Available online at www.sciencedirect.com
ScienceDirect
Behavior Therapy 55 (2024) 361–375
Behavior Therapy
www.elsevier.com/locate/bt

Tackling Costly Fearful Avoidance Using Pavlovian Counterconditioning

Anneloes M. Hulsman [1]
Donders Centre for Cognitive Neuroimaging, Radboud University
Behavioural Science Institute, Radboud University

Iris van de Pavert [1]
Donders Centre for Cognitive Neuroimaging, Radboud University
Behavioural Science Institute, Radboud University
KU Leuven

Karin Roelofs
Floris Klumpers
Donders Centre for Cognitive Neuroimaging, Radboud University
Behavioural Science Institute, Radboud University

Avoidance behavior constitutes a major transdiagnostic symptom that exacerbates anxiety. It hampers fear extinction and predicts poor therapy-outcome. Pavlovian counterconditioning with a reward could alleviate avoidance better than traditional extinction by reducing negative valence of the feared situation. However, previous studies are scarce and did not consider that pathological avoidance is often costly and typically evolves from an approach-avoidance conflict. Therefore, we used an approach-avoidance conflict paradigm to model effects of counterconditioning on costly avoidance (i.e., avoidance that leads to missing out on rewards). Results from our preregistered Bayesian Mixed Model analyses in 51 healthy participants (43 females) indicated that counterconditioning was more effective in reducing negative valuation and decreasing costly avoidance than traditional extinction. This study supports application of a simple counterconditioning technique, shows that its efficacy transfers to more complex avoidance situations, and suggests treatment may benefit from increasing reward drive in combination with extinction to overcome avoidance. Application in a clinical sample is a necessary next step to assess clinical utility of counterconditioning.

Keywords: avoidance; counterconditioning; anxiety; threat; reward

We thank Lukas Schmid and Nele Kettler for their valuable help in data collection.
This work was funded by a Netherlands Organization for Scientific Research (NWO) Research Talent Grant #06-18-540 awarded to Anneloes Hulsman, an NWO VICI Grant #453-12-0010 awarded to Karin Roelofs, and a European Research Council (ERC) Starting Grant #ERC_StG2012_313749 awarded to Karin Roelofs. Further, this paper was supported by funding from the Fonds Wetenschappelijk Onderzoek (FWO) awarded to Ilse van Diest (PI), Lukas van Oudenhove, and Jonas Zaman #ZKD3921/G07198N.
The authors declare there are no conflicts of interest.
Address correspondence to Floris Klumpers, Experimental Psychopathology & Treatment, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. e-mail: Floris.Klumpers@ru.nl.
[1] Equal author contribution.

0005-7894/

ANXIETY DISORDERS are among the most prevalent and costly disorders worldwide (Craske & Stein, 2016; Kessler et al., 2010). Although exposure-based cognitive-behavioral therapy (CBT) interventions have high success rates of around 50–60% (Loerinc et al., 2015), a considerable amount of patients profit insufficiently and/or show a return of symptoms. Avoidance behavior is a key transdiagnostic factor that not only severely impairs patients' everyday functioning but also reduces therapy efficacy (Craske et al., 2017; Kashdan et al., 2008; Pittig et al., 2018). Avoidance counteracts the effect of exposure therapies by protection from extinction (Lovibond et al., 2009; Rattel et al., 2017; Vervliet & Indekeu, 2015) and predicts clinical outcomes better than anxiety (Castriotta, 2013; Hendriks et al., 2013; Pittig et al., 2015). Particularly in patients, avoidance is often not only associated with avoiding threats and reducing anxiety, but also leads to significant costs (e.g., inability to sustain employment or maintain relationships). Therefore, treatments to tackle this "costly" avoidance behavior more successfully are sorely needed. In this study we used an approach-avoidance conflict paradigm to model effects of counterconditioning on so-called "costly avoidance" and compared it to "extinction alone" in reducing negative valuation and avoidance.

A promising avenue to innovate anxiety treatment has been shifting focus from mere exposure to actively changing the evaluation of the feared situation (Baeyens et al., 1992; De Houwer et al., 2000; Dirikx et al., 2004; Engelhard et al., 2014; Hermans et al., 2002). Traditional exposure therapy predominantly focuses on changing negative outcome expectations. This change is triggered by extinction procedures, where individuals are repeatedly exposed to the feared situation without the expected negative outcome to generate a newly learned safety association that inhibits the original fear response (Bouton, 2004; Pittig et al., 2018). However, whereas extinction often effectively changes the expectancy of the negative situation, it is typically less effective in changing the disliking of that situation (Gawronski et al., 2015; Kerkhof et al., 2011; van Dis et al., 2019; Vansteenwegen et al., 2006). Since the negative evaluation associated with feared situations predicts the return of fear and other debilitating symptoms such as avoidance, dampening remaining negative feelings could be key in enhancing treatment success (Dirikx et al., 2004).

An effective way to manipulate the evaluation of a stimulus is counterconditioning (CC). Ever since the seminal studies by Mary Cover Jones (Jones, 1924), the CC procedure is used to reverse negative valence of a disliked stimulus by associating it with a pleasant stimulus through Pavlovian conditioning. Whereas groundbreaking early studies provided clear evidence of its clinical potential (Bandura, 1961), counterconditioning has ever since been studied considerably less than extinction (see Keller et al., 2020, for review). This is surprising because indeed CC-based interventions also in more recent studies resulted in robust improvements in a variety of outcomes, sometimes even in direct comparison to traditional extinction-based interventions (Gatzounis et al., 2021; Hendrikx et al., 2021; Kang et al., 2018; Keller & Dunsmoor, 2020; Kerkhof et al., 2011; Newall et al., 2017; Raes & De Raedt, 2012; Reynolds et al., 2018; van Dis et al., 2019; Zenses et al., 2021).

Previous studies have suggested that CC may also help overcome avoidance (De Jong et al., 2000; Hendrikx et al., 2021; Jones, 1924; Reynolds et al., 2018) and thereby constitutes an attractive path for improving success in treating anxiety-related psychopathology (Dymond, 2019). However, these previous studies did not account for the fact that in psychopathology, avoidance is often costly, evolving from an approach-avoidance conflict caused by a mixed-outcome prospect (Bach, 2015; Pittig et al., 2018). That is, direct benefits of avoidance, such as reduced threat exposure, increased feelings of control (Boeke et al., 2017) and relief (Vervliet et al., 2017) usually co-occur in anxiety patients with significant negative long-term consequences. Indeed, what defines pathological avoidance is that it can lead to profound social isolation, impairments in professional life, and missed opportunities for personal development (Kashdan et al., 2008). Therefore, in severe anxiety not only do threat appraisals inform the decision to avoid but also appraisal of the foregone rewards typically weigh in to avoidance decisions (Aupperle et al., 2015; Ball & Gunaydin, 2022; Krypotos et al., 2018; Pittig, Brand, et al., 2014). To date, the vast majority of empirical and theoretical work, including work on CC, has however focused on how anxious individuals process threats, leaving the potential influence of reward processing in avoidance and anxiety understudied (Arnaudova et al., 2017; Emerson, 2018; Pittig et al., 2018). It therefore remains critically unclear whether CC could effectively reduce avoidance during the more representative and ecologically valid situation of approach-avoidance conflict generated by mixed-outcome prospects.

We investigated the effects of CC on costly fearful avoidance behavior, as assessed by our Fearful Avoidance Task (FAT). This instructed fear task probes instrumental avoidance of aversive electrical stimulation at the expense of monetary rewards, and shows external validity in that higher real-life anxiety predicts stronger avoidance behavior in the task (Hulsman et al., 2021). We hypothesized that CC is successful in reducing negative valence as well as avoidance under mixed outcome prospects. Additionally, we assessed whether CC produces changes in basic physiological emotional stance indexed by eye-blink startle. Amplification of the eye-blink startle reflex is a cross-species measure for defensive responses (Grillon & Baas, 2003; Lang et al., 1990) and could therefore be used to track changes in basis defensive physiology after the counterconditioning.

Methods

TRANSPARENCY AND OPENNESS

All hypotheses, methods, and analyses were preregistered on the Open Science Framework (https://osf.io/p8gyn). After the first preregistered analyses confirmed our hypotheses (see preprint v1: https://psyarxiv.com/h85ay), it became apparent that our analyses may be underpowered. Although we found the expected key results, to confirm the reliability of our findings we amended our preregistration with an expanded sample size based on power calculations. Importantly, all key findings from the original preregistration were replicated in the expanded sample. All data and code are available at the Donders Institute for Brain, Cognition and Behaviour repository at https://doi.org/10.34973/1tvv-wa16.

