Thought Conditioning: Inducing and Reducing Thoughts about the Aversive Outcome in a Fear Conditioning Procedure

Ann-Kathrin Zenses, Frank Baeyens, Tom Beckers

KU Leuven

Yannick Boddez

KU Leuven & Ghent University

Ann-Kathrin Zenses, Center for the Learning of Psychology and Experimental Psychopathology, KU Leuven; Frank Baeyens, Center for the Learning of Psychology and Experimental Psychopathology, KU Leuven; Tom Beckers, Center for the Learning of Psychology and Experimental Psychopathology, KU Leuven, and Leuven Brain Institute, KU Leuven; Yannick Boddez, Center for the Learning of Psychology and Experimental Psychopathology, KU Leuven, and Ghent University.

This research was supported by the grant G076015N of the Research Foundation – Flanders (FWO). Yannick Boddez is supported by Ghent University grant BOF16/MET_V/002 awarded to Jan De Houwer.

Correspondence concerning this article should be addressed to Yannick Boddez, Ghent University, Henri Dunantlaan 2 B-9000 Ghent Belgium. E-mail: yannick.boddez@ugent.be

This is a preprint of a manuscript accepted for publication in *Clinical Psychological Science*. It is not the version of record and may deviate from the final version as published.

Abstract

Human fear conditioning paradigm is a widely-used procedure to study anxiety. However, merely thinking about the aversive outcome is typically not measured in this procedure. This is surprising because thinking of an aversive event is of clinical relevance (e.g., in the form of intrusions) and of theoretical interest. We present two pre-registered studies that (1) included thinking of an aversive outcome as an additional dependent variable and (2) compared several interventions to reduce it. We found that mere thinking of an aversive outcome could be successfully conditioned. Among the participants who showed successful acquisition, extinction training was less successful in reducing it than counterconditioning. Presenting new additional outcomes also proved effective to reduce thoughts about the initial outcome when the new outcomes were positive stimuli. Including thinking of the aversive outcome as an additional dependent variable may serve to enhance the understanding of anxiety-related disorders and inform their treatment.

Keywords: human fear conditioning, extinction, association splitting, counterconditioning, intrusive thinking, anxiety disorders

Thought Conditioning: Inducing and Reducing Thoughts about the Aversive Outcome in a Fear Conditioning Procedure

Anxiety-, trauma- and stressor-related disorders[1] are among the most common psychiatric disorders (Wittchen et al., 2011) and often lead to severe impairments in quality of life (Olatunji, Cisler, & Tolin, 2007). As such, they have become an important target of research. The fear conditioning procedure is a widespread model to study anxiety disorders in the laboratory (Beckers, Krypotos, Boddez, Effting, & Kindt, 2013; Craske et al., 2009; Mineka & Oehlberg, 2008; Scheveneels, Boddez, & Hermans, 2019). In this procedure, the effect of contingently presenting a cue and an outcome on responding to the former is assessed. In layman terms, the outcome is often termed and selected to be "aversive" (e.g., an electric shock) and the responses are typically termed anticipatory "fear responses". Such fear responses are traditionally assessed by the measurement of physiology (e.g., skin conductance, fear-potentiated startle), and in humans also by verbal responses (e.g., subjective fear ratings, outcome expectancy ratings; Lonsdorf et al., 2017). Merely thinking of the aversive outcome is typically not assessed even though unpleasant thoughts of an aversive event are clinically relevant (e.g., in the form of intrusive thinking; Harvey, Watkins, Mansell, & Shafran, 2004) and of theoretical interest. For example, the sight of a hospital bed may make one think of how a loved one died in pain (Boddez, 2018).

Baeyens, Vansteenwegen, Hermans, and Eelen (2001) theorized that people may acquire both signal relations and referential relations when partaking in a conditioning procedure. Signal relations underlie the expectancy that an outcome will occur in the immediate future when

[1] For reasons of simplicity and readability, we refer to these different diagnostic categories as anxiety-related disorders in the remainder of the manuscript.

presented with the cue (i.e., the cue functions as a signal) and are captured by the outcome expectancy measure (Boddez et al., 2013). Referential relations, in contrast, merely make one think of the outcome when presented with the cue (i.e., the cue functions as a referent). Such referential relations may also occur without activating an expectancy that the outcome will be imminent (Baeyens et al., 2001; also see Jozefowiez, 2018). For instance, a soldier's old gas mask may make him think of a chemical attack that he experienced without making him expect an imminent chemical attack. Although this may be a sufficient basis for discomfort, merely thinking of the aversive outcome has never been included as an outcome measure in human fear conditioning research.

Importantly, there are theoretical and empirical reasons to assume that such thoughts, unlike expectancies, would survive extinction training (i.e., a laboratory model of exposure treatment; e.g., Eelen, Hermans, & Baeyens, 2001). Extinction training involves presenting the cue that previously contingently preceded the aversive outcome by itself (Bouton, 1988; Hermans, Craske, Mineka, & Lovibond, 2006; Milad & Quirk, 2012; Quirk & Mueller, 2008; for a discussion of the mechanisms mediating extinction effects see General Discussion). It has been theorized that the information provided in an extinction procedure counteracts signal relations, while leaving referential relations intact (Baeyens et al., 2001). As described above, a signal relation implies that the cue functions as a signal for the outcome occurring in the imminent future. Such signal relation, and the concomitant outcome expectancy, can be supported or refuted by the occurrence or nonoccurrence of the outcome. Because the cue is no longer followed by the outcome during extinction training, the individual can learn that this relation no longer holds. This, in turn, may result in a reduction in outcome measures indexing signal relations such as outcome expectancies. In case of a referential relation, however, the cue merely

refers to the outcome. Because this does not entail a (falsifiable) prediction of the outcome, nonoccurrence of the outcome does not invalidate a referential relation and should leave it unaffected (Baeyens et al., 2001; Hermans & Baeyens, 2012). Accordingly, thinking of the aversive outcome may be insensitive to extinction training.

In line with these theoretical considerations, there is a plethora of evidence showing a decline in outcome expectancies following extinction training (e.g., Hermans et al., 2006). There is also indirect evidence, stemming from evaluative conditioning research, suggesting that thinking of the aversive outcome might indeed be resistant to extinction training (Baeyens, Díaz, & Ruiz, 2005; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). In evaluative conditioning, the valence of a cue changes due to pairing of the cue with a positively or negatively valenced outcome. It has been proposed that this change in valence is driven by a merely referential relation between cue and outcome (Baeyens et al., 2001). An interesting finding is that evaluative shifts induced by conditioning have indeed been found to survive extinction training (Baeyens et al., 2005; Vansteenwegen et al., 2006). As such, it seems likely that other indices of referential relations, such as thinking of the aversive outcome, might also survive extinction learning.

There may be important clinical implications if thinking of an aversive outcome is unaffected by extinction. That is, a patient may still have unpleasant thoughts about the outcome following an otherwise successful exposure treatment. This calls for a search for alternative interventions that do successfully reduce thinking of a previously associated aversive outcome. Here we present two studies that were aimed at (1) testing thinking of the aversive outcome as an additional outcome variable in a fear conditioning procedure and (2) testing several interventions to successfully reduce it.

Study 1

A candidate for reducing thinking of an aversive outcome is a procedure termed association splitting. At a procedural level, this entails pairing the cue with additional "competing" outcomes (Moritz & Jelinek, 2011; Moritz, Jelinek, Klinge, & Naber, 2007). One potential mechanism is that increasing the number of associations that *originate* from a single cue decreases the strength or likelihood of retrieval of each individual cue-outcome association (i.e., the fan effect; Anderson, 1974; Moritz & Jelinek, 2011; Moritz et al., 2007). One can compare this to water running from a faucet. If the number of containers placed below the faucet increases, then the volume of water in each container will decrease. As such, association splitting can be seen as the mirror image of the more widely studied cue competition phenomena (Miller & Matute, 1998). In a standard cue competition procedure, the outcome is not preceded by just one, but by a compound of multiple cues that are presented simultaneously. Here, the dominant theory holds that increasing the number of associations that *end* in a single outcome decreases the strength or retrievability of the individual cue-outcome associations (Boddez, Haesen, Baeyens, & Beckers, 2014; Rescorla & Wagner, 1972). In line with this notion, cues that have been trained in compound with other cues indeed elicit less conditioned responding than cues trained separately. Translating the association splitting intervention to a fear conditioning procedure, pairing the cue with a compound of its initial outcome and additional novel outcomes should reduce thinking of the initial outcome when presented with the cue.

In the current study, participants indicated both outcome expectancies and the extent to which the cue made them think of the initial outcome during each trial of a fear conditioning task. Following a differential fear conditioning phase, we compared the effectiveness of an extinction intervention (i.e., repeatedly presenting the cue without the initial aversive outcome)

to an association splitting intervention (i.e., pairing the cue with a compound of the initial outcome and two novel outcomes).

First, we predicted that mere thinking of the aversive outcome and outcome expectancies would be successfully conditioned during the fear conditioning procedure across all participants. Second, we predicted that the association splitting intervention would be more successful in decreasing thinking of an aversive outcome than the extinction intervention. At the same time, we hypothesized that the extinction intervention would be more successful in decreasing outcome expectancies than the association splitting intervention because the initial outcome was still presented during the latter. We added a no extinction condition (i.e., continued pairing of the cue with the initial outcome) as a control group. This group only differed from the association splitting group with respect to the added outcomes and therefore isolates their effect on conditioned responding. We predicted that the association splitting intervention would decrease thinking of the aversive outcome as compared to the no extinction intervention. The extinction group was expected to be more successful in decreasing outcome expectancies than the no extinction group.

Method

Preregistration. The experimental procedures and statistical analyses of Study 1 were preregistered on Aspredicted.org (http://aspredicted.org/blind.php?x=bm5iy8). Additional analyses will be explicitly termed as "exploratory" in the results section.

Participants. Seventy-two participants (n female = 60, n male = 12; M age = 21.42, SD age = 4.84) were included in the present study, with each of the three experimental conditions (i.e., extinction, no extinction, association splitting) comprising twenty-four participants (See Supplemental Materials for additional participants' characteristics). One

participant discontinued his or her participation because of feeling unwell. The data of this participant were replaced. In total, 73 participants took part in the study. Participants were recruited through an online recruitment platform (Sona Systems, Ltd., Tallinn, Estonia), which is open to all KU Leuven students and the public. Exclusion criteria were based on self-report and were: presence of a clinical depression, anxiety disorder, posttraumatic stress disorder, panic disorder or another psychiatric disorder in the past or at the time of participation, medical advice to avoid stressful situations, or a history of fainting. Ethics approval was obtained from the Social and Societal Ethics Committee of KU Leuven and all participants gave written informed consent. Participation was reimbursed with partial course credits or 8 euros.

Apparatus and stimuli. Affect 5 software, which is an updated version of Affect 4.0 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2009), was used to program and administer the conditioning task.

Stimuli for the fear conditioning procedure. Images of a small and a big circle were used as the cues (allocation to either reinforced or unreinforced cue was counterbalanced within each experimental condition). The small circle had a diameter of 2 cm, while the large circle had a diameter of 5 cm. The outline of the circles was white and the inside of the circles was black. The images for the outcome were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2005). All outcome images were 6 cm wide and 4.5 cm high. The image of a mutilated body (i.e., image 3051) was used as the initial outcome for all groups during Condition phase 1. The images of a thoracotomy (i.e., image 3250) and an aggressive dog (i.e., image 1300; Lenaert et al., 2014) were used as the two novel outcomes during Conditioning phase 2 for the association splitting intervention. The two novel outcomes were chosen to be

aversive. This was done to distinguish the association splitting intervention from counterconditioning interventions where cues are typically paired with outcomes of opposite valence to the initial outcome (De Houwer, 2011; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). We reasoned that if the splitting of associations (i.e., increasing the number of associations) is the main underlying mechanism (Moritz & Jelinek, 2011; Moritz et al., 2007), then valence should not matter. All stimuli were presented against a black background and were presented slightly above the center of the screen.

