Behaviour Research and Therapy 108 (2018) 78–84

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Reduced return of threat expectancy after counterconditioning versus extinction

Sahaj Kang[a], Bram Vervliet[b], Iris M. Engelhard[a], Eva A.M. van Dis[a], Muriel A. Hagenaars[a,c,*]

[a] Department of Clinical Psychology, Utrecht University, The Netherlands
[b] Center for the Psychology of Learning and Experimental Psychopathology, Leuven University, Belgium
[c] Department of Clinical Psychology, Leiden University, The Netherlands

ARTICLE INFO

Keywords:
Anxiety disorders
Exposure therapy
Counterconditioning
Extinction
Evaluative learning
Return of fear

ABSTRACT

Exposure-based therapies are effective for anxiety disorders, but relapse remains a problem. One explanation might be that exposure therapy reduces threat expectancy but not related feelings of unpleasantness (negative valence of the conditioned stimulus; CS+), which may promote return of threat expectancy and associated fear. Laboratory research has indeed shown that fear extinction leaves negative valence of the conditioned stimulus (CS+) intact. Here, we tested whether adding positive consequences to the CS+ during extinction, a procedure known as counterconditioning, would change the valence of the CS+ and thereby prevent return of threat expectancy. Participants underwent Acquisition (day 1), Intervention (counterconditioning or extinction; day 2), and Spontaneous recovery and Reinstatement (day 3). As expected, threat expectancy ratings during the Spontaneous recovery and Reinstatement tests were lower after counterconditioning than after extinction, but counterconditioning did not reduce CS + negative valence more than extinction. Alternative mechanisms and clinical implications are discussed.

1. Introduction

Anxiety disorders are common with a life-time prevalence of up to 24.9% of the population (Baxter, Scott, Vos, & Whiteford, 2013; Kessler, McGonagle et al., 1994). An effective, evidence based treatment for anxiety disorders is exposure-based therapy (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; Norton & Price, 2007), in which individuals are exposed to fearful stimuli in order to diminish threat expectancy and associated fear responses. However, despite being highly successful in reducing fear, return of fear is a serious problem, with estimates ranging from 19% to 62% (Boschen, Neuman & Waters, 2009; Craske, 1999; Vervliet, Craske, & Hermans, 2013; Yonkers, Bruce, Dyck, & Keller, 2003).

Pavlovian fear conditioning is considered to be the basis for the development of pathological anxiety and fear (Lommen, Engelhard, Sijbrandij, van den Hout, & Hermans, 2013; VanElzakker, Dahlgren, Davis, Dubois, & Shin, 2014; Vervliet, 2008). Pavlovian fear conditioning constitutes associative learning: A neutral stimulus (conditioned stimulus; CS) is associated with an aversive event (unconditioned stimulus; US_{neg}). As a result, the neutral stimulus itself comes to elicit conditioned fear reactions (conditioned responses; CR). Consequently, exposure therapy is viewed as a form of extinction learning, in which the CS is presented repeatedly in the absence of the US_{neg}. Arguably, during extinction a new inhibitory memory is formed (CS → $noUS_{neg}$), which competes with the original memory (CS → US; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Jacoby & Abramowitz, 2016; Pavlov, 1927). The new memory becomes relatively stronger after extinction, but the original memory can still be triggered rather easily, for instance through unsignaled encounters with the US_{neg} (reinstatement), thereby allowing return of fear (Craske et al., 2014; Jacoby & Abramowitz, 2016; Vervliet, 2008).

Extinction-based procedures indeed successfully diminish threat expectancy, i.e., violating CS+/US_{neg} contingency typically results in reduced expectancy of the US_{neg} after CS + presentation (Lovibond, 2004; Vansteenwegen, Francken, Vervliet, Clerq & Eelen, 2006). However, by repeated association with the US_{neg}, the negative US_{neg} valence can be transferred to the CS+ (evaluative learning). Despite diminished threat expectancy, this conditioned negative CS + valence is not necessarily reduced by extinction-based procedures (Baeyens, Díaz & Ruis, 2005; Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Engelhard, Leer, Lange, & Olatuji, 2014; Luck & Lipp, 2015; Olatunji, Forsyth, & Cherian, 2007; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). Importantly, negative CS + valence was associated with higher spontaneous recovery and reinstatement in

* Corresponding author. Department of Clinical Psychology, Utrecht University, The Netherlands.
E-mail address: m.a.hagenaars@uu.nl (M.A. Hagenaars).

https://doi.org/10.1016/j.brat.2018.06.009
Received 17 January 2018; Received in revised form 10 May 2018; Accepted 25 June 2018
Available online 26 June 2018

previous studies (e.g. Dirikx et al., 2004; Hermans, Dirikx, Vansteenwegen, Baeyens, & Eelen, 2005; Vasey, Harbaugh, Buffington, Jones, & Fazio, 2012; Zbozinek, Hermans, Pernoveau, Liao & Craske, 2015). Dirikx et al. (2004) explained the relationship between valence and return of fear with Lang's emotion theory (1995), which proposes that all emotions consist of a valence and an arousal component. Fear includes high levels of arousal and negative valence. Although CS + related arousal is reduced after extinction, its valence may still be negative. When a negatively valenced CS+ is presented in an arousing context (e.g., a reinstatement procedure), the combination of negative CS + valence and arousal again elicits fear as well as the old CS-US_{neg} association (Dirikx et al., 2004; Hermans et al, 2005).

