KU LEUVEN

Acknowledgement	Clinical Psychological Science, 2, 336-343
Archived version	Author manuscript: post-print (post-refereeing)
Published version	http://dx.doi.org/10.1177/2167702613503139
Journal homepage (Copyright)	http://cpx.sagepub.com/ ISSN: 2167-7026, ESSN: 2167-7034

(article begins on next page)

Avoided by Association: Acquisition, Extinction, and Renewal of Avoidance Tendencies towards Conditioned Fear Stimuli

Angelos-Miltiadis Krypotos, Marieke Effting, Inna Arnaudova, Merel Kindt

University of Amsterdam

Tom Beckers

University of Amsterdam and KU Leuven

Preprint version, including supplemental materials. Final version published in *Clinical Psychological Science*, published online before print October 7, 2013, doi: 10.1177/2167702613503139.

The preprint version may deviate considerably from the version as published. Please use the published version for citation and reference.

Author Note

This research was supported by an Innovation Scheme (Vidi) Grant of the Netherlands Organisation for Scientific Research (NWO) awarded to TB. MK is supported by an NWO Innovation Scheme (Vici) Grant. AMK is a scholar of the Onassis Foundation.

Abstract

Traditional theoretical models hold that avoidance reflects the interplay of Pavlovian and instrumental learning. Here we suggest that avoidance tendencies to intrinsically neutral cues may be established by mere Pavlovian association. Following fear conditioning, in which pictures of one object were paired with shock (CS^+) whereas pictures of another object were not (CS^-), CS^+ pictures facilitated avoidance reactions and interfered with approach responses, relative to CS^- pictures, in a symbolic approach/avoidance reaction time task. This was achieved without any instrumental relation between responses and CS continuation or US presentation. Moreover, those avoidance tendencies were sensitive to Pavlovian extinction (they were reduced after repeated presentations of the CS^+ without shock) and renewal (recovery of conditioned responding upon returning to the initial conditioning context after extinction in a different context). The present results may help to understand the self-perpetuating nature of pathological fear and anxiety.

Avoided by Association: Acquisition, Extinction, and Renewal of Avoidance Tendencies towards Conditioned Fear Stimuli

Avoidance behavior is a core concomitant of normal anxiety. Nonetheless, excessive avoidance of essentially neutral situations and non-dangerous entities can impair individuals' functioning and prevent confrontation with anxiety-correcting information (Barlow, 2002). Accordingly, avoidance is a key feature of several anxiety disorders (APA, 2000) making it critical to understand the factors that control such behavior.

Historically, avoidance behavior has been assumed to reflect the interplay of Pavlovian and instrumental learning. According to two-factor theory (Mowrer, 1960), an initially neutral cue (conditioned stimulus; CS) that reliably precedes an aversive outcome (unconditioned stimulus; US) will come to elicit (conditioned) fear (Pavlovian component). Avoidance responses that lead to CS termination or omission will be positively reinforced by reduction of CS-elicited fear (instrumental component).

Subsequent animal research suggested that rather than CS termination, it is omission of the anticipated US that instrumentally reinforces avoidance (Bolles, Stokes, & Younger, 1966; Seligman & Johnston, 1973). Alternative cognitive theories of avoidance learning have since been postulated, stressing the role of explicit knowledge about the CS-US association on the one hand and between the avoidance response and US-omission on the other (Declercq, De Houwer, & Baeyens, 2008; Lovibond, Saunders, Weidemann, & Mitchell, 2008). Although, unlike two-factor theory, those newer accounts are mainly concerned with primary avoidance (responses aimed at avoiding the aversive US) rather than secondary avoidance (responses aimed at escaping the fearsome CS), a common characteristic of the aforementioned avoidance theories is their reliance on a combination of Pavlovian and instrumental processes.

Still, some findings suggest that explicit omission of an aversive outcome is not always involved in establishing avoidance behavior. Miller (1944) reported that rats, after being shocked in a maze, avoid that part of the maze in following trials, although avoidance is never explicitly paired with shock omission or termination. Such findings of CS-oriented or secondary avoidance have been invoked as an argument to retain two-factor theory's reinforcement-through-fear-reduction account of avoidance (e.g., Walker, 1987).

Here, we aim to test the possibility that, instead of relying on instrumental reinforcement through fear reduction, a CS-elicited avoidance tendency can be a direct consequence of *mere* Pavlovian learning. That is, we propose that once a CS has been established as a predictor of an aversive event and induces fear, it will automatically elicit avoidance *without* the involvement of an instrumental component (be it fear-reduction or outcome omission).

This hypothesis fits with contemporary emotion theories, which propose that fear, like any emotion, is a composite of subjective experience, physiological activity, and action disposition (Mauss & Robinson, 2009). For many theorists, emotions are even primarily action tendencies (Frijda, 2010; Lang, 1985). In such framework, a CS eliciting fear after pairing with a threatening US implies that it also elicits an avoidance tendency – to fear *equals* to have the urge to avoid. CS-presentation prompting avoidance *behavior* then merely means that the avoidance *tendency* installed through Pavlovian fear learning is translated into overt action.

In Experiment 1, we tested whether following Pavlovian differential fear learning in which a neutral picture (CS^+) was paired with an electric stimulus (shock US), the CS^+ would elicit an automatic avoidance tendency (relative to a stimulus never paired with shock; CS^-) in a speeded approach/avoid reaction time task (AAT). Importantly, the AAT precluded all possibility for differential reinforcement of approach or avoidance: Responses had no influence on CS-presentation or US-occurrence. Moreover, shock electrodes were removed before the AAT, excluding any US-centered basis for CS-elicited avoidance. Also, participants equally often approached and avoided CSs in the AAT, preventing the experience of differential approach/avoidance and US-omission associations. To evaluate the automaticity of CS-elicited avoidance, we manipulated whether the stimulus dimension relevant for responding in the AAT was the same or different as the dimension distinguishing the CS^+ from the CS^- (see below).

Experiment 1

Inherently aversive stimuli yield an avoidance tendency, revealed through symbolic approach/avoidance reaction time tasks (Krieglmeyer & Deutsch, 2010). In such tasks, reaction times (RTs) are influenced by the congruency between the hedonic nature of the stimulus (appetitive vs. aversive) and the to-be-emitted response (approach vs. avoidance). Specifically, participants react faster on congruent trials (approaching appetitive stimuli, avoiding aversive stimuli) than incongruent trials (avoiding appetitive stimuli, approaching aversive stimuli). The difference in RTs is considered a measure of automatic action tendencies.

Here we test the idea that neutral stimuli previously paired with an aversive US will elicit, by association, a similar avoidance tendency. First, participants underwent differential conditioning with pictures of one neutral object (CS^+) consistently paired with shock, whereas pictures of another object were never paired with shock (CS^-). In the subsequent AAT, the same pictures were presented one-by-one accompanied by a manikin. Participants had to move the manikin towards or away from the pictures as quickly as possible. We hypothesized that participants would move the manikin faster away from CS^+ pictures than towards them, relative to CS^- pictures.

To evaluate the automaticity of CS-elicited avoidance tendencies, we manipulated the stimulus dimension to which participants had to respond in the AAT. One group of participants had to make the manikin approach and avoid pictures on the basis of the same stimulus feature that discriminated the CSs (i.e., stimulus shape; relevant feature group). The other group had to respond on the basis of a feature that was irrelevant for CS^+/CS^- discrimination (i.e., orientation of the surrounding frame; irrelevant feature group). If avoidance is under voluntary control, it should be expressed more readily when task demands require processing of stimulus features that supposedly elicit avoidance. However, if avoidance tendencies are triggered automatically, they might be also observed when people do not explicitly have to process features that discriminate CS^+ and CS^- to execute the AAT (Moors & De Houwer, 2006).