Outcome measures. Participants rated the extent to which they expected the cue to be followed by the initial outcome (i.e., *To what extent do you expect the geometrical figure to be followed by the image of the mutilated body in the following seconds?*) on a scale ranging from 0 (*I do not expect the mutilated body at all*) to 10 (*I expect the mutilated body to a high extent*) during each cue presentation. Additionally, participants indicated the extent to which the cue made them think of the initial outcome (i.e., *To what extent does the geometrical figure make you think of the image of the mutilated body?)* on a scale ranging from 0 (*Does not make me think of the mutilated body at all*) to 10 (*Makes me think of the mutilated body to a high extent*) during each cue presentation. Both scales were presented simultaneously and below the cue. They appeared on screen for 12 s. The scale of the expectancy ratings was always presented above the scale of the thinking of ratings during all trials. Participants indicated their ratings with a single mouse click**.** The rating scales referred exclusively to the initial outcome used during Conditioning phase 1 on all trials of Conditioning phase 1 and 2. This was done to show the effect of the interventions on expectancy and thinking of the initial outcome.

Procedure. At the beginning of the experiment, the exclusion criteria were assessed, and participants gave written informed consent. Prior to the conditioning task, participants completed

the Depression Anxiety Stress Scales (DASS-21; Lovibond & Lovibond, 1995) and received general instructions concerning the task:

> You will get to see images of geometrical shapes. Some of these geometrical figures may be followed by an unpleasant image, namely an image of a mutilated body. When a geometrical figure appears on the screen, you have to complete two different scales.

Afterwards, participants received a description of the two above-mentioned rating scales. Additionally, participants were told:

> While *expecting* and *thinking of* can go hand in hand, this does not necessarily have to be so. For instance, a souvenir can make you think of a nice evening abroad without making you expect to experience such a nice evening imminently.

Then, the conditioning task commenced (see Figure 1). At the beginning of each trial, a fixation cross was presented for 2 s. It was followed by the presentation of the cue for 12 s. Both rating scales appeared and disappeared with cue onset and offset, respectively. Not answering in time resulted in an empty data cell. Immediately after cue offset, the outcome was displayed for 1 s on reinforced trials, while a fixation cross was presented for the same length of time on unreinforced cue trials. The intertrial-interval (ITI) varied between 4.3 s and 4.7 s (M_{ITI} = 4.5 s) and comprised a fixation cross presentation. Conditioning phase 1 was identical for all participants and consisted of 8 presentations of both the reinforced and unreinforced cue in random order (i.e., without any restrictions). The reinforced cue was always followed by the initial outcome, while the unreinforced cue was never followed by an outcome. Conditioning phase 2 also consisted of 8 presentations of both the reinforced and unreinforced cue in random order. Conditioning phase 2 was a prolongation of Conditioning phase 1 (i.e., there was no break between both phases). Participants were randomly assigned to one of three conditions (i.e., extinction, no extinction,

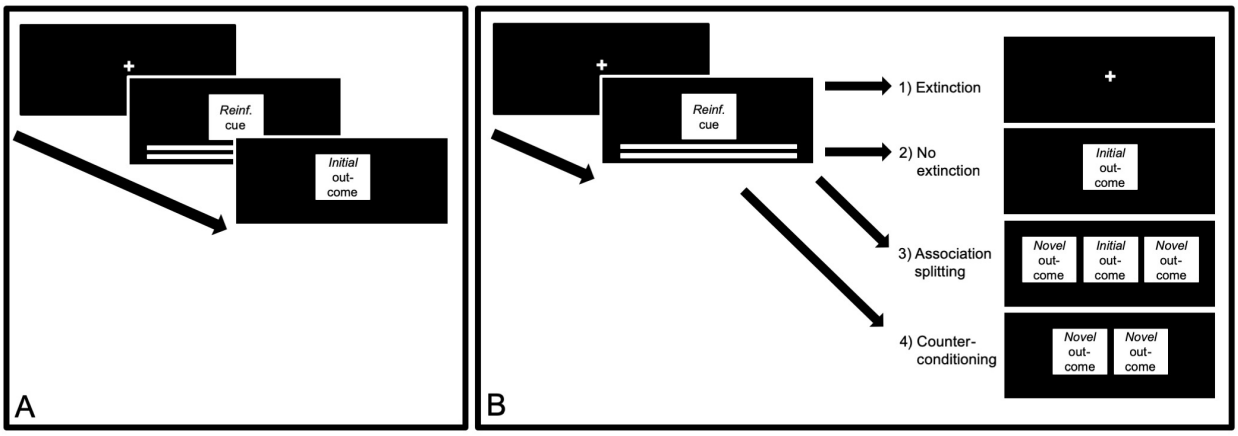


Figure 1. Panel A shows the setup of a reinforced cue trial during Conditioning phase 1, which was the same for all experimental groups. Panel B shows the setup of a reinforced cue trial during Conditioning phase 2 for the different experimental groups. Group 1-3 were part of Study 1. Group 1-4 were part of Study 2. Reinf. cue = Reinforced cue.

association splitting). In the extinction condition, the reinforced cue was no longer followed by the outcome. In the no extinction condition, the reinforced cue continued to be paired with the initial outcome. In the association splitting condition, the reinforced cue was followed by a compound consisting of the initial outcome and the two novel outcomes. The image of the initial outcome was always in the center of the compound, while the position (i.e., left or right) of the novel outcomes was randomly determined on each trial for each participant. The unreinforced cue was never followed by any outcome. At the end of the conditioning task, participants rated the cues in terms of valence and the outcome images in terms of valence, tension, and attention. In addition, they indicated the extent to which the outcome image made them look away and whether they had already seen the outcome images before the present study. These cue and outcome ratings are reported in the Supplemental Materials. Finally, participants completed the

Preservative Thinking Questionnaire (PTQ; Ehring et al., 2011) and the White Bear Suppression Inventory (WBSI; Wegner & Zanakos, 1994)[2] and were debriefed.

Data analyses and reduction. The outcome expectancy and thinking of the aversive outcome ratings were analyzed with mixed-design analyses of variance (ANOVAs). Group (i.e., extinction, no extinction, association splitting) was included as a between-subjects variable and stimulus (i.e., reinforced cue, unreinforced cue) and trial (i.e., trial 1-8) were included as within-subjects variables. To test the hypothesized differences between groups, we ran mixed-design ANOVAs with Group as between-subjects factor and stimulus and trial as within-subjects factors. Moreover, we conducted separate repeated-measures ANOVAs for each group with stimulus and trial as within-subjects factors as exploratory (i.e., non-preregistered) analyses. We report the Greenhouse-Geisser correction in case of violations of the sphericity assumption. The level of significance was fixed at $\alpha = 0.05$. Partial eta squared (η_p^2), with its corresponding 90% confidence interval (Smithson, 2001), is reported as the effect size. Failure to respond within the provided response window resulted in missing values in the analyses.

To test whether thinking of the aversive outcome can be successfully conditioned during Conditioning phase 1, the analyses were run on the data of all participants who had complete acquisition data ($N = 66$ and $N = 46$ for the expectancies and thinking of the aversive outcome, respectively). To compare the effectiveness of the different interventions during Conditioning phase 2, analyses were run with and without an acquisition criterion, which was determined and preregistered before the start of data collection. Based on this acquisition criterion, only those

[2] We preregistered additional exploratory analyses to examine whether the WBSI, PTQ and DASS-21 predict participants' performance on the thinking of the aversive rating scale. These results will not be reported here as they relate to a separate research question.

participants who scored higher on the reinforced cue than on the unreinforced cue on the last acquisition trial for both outcome expectancy and thinking of the aversive outcome ratings (N = 59) were included. Higher should be interpreted as arithmetically higher (i.e., reinforced cue > unreinforced cue). In case of missing values on the last trial, the value of the next to last trial was used. It was reasoned that only if the responses were successfully acquired, they could subsequently be reduced by our interventions. In the main text, we report the analyses on the data of all participants who had complete data and without applying the acquisition criterion (N = 72 and N = 69 for the expectancies and thinking of the aversive outcome, respectively). However, we also mention whenever a different conclusion was reached based on the analyses of the data with acquisition criterion. In addition, we mention all analyses on the data with applying the acquisition criterion (N = 59) for participants who did not have missing values in Conditioning phase 2 (N = 59 and N = 56 for the expectancies and thinking of the aversive outcome, respectively) in the Supplemental Materials. An overview, which specifies the number of participants in the different stages can be found in the Supplemental Materials as well.

Results

Conditioning phase 1.

Expectancies of the aversive outcome. There was a significant main effect of stimulus, $F(1, 63) = 265.14$, $p < .001$, $\eta_\rho^2 = .808$, 90% CI [.733, .849], and a significant Stimulus x Trial interaction, $F(4.01, 252.85) = 91.65$, $p < .001$, $\eta_\rho^2 = .592$, 90% CI [.526, .636], suggesting successful acquisition (Figures 2A-C). There were no significant group differences in acquisition, as indicated by the non-significant Stimulus x Group, $F(2, 63) = 2.98$, $p = .06$, $\eta_\rho^2 = .086$, 90% CI [.000, .192], and Stimulus x Trial x Group interactions, $F(8.03, 252.85) = 1.80$, $p = .08$, $\eta_\rho^2 = .054$, 90% CI [.000, .076]. Although these interactions were, as anticipated, non-

significant, it could, however, be critically noted that the *p*-values approached the common criterion for significance.

Thinking of the aversive outcome. The significant main effect of stimulus, $F(1, 43) = 60.15$, $p < .001$, $\eta_\rho^2 = .583$, 90% CI [.409, .684], and the significant Stimulus x Trial interaction, $F(3.27, 140.47) = 33.83$, $p < .001$, $\eta_\rho^2 = .441$, 90% CI [.330, .513], are indicative of successful acquisition (Figures 2D-F). There were no significant group differences in acquisition, as shown by the non-significant Stimulus x Group, $F(2, 43) = 1.93$, $p = .16$, $\eta_\rho^2 = .082$, 90% CI [.000, .205], and Stimulus x Trial x Group interactions, $F(6.53, 140.47) = 1.29$, $p = .26$, $\eta_\rho^2 = .057$, 90% CI [.000, .086].

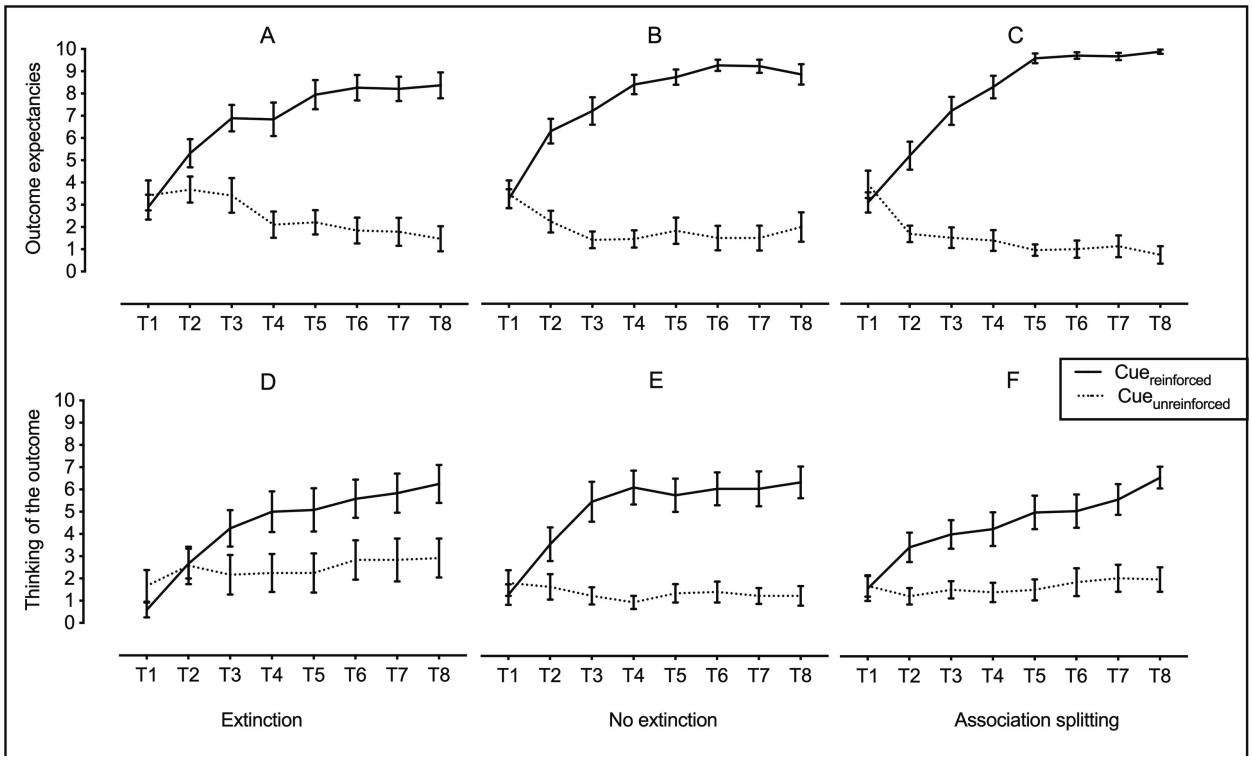


Figure 2. Panels A-C show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 1 of Study 1. Panels D-F show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 1 of Study 1. T = trial number.

