Paper:
Xia, W., Eyolfson, E., Lloyd, K., Vervliet, B. & Dymond, S. (2019). Living in fear: Low-cost avoidance maintains low-level threat. *Journal of Behavior Therapy and Experimental Psychiatry, 62*, 57-64.
http://dx.doi.org/10.1016/j.jbtep.2018.09.001

18 month embargo.

Running head: THREAT VALUE OF SAFETY

Living in Fear: Low-Cost Avoidance Maintains Low-Level Threat

Weike Xia[1], Eric Eyolfson[1], Keith Lloyd[2], Bram Vervliet[3,4] and Simon Dymond[1,5,*,]

[1] Experimental Psychopathology Lab, Department of Psychology, Swansea University, Park Campus, Swansea, SA2 8PP, United Kingdom.

[2] Swansea University Medical School, Park Campus, Swansea SA2 8PP, United Kingdom.

[3] Behavioral Neuroscience Program, Department of Psychiatry, Harvard Medical School, Massachusetts General Hospital, 149 13th St, Charlestown, MA 02129, USA.

[4] Center for Excellence on Generalization, Department of Psychology, KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium.

[5] Department of Psychology, Reykjavík University, Menntavegur 1, Nauthólsvík, 101 Reykjavík, Iceland.

* Corresponding author

Abstract

Background and objectives: Excessive avoidance of potential threat is a hallmark of anxiety and is thought to maintain fear by preserving the perceived high-threat value of avoided situations. Previous research has shown that the availability of avoidance maintains low-level threat. Here, we investigated whether an opportunity to engage in avoidance in the presence of a low-threat value safety cue would maintain its perceived threat value when avoidance was unavailable.

Methods: In a threat conditioning procedure, one conditional danger stimulus (CS+; A+) was followed by an aversive unconditioned stimulus (US; electric shock), and two safety stimuli (CS-; B- and C-) were never followed by the US. Next, clicking a button present during A+ avoided the scheduled US. Avoidance was then made available during C- for participants in the Experimental group but not in the Control group. In the test, all stimuli were presented without the opportunity to avoid. Threat expectancy, eyeblink startle electromyography (EMG), and skin conductance responses (SCRs) were measured.

Results: Findings showed an increase in threat expectancy for only C- in the Experimental group during the test phase following avoidance learning to similar levels as during threat conditioning. Compared to the Control group, threat expectancy for both B- and C- remained higher in Experimental group. SCR and startle EMG data did not corroborate these findings.

Limitations: Further research is needed to test the commonly held clinical assumption that avoidance can increase threat value.

Conclusions: Low-cost avoidance maintains low-threat value of safety cues.

1. Introduction

Learning to avoid potential threat is key to survival. Sometimes, however, individuals engage in excessive threat avoidance despite a relatively low probability of danger. Compared with driving a car, travelling by airplane is one of the safest forms of transportation (Shaver, 2015). Yet, following the terrorist attacks in the United States in September 2001, an increase in automotive fatalities soon after was attributed to an increased avoidance of flying (Gigerenzer, 2004). Similarly, a sustained reduction in train travel was noted after the terrorist bombings on the London Underground in July 2005 and attributed to increased perception of potential risk (Handley, Salkovskis, Scragg, & Ehlers, 2009; Prager, Asay, Lee & von Winterfeldt, 2011; Rubin, Brewin, Greenberg, Simpson & Wessely, 2007). Paradoxically, the increased perception of danger and the avoidance it motivates may in part be due to increased safety. Additional security at airports and train stations, visible police and army street patrols, combined with regular news reports about escalating national security threat levels, may actually serve as a reminder of threat (Renard, 2016; van de Veer, de Lange, van der Haar, & Karremans, 2012). In this way, a paradoxical outcome of safety behavior is enhanced threat perception in the presence of objective safety which can serve to strengthen excessive avoidance (Engelhard, van Uijen, Seters & Velu, 2015).

Avoidance is a central feature in anxiety and methods for minimizing avoidance behavior are vital elements in therapeutic success (Barlow, Raffa & Cohen, 2002; Blakey & Abramowitz, 2016). In exposure-based treatments of anxiety, clients are repeatedly presented with cues that have come to signal an aversive outcome or an impending catastrophe (Barlow et al., 2002). Through repeated presentation in the absence of these outcomes, anxious clients come to learn that these cues are in fact safe, which may then lead to a reduction in fear (Hermans, Craske, Mineka, & Lovibond, 2006). Avoidance is thought to directly interfere with the goals of exposure by preserving threat beliefs as anxious patients misattribute their reduced fear to the avoidance behavior (LeDoux, Moscarello, Sears, & Campese, 2016; Lovibond, Mitchell, Minard, Brady, & Menzies, 2009; Volders, Meulders, de Peuter, Vervliet, & Vlaeyen, 2012).