Method

Participants

Thirty-five adults participated for course credits or monetary reward (€7). Three participants were excluded[1] for lack of contingency awareness (see Exit Interview; Cornwell, Echiverri, & Grillon, 2007). The remaining 32 (23 females; mean age: 22.09 years; range: 18-31) were randomly assigned to the experimental groups.

Stimuli and apparatus

Pictures (50 mm × 50 mm) of a cube and a cylinder, depicted from four viewpoints against a white frame (100 mm × 100 mm), served as CS^+ or CS^- (counterbalanced across participants).

The US was a 2-ms electric stimulation delivered through two Ag electrodes to the wrist of the non-preferred hand (Effting & Kindt, 2007).

For the AAT, the same objects as for conditioning were used. However, the surrounding frame differed between groups. For the relevant feature group, each object was surrounded by the same frame as during conditioning. For the irrelevant feature group, objects were depicted against a landscape (100 mm × 55 mm) or portrait (55 mm × 100 mm) frame.

CS and US Measurements

During fear conditioning, US-expectancy ratings were recorded trial-by-trial, within 7 s after CS onset, on a scale from -5 (*certainly no electric stimulus*) to +5 (*certainly an electric stimulus*).

After the AAT, participants rated the valence of the CSs (from -5, *negative,* to +5, *positive*), US pleasantness (from -5, *unpleasant,* to +5, *pleasant*), US intensity (*weak, moderate*, *intense, enormous*, *unbearable*) and US startlingness (*not*, *light*, *moderate*, *strong*, *too strong*).

Procedure

For a schematic overview of the procedure, see Table S2[2].

Preparation. After informed consent, shock electrodes were attached and US intensity was individually set at a level deemed ‘uncomfortable, but not painful’.

Fear conditioning. On-screen and oral instructions indicated that pictures of two objects would be presented; one object always followed by an electric stimulus, the other never followed by an electric stimulus. Participants were asked to learn to predict after which object an electric stimulus would occur. They indicated their expectancy on a rating scale presented at the bottom of the screen.

Each CS was presented eight times (two repetitions per viewpoint, 16 trials in total) for 8 s. On CS^+ trials, the US appeared 7.5 s after CS onset. Inter-trial intervals (ITIs) varied between 15, 20, and 25 s with a 20-s mean. CS order was semi-random (no more than two consecutive CS^+ or CS^- trials).

Next, after a 3-min waiting period, shock electrodes were removed and the AAT started.

AAT. The AAT consisted of two blocks of 20 trials each (4 practice trials followed by 16 test trials). For practice trials, 2 CS^+ and 2 CS^- pictures were randomly selected for each participant and presented once. For test trials, each CS viewpoint (4 CS^+, 4 CS^-) was presented twice, in semi-random order (no more than two consecutive CS^+ or CS^- trials).

On each trial, a white manikin figure (71 mm × 95 mm) appeared centered to the bottom or top half of a black computer screen. After 1500 ms, a CS picture was presented, centered to the opposite half of the screen.

On-screen and oral instructions emphasized speed and accuracy. Instructions, counterbalanced across participants, differed according to group allocation and block number. Participants in the *relevant feature group* were instructed to move the manikin towards or away from each object according to its identity (i.e., cube or cylinder), with reversed instructions between blocks. Participants in the *irrelevant feature group* were asked to move the manikin according to the orientation of the white frame (i.e., towards landscape and away from portrait or vice versa), with instructions switched between blocks.

Participants moved the manikin upwards or downwards by pressing the ‘Y’ (labeled ‘ ↑ ’) or ‘B’ (labeled ‘ ↓ ’) key, respectively, on a standard computer keyboard. Upon a response, the manikin started moving, to disappear after 500 ms. In case of an incorrect response, a red cross followed at the manikin’s starting position for 500 ms; no feedback followed after a correct response. The ITI was 2000 ms. Time between CS picture onset and response was measured as dependent variable.

Exit Interview. Upon completion of both tasks, evaluative ratings of CSs and US were obtained and contingency awareness was assessed.

Statistical analyses

Ratings for US and CS characteristics were analyzed using one-way Analyses of Variance (ANOVAs). US-expectancy ratings were analyzed with a 2 (Stimulus: CS^+, CS^-) × 8 (Trial: 1-8) × 2 (Group: relevant feature, irrelevant feature) ANOVA, using Greenhouse-Geisser correction were necessary, with stimulus and trial as within-subject factors and group as between-subject factor. For the AAT, practice trials, test trials with incorrect responses (5.37%), and test trials above 3000 ms (0.10%) were omitted from the analyses. For each participant, median RTs were calculated for each stimulus (CS^+, CS^-) by response type (approach, avoid) combination. RTs were analyzed with a 2 (Stimulus: CS^+, CS^-) × 2 (Response: approach, avoid) × 2 (Group: relevant feature, irrelevant feature) ANOVA with stimulus and response as within-subject factors and group as between-subject factor.

Results and Discussion

No group differences were observed in US or CS evaluation or selected US intensity (Fs < 1.03; Table S1[2]). Figure S1[2] suggests the development of differential US-expectancy (CS+ versus CS-) during conditioning in both groups. This was confirmed by a significant Stimulus × Trial interaction, $F(3.84, 115.08) = 61.48$, $p < .001$, $\eta_p^2 = .67$, and a nonsignificant Stimulus × Trial × Group interaction, $F(3.84, 115.08) = 1.47$, $p = .34$, $\eta_p^2 = .04$.

Figure 1 shows that, in line with our predictions, participants were faster to approach the CS^- and avoid the CS^+ than vice versa, Stimulus × Response interaction, $F(1, 30) = 7.41$, $p = .01$, $\eta_p^2 = .20$, indicating that Pavlovian conditioning successfully induced avoidance tendencies. Paired t-tests indicated that participants were faster to avoid the CS^+ than the CS^-, $t(31) = 2.71$, $p = .01$; no differences were found for approaching both CSs, $t(31) = -1.51$, $p = .14$. Furthermore, participants were faster to approach the CS^- than avoid it, $t(31) = -3.78$, $p = .001$, whereas no differences were found for approaching versus avoiding the CS^+, $t < 1$.

However, unlike the interaction effect, those simple effects do not lend themselves to unambiguous interpretation[2].

Importantly, the RT pattern did not differ between groups, Stimulus × Response × Group interaction, $F < 1$. Apparently, the tendency to avoid a fear-conditioned CS is elicited also when task demands do not require processing of stimulus dimensions relevant for predicting threat, which attests to the automatic nature of such avoidance tendency.

The present results might bear on the origins of excessive avoidance in anxiety disorders. Our findings suggest that people may develop avoidance tendencies towards essentially innocuous cues by mere association of those cues with intrinsically unpleasant or dangerous events even if avoiding those cues does not serve a direct purpose in terms of dealing with the unpleasant event that they are associated with. This provides a mechanism through which an ever-increasing set of intrinsically harmless stimuli might become objects of avoidance. The automatic nature of such avoidance tendencies moreover implies that they might occur most readily in situations that allow little opportunity for conscious thought or in individuals with limited cognitive control, creating room for situational variability and individual differences in the degree of excessive avoidance.