Conditioning phase 2.

Expectancies of the aversive outcome. In line with our hypothesis, the Stimulus x Trial x Group interaction was significant, $F(7.08, 244.24) = 10.33$, $p < .001$, $\eta_p^2 = .230$, 90% CI [.138, .283], (Figures 3A-C). As suggested by Figures 3A-C and in line with our hypotheses, the extinction intervention was more successful in reducing outcome expectancies than the association splitting intervention and the no extinction group. This was confirmed by a significant Stimulus x Trial x Group interaction comparing extinction with association splitting, $F(3.41, 156.78) = 11.29$, $p < .001$, $\eta_p^2 = .197$, 90% CI [.099, .272], and comparing extinction with no extinction, $F(3.51, 161.60) = 12.67$, $p < .001$, $\eta_p^2 = .216$, 90% CI [.116, .290]. Explorative analyses showed that the Stimulus x Trial interaction in the extinction group was significant, $F(3.25, 74.66) = 11.08$, $p < .001$, $\eta_p^2 = .325$, 90% CI [.162, .426] (Figure 3A). In contrast, the Stimulus x Trial interactions in both the no extinction and association splitting group were non-significant, $F(2.50, 57.40) = 2.74$, $p = .06$, $\eta_p^2 = .107$, 90% CI [.000, 0.214], and, $F(2.43, 55.99) = 1.25$, $p = .30$, $\eta_p^2 = .051$, 90% CI [.000, .138], respectively. That is, we only found statistical evidence for successful extinction of the outcome expectancies in the extinction group.

Thinking of the aversive outcome. The Stimulus x Trial x Group interaction was significant, $F(7.23, 238.48) = 3.97$, $p < .001$, $\eta_p^2 = .107$, 90% CI [.032, .147], (Figures 3D-F). The Stimulus x Trial x Group interaction comparing the extinction and association splitting interventions was significant, $F(3.59, 154.26) = 6.15$, $p < .001$, $\eta_p^2 = .125$, 90% CI [.041, .193]. However, contrary to our hypotheses, comparing Figures 3D and 3F did not support that the association splitting intervention was more successful in reducing thinking of the aversive outcome than the extinction intervention. There were no significant differences between association splitting and no extinction as indicated by a non-significant Stimulus x Trial x Group

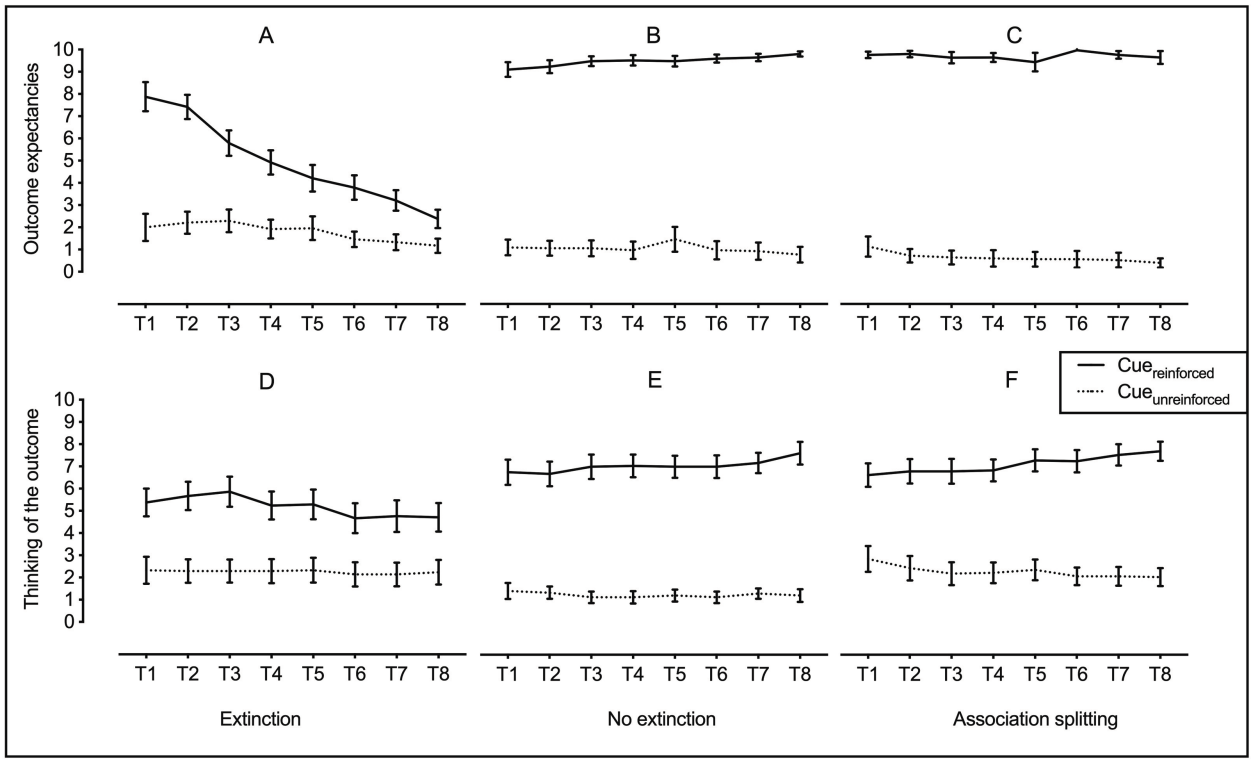


Figure 3. Panels A-C show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 1. Panels D-F show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 1. T = trial number.

interaction comparing both groups, $F(2.99, 137.39) = 1.38$, $p = .25$, $\eta_p^2 = .029$, 90% CI [.000, .071]. Although this Stimulus x Trial x Group interaction was significant in the analyses with the acquisition criterion, $F(2.76, 104.95) = 3.13$, $p = .03$, $\eta_p^2 = .076$, 90% CI [.003, .149], comparing Supplemental Figures 1E and 1F did not support that the association splitting intervention reduced thinking of the aversive outcome more than the no extinction control group (to the contrary, it seemed to increase thinking of).

We ran additional, explorative analyses to assess the reduction in each group separately. These analyses revealed that the Stimulus x Trial interaction in the extinction group was not

significant, $F(3.89, 77.84) = 2.07$, $p = .09$, $\eta_p^2 = .094$, 90% CI [.000, .168]. That is, we found no evidence that extinction training could significantly reduce thinking of the aversive outcome. The Stimulus x Trial interaction was significant in the no extinction group, $F(2.65, 60.87) = 3.19$, $p = .04$, $\eta_p^2 = .122$, 90% CI [.005, .228], and in the association splitting group, $F(2.72, 62.61) = 5.63$, $p < .001$, $\eta_p^2 = .197$, 90% CI [.048, .310]. However, Figures 3E and 3F again suggest that neither group was successful in decreasing thinking of the aversive outcome. In the analyses with the acquisition criterion, the Stimulus x Trial interaction in the no extinction group was not significant, $F(3.32, 63.14) = 2.23$, $p = .09$, $\eta_p^2 = .105$, 90% CI [.000, .196].

Discussion

Study 1 shows that thinking of an aversive outcome can be successfully acquired in a differential fear conditioning procedure. Moreover, extinction training seems to leave thinking of the aversive outcome relatively intact, while successfully reducing outcome expectancies. This stresses the clinical importance of including thinking of the aversive outcome as an additional variable and searching for an intervention that can successfully reduce it.

However, contrary to our predictions, association splitting was not successful in doing so. One reason for this could be that we used additional, aversive images rather than positive or neutral images during the association splitting intervention, which would be more in line with a typical clinical intervention (e.g., Jelinek, Hauschildt, Hottenrott, Kellner, & Moritz, 2014, 2018; Moritz & Jelinek, 2011; Moritz et al., 2007). If association splitting worked solely due to increasing the number of associations, then the valence of the additional outcomes should in principle not have mattered. However, the valence of the additional outcomes might still have played a role in an indirect way. For example, the competing aversive outcomes may have elicited thoughts about the initial outcome, because they were similar in valence and were

perhaps perceived as related in other ways (e.g., the aggressive dog could have caused the mutilation; Collins & Loftus, 1975).

Study 2

An intervention that may prove more successful in reducing thinking of an aversive outcome is counterconditioning. In a counterconditioning procedure, the original aversive outcome is replaced by an outcome of opposite valence (De Houwer, 2011; Hofmann et al., 2010). Research has shown that counterconditioning is successful in reducing previously acquired cue valence (e.g., Baeyens, Eelen, van den Bergh, & Crombez, 1989; Engelhard, Leer, Lange, & Olatunji, 2014). Given that cue valence and thinking of the aversive outcome are both hypothesized to stem from a referential relation between cue and outcome (Baeyens et al., 2001), it seems plausible that counterconditioning can reduce thinking of the aversive outcome as well. We also included a revised association splitting group with positive competing images instead of aversive competing images in this study to test whether the negative valence of the added outcomes was responsible for the failure of association splitting to reduce thinking of the aversive outcome in Study 1. In addition, we again included an extinction group to assess whether we would replicate our previous finding where we found no evidence for a successful reduction in thinking of the aversive outcome in the extinction group.

In summary, we compared a counterconditioning intervention (i.e., presenting the cue with a compound of two novel positive outcomes without the initial outcome), a revised associative splitting intervention (i.e., presenting the cue with a compound of the initial outcome and two novel, positive outcomes), and an extinction intervention (i.e., presenting the cue without an outcome) to each other. Moreover, we again included a no extinction control group.

As for thinking of the aversive outcome, we expected to replicate the findings of Study 1. That is, we expected that thinking of the aversive outcome could be successfully conditioned and that the treatment in the extinction and no extinction group would not significantly reduce it. In addition, we hypothesized that the counterconditioning and the revised association splitting intervention would not only be successful in decreasing thinking of the aversive outcome but that they would also be more successful in it than the extinction and no extinction group.

As for the outcome expectancies, we predicted that they could be successfully conditioned across all participants. The extinction and counterconditioning intervention were expected to significantly reduce outcome expectancies, while we expected the absence of such a significant reduction in the no extinction group. Moreover, we expected both the extinction and counterconditioning intervention to be more successful in decreasing outcome expectancies than the no extinction group.

Method

Preregistration. The experimental procedures and statistical analyses of Study 2 were preregistered on Aspredicted.org (http://aspredicted.org/blind.php?x=hy92e6). Additional analyses will be explicitly termed as "exploratory" in the results section.

Participants. Two-hundred first-year psychology students of KU Leuven (n_{female} = 178, n_{male} = 22; M_{age} = 18.38, SD_{age} = 0.84) were included in the data-set (see Supplemental Materials for additional participants' characteristics). Due to technical problems, eight participants could not complete the conditioning task and their data were replaced. In total, 208 participants took part in the study. Each of the four experimental conditions (i.e., extinction, no extinction, association splitting, counterconditioning) comprised fifty participants. No exclusion criteria were used because this experiment was part of a collective testing session, which should be open

to the entire first-year psychology student pool. However, participants could select the aversiveness level of the images (see Procedure) so that possibly vulnerable students could choose less aversive images. The study was approved by the Social and Societal Ethics Committee of KU Leuven. All participants gave written informed consent and received partial course credits for participation.

Apparatus and stimuli. The conditioning task was programmed and presented with Affect 5 software.

Stimuli for the fear conditioning procedure. Images of a small and big house (with black fill and white background) were used as the cues. The dimensions of the small house were 2 cm (width) x 2 cm (height) and the dimensions of the big house were 4 cm (width) x 5 cm (height). Allocation of the images to either reinforced or unreinforced cue was counterbalanced within experimental conditions. We did not use the same stimuli as in Study 1 because this pool of participants had been previously exposed to these stimuli in an unrelated experiment. Since we did not use exclusion criteria in Study 2, participants could select the aversiveness level (i.e., mild, moderate, high) of the outcome images (see Lenaert et al., 2014). The images for the initial outcome categorized as highly, moderately and mildly aversive included an image of a mutilated body (i.e., image 3071), a mutilated lip (i.e., image 9042), and a cockroach (i.e., image 7380), respectively. The images of a baby seal (i.e., image 1440) and a polar bear and her cub (i.e., image 1441) were used as the novel, positive outcome images (Engelhard et al., 2014) in the association splitting and counterconditioning conditions during Conditioning phase 2. All outcome images were taken from the IAPS (Lang et al., 2005) and were 6 cm wide and 4.5 cm high. An image of a star-shaped geometrical figure (with black fill and white background; outer

dimensions: 3 cm x 3 cm) was used as the cue for the practice trials. All stimuli were presented against a black background and were presented slightly above the center of the screen.