Counterconditioning has been proposed as a better alternative to extinction, as it would at the same time decrease threat expectancy (due to the absence of the previous US_{neg}) and enhance evaluative learning (due to the presence of a novel US_{pos}; Baeyens, Eelen, Van den Bergh, & Crombez, 1989; De Houwer, Thomas, & Baeyens, 2001). Counterconditioning constitutes pairing the CS+ with a US of a valence that is opposite to the original US to form an alternative association. The modality between the original US_{neg} and the new US_{pos} can be different (De Houwer, 2011). For example, the original US_{neg} may be a shock, and the new US_{pos} a positive image. Thus, the negative emotions elicited by the original US_{neg} are replaced with positive ones elicited by the new US_{pos} (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Paunovic, 2003). In addition to merely creating an inhibitory CS-noUS_{neg} memory, as in extinction, the new association with a pleasant stimulus changes evaluative features, while at the same time reducing US_{neg} expectancy due to the absence of the US_{neg} (animals: Brooks, Hale, Nelson, & Bouton, 1995; Bouton & Peck, 1992; Peck & Bouton, 1990; humans: Engelhard, Leer, Lange, & Olatunji, 2014; Raes & De Raedt, 2012; animals and humans: Kerkhof et al., 2009). Given that counterconditioning changes both threat expectancy and negative CS + valence, it may be more effective than standard extinction in reducing return of fear.

Just three studies examined return of fear after counterconditioning and extinction in rodents, finding that the effects were comparable for spontaneous recovery (Bouton & Peck, 1992) and reinstatement (Brooks et al., 1995), and inferior for counterconditioning regarding renewal (Holmes, Leung, & Westbrook, 2016). The effects of counterconditioning on return of fear has not been studied in humans. The few studies with humans that compared counterconditioning with extinction, found that counterconditioning was more effective in reducing explicit and implicit CS + negative valence (Engelhard, Leer, Lange & Olatuji, 2014; Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011; Raes & De Raedt, 2012), and in reducing avoidance, heart rate responses and fear beliefs in children (Dunne & Askew, 2013; Reynolds, Field, & Askew, 2016). However, some studies failed to find beneficial effects of counterconditioning over extinction on US_{neg} expectancy (Meulders, Karsdorp, Claes, & Vlaeyen, 2015), fear or threat valence (De Jong, Vorage, & Van den Hout, 2000). One study found that counterconditioning had a superior effect than extinction on implicit but not explicit valence (Raes & De Raedt, 2012). Yet, as mentioned, none of these studies tested return of fear after counterconditioning.

The present study therefore aimed to investigate the effects of counterconditioning on return of fear (spontaneous recovery and reinstatement). We used a 3-day paradigm to optimize memory consolidation of each phase: acquisition on day 1, intervention (counterconditioning or extinction) on day 2, and spontaneous recovery and reinstatement phases on day 3. We expected that counterconditioning would result in less spontaneous recovery and less reinstatement than extinction. We also expected that CS + valence would be more positive after counterconditioning than after extinction.

2. Methods

2.1. Participants

Fourty participants (28 females, 12 males) were recruited at Leiden University and randomly assigned to the extinction (EXT; $n = 20$) or counderconditioning (CC) group ($n = 20$). Participants were all Dutch speaking and caucasian. Mean age was 19.73 years (SD = 1.74). Exclusion criteria were: self-reported color blindness or uncorrected vision, past or current psychiatric or neurological treatment, severe medical diseases (e.g. heart conditions), and substance abuse. Participants received course credits for participation.

2.2. Measures

State and trait anxiety. The Spielberger state-trait anxiety inventory (STAI) was used to assess state (20 questions) and trait anxiety (20 questions). The state anxiety scale (STAI-S) measures anxiety feelings at the moment, and the trait anxiety scale (STAI-T) measures anxiety feelings in general on a scale from 1 (not at all/almost never) to 4 (very much so/almost always; Spielberger, Gorsuch, & Lushene, 1970).

Spider Phobia. The Spider Phobia Questionnaire (SPQ) was used to assess spider phobia. The SPQ has 31 items with a true/false answering scale, that involve interactions with spiders; for example, “I avoid going to parks or camping trips because there may be spiders about”. Internal consistency for the SPQ is high, ranging from .83 to .90 (Klorman, Weerts, Hastings, Melamed, & Lang, 1974). Higher scores indicate greater spider fear.

2.3. Stimuli

Conditioned stimulus (CSs): Two images of fear relevant stimuli (i.e., spiders) selected from the international affective picture system (Lang, Bradley, & Cuthbert, 1999; numbers 1200 and 1201) served as CS+ and CS- (counterbalanced). These specific pictures have been used frequently in order to facilitate threat learning in acquisition and prevent fast fear extinction (e.g., Kindt & Soeter, 2013).

Negative unconditioned stimulus (US_{neg}): The unconditioned stimulus (US_{neg}) was a 2 ms electrocutaneous shock, delivered to the wrist of the non-dominant hand using the Grass S48 stimulator (Volt regulator: 150 V). The intensity of the shock was tailored to participants individually using a shock workup procedure (see below).

Positive unconditioned stimuli (US_{pos}). Four cartoon images[1] were selected from the internet and used as US_{pos}. These were rated as the most pleasant and arousing in a pilot study with 12 pictures ($N = 10$).