Experiment 2

Avoidance *behavior* is notoriously persistent, due to its self-perpetuating nature: CS avoidance prevents people from experiencing that the CS will not be followed by the US, thereby withholding individuals from anxiety-correcting information and maintaining CS-elicited fear (protection-from-extinction; Lovibond, 2009). One way to break this self-perpetuating cycle is repeatedly presenting the CS in the absence of the opportunity for avoidance behavior, a technique known as *exposure treatment with response prevention*.

Secondary avoidance *tendencies* are arguably less self-perpetuating, given that they do not hinge on an actual contingency with CS removal or US omission (see Experiment 1).

In Experiment 2, we evaluated the persistence of conditioned CS-elicited avoidance tendencies by testing whether they are sensitive to Pavlovian extinction and renewal.

Fear extinction involves repeated presentation of a CS^+ without US. Such a procedure typically leads to attenuation of the conditioned fear response, providing ground for therapeutic interventions such as exposure (Hermans, Craske, Mineka, & Lovibond, 2006). However, conditioned fear may be renewed when after extinction the CS^+ is tested in a context different than extinction (e.g., Effting & Kindt, 2007). Such renewal suggests retention of the original CS^+-US memory throughout extinction and has been used as a theoretical model for explaining relapse of anxiety symptoms after successful exposure treatment (Bouton, 2000).

To test for extinction and renewal, we extended the conditioning procedure of Experiment 1 with an extinction phase that was performed in a different context (Context B) than fear acquisition (Context A; see Effting & Kindt, 2007). We assessed renewal by having one group (ABA) perform the AAT in the acquisition context (A), while the other group (ABB) performed the AAT in the extinction context (B). Similar to Experiment 1, shock electrodes were detached for the AAT. We expected avoidance tendencies to be weaker in the ABB group than the ABA group. As Experiment 1 revealed no between-group differences, we used the irrelevant-feature version of the AAT in Experiment 2, as it provides the more automatic assessment of avoidance tendencies and is least likely to be influenced by response strategies.

Method

Participants

Thirty-three adults took part. One participant was excluded[1] for lack of contingency awareness, leaving a sample of 32 (22 females; mean age: 21.94 years; range: 18-31). Participants were randomly assigned to Group ABA or ABB.

Stimuli and context manipulation

We used the same stimuli as in Experiment 1 (irrelevant feature group). Contexts were manipulated by switching the room lighting on or off.

Procedure

The procedure was similar to that of Experiment 1, with a few modifications (see Table S3[2]). Fear acquisition was followed by an extinction phase, during which the different projections of the CSs were presented five times each (40 trials total), without US. Instructions mentioned that one object would sometimes be followed by an electric stimulus whereas another object would never be. Extinction was performed in a different context (B) than acquisition (A). Group ABA then conducted the AAT in context A, group ABB in context B (contexts were counterbalanced across participants), both with shock electrodes detached.

Statistical analyses

Ratings for US and CS characteristics were analyzed as before. US-expectancy ratings for acquisition were analyzed with a 2 (Stimulus: CS^+, CS^-) × 8 (Trial: a1-a8) × 2 (Group: ABA, ABB) ANOVA with stimulus and trial as within-subject factors and group as between-subject factor. We tested for generalization of acquisition to the extinction context with a 2 (Stimulus: CS^+, CS^-) × 2 (Trial: a1, e1) × 2 (Group: ABA, ABB) ANOVA. Fear extinction was evaluated using a 2 (Stimulus: CS^+, CS^-) × 2 (Trial: e1-e20) × 2 (Group: ABA, ABB) ANOVA. Greenhouse-Geisser corrections were applied when necessary.

For the AAT, test trials with incorrect (6.25 %) and late (0.97 %) responses were excluded. Median RTs were subjected to a 2 (Stimulus: CS^+, CS^-) × 2 (Response: approach, avoid) × 2 (Group: ABA, ABB) ANOVA.

Results and Discussion

No significant group differences emerged for US and CS characteristics, *F*s < 1.12 (Table S1[2]).

During acquisition, differential US expectancies were established, Stimulus × Trial interaction, $F(3.42, 102.63) = 75.49$, $p < .001$, $\eta_p^2 = .72$, similarly in both groups, Stimulus × Trial × Group interaction, $F(3.42, 102.63) = 1.38$, $p = .25$, $\eta_p^2 = .04$ (see Figure S2[2]).

Switching context produced a decrease in differential ratings from the last acquisition trial to the first extinction trial, CS × Trial interaction, $F(1, 30) = 515.80$, $p < .001$, $\eta_p^2 = .89$. Paired samples *t*-tests revealed that this was due to an increase in CS^- ratings, $t(31) = -38.42$, $p < .001$, $d = -13.8$; the change for CS^+ was not significant, $t < 1$. Differentiation was still highly reliable on the first extinction trial, $t(31) = -19.95$, $p < .001$, $d = -7.16$.

Differential expectancy ratings were successfully extinguished, Stimulus × Trial interaction, $F(3.30, 98.98) = 34.55$, $p < .001$, $\eta_p^2 = .54$. Unexpectedly, extinction differed between groups, Stimulus × Trial × Group interaction, $F(3.30, 98.98) = 3.14$, $p = .03$, $\eta_p^2 = .10$. Although significant for both groups, extinction was somewhat stronger in the ABB than the ABA group (which, if anything, should work against our hypothesis for the AAT). Nevertheless, an additional 2 (Stimulus: CS^+, CS^-) × 2 (Group: ABA, ABB) ANOVA on ratings for the last extinction trial revealed a main effect of stimulus, $F(1, 30) = 8.68$, $p = .006$, $\eta_p^2 = .22$, indicating higher CS^+ than CS^- ratings at the end of extinction, but no group effects, *F*s < 1.

For the AAT (Figure 2), a significant Stimulus × Response × Group interaction was obtained, $F(1, 30) = 5.57$, $p = .03$, $\eta_p^2 = .16$. To decompose this interaction, we conducted separate 2 (Stimulus: CS^+, CS^-) × 2 (Response: approach, avoid) ANOVAs for each group. For the ABA group, the main effect of response was significant, $F(1, 15) = 4.95$, $p = .04$, η_p^2

= .25, as was the crucial Stimulus × Response interaction, $F(1, 15) = 4.92$, $p = .04$, $\eta_p^2 = .25$, indicating faster responses for approaching the CS^- and avoiding the CS^+ than vice versa (main effect of stimulus $F < 1$). Paired *t*-tests showed that participants were faster to approach than avoid the CS^-, $t(15) = -3.15$, $p = .007$; all other simple effects were non-significant, *t*s < 1.20 (but see supplementary materials[2]). For the ABB group, no significant effects were obtained, *F*s < 1.76. In combination, these results suggest that secondary avoidance tendencies established through Pavlovian learning can be extinguished, but will renew when conditioned fear cues are presented outside the extinction context.

Actual avoidance behavior can serve to maintain CS fear through protection-from-extinction (Lovibond et al., 2006), lending it a self-perpetuating nature. In clinical practice, the persistence of avoidance behavior can be disrupted through exposure treatment with response prevention, which allows for corrective experiences. The present results suggest, however, that even when persistent avoidance triggered by a cue or situation is disrupted and corrective information about that cue or situation is acquired (extinction learning), cue-elicited avoidance *tendencies* can readily recover upon a context switch. This may in turn lead to a recovery of overt CS-elicited avoidance behavior, which might provide an additional mechanism to account for maintenance of cue-elicited avoidance, in addition to the self-perpetuating mechanism based on preservation of CS fear through protection-from-extinction. Together, both mechanisms provide a particularly strong drive towards persistence of conditioned fear and avoidance.

