by Craske et al. (2014) involves *occasionally reinforced extinction*, meaning intentionally exposing clients to occasional presentations of the feared event (US) during exposure therapy. In the case of social anxiety, this may involve the delivery of rejection or "shame attacks" during exposure to social situations (Craske et al., 2014). While this idea appears counterintuitive as CS-US pairings are implicated in fear acquisition (Davey, 1992), extant literature suggests that occasional presentations of the US during extinction training may be superior to conventional, non-reinforced extinction in preventing recovery of extinguished responding (e.g. Bouton, Woods, & Pineño, 2004; Culver, Stevens, Fanselow, & Craske, 2018).

Specifically, experiments conducted with animal subjects demonstrated that partially reinforced extinction training, involving occasional delivery of CS-US pairings, interfered with the reacquisition of extinguished responding in appetitive (Bouton et al., 2004) and operant conditioning preparations (Woods & Bouton, 2007). Of particular interest was the observation that partially reinforced extinction slowed the reduction of responding during extinction, as would be expected from reinforced training, but protected against rapid reacquisition, relative to non-reinforced extinction. Additionally, compared to partially reinforced training, an unpaired extinction procedure, whereby reinforcers were not paired with the CS, but instead delivered in the intertrial interval, further reduced the rate of reacquisition (Bouton et al., 2004 [Experiment 2]). Replication attempts with humans, however, have yielded mixed results in an appetitive conditioning study (van den Akker, Havermans, & Jansen, 2015), showing reduced reacquisition of US expectancies, but not self-rated *conditioned desires* for chocolate mousse, subsequent to partially reinforced and unpaired extinction training. That being said, the authors also reported group differences at baseline and differential effects of acquisition training on verbal (e.g. US expectancy) and physiological indices of conditioned responding (i.e. participants' rate of salivation in anticipation of food), making the overall interpretation of findings difficult.

An extension of Bouton et al's. (2004) findings to human fear conditioning, on the other hand, has yielded more promising results, suggesting that partially reinforced extinction may successfully reduce the

reacquisition of extinguished fear responses (Culver et al., 2018). Following differential fear conditioning to neutral cues, participants underwent either non-reinforced or partially reinforced extinction training. Similar to Bouton and colleagues' work, a 2:8 reinforcement schedule was used during extinction in the partially reinforced group, translating to six reinforced and 18 non-reinforced CS+ trials and 24 non-reinforced CS- trials. Tests of fear recovery showed that partially reinforced extinction training, relative to non-reinforced extinction, interfered with subsequent reacquisition of conditioned fear, as indexed by electrodermal responding. An aspect requiring further investigation, however, is the effect of partially reinforced extinction on the spontaneous recovery of extinguished fear. While Culver and colleagues observed reduced recovery of electrodermal responding to the CS+ after partially reinforced extinction, relative to non-reinforced extinction, these results must be interpreted with caution, as conditioned responding failed to extinguish during partially reinforced extinction training and, consequently, could not "recover." Nevertheless, the results of the reacquisition test provide evidence for cross-species applicability of partially reinforced extinction. The aim of the present study was to replicate and extend previous findings (Bouton et al., 2004; Culver et al., 2018) to the spontaneous recovery, reinstatement, and reacquisition of extinguished conditioned responding in human fear conditioning, employing partially reinforced, unpaired, and non-reinforced extinction training. Furthermore, a direct comparison of occasionally paired and unpaired US presentations during extinction would also allow for examination of underlying mechanisms, which may differ across different types of reinforced extinction training (e.g. Bouton et al., 2004; Rauhut, Thomas, & Ayres, 2001; Rescorla & Skucy, 1969).

Several mechanisms have been proposed to account for the superior protection from fear recovery effects subsequent to reinforced and unpaired extinction training, compared to non-reinforced extinction, including: Weakening of the CS-US contingency through unpaired US presentations (Frey & Butler, 1977; Rescorla & Skucy, 1969; Vervliet, Vansteenwegen, & Hermans, 2010); US habituation (Rauhut et al., 2001; but see; Thomas, Longo, & Ayres, 2005); sequential learning (Bouton et al., 2004; Capaldi, 1966, 1994); and enhanced extinction learning through violation of expectancies, also referred to as prediction errors (Craske et al., 2014; Culver et al., 2018; Rescorla & Wagner, 1972; Vurbic & Bouton, 2014).

Prediction errors are implicated in the acquisition and extinction of conditioned responding (e.g. Pearce & Hall, 1980; Rescorla & Wagner, 1972; Vurbic & Bouton, 2014), whereby learning is proposed to cease when the CS reliably predicts the delivery of the US (or its absence, in the case of extinction learning). Extinction learning may be enhanced through the occasional presentation of the US during extinction training, due to the violation of expectancies regarding the frequency of US presentations or changes to the CS-US relationship (e.g. Craske et al., 2014). For instance, the omission of the US at the onset of extinction provides an opportunity for new learning due to the mismatch between current information (CS-no US) and past learning (CS-US), while the presentation of occasionally paired and unpaired USs on later trials would sustain learning through the presentation of novel information that needs to be reconciled with prior learning. Hence, the occasional presentation of USs during extinction would allow participants to learn about the likelihood of future threat encounters, such as the frequency of US presentations, relative to CS-only trials, or the relationship between the CS and the US (i.e. occasionally paired or unpaired). Subsequent fear recovery could be reduced because participants learned that the CS predicts the absence of the US (unpaired extinction; Rescorla & Skucy, 1969; Vervliet et al., 2010) or that occasional CS-US trials occur in the presence of many CS-no US trials (partially reinforced extinction). This proposition is also supported by Bouton et al's. (2004) adaptation of sequential theory (Capaldi, 1966, 1994).

Bouton et al. (2004) proposed that the key aspect learned during