PARTICIPANTS

Power calculations indicated a sample of $N \geq 45$ would be required to detect the critical effect of counterconditioning on avoidance (for >80% power as indicated by formal simulations using SIMR; Green & Macleod, 2016). Participants were recruited through a university recruitment platform, primarily comprising university students. In total, there were 57 participants (43 females, $M_{age} = 22.68$, $SD_{age} = 4.15$). However, one participant was unwilling to complete the test session due to their aversive reaction to the startle sound. In addition, six participants were excluded from the analyses based on our preregistered a priori exclusion criteria (see supplement), leading to a final sample of 51 participants (39 females, $M_{age} = 22.88$, $SD_{age} = 4.26$). Importantly, we confirmed that excluding these participants did not affect the main results. Further exclusion criteria included lifetime diagnosis of psychological, cardiological, or neurological disorders, non-Dutch language comprehension, hearing loss, insufficient (corrected) vision, use of psychoactive medication, epilepsy, and pregnancy. Additionally, participants were instructed to refrain from eating 3 hours prior to the experiment to enhance the effect of counterconditioning (Andreatta & Pauli, 2015). At the start of the experiment, all participants reported to have followed this instruction. Participants were financially compensated with €10 and received an additional bonus between €0-€5 depending on their performance in the FAT (described below). This study was carried out in compliance with the declaration of Helsinki and approved by a local medical-ethical committee (CMO Arnhem-Nijmegen). All participants provided written informed consent.

GENERAL PROCEDURE

First, electrodes for physiology recordings and shock administration were applied. Subsequently, participants underwent *titration procedures* to determine the threat and reward levels for the FAT at an individual level. This was followed by *CS valence ratings* in which participants rated the pleasantness of each CS-type included in the FAT prior to conditioning. Next, *baseline startle measurements* were taken to correct for individual differences in baseline startle response and to habituate subjects to the procedure. Subsequently, participants performed the FAT for the first time to assess baseline approach-avoidance behavior prior to the intervention (i.e., from here on referred to as acquisition). The FAT was followed by a second CS valence rating. Next, the *counterconditioning and extinction intervention* took place, in which one of the CS-types of the FAT was counterconditioned (CS + CC) and another was extinguished (CS + EXT). After the intervention, another CS valence rating was taken and participants completed the FAT a second time to assess the impact of the counterconditioning and extinction intervention on avoidance (i.e., from here on referred to as test). The study ended with a last CS valence rating and baseline startle measurement. The overall procedure is summarized in Figure 1.

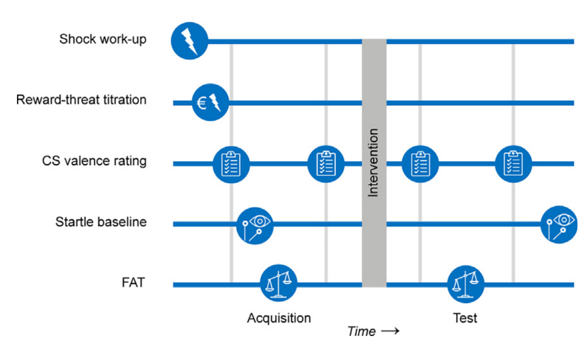

FIGURE 1 Study timeline. FAT = Fearful Avoidance Task. Intervention = counterconditioning and extinction intervention.

TITRATION PROCEDURES

Shock Work-Up

Participants underwent a standardized shock work-up procedure to set shock intensity to an individual level that was maximally uncomfortable without being painful (Klumpers et al., 2010). See supplement for full details of this procedure.

Reward-Threat Titration

Reward levels (low/high) of the FAT were titrated relative to the shock at an individual level to (a) ensure they had similar subjective value across participants and (b) ensure it would lead to sufficient variation in avoidance. In this procedure, participants were asked whether they would be willing to risk receiving electrical stimulation for different amounts of money (range: €0.20–€10.00, semirandom order). The participant's indifference value ($M = 1.44$, $SD = 1.97$) was subsequently used to calculate low and high reward values for the ensuing FAT. See supplement for full details of this procedure.

VALENCE RATINGS

The FAT consisted of four different CS-types. To investigate whether the counterconditioning and extinction intervention had an effect on the subjective valence of the CS-types, participants were asked to rate all four CS-types that were shown during the FAT on a scale ranging from −100 (*Very unpleasant*) to 100 (*Very pleasant*), where 0 was a neutral score. To account for order effects, participants rated the CS-types in a random order. Ratings were taken at four different timepoints throughout the experiment (see Figure 1).

FEARFUL AVOIDANCE TASK (FAT)

The FAT (Hulsman et al., 2021) was used to assess approach-avoidance behavior under competing reward and threat prospects. Participants received on-screen instructions for the task. Each trial in the FAT consisted of a cascade of three phases: i.e., the offer, anticipation, and outcome phase (Figure 2A). In the *offer* phase, a monetary amount and avatar were presented. Participants were instructed that the amount indicated the potential reward that could be earned on this trial, whereas the avatar indicated the threat level. There were two reward levels (low/high), which were set 40, 45, and 50% below and above the indifference point, respectively. The jitter within each reward level was intended to create variability over trials to keep subjects attentive. Threat levels (low/high) were indicated with four different avatars (representing the CS-types): one signaling absence of threat of shock (CS−) and three signaling threat of shock (CS + CC, CS + EXT, CS + Control, see Figure 2B). To enhance threat acquisition, participants were also provided with explicit instructions regarding the threat levels associated with the stimuli. Achieving robust acquisition was essential for the current study, as a lack of avoidance acquisition would impede proper assessment of the effects of counterconditioning due to floor effects. Thus, participants received explicit verbal and visual instructions on which avatar was associated with low threat (CS−, no shock) and which avatars were associated with high threat (CS+, threat of shock). Unbeknownst to the participants, one of these high threat avatars was predetermined to be later counterconditioned (CS + CC), one to be extinguished (CS + EXT), and one to serve as a nonextinguished control (CS + Control) during the subsequent intervention. Reward and CS-types were combined in a full factorial manner with each combination occurring equally often. After offer phase onset, participants had 4 seconds to decide to approach or avoid by pulling the joystick towards them or pushing the joystick away from them, respectively. As confirmation of a timely response (<4 s), a white square appeared around the reward. Next, a 2–4 s *anticipation* phase started. In all trials, startle probes (50 ms white noise, ±105 dB) were presented 1500 ms after the response to measure the eye-blink startle response during outcome anticipation. After the anticipation phase, participants received the *outcome* of their decision. The outcome could be positive (avatar offered money), negative (avatar drew a gun), or neutral (avatar disappeared; Figure 2B). Participants were explicitly instructed on the outcome probabilities: for approach = 40/40% chance of receiving the positive/negative outcome and a 20% chance on receiving the neutral outcome; for avoid = 10/10% chance of receiving the positive/negative outcome and an 80% chance of receiving the neutral outcome (Figure 2C). Only for high threat trials (CS+), electric shocks were delivered 50ms after visual feedback of the negative outcome (see supplement for shock settings). If participants did not give a timely response (<4 s), they always received the negative outcome. After receiving the outcome, a 5–7 s inter-trial interval occurred in which a fixation cross was presented. The FAT was administered twice: before (acquisition) and after (test) the counterconditioning and extinction intervention. For both administrations, the FAT consisted of 6 blocks, with each block comprising all eight 8 trial types, leading to a total of 48 trials per FAT administration (i.e., 12 trials for each CS-type). The order of the reward levels and CS-types was pseudorandomized so that