2.4. Ratings

Valence. Pictures (CS and US_{pos}) were rated on a scale from 0 (very unpleasant) to 10 (very pleasant); intensity of the US_{neg} (shock) was rated on a scale ranging from 0 (not at all discomforting) to 10 (extremely discomforting). CS valence was measured at the start of the experiment to rule out baseline group differences, and before and after the Intervention phase to test valence changes from pre to post EXT or CC. US_{neg} intensity was rated before the start of the experiment when the shock level was determined. US_{pos} valence was rated at the end of the experiment.

US_{neg} expectancy. US_{neg} expectancy was rated during each trial on a scale from −100 (convinced no shock will follow) to 100 (convinced a shock will follow), with 0 indicating “uncertain”.

[1] Available via the corresponding author.

2.5. Procedure

Participants were informed about the study and signed the consent form in the lab. They completed STAI-S, STAI-T and SPQ after which a shock workup procedure was started. The intensity of shock was gradually increased until the participant indicated the shock was "definitely uncomfortable but not painful". On each day, electrodermal activity and EMG devices were put in place, after which instructions were presented on the computer screen.[2] All phases took place on consecutive days to optimize consolidation of each phase. On day 1, participants started with CS valence ratings and then underwent Acquisition to learn which CS would be followed by US_{neg}. The Intervention phase, with pre and post CS valence ratings, took place on day 2. On day 3, participants engaged in Spontaneous recovery and Reinstatement phases, and gave US_{pos} ratings at the end. The spider pictures (CS + and CS-) were counterbalanced within each condition. On all days, the computer background was grey.

Day 1: Acquisition. The Acquisition phase consisted of 12 trials (6 CS+ and 6 CS-) with a 100% reinforcement rate. Each CS was presented for 8 s; US_{neg} expectancies were rated in the first 7 s. US_{neg} was delivered at CS off-set (startle probe at 7.5 s). Intertrial intervals (ITIs) were a blank grey screen and lasted 15, 20 or 25 s (M = 20 s). CSs were pseudo-randomized with a maximum of two sequential trials containing the same CS. All trials in all phases used this trial-format, except for the Intervention phase in the CC condition, in which the US_{pos} was presented. At the end of Acquisition, participants were instructed to memorize what they had learned.

Day 2: Intervention. Participants were instructed to recall what they had learned the first day. In the EXT condition both CSs were presented without US_{neg} (12 CS+ and 12 CS-). In the CC condition, the CS+ was now followed by the US_{pos}, with a 100% reinforcement rate. The US_{pos} was presented 500 ms after CS + off-set and remained on the screen for 3 s.

Day 3: Spontaneous recovery and Reinstatement. Spontaneous recovery was examined by presenting the CSs (1 CS+ and 1 CS-) without US_{neg} or US_{pos} on day 3. It was part of a Re-extinction procedure (i.e., another 11 CSs+ and 11 CSs-without reinforcement) in order to meet optimal conditions for establishing reinstatement. Nineteen seconds after this phase, three shocks were administered with a 18 s interval (without CSs) during which a grey computer screen was shown. Reinstatement trials started 18 s after the last shock. CSs were presented without US_{neg} or US_{pos} (6 CS+ and 6 CS-).

2.6. Statistical analysis

A 6 (Trial: 1–6) x 2 (Stimulus type: CS+, CS-) repeated measures (rm) ANOVA was used to test US_{neg} expectancy learning across conditions during Acquisition. A 12 (Trial: 1–12) x 2 (Stimulus type: CS+, CS-) rmANOVAs was done to test whether US_{neg} expectancy decreased across conditions during the Intervention phase. Condition (EXT, CC) was added to the model to test for potential pre-intervention group differences (Acquisition) and for explorative reasons (Intervention).

To test differences in spontaneous recovery between conditions, a 2 (Trial: trial 12 Intervention, trial 1 Spontaneous recovery) x 2 (Stimulus type: CS+, CS-) x 2 (Condition: EXT, CC) rmANOVA was done with US_{neg} expectancy as dependent variable. Similarly, Reinstatement was tested by a 2 (Trial: trial 12 Spontaneous recovery, trial 1 Reinstatement) x 2 (Stimulus type: CS+, CS-) x 2 (Condition: EXT, CC) rmANOVA.

Differential effects on CS Valence were tested by comparing conditions on changes in CS + valence from pre to post Intervention using a 2 (Time: pre and post Intervention) x 2 (Stimulus type: CS+, CS-) x 2 (Condition: EXT, CC) rmANOVA.

[2] Electromyographic activity (EMG, evoked by a 120 ms 100 dB tone through headphones) and electrodermal activity were also measured but due to technical problems, we were not able to analyse physiological data.

3. Results

3.1. Manipulation checks

Two participants dropped out after day 1 and were excluded from the analyses, leaving 20 participants for EXT and 18 for CC.

EXT and CC participants did not differ regarding STAI-S, STAI-T and SPQ (all $ts < 1.19$, all $ps > .243$). The US_{neg} was rated as moderately intense ($M = 6.17$, $SD = 1.77$), with no differences between EXT and CC groups ($t(36) = .651$, p = .519). Both CSs were rated as slightly unpleasant before Acquisition ($M = 4.80$ and 4.83, $SD = 2.15$ and 2.49), with no differences between EXT and CC groups (both $ts < |1.38|$, both ps > .193). The four US_{pos} were rated as pleasant (all $Ms > 7.33$, $SDs < 2.26$), and did not differ in pleasantness ($F(3, 111) = 9.12$, $p = .438$).