General Discussion

We tested whether initially neutral cues will elicit avoidance tendencies through their mere pairing with an aversive outcome, in the absence of instrumental reinforcement or an instrumental basis for avoidance. Experiment 1 showed that, following Pavlovian fear learning, participants responded slower when they symbolically approached a conditioned fear stimulus or avoided a conditioned safety stimulus than vice versa. Experiment 2 showed that although fear extinction resulted in the attenuation of avoidance tendencies in the extinction context, a switch to the acquisition context lead to their reappearance. Importantly, avoidance tendencies did not reduce contact with the CS^+ or affect US occurrence, something that past theories would posit as necessary for observing avoidance. In fact, in both experiments shock electrodes were detached during the AAT, removing any instrumental basis for avoidance behavior in terms of US anticipation.

We assessed avoidance tendencies in a reaction time task rather than overt avoidance behavior. Even though emotion theories regard action dispositions as central to emotions (Frijda, 2010; Lang, 1985), such tendencies need not translate into overt behavior. Dual process models, for instance, advocate that although emotions involve an automatic tendency to act (impulsive system), emotional impulses can be regulated by cognitive evaluation processes operating under cognitive control (reflective system; Frijda, 2010; Strack & Deutsch, 2004). Therefore, assessment of overt avoidance behavior might fail to find strong links between fear and avoidance (e.g., Mineka, 1979).

The observation that conditioned fear is accompanied by an automatic secondary avoidance tendency does not imply that avoidance behavior occurs only in the presence of fear. Evidence suggests that with prolonged instrumental reinforcement of avoidance responding, stable avoidance behavior can be maintained even when the fear for the CS has waned (Herrnstein, 1969; Rachman, 1977). Neither do our data suggest that *primary* avoidance ever occurs for other reasons than instrumental learning. As such, our findings neither contradict an expectancy-based account of primary avoidance, nor do they rule out expectancies as a possible source for secondary avoidance. They merely suggest that CS-avoidance *can* be established through mere Pavlovian association and expressed in the complete absence of either fear reduction or the expectation of a negative outcome.

We may not be able to completely rule out all instrumental basis for the observed avoidance tendencies. Shock electrodes were detached during the AAT and approach and avoidance responses had to be made equally often, so that participants got to experience that symbolic approach and avoidance responses had no differential contingencies with outcome omission. Also, moving the manikin towards or away from the pictures did not affect their presentation duration or size. This all notwithstanding, it is probably impossible to definitively rule out an instrumental basis for the observed effects; research has yielded remarkable examples of how behavior that was seemingly under Pavlovian control turned out to be instrumental in nature and vice versa (Hearst & Jenkins, 1974). Future studies could try to yet further disentangle the Pavlovian versus instrumental basis of the effects reported here, for instance by assessing the effect of instrumental punishment on automatic conditioned avoidance tendencies (see Coleman & Gormezano, 1971). Specifically, following Pavlovian conditioning, an aversive US could be presented upon avoidance of the CS^+ or approach of the CS^- in the AAT. If the avoidance tendency reported here is not instrumental, results should be similar to an unpunished version of the AAT. If it would be instrumental, RTs should reveal a main effect of approach versus avoidance rather than a CS by Response interaction as observed here.

A potential limitation of Experiment 2 concerns the lack of control for context familiarity that is often included in animal renewal studies. Such control is particularly important for establishing the cognitive mechanisms underlying renewal; it is perhaps less important for the present purpose of using renewal as a laboratory model for return of fear after exposure treatment.

On a methodological level, when aiming for a comprehensive assessment of fear in laboratory research, the inclusion of avoidance tendencies should perhaps be considered.

Finally, in terms of intervention, targeting conditioned avoidance tendencies, through retraining, could augment the therapeutic efficacy of exposure treatment for anxiety disorders. Therapeutic application of such intervention seems feasible, given that action tendency modification has recently been demonstrated to improve social behavior in individuals with elevated social anxiety symptoms (Taylor & Amir, 2012) and has therapeutic effects in other types of dysfunctional behaviors (e.g., heavy alcohol use; Wiers et al., 2011).

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: Author.

Barlow, D.H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic.* (2nd ed.). New York: Guilford.

Bolles, R.C., Stokes, L.W., & Younger, M.S. (1966). Does CS termination reinforce avoidance behavior? *Journal of Comparative and Physiological Psychology*, *62*, 201-207.

Bouton, M.E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychology, 19,* 57-63. doi: 10.1037/0278-6133.19.Suppl1.57

Coleman, S.R., & Gormezano, I. (1971). Classical conditioning of the rabbit's (Oryctolagus cuniculus) nictitating membrane response under symmetrical CS-US interval shifts. *Journal of Comparative and Physiological Psychology*, *77,* 447-455.

Cornwell, B.R., Echiverri, A.M., & Grillon, C. (2007). Sensitivity to masked conditioned stimuli predicts conditioned response magnitude under masked conditions. *Psychophysiology*, *44,* 403-406. doi: 10.1111/j.1469-8986.2007.00519.x

Declercq, M., De Houwer, J., & Baeyens, F. (2008). Evidence for an expectancy-based theory of avoidance behavior. *The Quarterly Journal of Experimental Psychology, 61,* 1803-1812. doi:10.1080/17470210701851214

Effting, M., & Kindt, M. (2007). Contextual control of human fear associations in a renewal paradigm. *Behaviour Research and Therapy, 45,* 2002-2018. doi: 10.1016/j.brat.2007.02.011

Frijda, N.H. (2010). Impulsive action and motivation. *Biological Psychology, 84,* 570-579. doi: 10.1016/j.biopsycho.2010.01.005

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

Herrnstein, R.J. (1969). Method and theory in the study of avoidance. *Psychological Review, 76,* 49-69. doi: 10.1037/h0026786

Hearst, E., & Jenkins, H.M. (1974). Sign tracking: The stimulus-reinforcer relation and directed action. Austin, TX: Psychonomic Society.

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition and Emotion*, *24*, 810-828. doi: 10.1080/02699930903047298

Lang, P.J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety. In A.H. Tuma & J. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131–170). Hillsdale, NJ: Erlbaum.

Lovibond, P.F. (2006). Fear and avoidance: An integrated expectancy model. In M.G. Craske, D. Hermans & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implications* (pp. 117-132). Washington, DC: American Psychological Association.

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,* 716-720.

Lovibond, P.F., Saunders, C., Weidemann, G., & Mitchell, C.J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *The Quarterly Journal of Experimental Psychology, 61,* 1199-1216. doi: 10.1080/17470210701503229

Mauss, I.B., & Robinson, M.D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*, 209-237. doi: 10.1080/02699930802204677

Miller, N.E. (1944). Experimental studies of conflict behavior. In J.McV. Hunt (Ed.), *Personality and the behaviour disorders* (pp. 431-465). New York: Ronald Press.

Mineka, S. (1979). The role of fear in theories of avoidance learning, flooding, and extinction. *Psychological Bulletin, 86,* 985-1010. doi: 10.1037/0033-295X.108.3.483

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297-326. doi: 10.1037/0033-2909.132.2.297

Mowrer, O.H. (1960). *Learning theories and behavior.* New York: Wiley.