Outcome measures. The rating scales were identical to the ones in Study 1 (i.e., always referring to the initial outcome for the respective aversiveness level) and were presented for 14 s. The scale of the thinking of the aversive outcome ratings was always presented above the scale of the outcome expectancy ratings on the screen.

Procedure. Participants always entered the lab in groups of two to nine people. Participants gave written informed consent before they were seated and tested in separate cubicles. Participants then started with the conditioning task. Participants first selected the aversiveness level for the initial outcome image based on a content description of the images. They were asked to select an outcome aversiveness level that would be unpleasant but not unbearable for them. This procedure is in line with fear conditioning experiments that rely on electric shocks as the aversive outcome and in which participants can select the intensity of the electric shock but are asked to select an intensity that is "*unpleasant but not painful*" (Lenaert et al., 2014). Frequencies for the selected aversiveness levels are reported in the Supplemental Materials.

The conditioning task was identical to the task in Study 1 with two major exceptions. First, the revised association splitting intervention comprised presentations of the reinforced cue followed by a compound of the initial outcome and two novel, positive outcomes. Second, the counterconditioning intervention comprised presentations of the reinforced cue followed by a compound of two novel outcomes (i.e., without the initial outcome). The position (i.e., left or right) of the novel outcomes in the counterconditioning and association splitting intervention was determined randomly on each trial for each participant.

There were also minor methodological differences in the conditioning task between Study 1 and Study 2. The time restriction for completing both ratings scales may have been responsible for the many missing values in the analyses in Study 1. We, thus, extended the duration of the cue and rating scale presentation to 14 s to provide participants with more time to respond. Moreover, we included two practice trials with an otherwise not used stimulus prior to Conditioning phase 1 so that participants could familiarize themselves with the rating scales. There was no outcome presentation during the practice trials. Finally, the PTQ (Ehring et al., 2011) was no longer administered.

Data reduction and analyses. We ran the same analyses as for Study 1. Failure to respond within the provided response window resulted in missing values in the analyses. To test whether thinking of the aversive outcome can be successfully conditioned during Conditioning phase 1, the analyses were run on the data of the participants who had complete acquisition data (N = 163 and N = 183 for the expectancies and thinking of the aversive outcome, respectively). To compare the effectiveness of the different interventions during Conditioning phase 2, we conducted the analyses with and without the acquisition criterion that was used in Study 1. Here we report the analyses on the data of all participants who had complete data and without applying the acquisition criterion (N = 163 and N = 164 for the expectancies and thinking of the aversive outcome, respectively). However, we mention in the main text whenever a different conclusion was reached based on the analyses of the data with acquisition criterion. In addition, we report all analyses on the data with acquisition criterion (N = 160) for participants without missing values in Conditioning phase 2 (N = 133 and N = 136 for the expectancies and thinking of the aversive outcome, respectively) in the Supplemental Materials. An overview, which specifies the number of participants in the different stages can be found in the Supplemental

Materials as well.

Results

Conditioning phase 1.

Expectancies of the aversive outcome. There was a significant main effect of stimulus, $F(1, 159) = 871.62$, $p < .001$, $\eta_\rho^2 = .846$, 90% CI [.811, .869], and a Stimulus x Trial interaction, $F(4.32, 686.97) = 244.67$, $p < .001$, $\eta_\rho^2 = .606$, 90% CI [.569, .634], indicating that acquisition of outcome expectancies was successful (Figures 4A-D). Groups did not significantly differ in acquisition as shown by the non-significant Stimulus x Group interaction, $F(3, 159) = 0.81$, $p = .49$, $\eta_\rho^2 = .015$, 90% CI [.000, .043], and the non-significant Stimulus x Trial x Group interaction, $F(12.96, 686.97) = 0.77$, $p = .69$, $\eta_\rho^2 = .014$, 90% CI [.000, .012].

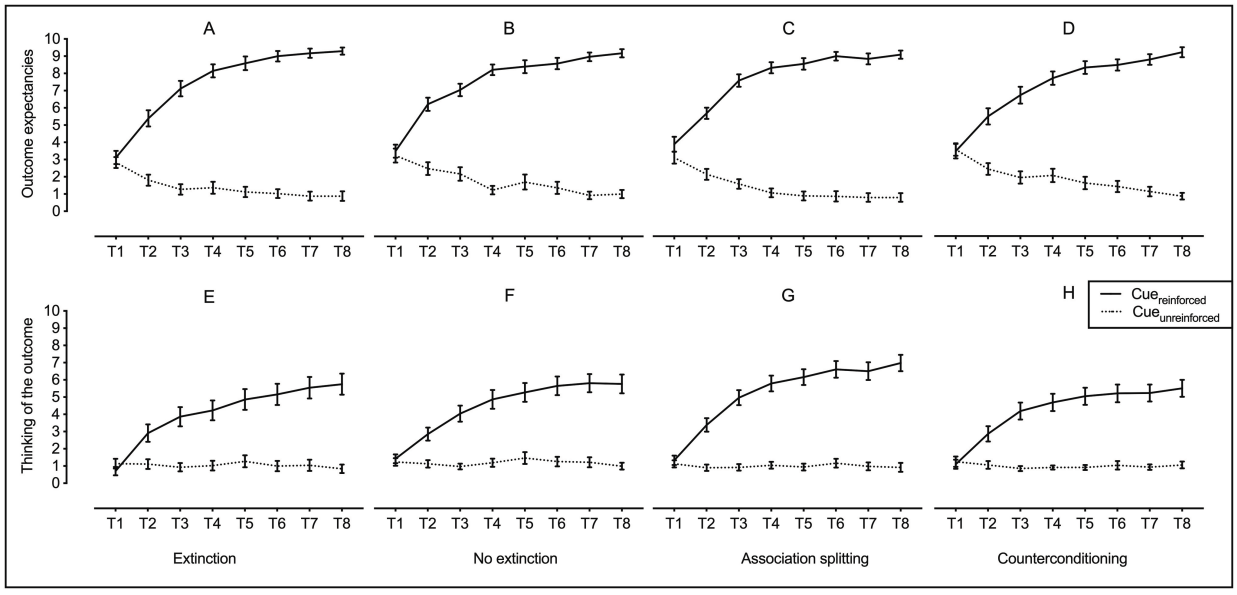


Figure 4. Panels A-D show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 1 of Study 2. Panels E-H show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 1 of Study 2. T = trial number.

Thinking of the aversive outcome. The main effect of stimulus, $F(1, 179) = 267.80$, $p < .001$, $\eta_p^2 = .599$, 90% CI [.526, .655], and the Stimulus x Trial interaction, $F(3.53, 631.97) = 117.56$, $p < .001$, $\eta_p^2 = .396$, 90% CI [.347, .436], were significant, suggesting that successful acquisition of thinking of the aversive outcome occurred (Figures 4E-H). There were no significant differences between groups in acquisition as shown by the non-significant Stimulus x Group, $F(3, 179) = 1.62$, $p = .19$, $\eta_p^2 = .026$, 90% CI [.000, .062], and Stimulus x Trial x Group interactions, $F(10.59, 631.97) = 0.62$, $p = .80$, $\eta_p^2 = .010$, 90% CI [.000, .008].

Conditioning phase 2.

Expectancies of the aversive outcome. There was a significant Stimulus x Trial x Group interaction, $F(11.66, 618.18) = 30.67$, $p < .001$, $\eta_p^2 = .366$, 90% CI [.307, .400], (Figures 5A-D). In line with our hypotheses, the extinction and counterconditioning groups differed from the no extinction group significantly, $F(3.70, 270.33) = 29.93$, $p < .001$, $\eta_p^2 = .291$, 90% CI [.211, .351], and, $F(4.01, 304.85) = 58.96$, $p < .001$, $\eta_p^2 = .437$, 90% CI [.364, .489], respectively (Figures 5A, 5B, and 5D). Explorative analyses revealed that the association splitting group did not significantly differ from the no extinction group, $F(3.48, 264.16) = 0.93$, $p = .43$, $\eta_p^2 = .012$, 90% CI [.000, .030].

Follow-up analyses showed a significant Stimulus x Trial interaction in the extinction group, $F(3.30, 131.85) = 34.82$, $p < .001$, $\eta_p^2 = .466$, 90% CI [.352, .537], and in the counterconditioning group, $F(3.31, 142.39) = 80.27$, $p < .001$, $\eta_p^2 = .651$, 90% CI [.569, .700], but not in the no extinction group, $F(2.87, 94.64) = 1.87$, $p = .14$, $\eta_p^2 = .054$, 90% CI [.000, .119]. Explorative analyses revealed that this was also not the case in the association splitting group, $F(3.19, 137.11) = 0.24$, $p = .88$, $\eta_p^2 = .006$, 90% CI [.000, .014]. While the extinction and counterconditioning interventions led to a significant reduction in outcome expectancies (Figures

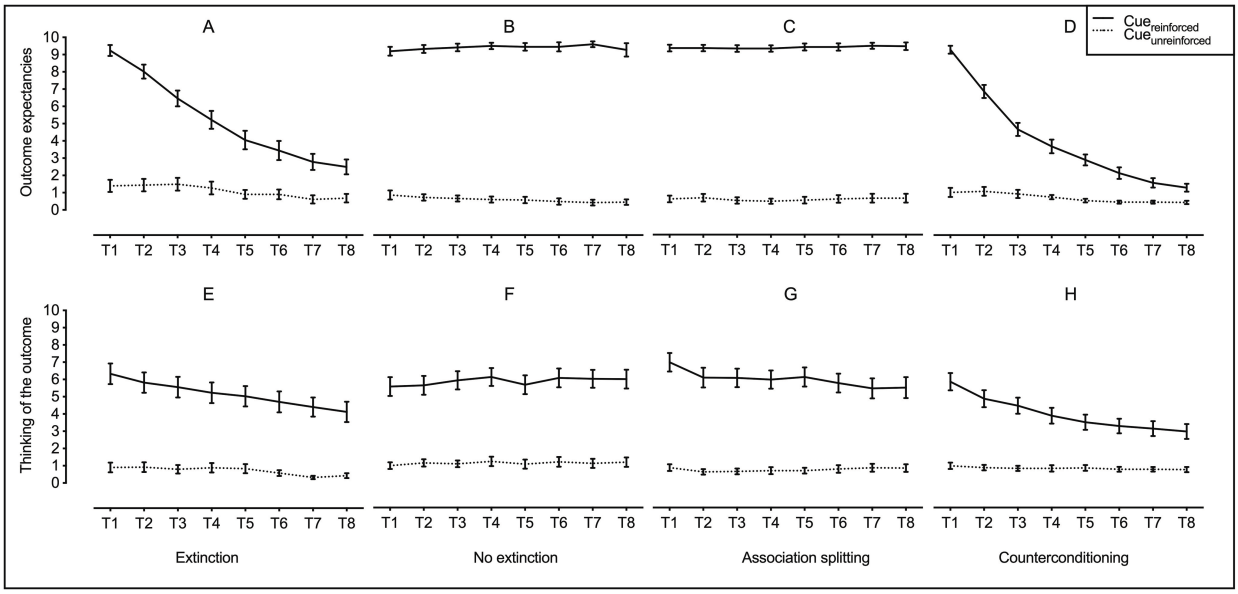


Figure 5. Panels A-D show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 2. Panels E-H show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 2. T = trial number.

5A and 5D), we obtained no evidence for this in the association splitting and no extinction group. In the analyses with the acquisition criterion, the Stimulus x Trial interaction in the no extinction group was significant, $F(3.18, 92.22) = 3.59$, $p = .01$, $\eta_p^2 = .110$, 90% CI [.013, .193]. However, as suggested by Supplementary Figure 3B, the no extinction group did not result in reducing outcome expectancies. Rather, there was a further differentiation between the reinforced and unreinforced cue.