The acquisition and extinction of avoidance is investigated using the fear or threat conditioning paradigm (LeDoux et al., 2016; Lonsdorf et al., 2017; Vervliet & Raes, 2013). During threat conditioning, a neutral stimulus is paired with an aversive unconditioned stimulus (US), such as an electric shock, and comes to elicit conditioned fear responses (i.e., CS+). Another neutral stimulus paired with the absence of the US (i.e., CS-) comes to be a learned safety cue. During extinction learning, the CS+ is presented repeatedly in the absence of the US (Bouton, 2004). Avoidance learning procedures typically entail the cancellation of the US through an active avoidance response, such as bar pressing (Krypotos, Effing, Kindt & Beckers, 2015).

Avoidance maintains perceived threat levels of a CS+ (Lovibond et al., 2009; Krypotos et al., 2015) and leads to a return of previously extinguished threat value (van Uijen, Leer & Engelhard, 2017; Vervliet & Indekeu, 2015). Avoidance may also increase the perceived threat level of low-threat safety cues (Engelhard, van Uijen, Seters & Velu, 2015). Engelhard et al. (2015) investigated whether shock expectancy for a low-threat value cue increased when the opportunity to avoid was removed. Participants first received threat conditioning with one CS+ (A) and two CS-s (B and C), followed by avoidance learning where an avoidance cue was available only during A+ trials. Later, the avoidance cue was available only during C trials in the Experimental group but not in the Control group. Finally, during the test phase, A, B and C were each presented once without the avoidance cue, and C was always presented last. Engelhard et al.

(2015) found that shock expectancy increased for C in the Experimental group only, indicating that the prior availability of avoidance for a previously safe stimulus increased its threat value. Notably, expectancy ratings of C did not increase in either group, suggesting that avoidance may have maintained the threat perception of a low-threat value cue.

The present study sought to replicate Engelhard et al. (2015), with minor methodological changes, to investigate whether avoidance increases, maintains, or decreases the threat value of a low-threat safety cue. Specifically, we explored the potential boundary conditions of the predicted effects on avoidance. In Engelhard et al., C was presented 50% less-often than B in both the fear/threat conditioning and avoidance learning phases (see also, Lovibond et al., 2009). As a result, the observed difference in threat expectancy may in fact reflect different levels of learned safety, rather than an avoidance-induced effect. Thus, we ensured there were equal numbers of presentations of the avoided, low-threat value stimulus (C) and the non-avoided, no-threat value stimulus (B). In addition, we collected measures of physiological reactivity (skin conductance responses and startle EMG), extended the critical test phase to investigate whether threat perception towards C would be maintained or decline over time, and used faces instead of geometrical figures as stimuli.

We predicted increased threat expectancy, SCR and startle for C- when avoidance was available and a decline in threat expectancy in the absence of avoidance for the Experimental group only.

2. Methods

2.1 Participants

One hundred and ten participants (73 female; M_{age} = 21.35, *SD* = 3.16) were recruited from Swansea University. Sample size was determined using G*Power (Erdfelder, Faul, &

Buchner, 1996) based on the *t*-test results from the critical test phase comparing shock expectancy ratings for C in Engelhard et al. (2015). Accordingly, the recommended sample size was 70, with 31 in the Experimental group and 39 in the Control group, for a power of 0.95. Due to our a priori exclusion criteria, the final sample size exceeded this figure; we randomly allocated $N = 59$ participants to the Experimental group and $N = 51$ to the Control group, respectively. Both groups were similar in age ($F_{(1, 75)} = 0.86$, $p > .05$) and gender, X^2 (2, N = 76) = 0.36, $p > .05$). Written consent was obtained at the outset and participants received either course credit or a £10 voucher as compensation at the end of the study. The Department of Psychology Ethics Committee, Swansea University approved the study.