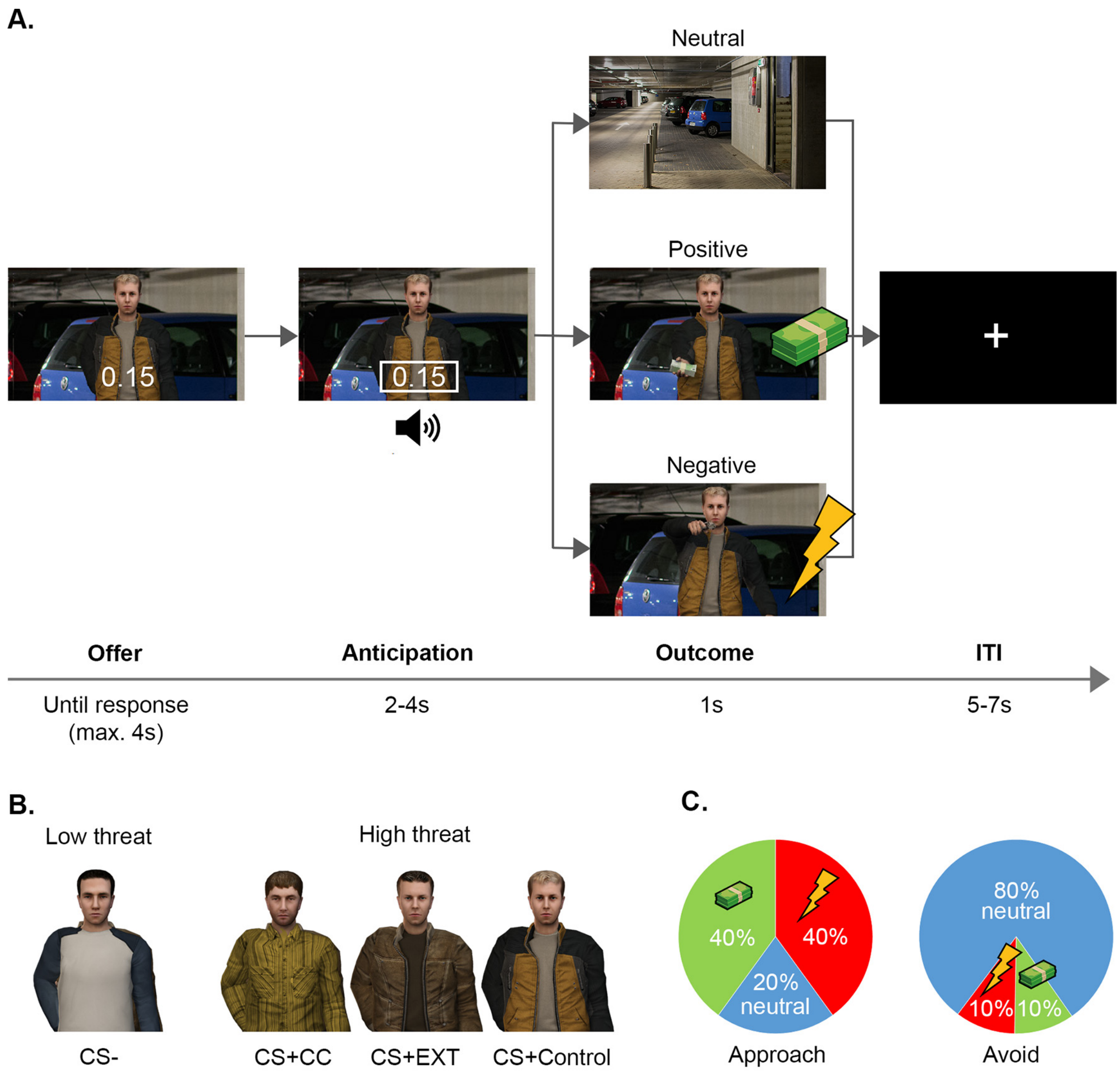

FIGURE 2 Fearful Avoidance Task (FAT) **A.** Example trial. In the offer phase, a monetary reward (low/high relative to the participant's indifference point) and an avatar indicating the shock threat level (low/high) were presented. The participant had to decide to approach or avoid by pulling the joystick towards themselves or away from them, respectively. After an anticipation phase, participants received a neutral, positive or negative outcome. Each trial ended with an inter-trial interval (ITI) during which a fixation cross was presented. **B.** One avatar indicated safety from receiving electrical shock (CS−); three avatars indicated threat of shock and were subsequently counterconditioned, extinguished or not shown during the intervention following the FAT (CS + CC, CS + EXT, CS + Control). **C.** Overview of the outcome probabilities for approach decisions (left panel) and avoidance decisions (right panel).

both could not be repeated more than once within a block. Before the first administration of the FAT, participants completed a practice session consisting of 8 trials (one of each trial type) without the risk of receiving an electric shock. The experimenter monitored their responses and verified that participants understood the task. To determine the bonus pay-out on the FAT, participants selected six random numbers prior to the first FAT. Each number was linked to a specific trial number by a mathematical formula that was unknown to the participants when they selected the numbers. At the end, participants received the total amount of rewards received on these six trials as a bonus with a maximum of €5.

COUNTERCONDITIONING AND EXTINCTION INTERVENTION

In this study, we aimed to compare the effect of the counterconditioning intervention directly with the effect of the extinction intervention (Keller et al., 2020). We based our counterconditioning intervention on studies by Kerkhof et al. (2011), who showed that a picture-taste paradigm was most effective in changing the valence of a negative stimulus. During the intervention, one avatar previously associated with risk of shock in the FAT was counterconditioned with a rewarding snack (CS + CC), while another was repeatedly presented without shock (extinguished) yet never paired with reward (CS + EXT). The third high threat avatar of the FAT was not included in the intervention to assess effects of repeating the FAT (CS + Control). Participants did not receive explicit instructions regarding the association between avatar and intervention type (i.e., which avatar would serve as CS + CC, CS + EXT, or CS + Control). Before the start of the intervention, a plate with small 1x1 cm cookies was placed in front of the participants, together with a glass of water. The cookies were made using 200 grams of oats, 150 grams of butter, and 100 grams of sugar. The plate and cookies were removed right after the intervention. Each trial started by showing the avatar (CS + CC or CS + EXT). After one second, a prompt appeared instructing participants to eat a cookie and drink a sip of water (CC trials) or to wait to continue (EXT trials; see Figure 3). Meanwhile, the avatar remained on the screen. In both conditions participants were able to continue after 20 s by pressing enter. After each trial there was a 2 s fixation cross. In total, participants completed 20 trials, 10 trials for each type of intervention (CC vs. EXT). The trial order was pseudorandomized and CS-types were never repeated more than once. Associations between specific avatar identities and conditions were counterbalanced across participants.

PHYSIOLOGICAL RECORDING

In this study the eye-blink startle and post-auricular reflex were measured. See supplement for full details of physiological recording and processing as well as post-auricular reflex results.

ANALYSES

Statistical analyses were carried out in R (Version 3.6.2; R Core team, 2019) in RStudio (Version 1.2.5033; RStudio Inc., 2019). Bayesian mixed effects models were used to estimate the effects for all dependent variables, using the package brms (Version 2.10.0; Bürkner, 2017). Bayesian mixed effects models are well-suited to account for data dependence, making them less prone to type I errors than conventional statistical analyses, such as ANOVAs (Aarts et al., 2014). In addition, Gelman et al. (2012) recommend using Bayesian mixed effects models in settings where multiple comparisons arise. They show that Bayesian mixed effects models shift point estimates and their corresponding intervals closer to each other (i.e., perform partial pooling), resulting in more reliable estimates. All models included a maximum random-effects structure. This consisted of a random intercept per participants and random slopes for the within-subject effects of the same variables that were fixed effects in all the models described below (both main effects and interactions). All continuous predictors were standardized and all categorical predictors were coded using sum-to-

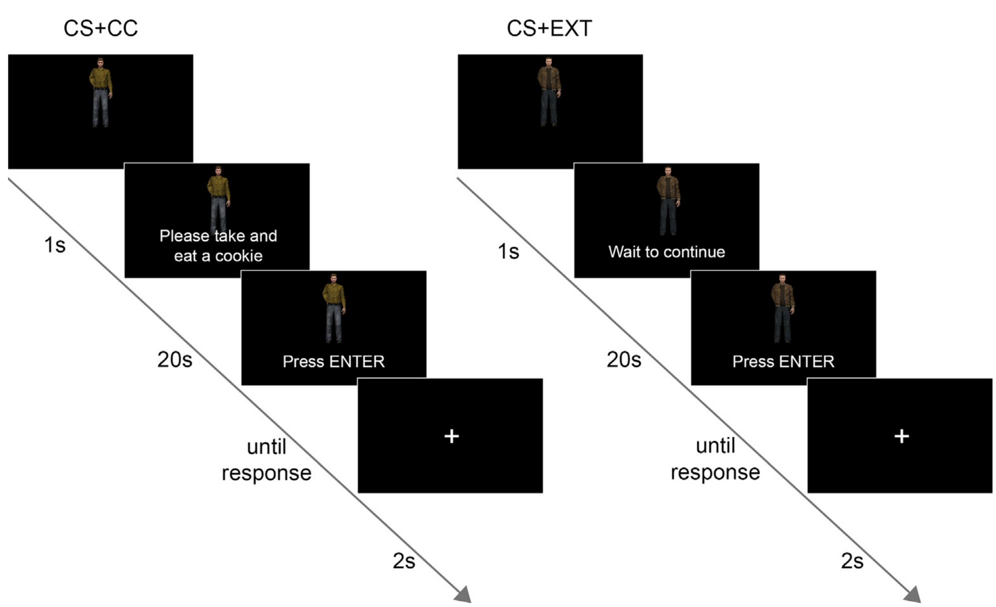

FIGURE 3 Counterconditioning and extinction intervention trial structure. Schematic overview with one example of a CS + CC and CS + EXT trial. The CS+Control was not presented during the CC-intervention.

zero contrasts. For all models we used the generic, default priors of the brms package to give fast and accurate model convergence, without making study-specific assumptions on the data (Bürkner, 2017). For the behavioral analyses, a binomial distribution was used. For the eye-blink startle and valence analyses a Gaussian distribution was used.