Thinking of the aversive outcome. The Stimulus x Trial x Group interaction was significant, $F(10.39, 554.25) = 3.52$, $p < .001$, $\eta_p^2 = .062$, 90% CI [.019, .080] (Figures 5E-H). Both the association splitting and the counterconditioning group were more successful in reducing thinking of the aversive outcome than the no extinction group, $F(3.21, 259.64) = 3.11$, p

= .02, η_p^2 = .037, 90% CI [.003, .072], and, $F(3.44, 282.18) = 9.74$, $p < .001$, η_p^2 = .106, 90% CI [.049, .156], respectively (Figures 5F, 5G, and 5H).

Follow-up analyses revealed that there was a significant Stimulus x Trial interaction in the extinction group $F(3.16, 123.30) = 5.89$, $p < .001$, η_p^2 = .131, 90% CI [.038, .209], the association splitting group $F(2.78, 108.45) = 3.45$, $p = .02$, η_p^2 = .081, 90% CI [.007, .155], and the counterconditioning group, $F(3.50, 140.04) = 14.52$, $p < .001$, η_p^2 = .266, 90% CI [.153, .346], but not in the no extinction control group, $F(2.97, 124.68) = 0.49$, $p = .68$, η_p^2 = .012, 90% CI [.000, .037]. This suggests that all three manipulations reduced thinking of the aversive outcome (Figures 5E, 5G, and 5H).

To test whether there were differences among the successful manipulations, we conducted separate mixed-design ANOVAs. Based on a non-significant Stimulus x Trial x Group interaction, we obtained no evidence that the counterconditioning group was more successful in reducing thinking of the aversive outcome than the extinction group, $F(3.56, 281.28) = 1.52$, p = .20, η_p^2 = .019, 90% CI [.000, .042]. Likewise, there was no significant difference between the association splitting and the extinction group, $F(3.15, 245.94) = 0.44$, $p = .73$, η_p^2 = .006, 90% CI [.000, .017].

Importantly, when applying the acquisition criterion, the counterconditioning group was more successful in reducing thinking of the aversive outcome than the extinction group as indicated by the significant Stimulus x Trial x Group interaction, $F(3.67, 230.96) = 3.04$, $p = .02$, η_p^2 = .046, 90% CI [.004, .084] (Supplemental Figures 2E and 2H).

Discussion

Study 2 replicates the finding of Study 1 that thinking of the aversive outcome can be successfully conditioned. Unlike in Study 1, extinction training significantly reduced thinking of

the aversive outcome. The same was the case for the revised association splitting intervention. However, the counterconditioning intervention was more successful in reducing thinking of the aversive outcome than the extinction intervention. It should be noted that these differences between interventions were only found for participants who showed successful acquisition during Conditioning phase 1 (i.e., for the data with acquisition criterion). When considering all participants (i.e., without acquisition criterion), the differences between the interventions were absent. It can be argued that it is only useful to compare interventions if the responses to be targeted have been successfully acquired though. Even then, we should critically note that the effect size of this comparison was small.

General Discussion

We showed in two separate studies that thinking of the aversive outcome could be successfully conditioned in a fear conditioning procedure. Thinking of the aversive outcome survived extinction training in Study 1. Although we did observe an extinction effect in Study 2, counterconditioning was still more successful in reducing thinking of the aversive outcome among the participants who showed successful acquisition. Association splitting with positive competing images (Study 2) but not with negative competing images (Study 1) successfully reduced thinking of the aversive outcome.

While it has been proposed that referential relations would be insensitive to extinction training (Baeyens et al., 2001), our studies found mixed results concerning the extinction of thinking of the aversive outcome. Interestingly, a meta-analysis concluded that conditioned valence may not be completely insensitive to extinction training either, as had been previously assumed, but may just extinguish at a very slow rate (Hofmann et al., 2010).

It is important to note that being asked to indicate the extent to which the cue made participants think of the outcome did not make them report high levels of thinking of the outcome in itself. This was evidenced by the successful differential acquisition in Study 1 and 2 as well as by the reductions in thinking of the aversive outcome following the interventions in Study 2. That is, participants' response to the unreinforced cue remained low in the former and their response to the reinforced cue declined in the latter even though they were asked if they were thinking of the aversive outcome.

Contrary to our predictions, the association splitting intervention with negative images was not successful in reducing thinking of the aversive outcome. In contrast, the revised association splitting intervention with positive images was able to successfully reduce thinking of the aversive outcome. Accordingly, merely increasing the number of associations may not be sufficient and a change in valence may be necessary for association splitting to work. This challenges the previously proposed working mechanism underlying association splitting (Moritz & Jelinek, 2011; Moritz et al., 2007). Future research should further investigate the boundary conditions of association splitting procedures, which, in turn, could inform treatment protocols. It should also be noted that the conclusion that association splitting with positive (rather than negative) additional outcomes is more effective should, as of yet, be treated with caution, because it is based on a comparison across the two studies.

By presenting the novel outcomes in compound with the initial outcome we designed our association splitting intervention in such a way that it mirrored cue competition treatment (Boddez et al., 2014). This design has several advantages. First, it can control for counterconditioning effects because novel outcomes are added to the initial outcome instead of merely replacing the initial outcome with novel outcomes (as would be the case in

counterconditioning procedures; De Houwer, 2011; Hofmann et al., 2010). Second, it can also control for extinction effects because the initial outcome was still present during the association splitting intervention procedure. Therefore, the absence of the initial outcome (as would be the case in extinction procedures; Bouton, 1988; Hermans et al., 2006) cannot account for the effects on our dependent variables.

Importantly, the rationale for using this design assumes that the compound was processed in an elemental way and not in a holistic way (i.e., the compound should be seen as consisting of different stimuli and not as a new stimulus in its own right; Pearce, 1987, 1994). Our finding that expectancies of the original aversive outcome survived the association splitting intervention, but not the counterconditioning intervention, suggests that this was indeed the case (i.e., if the original outcome had not been recognized as an element of the compound by participants, then the expectancies would not have stayed intact in the association splitting group). Moreover, this difference in results between the counterconditioning and association splitting groups suggests that both interventions, as used in the present study, are distinct.

When it comes to the mechanisms that mediate extinction effects, there is no consensus (De Houwer, 2020). Nonetheless, one often-cited theory is inhibitory learning theory. This theory takes many forms (for an extensive discussion see Boddez, Moors, Mertens, & De Houwer, 2020) but in its original formulation it holds that an inhibitory association is acquired during extinction training. This inhibitory association (characterized as a negative value; e.g., -0.5) is supposed to counteract the original association which drives the conditioned response (characterized as a positive value; e.g., +0.5), resulting in low to no responding when both these associations are activated (because of a summed associative value close to 0; e.g., Vervliet, Craske, & Hermans, 2013). In other words, when the outcome representation becomes activated

by the excitatory association – and therefore comes to mind – there will be conditioned responding. In contrast, when this activation is counteracted by the inhibitory association – and the outcome therefore does not come to mind – there will be no responding. To account for return of conditioned responding after initially successful extinction (Bouton, 2002), it is further assumed that the inhibitory association is only effective in the extinction context.

Although further research is needed, our findings might pose a challenge to this theory. Consider the results of Study 1 in which we found a significant reduction in outcome expectancies but not in thinking of the outcome due to extinction training. The latter finding is difficult to reconcile with a strict interpretation inhibitory learning theory: If the representation of the outcome remains completely inactive, it seems hard to imagine that there could be any thinking of the outcome. Nonetheless, as of yet, this possible challenge for inhibitory learning theory should be interpreted with caution, as the results concerning thinking of the outcome were mixed in Study 2.

Thinking of an aversive outcome can become clinically relevant when it takes the form of intrusions. Intrusions are commonly defined as automatic and recurrent recollections, for example, of an aversive event, and are a symptom of several psychiatric disorders (e.g., posttraumatic stress disorder, social phobia; Clark, 2005; Harvey et al., 2004). For example, a patient suffering from social phobia may have intrusions of a public speaking incident. Our finding that thinking of the aversive outcome can be successfully conditioned is in line with the literature suggesting that, although intrusions often appear to “come out of the blue”, they may actually be elicited by cues that are (semantically) related to the aversive event (e.g., watching someone else give a talk) or – as in our case – by cues that are spatiotemporally linked to the

aversive event (e.g., the color of the room in which the speech was given; Ehlers & Clark, 2000; Michael, Ehlers, & Halligan, 2005).

When it comes to the treatment of patients, our extinction results suggest that repeated confrontation with the object of fear, as applied in exposure therapy, may not be the best way of reducing these types of responses. This is cause for concern given that exposure-based techniques are one of the most commonly used treatments for anxiety-related disorders (Deacon & Abramowitz, 2004; Kaczkurkin & Foa, 2015). As shown in Study 2, interventions such as counterconditioning may be more suitable as they can more easily target a wider variety of responses (see also Engelhard et al., 2014). For these reasons, such interventions may also be superior in reducing treatment relapse, which has been frequently observed following exposure treatment. Relapse rates are estimated to range from 19% to 62% (Craske & Mystkowski, 2006). In future research, procedures aimed at modelling relapse in the laboratory (e.g., see Vervliet et al., 2013) should be applied to compare the effects of both interventions.

It should be noted that we did not specifically collect data on participants' ethnicity or culture, education or socioeconomic status. However, the vast majority of participants in Study 1 and all participants in Study 2 were university students of KU Leuven and proficient in Dutch. We therefore have fairly homogeneous samples and generalization to other populations might be problematic (Henrich, Heine, & Norenzayan, 2010). This is especially the case since some of the processes that might be at play in conditioning studies (e.g., inferential reasoning; Boddez, Bennett, van Esch, & Beckers, 2017) may differ across populations (Henrich et al., 2010).

To conclude, we demonstrated in two separate studies that thinking of an aversive outcome can be successfully conditioned in a fear conditioning procedure. Our studies provide evidence that including such an outcome variable is important because extinction training, the

standard intervention for anxiety-related disorders, may not be the most optimal treatment to target it. Other interventions, such as counterconditioning, may be more successful. Thus, including thinking of the aversive outcome as an additional outcome variable can be a valuable extension to the fear conditioning procedure, which, in turn, cannot only enhance the understanding of anxiety-related disorders but also inform their treatment.

Author contributions

A.-K. Z. and Y.B. developed the study concept. A.-K. Z. and Y.B. designed the study, with feedback of the other authors. Testing and data collection were performed by A.-K. Z. A.-K. Z. and Y.B. performed the data analysis and interpretation. A.-K. Z. drafted the manuscript in collaboration with Y.B., while F.B. and T.B. provided critical revisions. All authors approved the final version of the manuscript for submission.

Conflict of interest

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgement

This study was supported by grant G076015N of the Research Foundation – Flanders (FWO). Yannick Boddez is supported by Ghent University grant BOF16/MET_V/002 awarded to Jan De Houwer.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474. http://doi.org/10.1016/0010-0285(74)90021-8

Baeyens, F., Díaz, E., & Ruiz, G. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion*, *19*(2), 245–268. http://doi.org/10.1080/02699930441000300

Baeyens, F., Eelen, P., van den Bergh, O., & Crombez, G. (1989). Acquired affective-evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy*, *27*(3), 279–287. http://doi.org/10.1016/0005-7967(89)90047-8

Baeyens, F., Vansteenwegen, D., Hermans, D., & Eelen, P. (2001). Chilled white wine, when all of a sudden the doorbell rings: Mere reference and evaluation versus expectancy and preparation in human Pavlovian learning. In F. Columbus (Ed.), *Advances in psychology research* (pp. 241–277). Huntington, NY: Nova Science Publishers, Inc.