3.2. Acquisition and intervention across groups

Acquisition. US_{neg} expectancy ratings during all phases are depicted in Fig. 1. Across groups, expectancy learning was successful, as indicated by a significant Trial x Stimulus type interaction ($F(5, 185) = 23.63$, $p < .001$). Post hoc analyses showed that US_{neg} expectancy for CS+ and CS- did not differ on the first trial ($t(37) = .26$, $p = .800$), but did differ on the last trial of Acquisition ($t(37) = 8.79$, $p < .001$). From the first trial to the last trial of Acquisition US_{neg} expectancy increased for CS+ and decreased for CS- (both $ts > |4.12|$, both $ps < .001$). There were no differences between EXT and CC groups (main effect and all interactions with Condition: $F < 1.22$, $p > .303$).

Intervention. In the total sample, US_{neg} expectancy successfully declined in the Intervention phase, indicated by a significant Trial x Stimulus type interaction ($F(11, 407) = 19.28$, $p < .001$). Post hoc analyses showed that US expectancy for CS+ and CS- differed on the first trial ($t(37) = 6.12$, $p < .001$), but not on the last trial of the Intervention phase ($t(37) = 1.25$, p = .218). US_{neg} expectancy decreased from trial 1 to trial 12 for CS+ and CS- (both $Fs > 18.89$, both $ps < .001$). This decrease was larger for CS + than CS- ($t(37) = 6.08$, $p < .001$).

Interestingly, learning in the Intervention phase was different for EXT and CC groups, as shown by a significant Trial x Condition interaction ($F(11, 396) = 3.49$, $p < .001$, see Fig. 1), and a trend for the Trial x Stimulus type x Condition interaction ($F(11, 396) = 1.76$, $p = .058$). Post hoc analyses showed that decreases in US_{neg} expectancy from trial 1 to trial 12 were larger for CC relative to EXT for the CS+ ($F(11, 396) = 2.35$, $p = .008$) and smaller for the CS- ($F(11, 396) = 2.96$, $p = .001$). EXT and CC conditions did not differ on US_{neg} expectancy on trial 1 for CS+ ($t(36) = .13$, $p = .900$), so larger US_{neg} expectancy decreases for the CS + suggest faster expectancy learning for CC. Differences in US_{neg} expectancy for CS- may be explained by higher expectancy ratings on trial 1 in the EXT condition ($t(36) = 1.95$, $p = .059$).

Re-extinction. In the total sample, US_{neg} expectancy declined during Re-extinction, indicated by a significant Trial x Stimulus type interaction ($F(11, 396) = 4.41$, $p < .001$). Post hoc analyses showed that US expectancy for CS+ and CS- differed on the first trial ($t(37) = 3.95$, $p < .001$), but not on the last trial of Re-extinction ($t(37) = 1.33$, $p = .192$).

The Trial x Condition interaction was also significant ($F(11, 396) = 4.32$, $p < .001$), indicating larger decreases in US_{neg} expectancy in the EXT than in the CC condition (due to higher US_{neg} expectancy scores on the first trial; see Spontaneous recovery).