Models were fitted using 6 chains with 5,000 iterations each (1,000 warm-up). A coefficient is deemed "significant" when the associated 95% posterior credible intervals is non-overlapping with zero. In the case results did not reach "significance," 90% CIs are reported to investigate whether a marginally significant effect is observed with a narrower interval. When effects were significant at the 95% level, we additionally tested 99% and 99.9% CIs to explore in more detail the robustness of the effects (akin to reporting $p < .01$ and $p < .001$ respectively rather than only $p < .05$). Preregistered planned comparisons were conducted using the emmeans package (Version 1.4.5; Lenth et al., 2018). Emmeans incorporates the model's estimates of the variance and covariance of the model parameters into calculation of means and confidence intervals. When running separate models, each model is fit independently, which can result in inflated Type I errors due to multiple comparisons. By contrast, emmeans uses a single model to estimate the means and standard errors for each condition or combination of conditions, which can lead to more accurate results.

Baseline (Preconditioning)

First, three separate Bayesian mixed-effects models were performed to inspect the effects of threat and reward on valence, avoidance, and startle at baseline before the intervention. Threat (low: CS− vs. high: all CS+) was a predictor in all these models. Since the valence ratings were not acquired per reward level, reward (low, high) was only a predictor in the avoidance and eye-blink startle models.

Intervention Effects

Subsequently, the effects of the intervention were investigated. Again, three Bayesian mixed effects models were performed. CS+-type (CS + CC, CS + EXT, CS + Control) and time were predictors in all models. For the valence data, time had four levels (before acquisition, after acquisition, before test, after test). For the avoidance data and eye-blink startle data, time had two levels (all trials before CC and EXT-intervention, all trials after CC and EXT-intervention). Again, reward (low, high) was a predictor in the avoidance and eye-blink startle models.

Results

BASELINE EFFECTS (BEFORE INTERVENTION)

Valence (Pre-acquisition vs. Post-acquisition)

As expected, valence ratings for the CS+ stimuli were overall significantly more negative than for the CS− ($B = 0.47$, 99.9% CI [0.28, 0.67]). In addition, there was a significant decrease in valence after FAT administration ($B_{\text{pre-acquisition vs. post-acquisition}} = -0.11$, 99% CI [−0.21, −0.01]) that was further qualified by a significant interaction between threat and time ($B = 0.32$, 99.9% CI [−0.49, −0.14]). Before the first FAT administration, the valence of the CS+ stimuli was already more negative than the low threat (CS−) stimulus ($B = 0.29$, 95% CI [0.07, 0.52]), likely due to instructed CS-type contingencies. This difference was further enhanced after FAT administration and accompanying shock exposure ($B = 1.58$, 99.9% CI [0.98, 2.20]).

Indeed, the CS− was rated more positively over time ($B = -.88$, 99.9% CI [−1.38, −0.34]), while CS+ ratings decreased ($B = 0.42$, 99.9% CI [0.07, 0.76]; see Figure 4A).

Although no distinction was made between the CS+-types at baseline (i.e., before the intervention), we unexpectedly found a marginally significant interaction between CS-type and time for CS + CC vs. CS + Control ($B = 0.11$, 90% CI [0.01, 0.21]). No such effect was found for CS + EXT vs. CS + CC ($B = -0.03$, 90% CI [−0.13, 0.07]). Post-hoc analyses showed that before the first FAT administration, valence for the CS + Control was more positive than for CS + CC and CS + EXT ($B_{\text{CS+CC vs. CS+Control}} = 0.39$, 95% CI [0.08, 0.69]; $B_{\text{CS+EXT vs. CS+Control}} = 0.33$, 95% CI [0.02, 0.63]). Importantly however, after the first FAT administration, these differences were no longer apparent ($B_{\text{CS+CC vs. CS+-Control}} = 0.001$, 90% CI [−0.26, 0.25]; $B_{\text{CS+EXT vs. CS+Control}} = 0.05$, 90% CI [−0.20, 0.31]). Thus, before the CC and EXT intervention took place, all CS+-types had similar valence ratings ($M_{\text{CS+CC}} = -30.61$; $M_{\text{CS+EXT}} = -32.33$; $M_{\text{CS+Control}} = -30.59$; see Figure 4A).

Avoidance Behavior (Acquisition)

The task elicited the expected behavior, as evidenced by opposing effects of threat ($B_{\text{high vs. low threat}} = 0.21$, 99.9% CI [0.15, 0.28]) and reward ($B_{\text{high vs. low reward}} = 0.16$, 99.9% CI [0.11, 0.21]), with more avoidance in high threat and low reward conditions. Because intervention did not interact with reward, results in Figure 4 are averaged over reward levels. However, reward effects are shown in the supplement (Figure S2). As expected, there