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

Boddez, Y. (2018). The presence of your absence: A conditioning theory of grief. *Behaviour Research and Therapy*, *106*, 18–27. https://doi.org/10.1016/j.brat.2018.04.006

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 201–206. http://doi.org/10.1016/j.jbtep.2012.08.003

Boddez, Y., Bennett, M. P., van Esch, S., & Beckers, T. (2017). Bending rules: the shape of the

perceptual generalisation gradient is sensitive to inference rules. *Cognition and Emotion*, *31*(7), 1444–1452. https://doi.org/10.1080/02699931.2016.1230541

Boddez, Y., Haesen, K., Baeyens, F., & Beckers, T. (2014). Selectivity in associative learning: A cognitive stage framework for blocking and cue competition phenomena. *Frontiers in Psychology*, *5*, 1–13. http://doi.org/10.3389/fpsyg.2014.01305

Boddez, Y., Moors, A., Mertens, G., & De Houwer, J. (2020). Tackling fear : Beyond associative memory activation as the only determinant of fear responding. *Neuroscience and Biobehavioral Reviews*, *112*, 410–419. https://doi.org/10.1016/j.neubiorev.2020.02.009

Bouton, M. E. (1988). Context and ambiguity in the extinction of emotional learning: Implications for exposure therapy. *Behaviour Research and Therapy*, *26*(2), 137–149. http://doi.org/10.1016/0005-7967(88)90113-1

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. http://doi.org/10.1016/S0006-3223(02)01546-9

Clark, D. A. (2005). *Intrusive thoughts in clinical disorders: Theory, research, and treatment.* New York: Guilford Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Craske, M. G., & Mystkowski, J. L. (2006). Exposure therapy and extinction: Clinical studies. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and Learning: From Basic Processes to Clinical Implications* (pp. 217–33). Washington, DC: Am. Psychiatr. Assoc.

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

https://doi.org/10.1002/da.20633

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

De Houwer, J. (2020). Conditioning is more than association formation : On the different ways in which conditioning research is valuable for clinical psychology. *Collabra: Psychology*, *6*, 1–9. https://doi.org/DOI: https://doi.org/10.1525/collabra.239

Deacon, B. J., & Abramowitz, J. S. (2004). Cognitive and behavioral treatments for anxiety disorders: A review of meta-analytic findings. *Journal of Clinical Psychology*, *60*(4), 429–441. http://doi.org/10.1002/jclp.10255

Eelen, P., Hermans, D., & Baeyens, F. (2001). Learning perspectives on anxiety disorders. In E. J. L. Griez, C. Faravelli, D. Nutt, & J. Zohar (Eds.), *Anxiety disorders: An introduction to clinical management and research* (pp. 249–264). London: John Wiley and Sons.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*, 319–345.

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 225–232. http://doi.org/10.1016/j.jbtep.2010.12.003

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. http://doi.org/10.1016/j.beth.2014.04.003

Harvey, A. G., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes*

across psychological disorders: A transdiagnostic approach to research and treatment. New York: Oxford University Press, Inc. http://doi.org/10.1002/erv.647

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 2010. https://doi.org/10.1017/S0140525X0999152X

Hermans, D., & Baeyens, F. (2012). Expectancy learning and evaluative learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1203–1205). Berlin: Springer.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry*, *60*(4), 361–368. http://doi.org/10.1016/j.biopsych.2005.10.006

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. http://doi.org/10.1037/a0018916

Jelinek, L., Hauschildt, M., Hottenrott, B., Kellner, M., & Moritz, S. (2014). Further evidence for biased semantic networks in obsessive-compulsive disorder (OCD): When knives are no longer associated with buttering bread but only with stabbing people. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(4), 427–434. http://doi.org/10.1016/j.jbtep.2014.05.002

Jelinek, L., Hauschildt, M., Hottenrott, B., Kellner, M., & Moritz, S. (2018). "Association splitting" versus cognitive remediation in obsessive-compulsive disorder: A randomized controlled trial. *Journal of Anxiety Disorders*, *56*(February), 17–25. http://doi.org/10.1016/j.janxdis.2018.03.012

Jozefowiez, J. (2018). Associative versus predictive processes in Pavlovian conditioning. *Behavioural Processes*, *154*, 21–26. http://doi.org/10.1016/j.beproc.2017.12.016

Kaczkurkin, A. N., & Foa, E. B. (2015). Cognitive-behavioral therapy for anxiety disorders: An update on the empirical evidence. *Dialogues in Clinical Neuroscience*, *17*(3), 337–346. http://doi.org/10.4088/JCP.12r07757

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6.* University of Florida: Gainesville, FL.

Lenaert, B., Boddez, Y., Griffith, J. W., Vervliet, B., Schruers, K., & Hermans, D. (2014). Aversive learning and generalization predict subclinical levels of anxiety: A six-month longitudinal study. *Journal of Anxiety Disorders*, *28*(8), 747–753. http://doi.org/10.1016/j.janxdis.2014.09.006

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews*, *77*, 247–285. http://doi.org/10.1016/j.neubiorev.2017.02.026

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states : Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Depression*, *33*(3), 335–343.

Michael, T., Ehlers, A., & Halligan, S. L. (2005). Enhanced priming for trauma-related material in posttraumatic stress disorder. *Emotion*, *5*(1), 103–112. https://doi.org/10.1037/1528-3542.5.1.103

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience : Ten years of progress. *Annual Review of Psychology*, *63*, 129–151. https://doi.org/10.1146/annurev.psych.121208.131631

Miller, R. R., & Matute, H. (1998). Competition between outcomes. *Psychological Science*, *9*(2), 146–149. http://doi.org/10.1111/1467-9280.00028

Mineka, S., & Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychologica*, *127*, 567–580. https://doi.org/10.1016/j.actpsy.2007.11.007

Moritz, S., & Jelinek, L. (2011). Further evidence for the efficacy of association splitting as a self-help technique for reducing obsessive thoughts. *Depression and Anxiety*, *28*(7), 574–581. http://doi.org/10.1002/da.20843

Moritz, S., Jelinek, L., Klinge, R., & Naber, D. (2007). Fight fire with fireflies! Association splitting: A novel cognitive technique to reduce obsessive thoughts. *Behavioural and Cognitive Psychotherapy*, *35*(05), 1–5. http://doi.org/10.1017/S1352465807003931

Olatunji, B. O., Cisler, J. M., & Tolin, D. F. (2007). Quality of life in the anxiety disorders: A meta-analytic review. *Clinical Psychology Review*, *27*(5), 572–581. http://doi.org/10.1016/j.cpr.2007.01.015

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, *94*(1), 61–73. http://doi.org/10.1037/0033-295X.94.1.61

Pearce, J. M. (1994). Similarity and discrimination : A selective review and a connectionist model. *Psychological Review*, *101*(4), 587–607.

Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, *33*, 56–72. https://doi.org/10.1038/sj.npp.1301555

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York:

Appleton-Century-Crofts.

Scheveneels, S., Boddez, Y., & Hermans, D. (2019). Learning mechanisms in fear and anxiety: It is still not what you think it is. In B. O. Olatunji (Ed.), *The Cambridge handbook of anxiety and related disorders* (pp. 13–40). Cambridge: Cambridge University Press.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, *61*(4), 605–632.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology*, *57*(1), 36–45. http://doi.org/10.1027/1618-3169/a000005

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. http://doi.org/10.1037/0097-7403.32.1.71

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

Wegner, D. M., & Zanakos, S. (1994). Chronic Thought Suppression. *Journal of Personality*, *62*(4), 615–640. http://doi.org/10.1111/j.1467-6494.1994.tb00311.x

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., … Salvador-carulla, L. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655–679. https://doi.org/10.1016/j.euroneuro.2011.07.018

Supplemental Materials

Study 1

Below we report the analyses on the data of all participants who met the acquisition criterion (i.e., higher score on the reinforced cue than on the unreinforced cue on the last acquisition trial for both outcome expectancy and thinking of the aversive outcome ratings[1]; $N = 59$) and who had complete data ($N = 59$ and $N = 56$ for the expectancies and thinking of the aversive outcome, respectively) for Conditioning phase 2 of Study 1.

Conditioning phase 2.

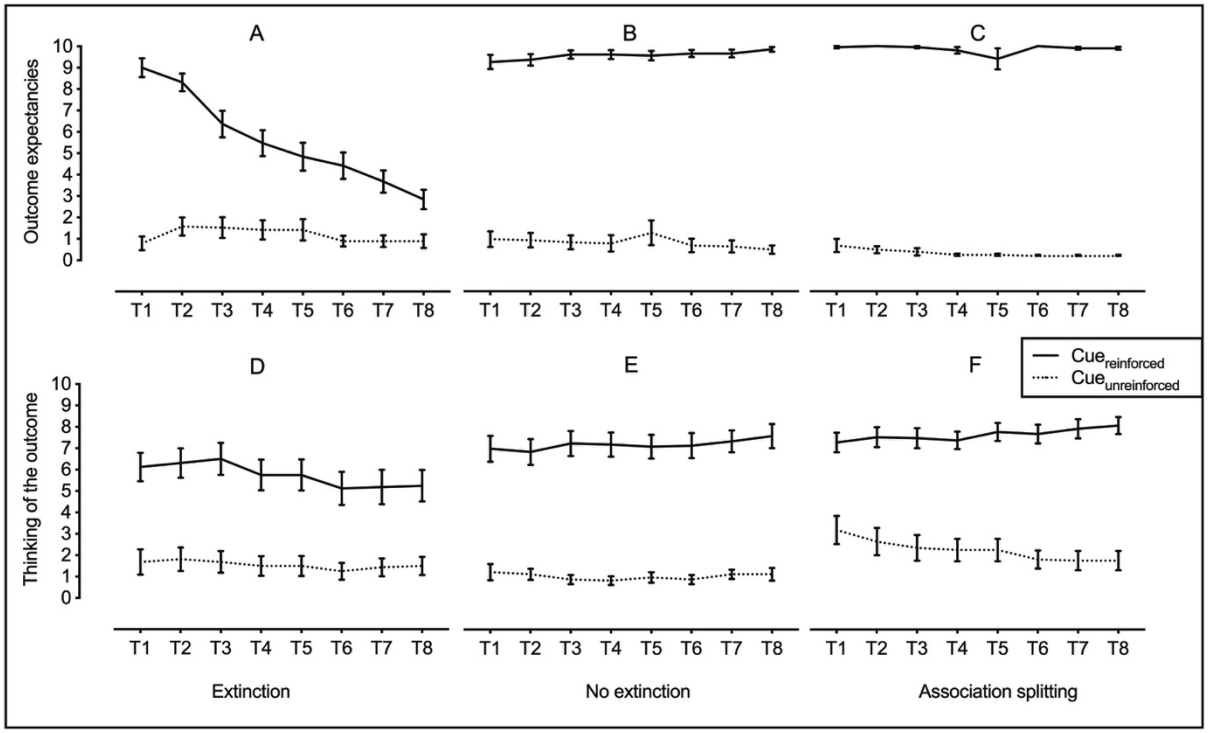
Expectancies of the aversive outcome. In accordance with our hypothesis, the Stimulus x Trial x Group interaction was significant, $F(6.82, 190.96) = 14.30$, $p < .001$, $\eta_p^2 = .338$, 90% CI [.229, .397]. As suggested by Supplemental Figures 1A-C and in line with our hypotheses, the extinction intervention was more successful in reducing outcome expectancies than the association splitting intervention and the no extinction group. This was confirmed by a significant Stimulus x Trial x Group interaction comparing extinction with association splitting, $F(3.21, 118.78) = 16.14$, $p < .001$, $\eta_p^2 = .304$, 90% CI [.179, .390], and comparing extinction with no extinction, $F(3.32, 122.81) = 18.00$, $p < .001$, $\eta_p^2 = .327$, 90% CI [.204, .411]. Explorative analyses showed that the Stimulus x Trial interaction in the extinction group was significant, $F(2.88, 51.75) = 16.67$, $p < .001$, $\eta_p^2 = .481$, 90% CI [.288, .581] (Supplemental Figure 1A). In contrast, the Stimulus x Trial interactions in both the no extinction and association splitting group were non-significant, $F(2.45, 46.59) = 2.29$, $p = .10$, $\eta_p^2 = .107$, 90% CI [.000, .225], and, $F(2.09, 39.69) = 0.97$, $p = .39$, $\eta_p^2 = .049$, 90% CI [.000, .155], respectively. That is, we only found evidence for successful extinction of the outcome expectancies in the extinction group.

Thinking of the aversive outcome. The Stimulus x Trial x Group interaction was significant, $F(6.92, 183.33) = 4.37$, $p < .001$, $\eta_p^2 = .142$, 90% CI [.047, .191] (Supplemental Figures 1D-F). The Stimulus x Trial x Group interaction comparing the extinction and the association splitting group was significant, $F(3.22, 109.60) = 6.67$, $p < .001$, $\eta_p^2 = .164$, 90% CI [.055, .249]. However, contrary to our hypotheses, comparing Supplemental Figures 1D and 1F did not support that the association splitting group was more successful in reducing thinking of the aversive outcome than the extinction group. Similarly, comparing Supplemental Figures 1E and 1F did not support that the association splitting intervention reduced thinking of the aversive outcome more than the no extinction control group, although the Stimulus x Trial x

[1] In case of missing values on the last trial, the value of the next to last trial was used.