2.2 Apparatus and Stimuli

Three male, neutral Caucasian faces selected from the NimStim set of facial expressions (Tottenham et al., 2009) presented against a white background in the middle of the screen were used as stimuli and counterbalanced across participants. An unlit image of a light bulb was presented next to the CS in the lower right corner of the screen. The avoidance cue was an illuminated version of the light bulb which replaced the unlit image. Acoustic startle probes were 50 ms white noise bursts of 95 dB presented through headphones. All stimuli were presented on a 17-inch computer screen with a 60Hz refresh rate and all tasks were programmed in *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012).

The US was a 250 ms electric shock generated using a STM200 stimulator (BIOPAC Systems, Santa Barbara, USA) and administered through a surface electrode (MLADDF30 bar electrode with two 9 mm contacts spaced 30 mm apart). Skin conductance was measured through two Ag/AgCl electrodes coated with non-hydrating gel attached to the middle phalanges on the index and middle fingers. Startle response was measured through eyeblink EMG in response to the startle probes. The eyeblink reflex was measured through 2, 4 mm Ag/AgCl electrodes coated with electrode gel. Electrodes were placed approximately 1 cm under the right pupil and 1 cm below the lateral canthus, while a ground electrode was placed on the forehead. Prior to electrode placement, the skin was cleansed with non-alcoholic cleanser. Both SCR and EMG signals were recorded with the MP150 (BIOPAC Systems, Santa Barbara, USA) and sampled at 1000Hz with a notch filter of 10Hz.

2.3 Procedure

Participants first had the SCR, EMG and shock electrodes attached prior to shock calibration. The current was initially set at 35 mV and increased or decreased in steps of 2.5 mV depending on participants' ratings (the maximum was 100 mV). Participants were instructed to rate the intensity of the shock on a scale from 1 (*'not at all uncomfortable'*) to 5 ('*very uncomfortable'*). When a shock level was rated as 'uncomfortable but not painful' twice consecutively, it was selected for that participant.

Following shock calibration, participants were instructed that on each trial one of the three stimuli would be presented and that if they paid attention, they could work out which one was followed by shock. Participants then received instructions about making the avoidance response and were given two practice trials. Finally, participants were instructed to rate shock expectancy using a visual analogue scale from 0 ('*certain no shock'*) to 100 ('*certain shock*'). The procedure was adapted from Engelhard et al. (2015), with several important differences (Table 1). First, during threat conditioning, B- and C- were presented in equal numbers of trials (four times each). Second, during the Test phase, all three stimuli were presented for 4 trials each. Third, trial durations were extended to 15 s to accommodate both the SCR first interval response (measured from 0.5-5 s after CS presentation) as well as startle EMG response (which occurred 9-10 s after CS presentation) and trial-by-trial threat expectancy ratings.

Insert Table 1 About Here

The experiment consisted of five phases: *habituation, threat conditioning, avoidance learning, avoidance shift learning* and *test* (Table 1). For all trials in all phases, participants were presented with one CS per trial for 15 s. Startle probes were presented in three out of four trials per CS approximately 9-10 s after CS presentation, while threat expectancy ratings appeared after 11 s of CS presentation and lasted until either a response was made or until CS termination (Figure 2). The ratings scale (‘*How likely do you think a shock will follow this trial?’*) consisted of a slider bar with markers, ‘*Not at all’* and ‘*Definitely*’. Participants responded by pressing the left mouse button. Inter-trial intervals (ITIs) ranged from 6-10 s, with inter-startle intervals approximately 15 s. Throughout the experiment, A+, B- and C- trials were presented pseudo-randomly to ensure that any trial type was never presented more than twice consecutively.

During *habituation*, eight startle probes were presented on a white background over a period of two minutes. Participants viewed the CSs passively, each once for 10 s and without the US and were then given two practice trials with an image of a sofa to acclimatize to the timing of the light bulb. The remaining phases occurred without interruption.

During *threat conditioning*, participants viewed A+, B-, and C- four times in a pseudo-random order, with each CS presented no more than twice consecutively. All A+ trials were coupled with the US (100% reinforcement), which occurred at CS termination. B- and C- were never followed by the US.

During *avoidance learning*, A+ was presented seven times, while both B- and C- were presented twice in a pseudo-random order. In six out of seven A+ trials, the avoidance cue was presented 5-6 s after CS presentation and was displayed for 3 s. If participants pressed the left mouse button within 3 s, the US was omitted. For one A+ trial, participants were unable to avoid the US to affirm the A+ and US contingency.