Rachman, S. (1977). The conditioning theory of fear-acquisition: A critical examination. *Behaviour Research and Therapy, 15,* 375-387. doi: 10.1016/0005-7967(77)90041-9

Seligman, M.E., & Johnston, J.C. (1973). A cognitive theory of avoidance learning. In F.J. McGuigan & D.B. Lumsen (Ed.). *Contemporary approaches to conditioning and avoidance learning* (pp. 69-110). Washington, DC: Winston.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220-247. doi: 10.1207/s15327957pspr0803_1

Taylor, C.T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy*, *50*, 529-536. doi:10.1016/j.brat.2012.05.004

Walker, S.F. (1987). *Animal learning: An introduction*. London: Routledge.

Wiers, R.W., Eberl, C., Rinck, M., Becker, E.S., & Lindenmeyer (2011). Retraining automatic action tendencies changes alcoholic patients’ approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*, 490-497. doi: 10.1177/0956797611400615

Footnotes

[1]See online Supplemental Material for full sample results.

[2]See online Supplemental Material.

Figures

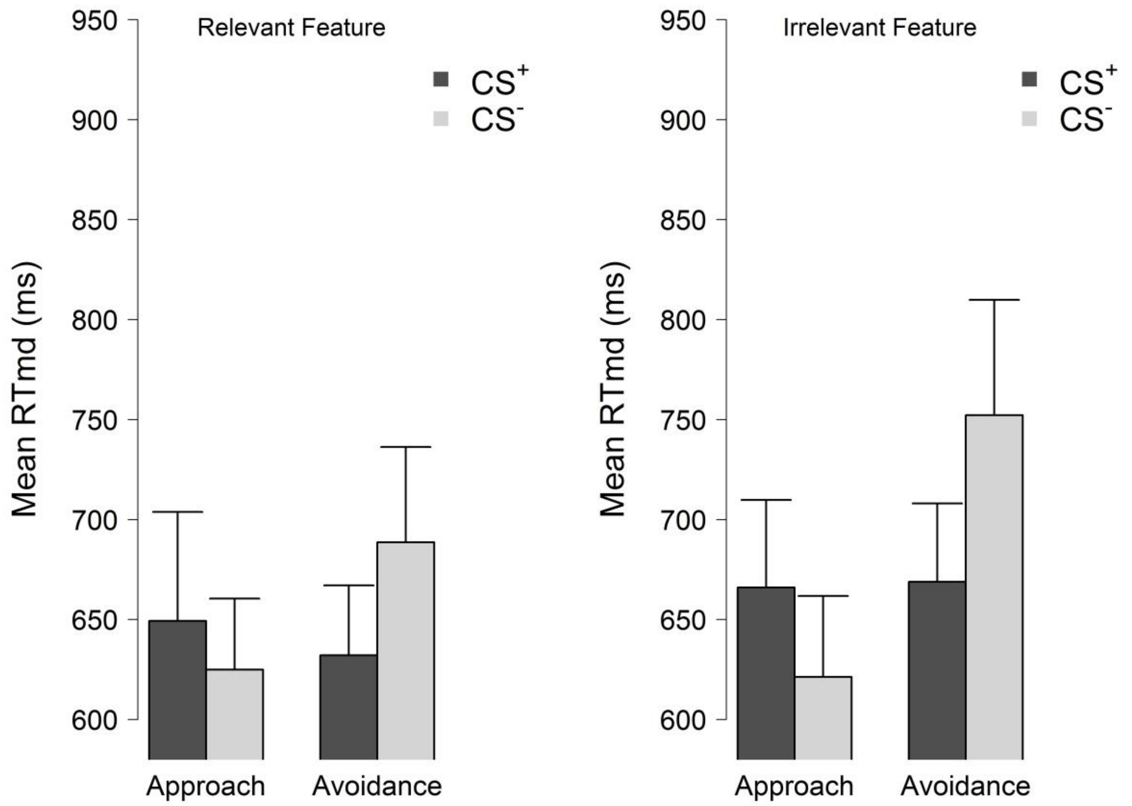


Figure 1. Experiment 1: Mean median RT in ms for groups Relevant Feature (left panel) and Irrelevant Feature (right panel) (error bars represent standard errors of the means).

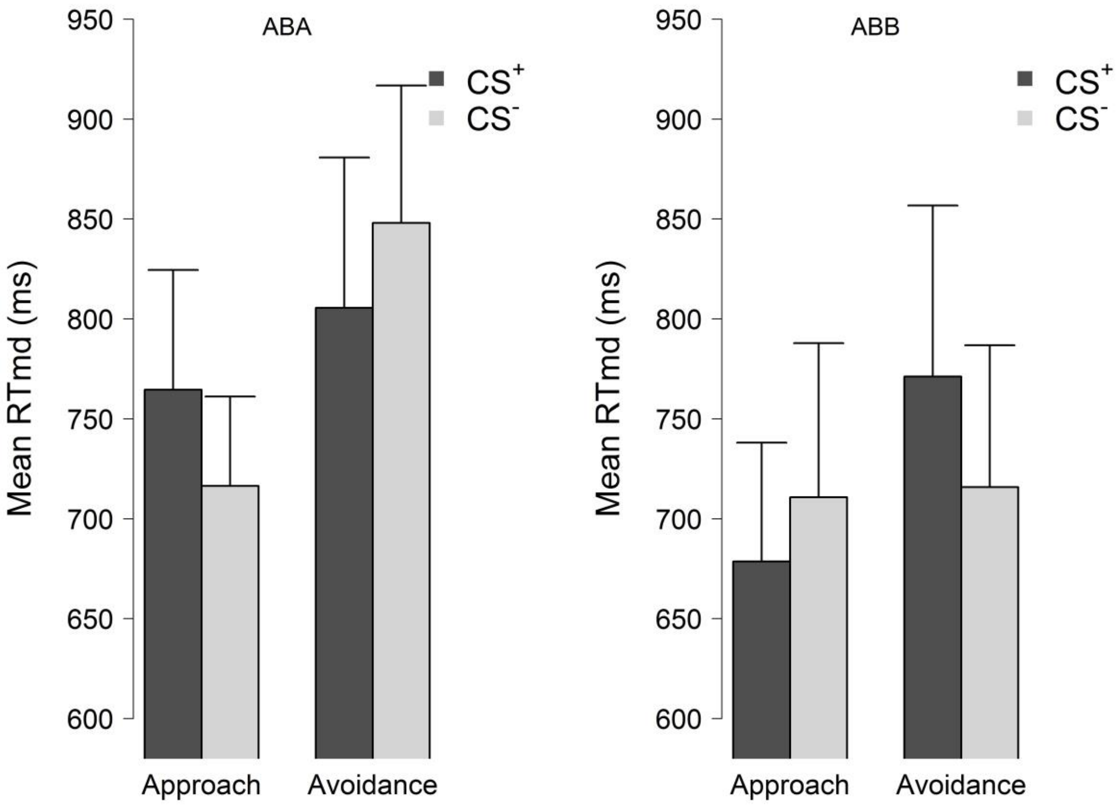


Figure 2. Experiment 2: Mean median RT in ms for groups ABA (left panel) and ABB (right panel) (error bars represent standard errors of the means).