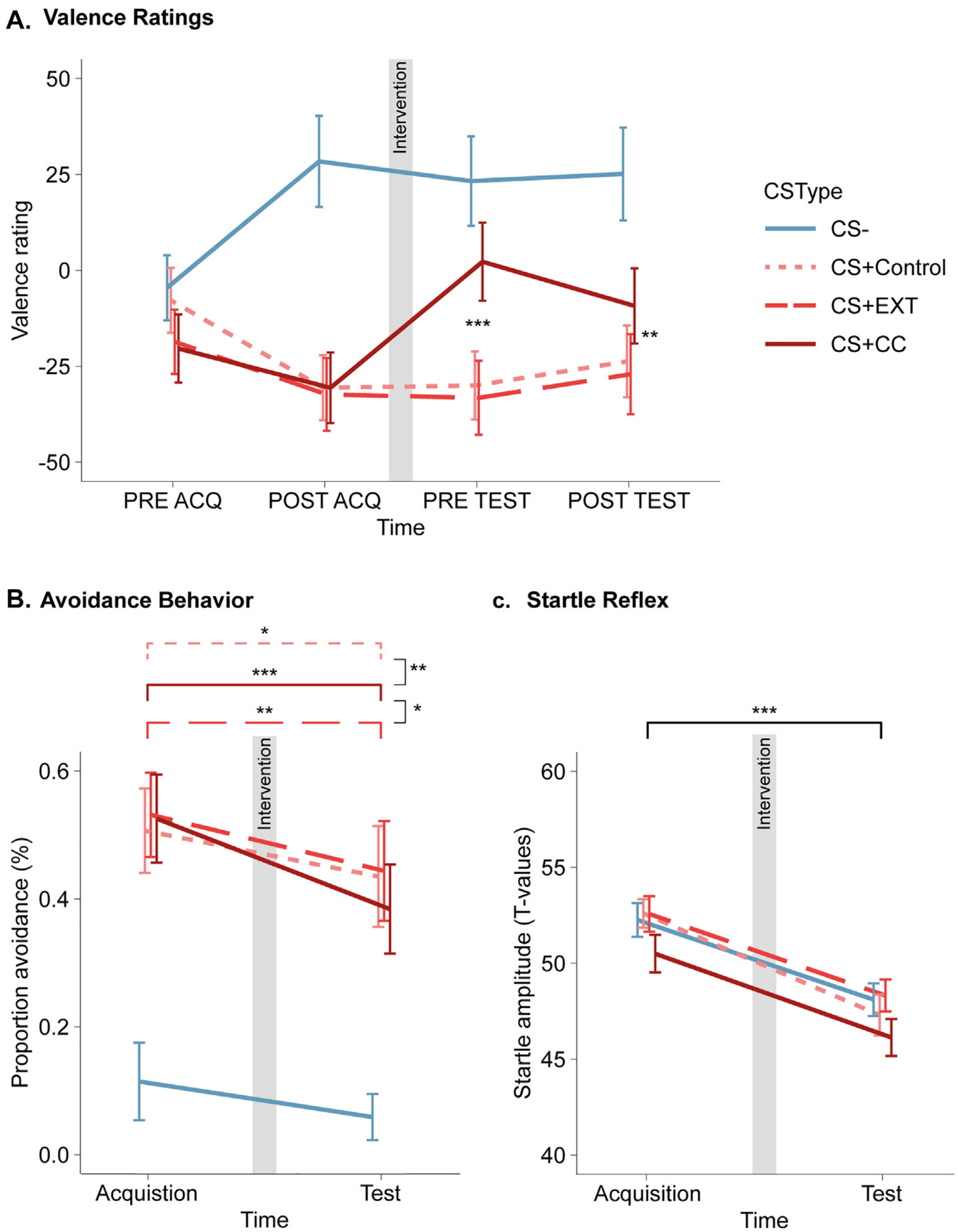

FIGURE 4 **A.** Average valence rating per CS-type as a function of time: pre-acquisition, post-acquisition (before intervention), pre-test (after intervention), and post-test. Valence ratings show a positive effect of counterconditioning. **B.** Average proportion avoidant decision per CS-type over time (collapsed across reward level given no interaction between CS+-type and reward). Avoidance decreased more for the CS + CC compared to the CS + EXT and CS + Control. **C.** Average eye-blink startle magnitude *T*-scores per CS-type over time (collapsed across reward levels given no interaction between reward and time). Startle showed a reduction over time across all CS+-types that was statistically indistinguishable. *95%, **99%, ***99.9% posterior credible intervals non-overlapping with zero.

was also an interaction between threat and reward ($B = 0.13$, 99.9% CI [0.08, 0.17]). For both low threat (CS−) and high threat (CS+) stimuli, participants avoided more in the low than high reward conditions ($B_{\text{low threat}} = 0.07$, 95% CI [0.01, 0.12]; $B_{\text{high threat}} = 0.58$, 99.9% CI [0.42, 0.74]). However, this effect of reward was stronger for high threat (CS+) stimuli (Figure S2).

Importantly, there was no difference in overall avoidance rates for the CS+-types before the intervention ($B_{\text{CS+CC vs. CS+Control}} = -0.15$, 90% CI [−0.38, 0.08], $B_{\text{CS+CC vs. CS+EXT}} = 0.13$, 90% CI [−0.09, 0.36]), and there was also no interaction between CS+-type and reward ($B_{\text{CS+CC vs. CS+Control}} = -0.11$, 90% CI [−0.34, 0.12], $B_{\text{CS+CC vs. CS+EXT}} = 0.04$, 90% CI [−.19, .26]).

Eye-Blink Startle (Acquisition)

In line with our hypotheses, startle magnitude was increased for high threat (CS+) stimuli compared to low threat (CS−) ($B = -0.11$, 99.9% CI [−0.21, −0.01]). There was no significant main effect of reward ($B = 0.01$, 90% CI [−0.04, 0.05]). There was a significant interaction between threat and reward ($B = -0.10$, 99% CI [−0.18, −0.02]; see supplement Figure S2). We found threat-potentiated startle in high reward conditions ($B = -0.42$, 99.9% CI [−0.71, −0.13]) but not in low reward conditions ($B = 0.21$, 90% CI [−0.15, 0.12]) where most avoidance occurred and therefore shock threat was low.

Importantly, before the CC and EXT intervention, there was no difference in startle magnitudes between CS+-types ($M_{CS+CC} = 52.24$; $M_{CS+\text{-}EXT} = 52.59$; $M_{CS+Control} = 52.64$; $B_{CS+CC\ vs.\ CS+\text{-}Control} = 0.02$, 90% CI [−0.05, 0.08]; $B_{CS+CC\ vs.\ CS+EXT} = 0.02$, 90% CI [−0.05, 0.08]).

Summarizing, prior to the CC and EXT intervention the task overall produced the expected results, and importantly there was no significant difference between the to be-conditioned CS-types in avoidance rates, valence ratings or startle directly before the intervention commenced.

Intervention Effects (Time × CS-Type: CS-Type Comparisons Before and After Intervention)

Stimulus Valence

As expected, compared to the CS + EXT and CS + Control, the CS+CC showed a significant increase in valence after intervention (i.e., post-acquisition vs. pre-test; $B_{CS+CC\ vs.\ CS+Control} = 0.14$, 99% CI [0.01, 0.26]; $B_{CS+CC\ vs.\ CS+EXT} = 0.17$, 95% CI [0.04, 0.29]). This difference decreased again after administration of the FAT during test (i.e., pre-test vs. post-test; $B_{CS+CC\ vs.\ CS+Control} = -0.15$, 99% CI [−0.29, −0.02]; $B_{CS+CC\ vs.\ CS+EXT} = -0.17$, 99% CI [−0.32, −0.02]).

Post-hoc analyses revealed that only for the CS +CC there was a significant effect of time ($B = -0.93$, 99.9% CI [−1.35, −0.49]), reflecting a more positive valence after the CC intervention ($M = 2.25$, $SD = 36.17$) compared to before ($M = -30.61$, $SD = 32.70$). After shock threat was reinstated in the FAT during test, the CS + CC valence rating decreased again ($B = 0.33$, 95% CI [0.07, 0.59]; $M = -9.64$, $SD = 35.10$). For both the CS + EXT and CS + Control there were no significant changes in valence following the intervention (all 90% CI's overlapping with zero; see Figure 4A).

Comparing conditions directly per timepoint, subjects reported significantly higher valence for CS + CC compared to both other CS+-types after the intervention (i.e., pre-FAT2; CS + CC vs. CS + EXT: $B = -1.00$, 99.9% CI [−1.68, −0.36]; CS + CC vs. CS + Control: $B = -0.90$, 99.9% CI [−1.45, −0.35]). Although the CS + CC decreased in valence after the FAT during test, it was still rated significantly more positive in the last rating than CS + EXT ($B = -0.50$, 95% CI [−0.88, −0.11]), and CS + Control ($B = -0.41$, 99% CI [−0.81, −0.01]). The CS+Control and the CS + EXT did not differ at any timepoint following the intervention ($B = .09$, 90% CI [−0.16, 0.34] and $B = 0.09$, 90% CI [−0.16, 0.34], respectively).

Thus, results indicated that the CC and EXT intervention produced a robust and specific increase in positive valence of the CS + CC with incomplete extinction of this effect upon shock threat re-exposure in the FAT during test.

Avoidance Behavior

There was a significant main effect of time indicating overall avoidance was lower after the CC and EXT intervention than before ($B = 0.72$, 99.9% CI [0.19, 1.39]). There were trends for overall reduced avoidance for the CS + CC compared to the other high threat (CS+) conditions across both timepoints ($B_{CS+CC\ vs.\ CS+EXT} = 0.24$, 90% CI [0.03, 0.47]; $B_{CS+CC\ vs.\ CS+Control} = 0.47$, 90% CI [0.001, 0.98]). Importantly, there was also a significant interaction between CS-type and Time for CS+CC versus CS + Control ($B = -0.25$, 95% CI [−0.48, −0.04]), reflecting stronger reductions in avoidance after CC-intervention compared to before (Figure 4B). EXT intervention showed intermediate success in reducing avoidance as it did not produce significantly stronger avoidance reductions than the passive control condition ($B_{CS+EXT\ vs.\ CS+Control} = -0.34$, 90% CI [−0.93, 0.29]) but the reduction in avoidance was also not significantly different compared to the counterconditioned stimulus ($B_{CS+CC\ vs.\ CS+EXT} = -$ −0.08, 90% CI [−0.32, 0.15]). Planned comparisons in fact showed significant decreases in avoidance after the intervention for all conditions ($B_{CS+CC} = 2.13$, 99.9% CI [0.70, 3.79]; $B_{CS+\text{-}EXT} = 1.28$, 99% CI [0.13, 2.53]; $B_{CS+Control} = 0.94$, 95% CI [0.18, 1.74]). Critically, while there were no differences in avoidance before intervention between CS-types, after the intervention avoidance on the CS + CC ($M = .40$, $SD = .49$) was significantly lower compared to both the CS + EXT ($M = 0.46$, $SD = 0.50$, $B = 1.00$, 95% CI [0.15, 1.85]), and the CS+Control ($M = 0.45$, $SD = 0.50$, $B = 1.01$, 99% CI [0.11, 1.98]). There was no significant difference between the CS + EXT and the CS + Control after the intervention (B = 0.01, 90% CI [−0.50, 0.55]).