Group interaction was significant, $F(2.76, 104.95) = 3.13$, $p = .03$, $\eta_\rho^2 = .076$, 90% CI [.003, .149].

To shed further light on this, we ran additional, explorative analyses to assess the reduction in each group separately. These analyses revealed that the Stimulus x Trial interaction in the extinction group was not significant, $F(3.67, 55.01) = 1.49$, $p = .22$, $\eta_\rho^2 = .090$, 90% CI [.000, .174]. That is, we found no evidence that extinction training could significantly reduce thinking of the aversive outcome. The Stimulus x Trial interaction was likewise non-significant in the no extinction group, $F(3.32, 63.14) = 2.23$, $p = .09$, $\eta_\rho^2 = .105$, 90% CI [.000, .196]. In contrast, the Stimulus x Trial interaction in the association splitting group was significant, $F(2.18, 41.46) = 7.44$, $p < .001$, $\eta_\rho^2 = .281$, 90% CI [.081, .420]. However, Supplemental Figure 1F suggests that association splitting was not successful in decreasing thinking of the aversive outcome as initially hypothesized, but rather increased it.



Supplemental Figure 1. Panels A-C show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 1 when the acquisition criterion was applied. Panels D-F show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 1 when the acquisition criterion was applied. T = trial number.

Supplementary Table 1

Participants' characteristics for each condition in Study 1

	Extinction (n = 24, 20 females)	No extinction (n = 24, 21 females)	Association Splitting (n = 24, 19 females)			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Age (range)	19.88 (2.95) (17-29)	22.04 (5.32) (18-42)	22.33 (18-47)	2,69	1.894	.158
DASS-D	2.67 (2.93)	4.67 (5.68)	5.00 (6.19)	2, 69	1.450	.242
DASS-A	4.83 (4.49)	3.25 (3.53)	3.67 (3.10)	2, 69	1.149	.323
DASS-S	6.83 (5.14)	8.33 (7.75)	8.75 (5.71)	2, 69	0.614	.544
PTQ	22.25 (6.77)	26.21 (9.48)	22.83 (8.26)	2, 69	1.613	.207
WBSI	46.12 (11.91)	49.17 (10.47)	45.83 (8.82)	2, 69	0.745	.478

Note. DASS-21, Depression Anxiety Stress Scales – 21 Items; D, depression subscale; A, anxiety subscale; S, stress subscale; the DASS sum scores were multiplied by 2 (Lovibond & Lovibond, 1995); PTQ, Preservative Thinking Questionnaire (Ehring et al., 2011); WBSI, White Bear Suppression Inventory (Wegner & Zanakos, 1994).

Supplementary Table 2
Ratings for the outcome images for each condition in Study 1

		Extinction	No extinction	Association splitting			
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Tension	Initial outcome	4.04 (2.63)	5.38 (2.60)	5.33 (1.95)	2, 69	2.368	.101
	Novel outcome1	NA	NA	4.75 (2.49)	-	-	-
	Novel outcome 2	NA	NA	5.88 (2.35)	-	-	-
Attention	Initial outcome	6.71 (2.46)	7.00 (1.72)	6.54 (1.77)	2, 69	0.320	.728
	Novel outcome1	NA	NA	6.33 (2.01)	-	-	-
	Novel outcome 2	NA	NA	7.58 (2.43)	-	-	-
Valence	Initial outcome	6.25 (1.89)	6.58 (1.38)	6.83 (1.52)	2, 69	0.789	.458
	Novel outcome1	NA	NA	6.17 (1.93)	-	-	-
	Novel outcome 2	NA	NA	8.04 (1.76)	-	-	-
Looking away	Initial outcome	3.37 (2.84)	5.33 (2.97)	4.00 (2.06)	2, 69	3.403	.039*
	Novel outcome1	NA	NA	3.29 (2.22)	-	-	-
	Novel outcome 2	NA	NA	5.83 (2.66)	-	-	-

Note. NA, not applicable (i.e., these images were not shown in this condition). Tension from 0 (*no tension*) to 10 (*much tension*); attention from 0 (*took no attention*) to 10 (*took much attention*); valence (*How pleasant/ unpleasant do you find the image?)* from 0 (*extremely pleasant*) to 10 (*extremely unpleasant*); looking away (*To what extent does the image make you look away?*) from 0 (*does not make me look away*) to 10 (*does make me look away*). The outcome images were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2005). Initial outcome = image of a mutilated body (image 3051),

novel outcome 1 = an aggressive dog (image 1300), novel outcome 2 = thoracotomy (image 3250). * $p < .05$.

Supplementary Table 3

Number of participants who had already seen the outcome images before the present study took place for each condition in Study 1

	Extinction		No extinction		Association splitting	
	Yes	No	Yes	No	Yes	No
Initial outcome	1	23	0	24	5	19
Novel outcome 1	NA	NA	NA	NA	1	23
Novel outcome 2	NA	NA	NA	NA	0	24

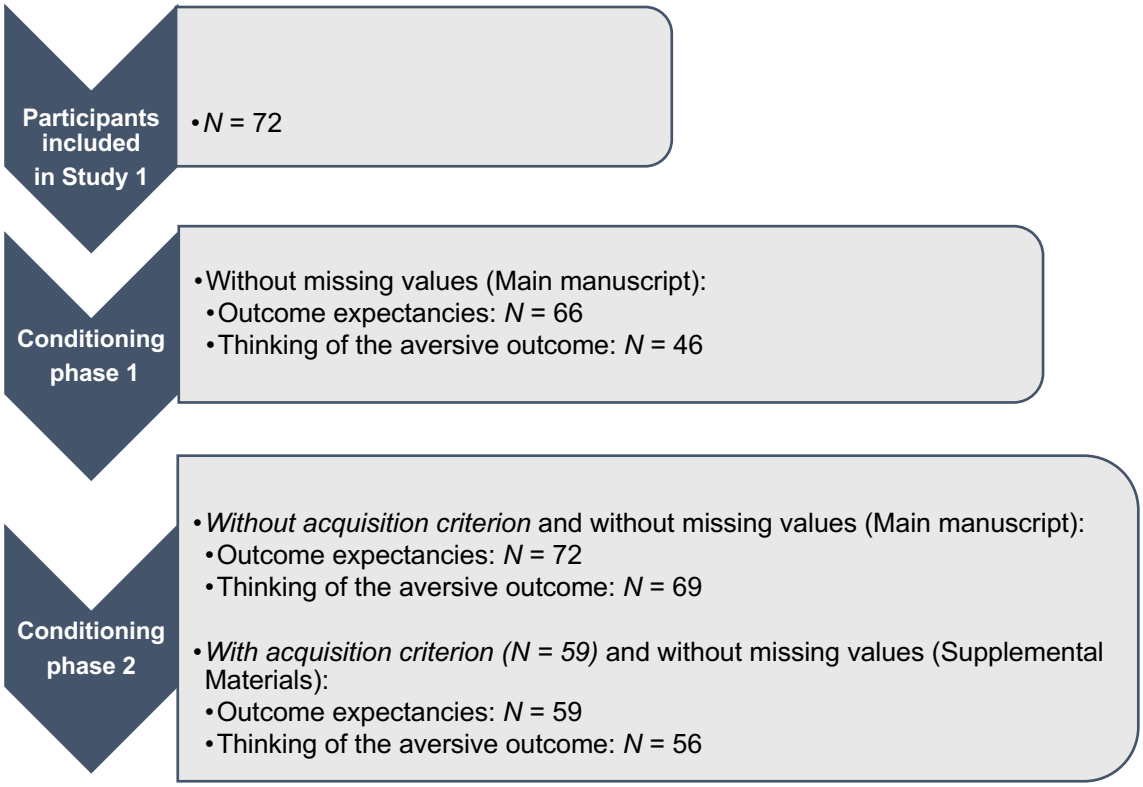
Note. NA, not applicable (i.e., these images were not shown in this condition).The outcome images were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2005). Initial outcome = image of a mutilated body (image 3051), novel outcome 1 = an aggressive dog (image 1300), novel outcome 2 = thoracotomy (image 3250). * $p < .05$.

Supplementary Table 4

Valence ratings for the cues for each condition in Study 1

	Extinction	No extinction	Association splitting			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Unreinforced cue	2.92 (2.17)	2.08 (2.24)	2.50 (1.89)	*2, 69*	0.941	.395
Reinforced cue	3.88 (2.40)	5.00 (2.59)	4.54 (2.38)	2, 69	1.272	.287

Note. Valence (*How pleasant/ unpleasant do you find this image?*) from 0 *(extremely pleasant)* to 10 *(extremely unpleasant).*



Supplementary Figure 2. Overview of the number of participants that were reported in the analyses for each conditioning phase for both outcome measures in Study 1.

Study 2

Below we report the analyses on the data of all participants who met the acquisition criterion (i.e., higher score on the reinforced cue than on the unreinforced cue on the last acquisition trial for both outcome expectancy and thinking of the aversive outcome ratings[2]; $N = 160$) and who had complete data ($N = 133$ and $N = 136$ for the expectancies and thinking of the aversive outcome, respectively) for Conditioning phase 2 of Study 2.

Conditioning phase 2.

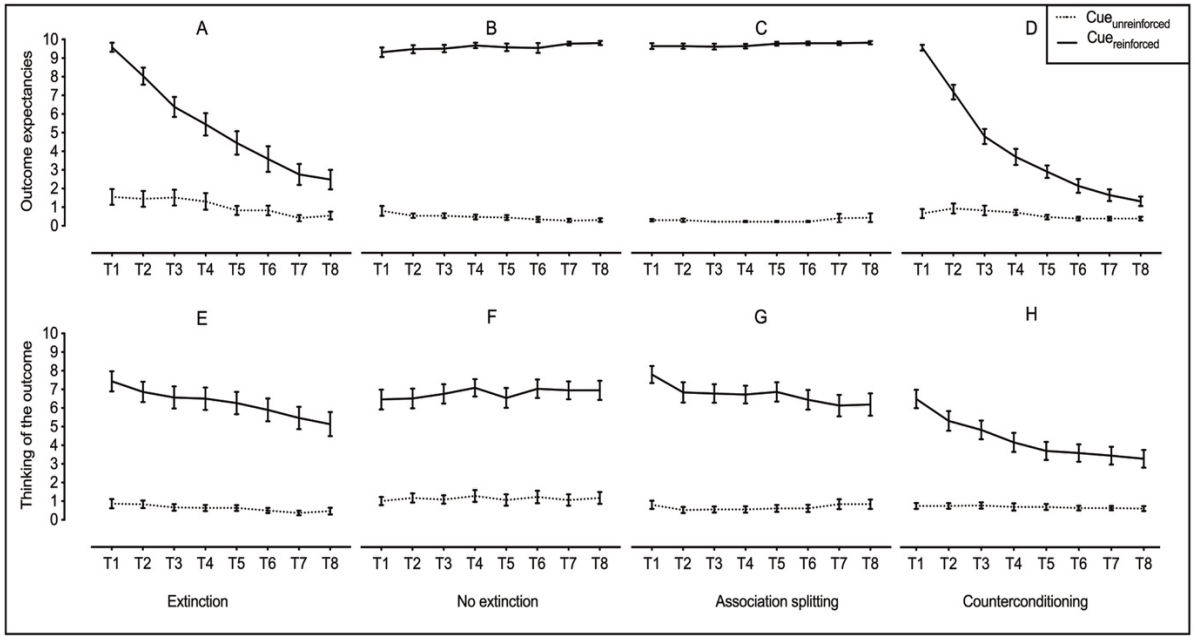
Expectancies of the aversive outcome. There was a significant Stimulus x Trial x Group interaction, $F(12.51, 537.76) = 33.45$, $p < .001$, $\eta^2_p = .438$, 90% CI [.376, .470], (Supplemental Figures 2A-D). In line with our hypotheses, the extinction and the counterconditioning group differed from the no extinction group significantly, $F(4.00, 227.73) = 28.92$, $p < .001$, $\eta^2_p = .337$, 90% CI [.247, .400], and, $F(3.99, 255.09) = 71.18$, $p < .001$, $\eta^2_p = .527$, 90% CI [.453, .576], respectively (Supplemental Figure 2A, 2B and 2D).

[2]In case of missing values on the last trial, the value of the next to last trial was used.