Supplementary analyses (including participants that failed to demonstrate contingency awareness in the exit interview)

Experiment 1

No significant group differences were observed in terms of US or CS evaluation or selected US intensity (all *F*s < 1.11, except US startlingness, $F(1, 33) = 2.19$, $p = .15$).

Regarding expectancy ratings for CS^+ and $CS^−$ across trials for both groups, statistical analyses revealed successful discrimination learning, that is, stronger US expectancy after CS^+ than after CS^- over the course of conditioning, Stimulus × Trial interaction, $F(3.99, 131.61) = 52.01$, $p < .001$, $\eta_p^2 = .61$. Discrimination learning was comparable between groups, Stimulus × Trial × Group interaction, $F < 1$.

Results for the reaction time task showed a significant Stimulus × Response interaction, $F(1, 33) = 10.54$, $p = .003$, $\eta_p^2 = .24$, indicating that participants were faster to approach the CS^- and avoid the CS^+ than vice versa. The reaction time pattern did not differ between groups, as indicated by a non-significant Stimulus × Response × Group interaction, $F < 1$.

Experiment 2

No significant group differences emerged for US and CS characteristics (all *F*s < 1.25, except CS$^-$ pleasantness, $F(1, 31) = 2.99$, $p = .09$).

During acquisition, differential US expectancies were established as indicated by a significant Stimulus × Trial interaction, $F(3.83, 118.73) = 69.41$, $p < .001$, $\eta_p^2 = .69$. The acquisition pattern was similar between groups, Stimulus × Trial × Group interaction, $F(3.83, 118.73) = 1.36$, $p = .22$, $\eta_p^2 = .04$.

Differential expectancy ratings were successfully extinguished, Stimulus × Trial interaction, $F(3.71, 115.16) = 31.63$, $p < .001$, $\eta_p^2 = .51$. However, the extinction pattern differed across groups, Stimulus × Trial × Group interaction, $F(3.71, 115.16) = 2.68$, $p = .04$, $\eta_p^2 = .08$. Nevertheless, an additional 2 (Stimulus: CS$^+$ vs. CS$^-$) × 2 (Group: ABA vs. ABB) ANOVA comparing ratings on the last extinction trial revealed only a main effect of stimulus, $F(1, 31) = 9.97$, $p = .004$, $\eta_p^2 = .24$, indicating higher CS$^+$ than CS$^-$ ratings at the end of extinction, but no effects of group, *F*s < 1.

In the reaction time task, participants were overall faster in approaching than avoiding stimuli, $F(1, 30) = 5.02$, $p = .03$, $\eta_p^2 = .14$. No main effect of stimulus nor interaction between stimulus and response emerged, *F*s < 1. A significant Stimulus × Response × Group interaction was obtained, $F(1, 31) = 6.70$, $p = .02$, $\eta_p^2 = .18$. For the ABB group, no significant effects were obtained for either stimulus or the Stimulus × Response interaction, *F*s < 1.76. For the ABA group, the main effect of stimulus was not significant, $F < 1$, the main effect of Response was marginally significant, $F(1, 16) = 4.12$, $p = .06$, $\eta_p^2 = .20$. The Stimulus × Response interaction revealed a significant conditioned avoidance tendency, $F(1, 16) = 6.58$, $p = .02$, $\eta_p^2 = .29$.

Supplementary Table S1

Mean Values (SD) for US Level and Rated Characteristics for US and CSs in Experiments 1 and 2.

Exp	US (mA)	US Pleasantness	US Intensity	US Startlingness	CS^+ Valence	CS^- Valence
1	22.94 (11.03)	-6.13 (2.73)	3.09 (0.47)	3.72 (0.52)	-3.02 (2.20)	3.05 (2.05)
2	28.69 (18.63)	-6.19 (1.71)	3.09 (0.53)	3.75 (0.67)	-1.91 (2.27)	3.61 (1.66)

Note. Exp = Experiment; US = Unconditioned Stimulus; mA = microampere; CS = Conditioned Stimulus.

Supplementary Table S2
Experiment 1: Procedure and measurements.

Differential Fear Acquisition →	Reaction Time Task →	CS Evaluation →	US Evaluations & Exit Interview
US Expectancies	Reaction times	CS ratings	US ratings, contingency awareness

Note. US = Unconditioned Stimulus; CS = Conditioned Stimulus.

Supplementary Table S3

Experiment 2: Procedure, with measurements and context for each phase.

Fear Acquisition	Fear Extinction	Reaction Time Task	CS Evaluation	US Evaluations & Exit Interview
US Expectancies	US Expectancies	Reaction times	CS ratings	US ratings, contingency awareness
A	B	B	B	N
A	B	A	A	N

Note. US = Unconditioned Stimulus; CS = Conditioned Stimulus; A = Context A; B = Context B; N = Neutral Context.

Supplementary Figure S1

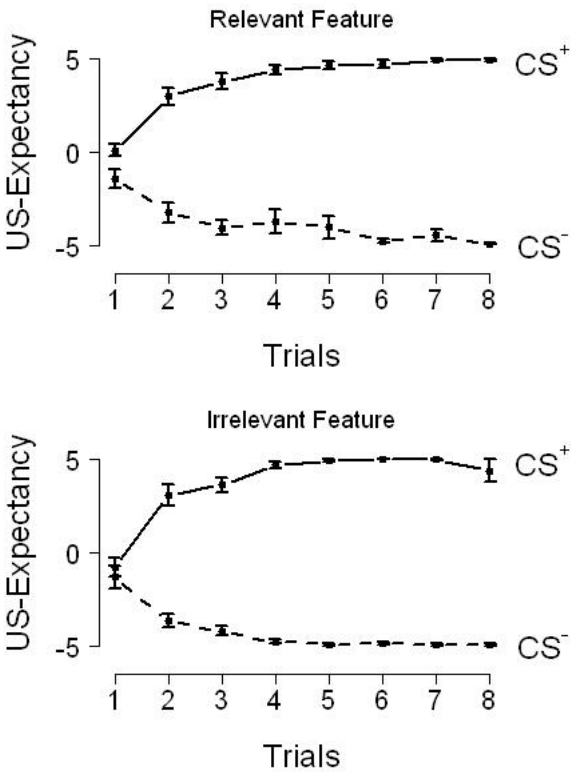


Figure S1. Mean US expectancy ratings for the CS^+ and CS^- across acquisition trials for groups relevant feature (top panel) and irrelevant feature (bottom panel) in Experiment 1. Error bars represent standard error of the mean