While we hypothesized that the effect of counterconditioning might be strongest in the low reward condition where there was potentially most room for improvement, the three-way interaction between CS+-type, time and reward returned non-significant ($B_{CS+CC\ vs.\ CS+Control} = 0.14$, 90% CI [−.04, .32]; $B_{CS+CC\ vs.\ CS+EXT} = -0.08$, 90% CI [−0.32, 0.12]; full descriptives per reward level can be found the supplement Table S1).

Together, results demonstrated that CC intervention reduced avoidance compared to the passive control, while EXT intervention did not. While the reduction over time was not significantly stronger for the counterconditioned stimulus than for traditional extinction, subjects reached a significantly lower level of avoidance after counterconditioning (Figure 4B).

Eye-Blink Startle

In contrast to subjective valence and avoidance, startle was not affected by CC. There was a significant main effect of time (B = −.05, 99.9% CI [−0.06, −0.03]), with eye-blink magnitude lower after the intervention (M = 47.92, SD = 8.95), than before (M = 52.49, SD = 9.48; Figure 4C). There was also a main effect of reward, with amplitudes for high reward larger (M = 50.92, SD = 9.66), than for the low reward condition where most avoidance occurred (M = 49.51, SD = 9.29, B = 0.01, 99.9% CI [0.00, 0.03]).

More importantly, there was no significant main effect of CS+-type ($B_{CS+CC\ vs.\ CS+Control}$ = 0.01, 90% CI [−0.00, 0.01]; $B_{CS+CC\ vs.\ CS+EXT}$ = −0.00, 90% CI [−0.01, 0.00]) nor interactions between CS-type and time ($B_{CS+CC\ vs.\ CS+Control}$ = 0.00, 90% CI [−0.01, 0.01]; $B_{CS+CC\ vs.\ CS+EXT}$ = −0.01, 90% CI [−0.02, .00]). The three-way interaction between CS-type, time and reward was also not significant ($B_{CS+CC\ vs.\ CS+Control} = 0.00$, 90% CI [−0.01, 0.01]; $B_{CS+CC\ vs.\ CS+EXT} = -0.00$, 90% CI [−0.01, 0.01]). These results indicate that the CC-intervention, while effective in changing subjective valence and avoidance did not influence the startle reflex amplitude and startle simply seemed to habituate similarly for all CS-types.

Discussion

This study investigated if an intervention based on counterconditioning is more effective in reducing negative valence of threatening stimuli and reducing avoidance behavior during approach-avoidance conflicts than extinction. We found that the CC-intervention successfully reduced the negative valence and more importantly led to significantly decreased avoidance for the appetitively conditioned stimulus compared to an extinction control. Contrary to our expectations, there was no concomitant change in startle potentiation. Our findings (a) provide a proof-of-principle demonstration of the effectiveness of counterconditioning as a pathway for reducing costly avoidance, (b) thereby demonstrate a causal role for reward processing in costly fearful avoidance under mixed outcome possibilities, and (c) suggest that reductions in avoidance may take place in the absence of changes in basic reflex physiology indexed by startle.

Our FAT paradigm produced the intended negative valence for stimuli associated with shock. We showed that only counterconditioning reduces the negative valence of the CS + CC whereas an extinguished stimulus (CS + EXT) or a passive control (CS + Control) did not show such changes. Although such effects are not universal and likely dependent on exact study procedures (Gatzounis et al., 2021; Kang et al., 2018), these findings are in line with the majority of previous counterconditioning (Keller & Dunsmoor, 2020; Kerkhof et al., 2011; Newall et al., 2017; van Dis et al., 2019) and evaluative conditioning studies (Hofmann et al., 2010). We demonstrate that this reduction in negative valence generalizes to acute approach-avoidance conflicts, showcasing its robustness even in the presence of acute threat of shock. This is noteworthy as previous studies have typically examined this under extinction conditions or following reinstatement.

Our paradigm also successfully elicited costly avoidance behavior, with avoidance levels varying dependent on the combination of threat and reward levels as expected (Hulsman et al., 2021; Pittig, Pawlikowski, et al., 2014). After the intervention we found reduced avoidance behavior for all CS-Types. Counterconditioning however led to a significantly lower level of avoidance than extinction. There was also a smaller reduction in avoidance over time for the CS+Control stimulus that was neither counterconditioned nor extinguished, likely due to habituation to the shock-US (Hall & Rodríguez, 2017) although generalization of CC or extinction cannot be ruled out. The observed superiority of counterconditioning is in line with recent work by Newall et al. (2017) and Reynolds et al. (2018) who found a reduction in avoidance behavior after CC in children. We extend these previous findings by including (a) a different population (i.e., adults), (b) subjective valence scores, and (c) a potentially more ecologically valid avoidance measure (Krypotos et al., 2018).

While our study showed the expected benefits of CC in reducing negative valence and avoidance, we in addition tested whether the effects of the CC-intervention would be visible on a measure that is not dependent on contingency awareness: eye-blink startle. Startle reflex modulation has been shown to provide an index of subconscious conditioned emotional responding (Hamm & Vaitl, 1996; Sevenster et al., 2014) and we added this measure to investigate effects on a more basic physiological level. Startle did show the expected potentiation under high threat on high reward trials, where conflict was highest and avoidance lowest. However, the CC-intervention did not decrease the eye-blink magnitude for the CS+CC compared to the control conditions, as would be expected based on previous studies that suggested startle potentiation reflects hedonic impact or "liking" (Hebert et al., 2015). Credible intervals were centered closely around zero for the CS effects, which suggests that our null finding is not a result of being underpowered to detect small effects. While care is required in interpreting the null finding, this is in line with a recent study (Stussi et al., 2018) that found that counterconditioning with a pleasant odor also resulted in subjective effects in the absence of eye-blink effects. Thus, a difference in subjective valence and choice behavior does not necessarily imply a decrease in more automatic defensive emotional stance reflected by startle. Together, these findings may suggest that the behavioral effects of counterconditioning take place at a neural implementation level that is, at least to some extent, independent of the neural circuits driving startle potentiation. Such dissociations provide support for recommendations to study affective processes at subjective, physiological and behavioral levels simultaneously (Beckers et al., 2013; Lang, 1994).

While our lack of a reduction of startle might lead to speculation that counterconditioning may not be of added benefit to change basic physiological aspects of anxiety, several recent studies have found superior effects of counterconditioning on skin conductance responding (Keller & Dunsmoor, 2020; Raes & De Raedt, 2012). An alternative hypothesis to be considered for the discrepancy between startle physiology on the one hand and choice behavior and ratings on the other, is that the reduction in avoidance rates after counterconditioning led to higher shock exposure for the CS + CC. These effects could have led to a quick reinstatement of conditioned startle compensating any reductions in startle due to the counterconditioning. There are two reasons why we think such an interpretation is unlikely. First, the probabilistic nature of the task leads to an increase in uncertainty about threat even when avoiding and thereby dampens any impact of an avoidant strategy on the perceived threat (Krypotos et al., 2018). Second, post-hoc analyses did not find evidence for an association between startle and avoidance levels during the baseline (precounterconditioning) phase. What then drives the change in avoidance observed after counterconditioning? This is an important question for future research as more knowledge of the working mechanisms could help understand for whom counterconditioning-based treatments may be most successful and how they can be improved. Recent studies seem to indicate that changes in subjective valence are not required (Kang et al., 2018) nor sufficient (van Dis et al., 2019) for robust changes after counterconditioning. A potential working mechanism besides the observed change in subjective valence could be that counterconditioning facilitates attention towards the CS + CC, which could boost learning and provide a robust and more resistant inhibitory memory trace (Keller et al., 2020). Indeed, recent studies indicate that counterconditioning may boost explicit memory compared to extinction (Keller et al., 2020; Keller & Dunsmoor, 2020).