During *avoidance shift learning*, C- was presented six times, while both A+ and B- were presented twice in a pseudo-random order. For the Experimental group, the avoidance cue was only presented during C- trials, 5-6 s after stimulus onset and was displayed for 3 s. No shock was presented if participants did not make an avoidance response. For the Control group, no avoidance cue was presented. Thus, the groups differed on the basis of the availability of avoidance for a previously safe cue (C-).

During the *test* phase, A+, B- and C- were presented four times each, with C- shown only after A+ and B- had been presented once. Presentation of either A+ or B- first was random, after which the remaining trials were presented pseudo-randomly.

2.4 Scoring criteria and data analysis

Following Engelhard et al. (2015), contingency awareness was defined as a higher threat expectancy rating to A+ than B- in the test phase. Unlike Engelhard et al., where separate criteria were applied during avoidance learning and avoidance shift learning, in the current study mastery criterion for both of these phases was set at four successful avoidance responses out of six trials. Participants were excluded from analysis if they were contingency unaware or it they did not meet criterion.

Electrophysiological data was processed using *AcqKnowledge* (BIOPAC Systems, Santa Barbara, USA). Skin conductance responses were calculated as the first peak within 0.5 to 5 s after CS onset. SCR threshold was set at 0.02 μS and responses under the threshold were scores as null responses. Prior to data analysis, SCRs were range-corrected per participant and square root transformed across all phases. Eyeblink data were processed according to established procedures (Blumenthal et al., 2005). The facial EMG signal was band pass filtered (28-500 Hz), rectified and smoothed by applying a 40Hz low-pass filer. Startle magnitude was defined as the amplitude of the highest peak to occur within 25-100 ms after startle probe. Startle data were checked for artefacts such as spontaneous blinks and movement during the analysis window (Klumpers, Morgan, Terburg, Stein, & van Honk, 2015). Baseline artefacts were identified as trials with a standard deviation of baseline activity (75 ms to 0 ms before probe onset) larger than two times the mean standard deviation and artefacts were reported as missing data. Trials that showed a less than 55% increase in standard deviation relative to the 75 ms baseline immediately preceding the response were scored as null responses. Startle data was then converted into *T*-scores ([Z-transformed data x 10] + 50) to control for individual differences in baseline reactivity.

For statistical analyses, threat expectancy ratings and SCR were blocked per two trials to reduce noise, while startle data was averaged per phase. Single trial analyses of shock expectancy ratings were performed for the first trial during test similar to Engelhard et al. (2015) and for the comparison between the unavoidable A+ trial and the first avoidable A+ trial during avoidance learning. Outcome measures were analyzed for each phase using mixed ANOVAs and Greenhouse-Geisser corrections were applied where sphericity was not met. Bonferroni correction was applied to all post-hoc tests and planned comparisons (alpha was set at .05 throughout). Follow-up comparisons consisted of paired-samples and independent-sample *t*-tests with corrected statistics and degrees of freedom for violations of homogeneity.

3. Results

3.1 Inclusion analysis

Data from 34 participants were excluded as, nine were unaware of the contingency, 11 did not meet avoidance learning criteria and 14 did not meet avoidance shift learning criteria. Further threat expectancy analyses were conducted on the remaining 76 participants (Experimental group: N=37; Control group: N=39). Of the 37 Experimental participants, only 5 did not press on all 6 trials, 1 pressed 4 times, and 4 pressed 5 times during the avoidance shift phase. Engelhard et al. (2015) used a 6/6 mastery criterion to prevent participants from potentially learning that C- did not bear threat. To account for this, we compared mean ratings (and standard deviation) between the two groups (6/6: M=23.65, SD=29.70; 4 or 5/6: M=30.77, SD=33.87) and found similar ratings. Thus, all threat expectancy data was included for analysis.

SCR analyses were conducted on 71 participants (Experimental group: *N*= 36; Control group: *N*= 35) due to technical problems with the physiological data. Finally, eight participants were excluded from startle EMG analyses due to excessive null responses (>95% of all trials, including habituation trials) leaving 63 participants (Experimental group: *N*= 32; Control group: *N*= 31).