Supplementary Figure S2

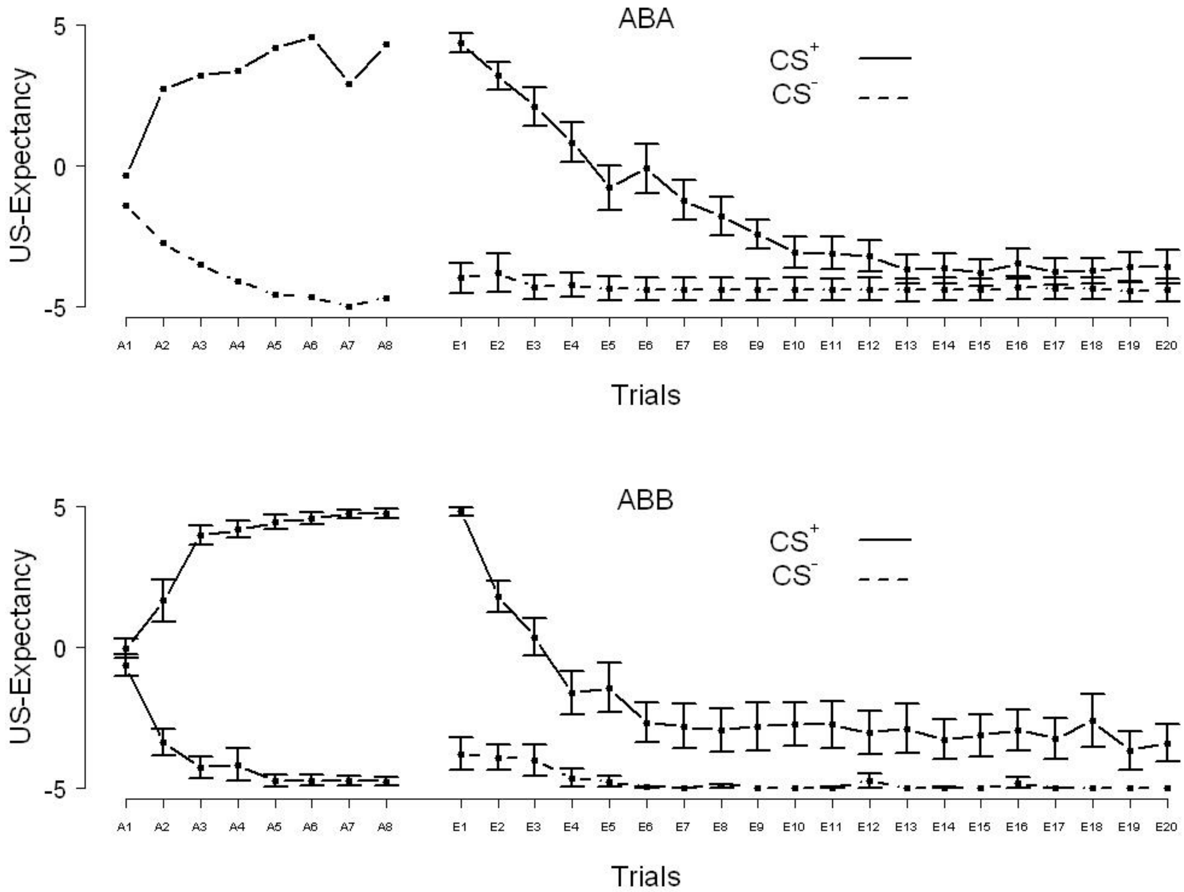


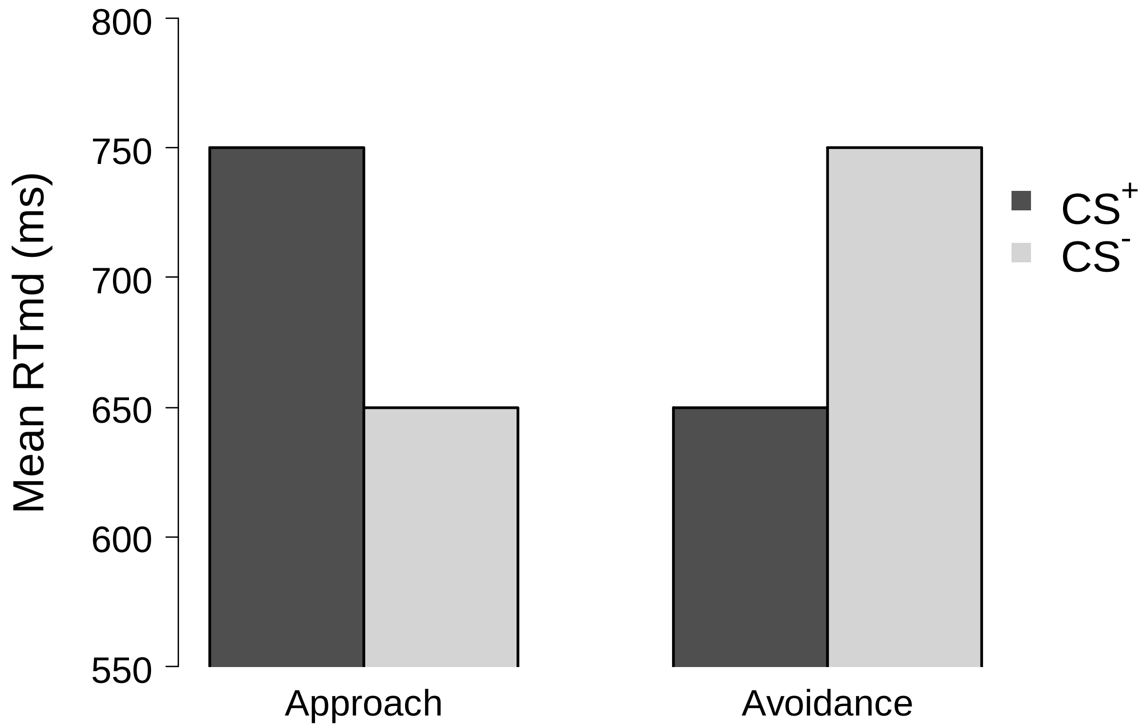
Figure S2. Mean US expectancy ratings for the CS^+ and CS^- across acquisition trials (A1 – A8) and extinction trials (E1 – E20) for Groups ABA (top panel) and ABB (bottom panel) in Experiment 2. Error bars represent standard error of the mean.

Interpretation of the Simple Effects in Approach-Avoidance Reaction Time Tasks

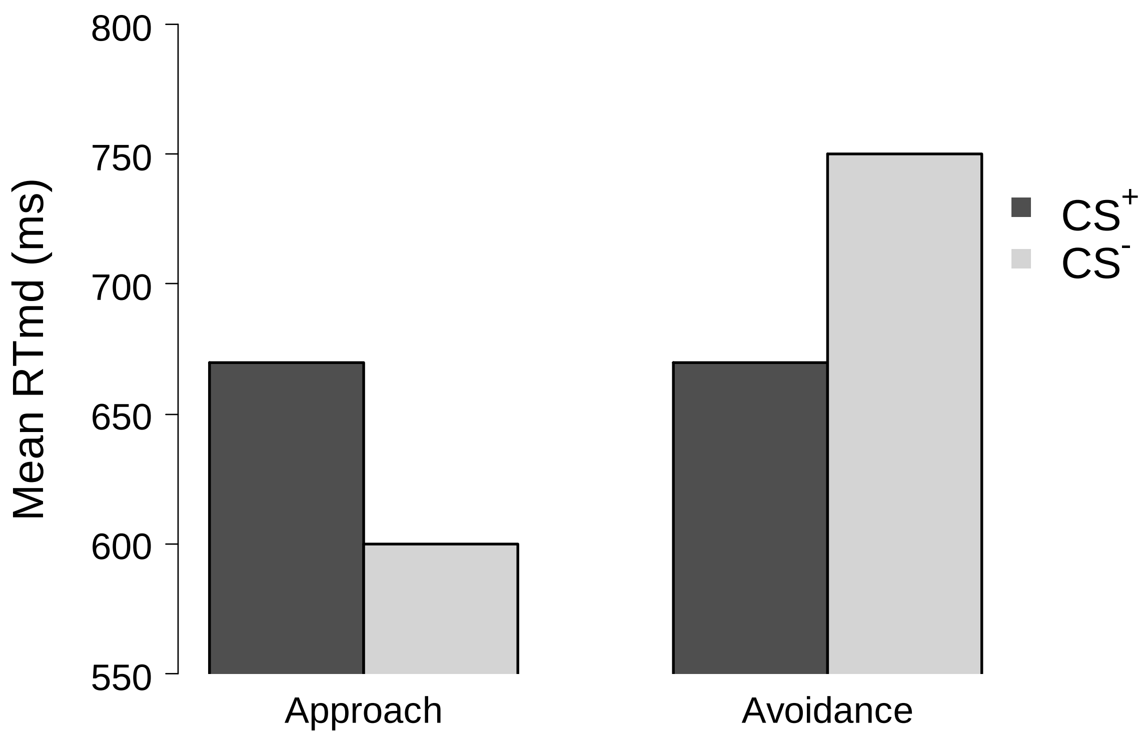
This is supplementary text regarding the interpretation of the statistically significant interactions reported in the manuscript 'Avoided by Association: Acquisition, Extinction, and Renewal of Avoidance Tendencies towards Conditioned Fear Stimuli' by Krypotos, Effting, Arnaudova, Kindt and Beckers (submitted). Since the issue is quite technical and not connected to the main point of the paper, we present our arguments here.