Explorative analyses revealed that the association splitting group did not differ significantly from the no extinction group, $F(3.19, 210.51) = 2.10$, $p = .10$, $\eta_\rho^2 = .031$, 90% CI [.000, .066]. Follow-up analyses showed a significant Stimulus x Trial interaction in the extinction group, $F(3.68, 102.95) = 24.98$, $p < .001$, $\eta_\rho^2 = .472$, 90% CI [.339, .548], the counterconditioning group, $F(3.37, 117.85) = 84.77$, $p < .001$, $\eta_\rho^2 = .708$, 90% CI [.628, .752], and in the no extinction group, $F(3.18, 92.22) = 3.59$, $p = .01$, $\eta_\rho^2 = .110$, 90% CI [.013, .193]. While the extinction and counterconditioning groups led to a reduction in outcome expectancies (Supplemental Figures 2A and 2D), this was not the case in the no extinction group as suggested by Supplemental Figure 2B. Explorative analyses revealed that the Stimulus x Trial interaction was not significant in the association splitting group, $F(2.42, 89.44) = 0.45$, $p = .68$, $\eta_\rho^2 = .012$, 90% CI [.000, .049]. That is, no evidence was obtained that the association splitting group could successfully reduce outcome expectancies.

Thinking of the aversive outcome. The Stimulus x Trial x Group interaction was significant $F(10.17, 447.37) = 4.07$, $p < .001$, $\eta_\rho^2 = .085$, 90% CI [.031, .108], (Supplemental Figures 2E-H). Both the association splitting and the counterconditioning group were more successful in reducing thinking of the aversive outcome than the no extinction group, $F(3.14, 216.62) = 3.19$, $p = .02$, $\eta_\rho^2 = .044$, 90% CI [.003, .085], and, $F(3.28, 226.57) = 10.31$, $p < .001$, $\eta_\rho^2 = .130$, 90% CI [.061, .189], respectively (Supplemental Figures 2F, 2G, and 2H).

Follow-up analyses revealed that there was a significant Stimulus x Trial interaction in the extinction group $F(3.61, 104.67) = 7.87$, $p < .001$, $\eta_\rho^2 = .213$, 90% CI [.088, .301], the association splitting group $F(2.72, 92.32) = 3.41$, $p = .02$, $\eta_\rho^2 = .091$, 90% CI [.007, .173], and the counterconditioning group, $F(3.39, 115.40) = 15.89$, $p < .001$, $\eta_\rho^2 = .318$, 90% CI [.190, .404], but not in the no extinction control group, $F(2.88, 100.88) = 0.57$, $p = .63$, $\eta_\rho^2 = .016$, 90% CI [.000, .051]. This suggests that all three manipulations reduced thinking of the aversive outcome (Supplemental Figures 2E, 2G, and 2H). To test whether there were differences among the successful manipulations, we conducted separate mixed-design ANOVAs. A significant Stimulus x Trial x Group interaction revealed that the counterconditioning group was more successful in reducing thinking of the aversive outcome than the extinction group, $F(3.67, 230.96) = 3.04$, $p = .02$, $\eta_\rho^2 = .046$, 90% CI [.004, .084] (Supplemental Figures 2E and 2H). There was no significant difference between the association splitting and the extinction group, $F(3.01, 189.67) = 0.20$, $p = .90$, $\eta_\rho^2 = .003$, 90% CI [.000, .008].



Supplemental Figure 3. Panels A-D show the expectancies of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 2 when the acquisition criterion was applied. Panels E-H show thinking of the aversive outcome (mean ± SEM) for each condition during Conditioning phase 2 of Study 2 when the acquisition criterion was applied. T = trial number

Supplementary Table 5

Participants' characteristics for each condition in Study 2

	Extinction (n = 50, 41 females)	No extinction (n = 50, 49 females)	Association Splitting (n = 50, 42 females)	Counter-conditioning (n = 50, 46 females)			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Age (range)	18.30 (0.58) (18-20)	18.26 (1.03) (18-25)	18.46 (0.89) (18-23)	18.48 (0.79) (18-22)	3, 196	0.884	.450
DASS-D	7.00 (9.70)	6.60 (8.00)	8.92 (8.78)	7.40 (6.29)	3, 196	0.748	.524
DASS-A	7.96 (8.17)	7.16 (7.30)	6.76 (5.76)	5.52 (6.00)	3, 196	1.095	.352
DASS-S	11.12 (8.40)	12.16 (9.14)	12.56 (7.51)	10.56 (8.09)	3, 196	0.615	.606
WBSI	49.88 (12.66)	51.22 (10.23)	50.70 (12.14)	48.14 (13.85)	3, 196	0.602	.615

Note. DASS-21, Depression Anxiety Stress Scales – 21 Items; D, depression subscale; A, anxiety subscale; S, stress subscale; the DASS sum scores were multiplied by 2 (Lovibond & Lovibond, 1995); WBSI, White Bear Suppression Inventory (Wegner & Zanakos, 1994).

Supplementary Table 6

Frequencies for the selected aversiveness levels of the initial outcome images for each condition in Study 2

		Extinction	No extinction	Association splitting	Counter-conditioning	Total
Aversive-ness level	mild	4	5	4	5	18
	moderate	26	28	22	25	101
	high	20	17	24	20	81

Note. All outcome images were taken from the International Affective Picture System (IAPS; Lang et al., 2005); mild = image of a cockroach (image 7380), moderate = image of a mutilated lip (image 9042), high = image of a mutilated body (image 3071).

Supplementary Table 7

Ratings for the outcome images for each condition in Study 2

		Extinction	No extinction	Associa-tion splitting	Counter-condition-ing			
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Tension	Initial Outcome	3.94 (2.45)	4.86 (2.35)	4.66 (2.80)	4.86 (2.11)	3, 195	1.586	.194
	Novel outcome1	NA	NA	0.82 (1.98)	1.02 (2.25)	-	-	-
	Novel outcome 2	NA	NA	1.72 (3.24)	0.80 (2.02)	-	-	-
Attention	Initial outcome	5.12 (2.43)	5.34 (1.97)	5.78 (2.19)	6.20 (2.15)	3, 195	2.391	.070
	Novel outcome1	NA	NA	5.26 (3.14)	6.20 (2.78)	-	-	-
	Novel outcome 2	NA	NA	5.46 (2.92)	6.14 (2.29)	-	-	-
Valence	Initial outcome	5.94 (1.83)	5.90 (1.75)	6.62 (1.77)	7.00 (1.34)	3, 196	5.073	.002*
	Novel outcome1	NA	NA	1.48 (2.89)	1.48 (3.09)	-	-	-
	Novel outcome 2	NA	NA	2.06 (3.40)	1.28 (2.79)	-	-	-
Looking away	Initial outcome	3.98 (3.11)	4.78 (2.81)	4.70 (3.07)	5.48 (2.14)	3, 196	2.386	.070
	Novel outcome1	NA	NA	0.86 (2.09)	0.20 (0.64)	-	-	-
	Novel outcome 2	NA	NA	0.52 (1.34)	0.22 (0.68)	-	-	-

Note. NA, not applicable (i.e., these images were not shown in this condition). Tension from 0 (*no tension*) to 10 (*much tension*); attention from 0 (*took no attention*) to 10 (*took much attention*); valence (*How pleasant/ unpleasant do you find the image?)* from 0 (*extremely pleasant*) to 10 (*extremely unpleasant*); looking away (*To what extent does the image make you look away?*) from 0 (*does not make me look away*) to 10 (*does make me look away*). All outcome images were taken from the International Affective Picture System (IAPS; Lang et al., 2005); Initial outcome, depending on the aversiveness level = image of a cockroach (image 7380, mild), a mutilated lip (i.e., image 9042, moderate) or mutilated body (i.e., image 3071, high), novel outcome 1 = image of a baby seal (image 1440), novel outcome 2 = a polar bear and her cub (image 1441). * $p < .05$.

Supplementary Table 8

Number of participants who had already seen the outcome images before the present study took place for each condition in Study 2

	Extinction		No extinction		Association splitting		Counter-conditioning	
	Yes	No	Yes	No	Yes	No	Yes	No
Initial outcome	4	46	3	47	4	46	0	50
Novel outcome 1	NA	NA	NA	NA	12	38	6	44
Novel outcome 2	NA	NA	NA	NA	5	45	5	45

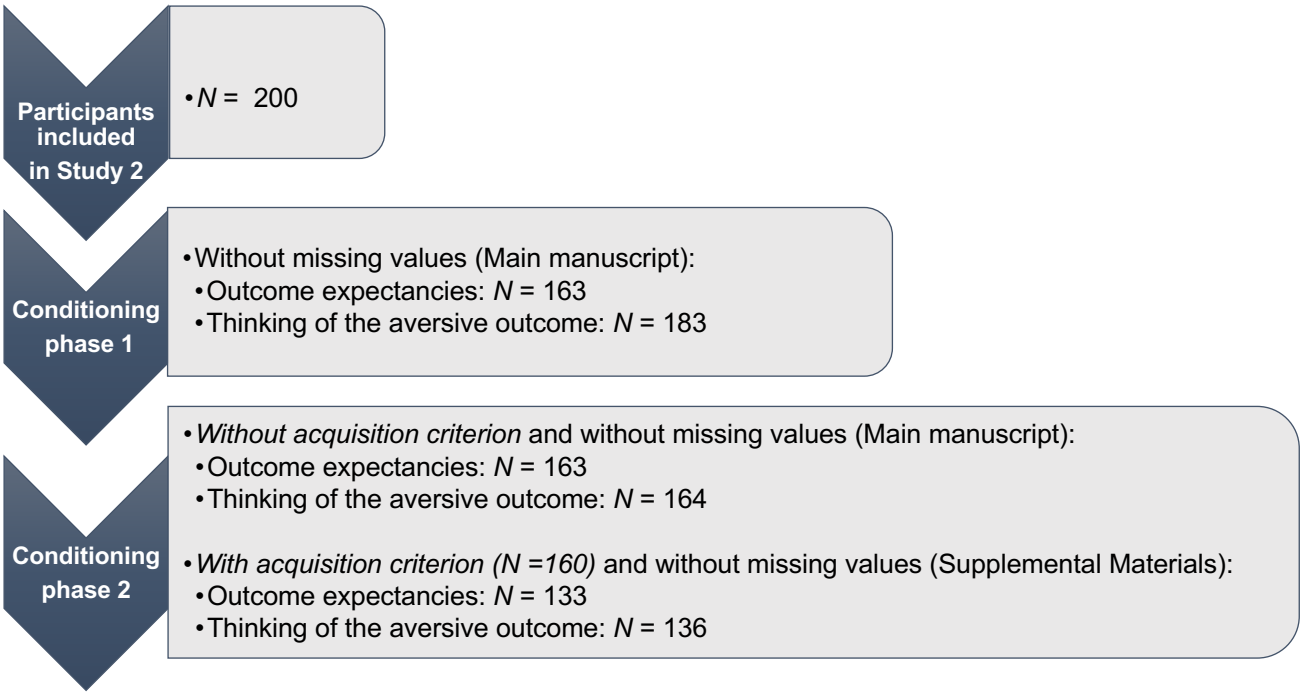
Note. NA, not applicable (i.e., these images were not shown in this condition). All outcome images were taken from the International Affective Picture System (IAPS; Lang et al., 2005); Initial outcome, depending on the aversiveness level = image of a cockroach (image 7380, mild), a mutilated lip (i.e., image 9042, moderate) or mutilated body (i.e., image 3071, high), novel outcome 1 = image of a baby seal (image 1440), novel outcome 2 = a polar bear and her cub (image 1441).

Supplementary Table 9

Valence ratings for the cues for each condition in Study 2

	Extinc-tion	No extinction	Associa-tion splitting	Counter-condition-ing			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*df*	*F*	*p*
Unreinforced cue	1.66 (2.06)	1.43 (1.84)	3.74 (1.80)	3.86 (2.36)	3, 195	20.661	.000**
Reinforced cue	2.96 (2.44)	3.59 (2.75)	4.62 (2.04)	4.02 (1.90)	3, 194	4.606	.004*
Practice cue	2.46 (2.31)	2.28 (2.00)	4.02 (1.70)	4.22 (2.12)	3, 196	12.374	.001**

Note. Valence *(How pleasant/ unpleasant do you find this image?)* from 0 *(extremely pleasant)* to 10 *(extremely unpleasant).* The practice cue was only presented during the two practice trials prior to Conditioning phase 1. * $p < .05$, ** $p < .001$.



Supplementary Figure 4. Overview of the number of participants that were reported in the analyses for each conditioning phase for both outcome measures in Study 2.