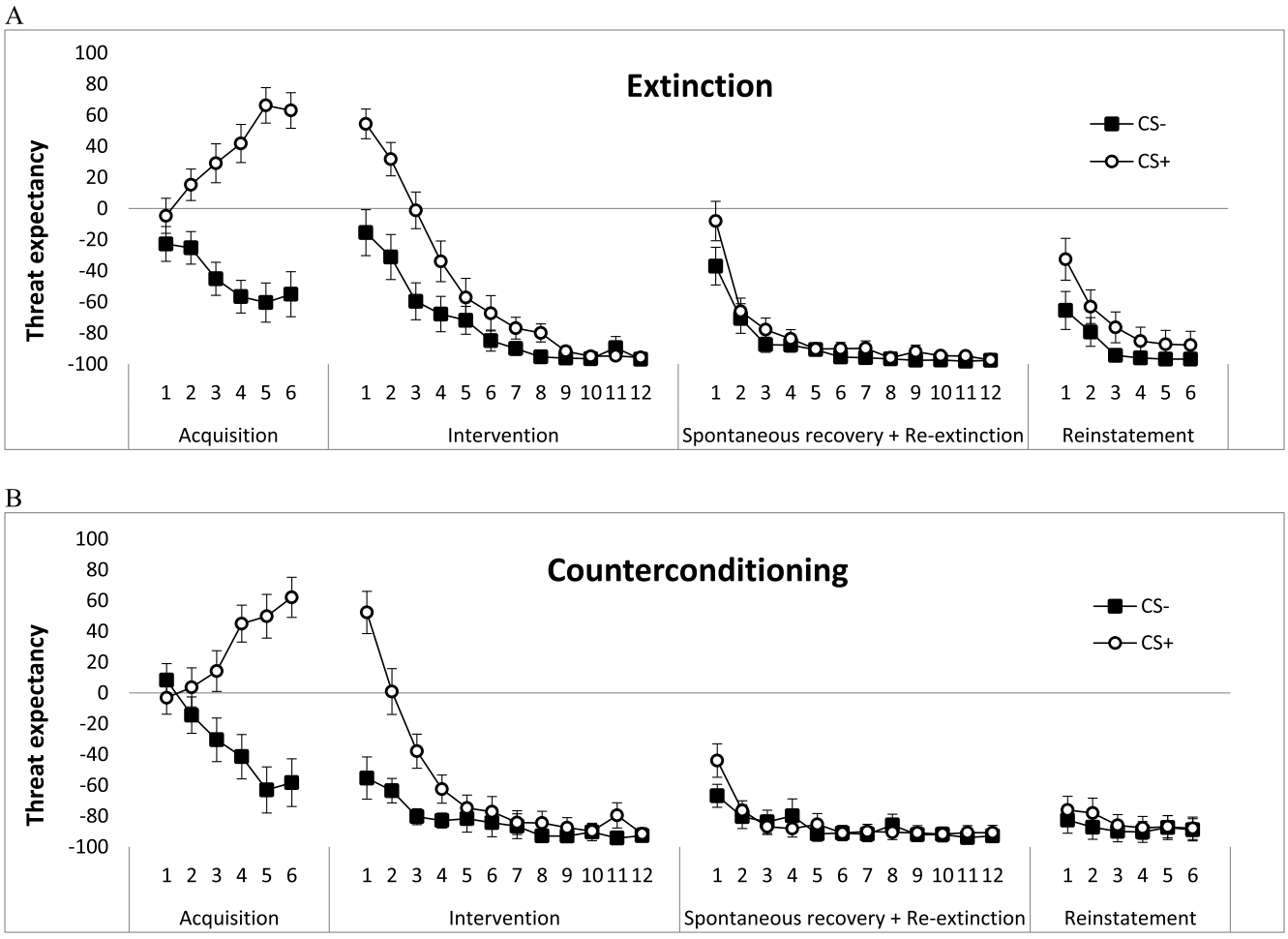

Fig. 1. US_{neg} expectancy during Acquisition, Intervention, Spontaneous recovery (and Re-extinction) and Reinstatement for Extinction (Panel A) and Counterconditioning (Panel B) groups (error bars represent the standard error of the mean).

3.3. Return of fear

Spontaneous recovery. The rmANOVA yielded a significant Trial x Stimulus type interaction. Post hoc analyses confirmed that US_{neg} expectancy increased from the last Intervention trial to the first Spontaneous recovery trial, with larger increases for CS + than CS- ($t(37) = 3.86$, $p < .001$), indicating differential spontaneous recovery in both conditions.

Crucially, the Trial x Condition interaction was also significant ($F(1,36) = 6.41$, $p = .016$). Post hoc analyses (using difference scores from the Intervention [last trial] to the Spontaneous recovery phase [first trial]) revealed larger increases in US_{neg} expectancy in the EXT group than in the CC group, for both CS+ and CS- (both $ts > 2.26$, both $ps < .027$; see Fig. 1). Stimulus type x Condition and Trial x Stimulus type x Condition interactions were not significant (both $Fs < .21$, both $ps > .646$). Thus, non-differential spontaneous recovery (for CS+ and CS-) was larger for EXT than CC participants.

Reinstatement. The rmANOVA yielded a significant Trial x Stimulus type interaction. Post hoc analyses confirmed that US_{neg} expectancy increased from the last Re-extinction trial to the first Reinstatement trial, with larger increases for CS + than CS- ($t(37) = 2.96$, $p = .005$), indicating differential reinstatement in both conditions.

Trial x Stimulus type x Condition and Trial x Condition interactions were significant ($F(1, 36) = 5.22$, $p = .028$ and $F(1, 36) = 7.41$, $p = .010$, respectively), and the Stimulus type x Condition interaction was marginally significant ($F(1, 36) = 3.84$, $p = .058$), suggesting that reinstatement for CS+ and CS- differed for EXT and CC groups (see Fig. 1). Post hoc analyses showed that both conditions showed reinstatement for CS+ ($|t|s > 2.36$, $ps < .04$), but it was larger for EXT than CC ($t(36) = 3.28$. $p = .002$). Reinstatement for CS- did not differ between conditions ($t(25.963) = 1.63$. $p = .114$).

3.4. CS + valence

There were no significant main or interaction effects for Condition, indicating that changes in subjective valence did not differ between groups (all $Fs < 2.43$, all $ps > .127$; see Fig. 2). The Time x Stimulus type interaction was significant ($F(1,36) = 10.12$, $p = .003$). Post hoc analyses revealed a trend in CS + being rated as more pleasant from pre to post Intervention ($t(37) = -1.74$, $p = .089$), but CS- valence did not change ($t(37) = 1.26$, $p = .217$).

4. Discussion

The present study was set up to compare the effects of standard extinction and counterconditioning on return of fear in humans. While extinction involves the repeated exposure of a CS without the aversive US it was previously paired with, counterconditioning additionally involves pairing the CS with a novel, positively valenced US. The results clearly showed that counterconditioning outperformed standard extinction by weakening the return of US_{neg} expectancy as measured during tests for spontaneous recovery and reinstatement. This suggests that creating a positive alternative association to a feared stimulus may

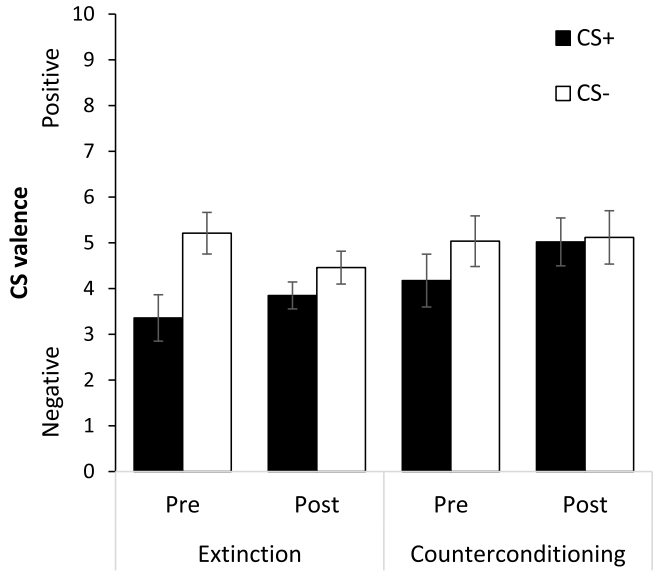

Fig. 2. Self-reported CS valence before and after the Intervention phase (error bars represent the standard error of the mean).