The current study has several strengths and limitations. As far as the strengths, we measured avoidance in a more representative context of approach-avoidance conflict, using a well-established task that was demonstrated to be sensitive to individual differences (Hulsman et al., 2021), a well powered sample and consideration of conditioning effects using subjective, behavioral as well as physiological measures. Regarding limitations, it is worth noting that consistent with our pre-registration, the observed decrease in avoidance after counterconditioning compared to extinction (i.e., during test) was statistically significant but relatively small. Additionally, the interaction between time and intervention type (CS +CC vs. CS + EXT) did not reach significance. However, it is important to point out that both renewal (due to removal of the plate of cookies) and reinstatement (due to shock exposure) likely counteracted the effects of the counterconditioning and extinction in our setup. Nevertheless, that counterconditioning was superior to extinction under these circumstances might underline the potential relevance for real life, where continued positive reinforcement after treatment is not always feasible and exposure to reminders of the threat may be a problem. In addition, the unconditioned stimuli in the current study were from different domains (US1 = tactile, US2 = taste). Prior research has indicated that USs do not need to belong to the same domain in order to have an effect on physiological, behavioral, or subjective responses (Gatzounis et al., 2021; Meulders et al., 2015; Raes & De Raedt, 2012). Nevertheless, it remains to be investigated which USs are most effective in counterconditioning. In addition, despite participants being unaware of the study's goal and not reporting any awareness afterwards, it is important to consider the potential influence of demand characteristics. However, we expect that demand characteristics are less likely to impact avoidance behavior compared to subjective ratings, given that a reduction in avoidance entails increased shock exposure. Further, while there is evidence that counterconditioning effects can last up to month (Keller et al., 2020; Kerkhof et al., 2011), it would for the clinical potential be important in future work to also assess the long term impact of counterconditioning on avoidance by using multi-day protocols (Kang et al., 2018; van Dis et al., 2019). Finally, we did not formally assess gender or ethnic, cultural, and socioeconomic backgrounds of our participants. Our study population consisted predominantly of a young, White, highly educated female population and therefore the findings of this study require exploration in more diverse populations. For instance, it has been found in multiple studies that females generally show more or longer avoidance behavior than males (Hulsman et al., 2021; McLean & Anderson, 2009; Pittig et al., 2018; Sheynin et al., 2015). Since the incidence of anxiety disorders is overall higher in females, the current results are promising, but more work in larger more diverse samples is needed to understand the impact of such individual differences on the efficacy of counterconditioning intervention.

Besides theoretical implications, this study may have implications for the treatment of anxiety disorders. Our finding should be seen as a promising proof-of-concept which requires more exploration to investigate clinical utility. Current standard treatment of anxiety disorders does not typically involve an explicit counterconditioning component (Craske & Stein, 2016; Stein & Craske, 2017). Translation of counterconditioning to treatment may be especially challenging also because higher levels of anxiety would likely require a sufficiently potent appetitive stimulus to counter the higher levels of negative valence. That said, therapeutic presence and feedback might already provide some implicit counterconditioning of threat. Our study, while in healthy participants, suggests together with recent studies (Newall et al., 2017; Reynolds et al., 2018) and promising work in spider phobics (De Jong et al., 2000) that counterconditioning strategies might be used more explicitly in clinical treatment of anxiety and avoidance. One consideration for its application would be that for the inhibitory extinction learning to be effective, the appetitive stimulus should not distract from the exposure and may best be presented, as was done here, in the background or immediately following exposure so that no competition arises. Future studies are needed to investigate such boundary conditions more. Also, it would be highly relevant for future work to study how a CC-intervention approach compares to extinction in the reduction of avoidance behavior in a population with clinical anxiety levels.

In conclusion, our intervention based on counterconditioning was successful in reducing negative valence and avoidance behavior in a mixed-outcome fearful approach-avoidance task. As excessive avoidance is one of the key predictors of clinical outcomes, this study should be an impetus for further investigation of counterconditioning as a potential treatment for anxiety disorders.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.beth.2023.07.013.

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience, 17*(4), 4. https://doi.org/10.1038/nn.3648.

Andreatta, M., & Pauli, P. (2015). Appetitive vs. Aversive conditioning in humans. *Frontiers in Behavioral Neuroscience, 9*(MAY). https://doi.org/10.3389/FNBEH.2015.00128.

Arnaudova, I., Kindt, M., Fanselow, M., & Beckers, T. (2017). Pathways towards the proliferation of avoidance in anxiety and implications for treatment. *Behaviour Research and Therapy, 96*, 3–13. https://doi.org/10.1016/j.brat.2017.04.004.

Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., & Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Human Brain Mapping, 36*(2), 449–462. https://doi.org/10.1002/hbm.22639.

Bach, D. R. (2015). A cost minimisation and Bayesian inference model predicts startle reflex modulation across species. *Journal of Theoretical Biology, 370*, 53–60. https://doi.org/10.1016/j.jtbi.2015.01.031.

Baeyens, F., Eelen, P., Crombez, G., & van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133–142. https://doi.org/10.1016/0005-7967(92)90136-5.

Ball, T. M., & Gunaydin, L. A. (2022). Measuring maladaptive avoidance: From animal models to clinical anxiety. *Neuropsychopharmacology, 47*(5), 978–986. https://doi.org/10.1038/s41386-021-01263-4.

Bandura, A. (1961). Psychotherapy as a learning process. *Psychological Bulletin, 58*(2), 143–159. https://doi.org/10.1037/H0040672.

Beckers, T., Krypotos, A.-M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology, 92*(1), 90–96. https://doi.org/10.1016/j.biopsycho.2011.12.015.

Boeke, E. A., Moscarello, J. M., LeDoux, J. E., Phelps, E. A., & Hartley, C. A. (2017). Active avoidance: Neural mechanisms and attenuation of pavlovian conditioned responding. *The Journal of Neuroscience, 37*(18). https://doi.org/10.1523/JNEUROSCI.3261-16.2017.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory (Cold Spring Harbor N. Y.), 11*(5), 485–494. https://doi.org/10.1101/LM.78804.

Bürkner, P. (2017). *Advanced Bayesian Multilevel Modeling with the R Package brms*. 1–18.

Castriotta, N. N. (2013). *Anxiety and avoidance: The relationship, function, and measurement*.

Craske, M. G., & Stein, M. B. (2016). Anxiety. *The Lancet, 388*(10063), 3048–3059. https://doi.org/10.1016/S0140-6736(16)30381-6.

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H. U. (2017). Anxiety disorders. *Nature Reviews Disease Primers, 3*(1), 1–19. https://doi.org/10.1038/nrdp.2017.24.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 26*(2), 237–242. https://doi.org/10.1037/0097-7403.26.2.237.

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*(11), 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 (Cold Spring Harbor, N.Y.), 11*(5), 549–554. https://doi.org/10.1101/LM.78004.

Dymond, S. (2019). Overcoming avoidance in anxiety disorders: The contributions of Pavlovian and operant avoidance extinction methods. *Neuroscience and Biobehavioral Reviews, 98*, 61–70. https://doi.org/10.1016/J.NEUBIOREV.2019.01.007.

Emerson, R. W. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Publishing Group, 19*(5), 269–282. https://doi.org/10.1038/nrn.2018.22.

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*(5), 708–719. https://doi.org/10.1016/J.BETH.2014.04.003.

Gatzounis, R., De Bruyn, S., Van de Velde, L., & Meulders, A. (2021). No differences in return of pain-related fear after extinction and counterconditioning. *Emotion*. https://doi.org/10.1037/emo0000960.

Gawronski, B., Gast, A., & De Houwer, J. (2015). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition & Emotion, 29*(5), 816–830. https://doi.org/10.1080/02699931.2014.947919.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504.

Grillon, C., & Baas, J. (2003). *A review of the modulation of the startle reflex by affective states and its application in psychiatry*. *114*, 1557–1579. https://doi.org/10.1016/S1388-2457(03)00202-5.

Hall, G., & Rodríguez, G. (2017). Habituation and conditioning: Salience change in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 43*(1), 48–61. https://doi.org/10.1037/XAN0000129.

Hamm, A. O., & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology, 33*(6), 698–710. https://doi.org/10.1111/J.1469-8986.1996.TB02366.X.

Hebert, K. R., Valle-Inclán, F., & Hackley, S. A. (2015). Modulation of eyeblink and postauricular reflexes during the anticipation and viewing of food images. *Psychophysiology, 52*(4), 509–517. https://doi.org/10.1111/psyp.12372.

Hendriks, S. M., Spijker, J., Licht, C. M. M., Beekman, A. T. F., & Penninx, B. W. J. H. (2013). Two-year course of anxiety disorders: Different across disorders or dimensions? *Acta Psychiatrica Scandinavica, 128*(3), 212–221. https://doi.org/10.1111/acps.12024.