In our manuscript, we base the interpretation of all statistical significant interactions on the reported omnibus effects, without referring to the reported simple effects. Yet, it is tempting to interpret the simple effects for the CS^+ and CS^- separately in follow-up to the significant overall interactions, for instance along the lines that a significant simple effect for the CS^- in the absence of a main effect for the CS^+ would indicate that our acquisition procedure resulted in the CS^- becoming a Pavlovian safety cue rather than the CS^+ becoming a danger cue. However, we maintain that any interpretation of the simple effects in our reaction time procedure (as in any Stroop- or Simon-like procedure, including the IAT) is unwarranted, an issue that we will expand upon below:

Suppose that, theoretically and all other things being equal, we would anticipate a fully crossed interaction between CS-type and response, as follows:

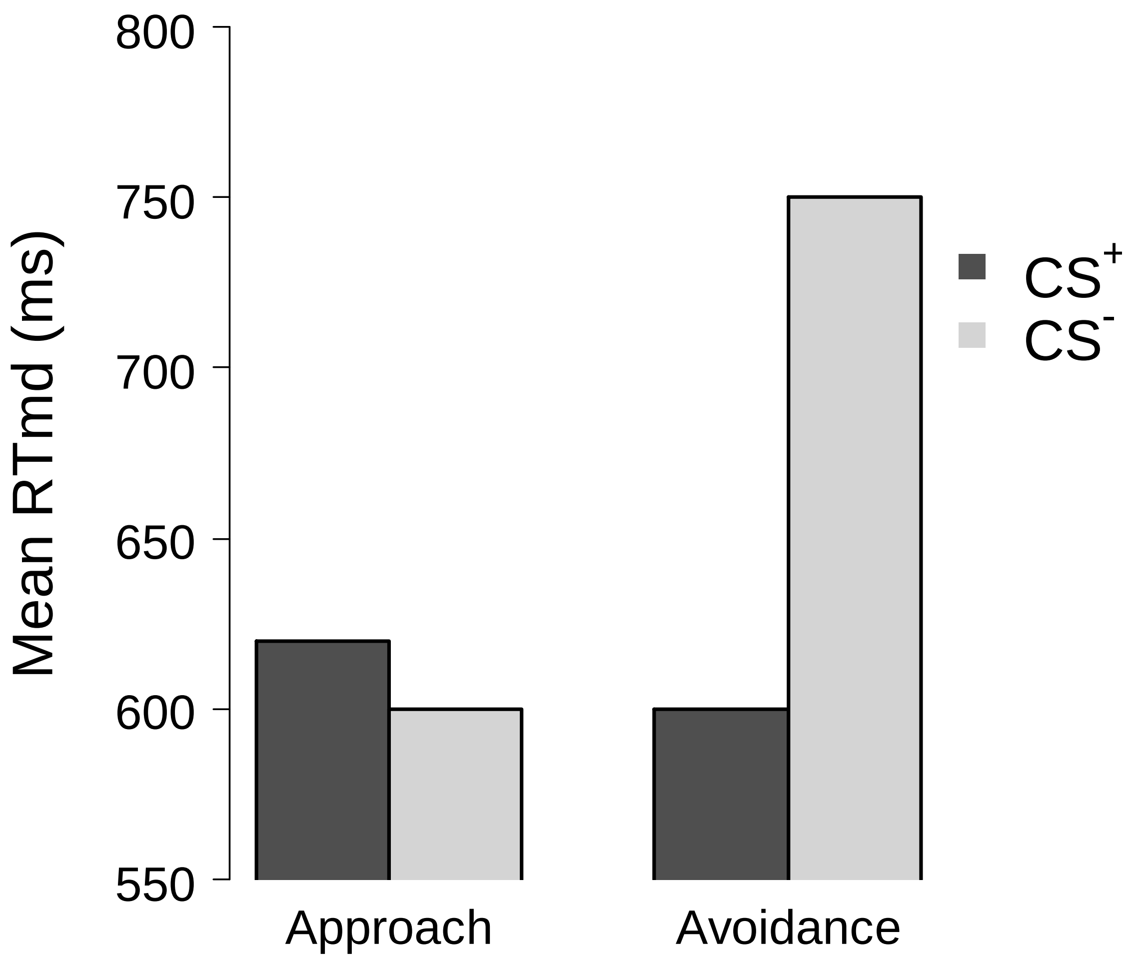


Suppose now that, orthogonally to this theoretical interaction, there would be a general tendency for people to respond quicker on approach than avoidance trials (a common observation, perhaps related to the fact that approach is coded as the target response and avoidance as the alternative response; much like in affective priming studies, people are generally faster to say “positive” than “negative”, irrespective of the specific stimulus they are responding to). As a result, both left-hand-side bars in the graph would be lowered to the same extent, possibly resulting in a graph like this:



Here, the simple effects for CS^+ and CS^- are very different, whereas statistically the interaction is basically identical. Notice that this graph looks much like the graphs in our Figure 1. So, while at first this second graph, like Figure 1 in our manuscript, might suggest that the effect is to be situated mainly with the CS^-, it may just as well reflect the combination of an effect that is mainly situated with the CS^+ with a general bias to be faster to approach than to avoid any given stimulus, so that the in the presence of the CS^+, the natural tendency to be quicker to approach than to avoid is greatly reduced, whereas the CS^- has no effect whatsoever on top of the general response bias; or that the CS^+ serves to reduce the natural response bias while the CS^- further enhances the natural response bias.

The issue is further complicated by the fact that there might be good reasons to suspect that people generally process the CS^+ faster than the CS^-. As a result of being paired with shock, the CS+ is probably more arousing/more attention-grabbing than the CS-. This could result in an overall tendency for shorter reaction times to the CS^+ than the CS^-, again orthogonal to the theoretical interaction between CS-type and response, which might further affect the shape of the interaction by reducing both blue bars, to result in something like this:



The latter graph might at first sight be taken to indicate not only that the effect is mainly driven by the CS^- but moreover that the effect is mainly due to a slowing of the speed with which the CS- can be avoided. Again, however, such interpretation is statistically unwarranted. All that can be said on the basis of any of the preceding graphs is that, relative to the CS$^-$, participants are faster to avoid the CS$^+$ than to approach it; **the simple effects are not a valid basis for inference**.