strengthen the fear extinction process in exposure treatment and thereby boost its long-term efficacy. Note that at the end of the Intervention phase, US_{neg} expectancy levels were similar for counterconditioning and extinction. Hence, benefits of counterconditioning-based interventions may only be found in clinical studies that include follow-up assessments, rather than mere post-treatment fear assessments (see also Craske et al., 2008). Although the current findings are based on self-reported expectancies of US_{neg} in a healthy population, they provide promising preliminary support for counterconditioning as a candidate for boosting long-term effects of exposure treatments for anxiety disorders. For example, Competitive Memory Training (COMET) could be a means to effectively establish new associations with positive representations in patients suffering from an anxiety disorder (Staring et al., 2016) and might better prevent relapse relative to mere exposure therapy.

We found that the counterconditioning procedure weakened return of fear without changing the conditioned negative value of the CS+. This was unexpected, because counterconditioning is proposed to affect evaluative learning and thus specifically reduce CS + valence (De Houwer, 2011; De Houwer et al., 2001). However, previous findings on the relevance of valence are mixed: Several studies found a more positive CS + valence after counterconditioning relative to extinction (e.g., Engelhard et al., 2014; Kerkhof et al., 2011), but other studies did not (Meulders et al., 2015), or only on an indirect measure (Raes & De Raedt, 2012). This may indicate that changes in CS + valence after counterconditioning are small or specific, showing at some levels (implicit or automatic behavior) but not at others (explicit or self-reports).

Alternatively, CS + valence may not be relevant for effects of counterconditioning on return of fear. For example, Luck and Lipp (2017) found reduced CS + negativity (after providing positive information) but not reduced reinstatement. Possibly, the added US (whether positive or neutral) enhances extinction by augmenting the prediction error (surprise), which is generally viewed as the motor for inhibitory learning during extinction (see also Dunsmoor, Campese, Ceceli, LeDoux, & Phelps, 2015). The use of positive stimuli may enhance this effect (Erez & Isen, 2002; Isen & Shalker, 1982; Kiefer, Schuch, Schenck, & Fiedler, 2007; Ludvik, Boschen, & Neumann, 2015). Coupling the CS + immediately to an novel outcome (US_{pos}; versus no outcome in the extinction group) may also have increased CS + salience, which may have enhanced learning. In sum, the exact mechanism for the return-of-fear reducing effects of counterconditioning is not known. Although a larger reduction of CS + negative valence is a theoretically plausible mechanism for reduced return of fear, it is not supported by our data and there are other candidate mechanisms. In our case, the humorous pictures may have been very adequate in introducing surprise as well as promoting a strong new association. Future studies might include the assessment of CS arousal (next to CS valence) to disentangle learning mechanisms during counterconditioning.

Interestingly, we found a faster reduction of US_{neg} expectancy in the intervention phase for participants in the counterconditioning group compared to the extincion group. This is surprising because an extinction procedure directly targets US_{neg} expectancy learning (Lovibond, 2004; Vansteenwegen et al., 2006). This faster reduction of US_{neg} expectancy might be explained by factors that were mentioned previously (prediction error and reduced uncertainty). This faster reduction is quite unique: Most interventions that strengthen extinction in the long-term, also make it slower in the short term (Craske et al., 2014). Counterconditioning may thus be faster but at the same time reduce return of fear. Nevertheless, US expectancy was similar in counterconditioning and extinction groups at the end of the intervention phase.

Our study has several limitations. First, self-report fear or physiological indicators of fear and anxiety were not included, nor were indirect measures of CS+ and CS- valence (e.g., affective priming, see Engelhard et al., 2014). Although our reinstatement and spontaneous recovery effects were similar, reinstatement effects should be replicated without a prior re-extinction procedure, because our counterconditioning group received counterconditioning and extinction and the extinction group received extinction twice. Future research may also include larger samples and long-term assessments (e.g., after 1 week). Our pictures may not have been the optimal stimuli to change CS + valence. The application of other -perhaps idiosyncratic-positive USs may also be useful in order to test the optimal conditions for counterconditioning. For example, film clips elicit specific emotional responses (Arnaudova & Hagenaars, 2017), but their use as counterstimuli is unknown. Relatedly, our findings may point towards the use of humor during exposure therapy, as it strongly promotes surprise (prediction error) as well as the formation of new associations. Note that spontaneous recovery and reinstatement can be interpreted as contextual learning (with the visit to the lab and the unsignaled US_{neg} as reminders of the acquisition memory, respectively), so the findings are likely to expand to reacquisition and renewal. All in all, counterconditioning procedures seem to be a promising alternative approach, but mediating and moderating factors largely merit further exploration. Finally, counterconditioning reduced spontaneous recovery for both the CS+ and the CS-. Future studies are needed to examine the specificity of the effects of the positive US.