Hendrikx, L. J., Krypotos, A. M., & Engelhard, I. M. (2021). Enhancing extinction with response prevention via imagery-based counterconditioning: Results on conditioned avoidance and distress. *Journal of Behavior Therapy and Experimental Psychiatry, 70*(November 2019), 101601. https://doi.org/10.1016/j.jbtep.2020.101601.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy, 40*(3), 217–234. https://doi.org/10.1016/S0005-7967(01)00006-7.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*(3), 390–421. https://doi.org/10.1037/A0018916.

Hulsman, A. M., Kaldewaij, R., Hashemi, M. M., Zhang, W., Koch, S. B. J., Figner, B., ... Klumpers, F. (2021). Individual differences in costly fearful avoidance and the relation to psychophysiology. *Behaviour Research and Therapy, 137*, 103788. https://doi.org/10.1016/j.brat.2020.103788.

Jones, M. C. (1924). A laboratory study of fear: The case of Peter. *The Journal of Genetic Psychology, 31*, 308–315.

Kang, S., Vervliet, B., Engelhard, I. M., van Dis, E. A. M., & Hagenaars, M. A. (2018). Reduced return of threat expectancy after counterconditioning verus extinction. *Behaviour Research and Therapy, 108*, 78–84. https://doi.org/10.1016/J.BRAT.2018.06.009.

Kashdan, T. B., Elhai, J. D., & Breen, W. E. (2008). Social anxiety and disinhibition: An analysis of curiosity and social rank appraisals, approach–avoidance conflicts, and disruptive risk-taking behavior. *Journal of Anxiety Disorders, 22*(6), 925–939. https://doi.org/10.1016/J.JANXDIS.2007.09.009.

Keller, N. E., & Dunsmoor, J. E. (2020). The effects of aversive-to-appetitive counterconditioning on implicit and explicit fear memory. *Learning & Memory, 27*, 12–19 https://learnmem.cshlp.org/content/27/1/12.

Keller, N. E., Hennings, A. C., & Dunsmoor, J. E. (2020). Behavioral and neural processes in counterconditioning: Past and future directions. *Behaviour Research and Therapy, 125*(103532), 1–12. https://doi.org/10.1016/j.brat.2019.103532.

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology, 58*(1), 31–38. https://doi.org/10.1027/1618-3169/A000063.

Kessler, R. C., Ruscio, A. M., Shear, K., & Wittchen, H. -U. (2010). Epidemiology of Anxiety Disorders. In M. Geyer, B. Ellenbroek, & C. Marsden (Eds.), *Current topics in behavioral neurosciences*. Springer. https://doi.org/10.1007/7854_2009_7.

Klumpers, F., Van Gerven, J. M., Prinssen, E. P. M., Niklson Roche, I., Roesch, F., Riedel, W. J., Kenemans, J. L., & Baas, J. M. P. (2010). Method development studies for repeatedly measuring anxiolytic drug effects in healthy humans. *Journal of Psychopharmacology (Oxford, England), 24*(5), 657–666. https://doi.org/10.1177/0269881109103115.

Krypotos, A.-M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy, 111*, 99–105. https://doi.org/10.1016/j.brat.2018.10.011.

Lang, P. J. (1994). The Varieties of Emotional Experience: A Meditation on James-Lange Theory. *Psychological Review, 101*(2), 211–221. https://doi.org/10.1037/0033-295X.101.2.211.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377–395.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans. *R Package Version 1.15-15, 34*(1), 216–221. https://doi.org/10.1080/00031305.1980.10483031.

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review, 42*, 72–82. https://doi.org/10.1016/j.cpr.2015.08.004.

Lovibond, P. F., Mitchell, C. J., Minard, E., Brady, A., & Menzies, R. G. (2009). Safety behaviours preserve threat beliefs: Protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy, 47*(8), 716–720. https://doi.org/10.1016/J.BRAT.2009.04.013.

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review, 29*(6), 496–505. https://doi.org/10.1016/J.CPR.2009.05.003.

Meulders, A., Karsdorp, P. A., Claes, N., & Vlaeyen, J. W. S. (2015). Comparing counterconditioning and extinction as methods to reduce fear of movement-related pain. *The Journal of Pain, 16*(12), 1353–1365. https://doi.org/10.1016/j.jpain.2015.09.007.

Newall, C., Watson, T., Grant, K. A., & Richardson, R. (2017). The relative effectiveness of extinction and counterconditioning in diminishing children's fear. *Behaviour Research and Therapy, 95*, 42–49. https://doi.org/10.1016/J.BRAT.2017.05.006.

Pittig, A., Alpers, G. W., Niles, A. N., & Craske, M. G. (2015). Avoidant decision-making in social anxiety disorder: A laboratory task linked to in vivo anxiety and treatment outcome. *Behaviour Research and Therapy, 73*, 96–103. https://doi.org/10.1016/j.brat.2015.08.003.

Pittig, A., Brand, M., Pawlikowski, M., & Alpers, G. W. (2014). The cost of fear: Avoidant decision making in a spider gambling task. *Journal of Anxiety Disorders, 28*(3), 326–334. https://doi.org/10.1016/J.JANXDIS.2014.03.001.

Pittig, A., Pawlikowski, M., Craske, M. G., & Alpers, G. W. (2014). Avoidant decision making in social anxiety: The interaction of angry faces and emotional responses. *Frontiers in Psychology, 5*, 1–11. https://doi.org/10.3389/fpsyg.2014.01050.

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience and Biobehavioral Reviews, 88*, 117–140. https://doi.org/10.1016/j.neubiorev.2018.03.015.

Posit team (2019). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/.

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*(4), 757–767. https://doi.org/10.1016/J.BETH.2012.03.012.

Rattel, J. A., Miedl, S. F., Blechert, J., & Wilhelm, F. H. (2017). Higher threat avoidance costs reduce avoidance behaviour which in turn promotes fear extinction in humans. *Behaviour Research and Therapy, 96*, 37–46. https://doi.org/10.1016/J.BRAT.2016.12.010.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/.

Reynolds, G., Field, A. P., & Askew, C. (2018). Reductions in children's vicariously learnt avoidance and heart rate responses using positive modeling. *Journal of Clinical Child and Adolescent Psychology, 47*(4), 555–568. https://doi.org/10.1080/15374416.2016.1138410.

Sevenster, D., Beckers, T., & Kindt, M. (2014). Fear conditioning of SCR but not the startle reflex requires conscious discrimination of threat and safety. *Frontiers in Behavioral Neuroscience, 8*(FEB), 32. https://doi.org/10.3389/FNBEH.2014.00032/BIBTEX.

Sheynin, J., Moustafa, A. A., Beck, K. D., Servatius, R. J., & Myers, C. E. (2015). Testing the role of reward and punishment sensitivity in avoidance behavior: A computational modeling approach. *Behavioural Brain Research, 283*, 121–138. https://doi.org/10.1016/j.bbr.2015.01.033.

Stein, M. B., & Craske, M. G. (2017). Treating anxiety in 2017: optimizing care to improve outcomes. *JAMA, 318*(3), 235–236. https://doi.org/10.1001/JAMA.2017.6996.

Stussi, Y., Delplanque, S., Coraj, S., Pourtois, G., & Sander, D. (2018). Measuring Pavlovian appetitive conditioning in humans with the postauricular reflex. *Psychophysiology, 55*(8), e13073.

van Dis, E. A. M., Hagenaars, M. A., Bockting, C. L. H., & Engelhard, I. M. (2019). Reducing negative stimulus valence does not attenuate the return of fear: Two counterconditioning experiments. *Behaviour Research and Therapy, 120*(May), 103416. https://doi.org/10.1016/j.brat.2019.103416.

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*(1), 71–79. https://doi.org/10.1037/0097-7403.32.1.71.

Vervliet, B., & Indekeu, E. (2015). Low-cost avoidance behaviors are resistant to fear extinction in humans.

Frontiers in Behavioral Neuroscience, 9, 1–12. https://doi.org/10.3389/fnbeh.2015.00351.

Vervliet, B., Lange, I., & Milad, M. R. (2017). Temporal dynamics of relief in avoidance conditioning and fear extinction: Experimental validation and clinical relevance. *Behaviour Research and Therapy, 96*, 66–78. https://doi.org/10.1016/J.BRAT.2017.04.011.

Zenses, A. K., Baeyens, F., Beckers, T., & Boddez, Y. (2021). Thought conditioning: Inducing and reducing thoughts about the aversive outcome in a fear-conditioning procedure. *Clinical Psychological Science, 9*(2), 252–269. https://doi.org/10.1177/2167702620954222.

Received: February 9, 2023
Revised: July 14, 2023
Accepted: July 16, 2023
Available online: 27 July 2023