In conclusion, although there is a need for further testing, this is an important first step in exploring counterconditioning as an approach to reduce treatment relapse. We provide preliminary evidence that expectancy learning does not differ for counterconditioning and extinction, but return of fear is lower after counterconditioning. Given that relapse rates are high after exposure therapy, this is a promising finding. Moreover, our counterconditioning procedure resulted in faster extinction rates. Interestingly, counterconditioning and extinction did not differ in their effect on CS + valence, so changes in CS + valence did not drive the weakened return of fear effects. Rather, novelty of the US_{pos} may have facilitated extinction learning. The findings are highly relevant clinically and contribute to theories of expectancy and evaluative learning.

Author contributions

Muriel Hagenaars developed the study concept and Muriel Hagenaars and Bram Vervliet contributed to the study design. Muriel Hagenaars performed the data analysis and interpretation. Sahaj Kang drafted the paper, and Bram Vervliet, Iris Engelhard, Evi-Anne van Dis and Muriel Hagenaars provided critical revisions. All authors approved

the final version of the paper for submission.

Declaration of conflicting interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Acknowledgements

Iris Engelhard is supported with a Vici grant (453-15-005) from the Netherlands Organization for Scientific Research (NWO).

References

Arnaudova, I., & Hagenaars, M. A. (2017). Lights … action: Comparison of trauma films for use in the trauma film paradigm. *Behavior Research and Therapy, 93*, 67–77. https://doi.org/10.1016/j.brat.2017.02.007.

Baeyens, F., Díaz, E., & Ruiz, G. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*, 245–268. https://doi.org/10.1080/02699930441000300.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective-evaluative value: Conservative but not unchangeable. *Behavior Research and Therapy, 27*, 279–287. https://doi.org/10.1016/0005-7967(89)90047-8.

Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychological Medicine, 43*, 897–910. https://doi.org/10.1017/S003329171200147X.

Boschen, M. J., Neumann, D. L., & Waters, A. M. (2009). Relapse of successfully treated anxiety and fear: Theoretical issues and recommendations for clinical practice. *Australian and New Zealand Journal of Psychiatry, 43*, 89–100. https://doi.org/10.1080/00048670802607154.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). *Learning & Behavior, 20*, 313–321. https://doi.org/10.3758/BF03197954.

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior, 23*, 383–390. https://doi.org/10.3758/BF03198938.

Craske, M. G. (1999). *Anxiety disorders: Psychological approaches to theory and treatment.* Boulder, CO: Westview Press.

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46*, 5–27. https://doi.org/10.1016/j.brat.2007.10.003.

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

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry, 15*, 245–258. https://doi.org/10.1002/wps.20346.

De Houwer, J. (2011). Evaluative conditioning: A review of functional knowledge and mental process theories. In T. R. Schachtman, & S. Reilly (Eds.). *Associative learning and conditioning theory* (pp. 399–416). Oxford University Press.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853–869. https://doi.org/10.1037//0033-2909.127.6.853.

De Jong, P. J., Vorage, I., & Van den Hout, M. A. (2000). Counterconditioning in the treatment of spider phobia: Effects on disgust, fear and valence. *Behaviour Research and Therapy, 38*, 1055–1069. https://doi.org/10.1016/S0005-7967(99)00135-7.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*, 549–554. https://doi.org/10.1101/lm.78004.

Dunne, G., & Askew, C. (2013). Vicarious learning and unlearning of fear in childhood via mother and stranger models. *Emotion, 13*, 974–980. https://doi.org/10.1037/a0032994.

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*, 203–209.

Engelhard, I. M., Leer, A., Lange, E., & Olatunji, B. O. (2014). Shaking that icky feeling: Effects of extinction and counterconditioning on disgust-related evaluative learning. *Behavior Therapy, 45*, 708–719. https://doi.org/10.1016/j.beth.2014.04.003.

Erez, A., & Isen, A. M. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology, 87*, 1055–1067. https://doi.org/10.1037/0021-9010.87.6.1055.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*, 533–551. https://doi.org/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, 130*, 39–42. https://doi.org/10.1037/a0018916.

Holmes, N. M., Leung, H. T., & Westbrook, R. F. (2016). Counterconditioned fear responses exhibit greater renewal than extinguished fear responses. *Learning & Memory, 23*(4), 141–150. https://doi.org/10.1101/lm.040659.115.

Isen, A. M., & Shalker, T. E. (1982). The effect of feeling state on evaluation of positive, neutral, and negative stimuli: When you "accentuate the positive", do you "eliminate the negative"? *Social Psychology Quarterly, 45*, 58–63. https://doi.org/10.2307/3033676.

Jacoby, R. J., & Abramowitz, J. S. (2016). Inhibitory learning approaches to exposure therapy: A critical review and translation to obsessive-compulsive disorder. *Clinical Psychology Review, 49*, 28–40. https://doi.org/10.1016/j.cpr.2016.07.001.

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology, 58*, 31–38. https://doi.org/10.1027/1618-3169/a000063.

Kerkhof, I., Vansteenwegen, D., Beckers, T., Dirikx, T., Baeyens, F., D'Hooge, R., et al. (2009). The role of negative affective valence in return of fear. In H. D. Friedman, & P. K. Revera (Eds.). *Abnormal Psychology: New research* (pp. 121–138). .

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the national comorbidity survey. *Archives of General Psychiatry, 51*, 8–19. https://doi.org/10.1001/archpsyc.1994.03950010008002.

Kiefer, M., Schuch, S., Schenck, W., & Fiedler, K. (2007). Mood states modulate activity in semantic brain areas during emotional word encoding. *Cerebral Cortex, 17*, 1516–1530. https://doi.org/10.1093/cercor/bhl062.

Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology, 92*, 43–50. https://doi.org/10.1016/j.biopsycho.2011.09.016.

Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific fear questionnaires. *Behavior Therapy, 5*, 401–409. https://doi.org/10.1016/S0005-7894(74)80008-0.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist, 50*, 372–385. https://doi.org/10.1037/0003-066X.50.5.372.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings.* Gainesville, Fl: The Center for Research in Psychophysiology, University of Florida. Retrieved from https://www2.unifesp.br/dpsicobio/adap/instructions.pdf.

Lommen, M. J. J., Engelhard, I. M., Sijbrandij, M., van den Hout, M. A., & Hermans, D. (2013). Pre-trauma individual differences in extinction learning predict posttraumatic stress. *Behaviour Research and Therapy, 51*, 63–67. https://doi.org/10.1016/j.brat.2012.11.004.

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory, 11*, 495–500. https://doi.org/10.1101/lm.79604.

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. https://doi.org/10.1016/j.brat.2015.01.001.

Luck, C. C., & Lipp, O. V. (2017). Verbal instructions targeting valence alter negative conditional stimulus evaluations (but do not affect reinstatement rates). *Cognition & Emotion, 31*, 1–20. https://doi.org/10.1080/02699931.2017.1280449.

Ludvik, D., Boschen, M. J., & Neumann, D. L. (2015). Effective behavioural strategies for reducing disgust in contamination-related OCD: A review. *Clinical Psychology Review, 42*, 116–129. https://doi.org/10.1016/j.cpr.2015.07.001.

Meulders, A., Karsdorp, P. A., Claes, N., & Vlaeyen, J. W. (2015). Comparing counter-conditioning and extinction as methods to reduce fear of movement-related pain. *The Journal of Pain, 16*, 1353–1365. https://doi.org/10.1016/j.jpain.2015.09.007.

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *The Journal of Nervous and Mental Disease, 195*, 521–531. https://doi.org/10.1097/01.nmd.0000253843.70149.9a.

Olatunji, B. O., Forsyth, J. P., & Cherian, A. (2007). Evaluative differential conditioning of disgust: A sticky form of relational learning that is resistant to extinction. *Journal of Anxiety Disorders, 21*, 820–834. https://doi.org/10.1016/j.janxdis.2006.11.004.

Paunović, N. (2003). Prolonged exposure counterconditioning as a treatment for chronic posttraumatic stress disorder. *Journal of Anxiety Disorders, 17*, 479–499. https://doi.org/10.1016/S0887-6185(02)00233-5.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Oxford University Press.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21*, 1–31. https://doi.org/10.1016/0023-9690(90)90002-6.

Raes, A. K., & De Raedt, R. (2012). The effect of counterconditioning on evaluative responses and harm expectancy in a fear conditioning paradigm. *Behavior Therapy, 43*, 757–767. https://doi.org/10.1016/j.beth.2012.03.012.

Reynolds, G., Field, A. P., & Askew, C. (2016). Reductions in children's vicariously learnt avoidance and heart rate responses using positive modeling. *Journal of Clinical Child and Adolescent Psychology, 23*, 1–14. https://doi.org/10.1080/15374416.2016.1138410.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the stait-trait anxiety inventory.* Palo Alto, CA: Consulting Psychologists Press.

Staring, A. B. P., van den Berg, D. P. G., Cath, D. C., Schoorl, M., Engelhard, I. M., & Korrelboom, C. W. (2016). Self-esteem treatment in anxiety: A randomized controlled crossover trial of eye movement desensitization and Reprocessing (EMDR) versus competitive memory training (COMET) in patients with anxiety disorders. *Behaviour Research and Therapy, 82*, 11–20. https://doi.org/10.1016/j.brat.2016.04.002.

VanElzakker, M. B., Dahlgren, M. K., Davis, F. C., Dubois, S., & Shin, L. M. (2014). From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiology of Learning and Memory, 113*, 3–18. https://doi.org/10.1016/j.nlm.2013.11.014.

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental*

Psychology: Animal Behavior Processes, 32, 71–79. https://doi.org/10.1037/0097-7403.32.1.71.

Vasey, M. W., Harbaugh, C. N., Buffington, A. G., Jones, C. R., & Fazio, R. H. (2012). Predicting return of fear following exposure therapy with an implicit measure of attitudes. *Behaviour Research and Therapy, 50*, 767–774. https://doi.org/10.1016/j.brat.2012.08.007.

Vervliet, B. (2008). Learning and memory in conditioned fear extinction: Effects of D-cycloserine. *Acta Psychologica, 127*, 601–613. https://doi.org/10.1016/j.actpsy.2007.07.001.

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9*, 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542.

Yonkers, K. A., Bruce, S. E., Dyck, I. R., & Keller, M. B. (2003). Chronicity, relapse, and illness-course of panic disorder, social phobia, and generalized anxiety disorder: Findings in men and women from 8 years of follow-up. *Depression and Anxiety, 17*, 173–179. https://doi.org/10.1002/da.10106.

Zbozinek, T. D., Hermans, D., Prenoveau, J. M., Liao, B., & Craske, M. G. (2015). Post-extinction conditional stimulus valence predicts reinstatement fear: Relevance for long-term outcomes of exposure therapy. *Cognition & Emotion, 29*, 654–667. https://doi.org/10.1080/02699931.2014.930421.