Insert Table 2 and Table 3 About Here

3.2 Threat conditioning

As expected, 2 (Group) x 3 (CS) x 2 (Trial) mixed ANOVAs confirmed that threat expectancy ratings ($F_{(1.61,\ 112)} = 153.7$, $p<.001$, $\eta_p^2= .69$) and SCR ($F_{(2,\ 138)} = 17.39$, $p<.001$, $\eta_p^2= .20$) were significantly different per CS. Threat expectancy (Table 2) and SCR (Table 3) for A+ were significantly higher than for both B- and C-, *p'*s <.001, while C- and B- also did not differ (Figure 1), $p>.05$. The CS x Trial x Group interaction was not significant for either expectancy (*p*=.45) or SCR (*p*=.27). Startle EMG analyses (Table 3 and Figure 2) revealed differential startle responses ($F_{(2, 122)}$ = 9.94, *p*<.001, η_p^2= .14) for A+, A+ vs. B-, and C-, (*p*'s <.01), B- vs. C- (*p*>.05). There were no significant group differences (*F*'s<1, *p*'s >.05) and the CS x Group interaction was not significant, *p*=.71. Together, these results indicate successful threat conditioning in both groups.

Insert Figure 1 and Figure 2 About Here

3.3 Avoidance learning

Threat expectancy for avoided trials was immediately significantly lower in the single avoidable trial compared to the unavoidable A+ trial ($F_{(1, 64)}$ =32.03, *p*<.001, η_p^2= .33) and continued to decline in a linear trend ($F_{\text{Trial} (1, 74)}$ =63.97, *p*<.001, η_p^2= .46) (see Figure 1A and Table 2). SCR for A+ also declined over time ($F_{\text{Trial} (1, 69)}$ =16.06, *p*<.001, η_p^2= .19) (see Figure 1B). Groups did not differ in A+ threat expectancy, SCR and startle EMG (*F*'s <1, *p*'s>.05) and there were no significant interactions (*p*'s >.05). Furthermore, pairwise comparisons showed that there were no significant group differences for B- and C- in threat expectancy (B-: *t*(74) = 0.06, *p*>.05; C-: *t*(61.67)= 1.99, *p*>.05) and SCR (B-: *t*(69) = 1.11, *p*>.05; C-: *t*(69)= 0.34, *p*>.05), showing similar avoidance learning across both groups.

3.4 Avoidance shift learning

A 2 (Group) x 3 (Trial) mixed ANOVA for threat expectancy of C- revealed a significant main effect of Trial ($F_{(1.35, 99.70)}$ =11.67, *p*<.001, η_p^2= .14), but only a trend level effect of Group ($F_{(1, 74)}$ =3.35, *p*=.07) (Figure 1 and Table 2). There were no significant trial x group interactions (*p*'s >.05). Planned comparisons revealed non-significant but trend level higher threat expectancy for C- during the first trial block for the Experimental group compared to the Control group, (*p*=.06). Skin conductance analyses for C- revealed a significant main effect of Trial (*F*

$_{(1.60,\ 110.5)}$ =9.75, p<.001, η_p^2=.12), but no effect of Group (F>1, p<.05). Planned comparisons revealed only a marginal difference during the last trial block between the Experimental group and the Control group in the avoidance shift learning phase (p=.05). There were no differences in startle EMG for C- (t(61) =1.24, p>.05) (Figure 2).

3.5 Test

A 2 (Group) x 3 (CS) mixed ANOVA was conducted using only the first A+, B- and C- trials during the test phase. Crucially, there was a significant Group x CS interaction ($F_{(1.55,\ 109.7)}$ =6.04, p<.01, η_p^2=.08) as well as significant main effects of Group ($F_{(1,\ 71)}$ =5.71, p=.02, η_p^2=.07) and CS ($F_{(1.55,\ 109.7)}$ =372.1, p<.001, η_p^2=.84) in threat expectancy (Figure 1 and Table 2). Planned comparisons further revealed that expectancy for A+ was higher than both B- and C-. The Experimental group had higher expectancy for C- and B- than the Control group (p's<.01), while the Control group made similar ratings for B- and C- (p>.05).

The 2 (Group) x 3 (CS) x 2 (Trial) mixed ANOVA using two binned trials of A+, B- and C- revealed similar results. There was a significant Group x CS interaction ($F_{(1.31,\ 95.81)}$ =3.97, p=.02, η_p^2=.05) as well as significant main effects of Group ($F_{(1,\ 73)}$ =6.71, p=.01, η_p^2=.08) and CS ($F_{(1.31,\ 95.81)}$ =555.8, p<.001, η_p^2=.88). The Group x CS x Trial interaction for threat expectancy, however, approached significance ($F_{(1.71,\ 125.19)}$ =3.00, p=.06), while the trial x group and CS by trial interactions were non-significant (p's>.05). Planned comparisons revealed similar results during trial block one to the single trial comparison. That is, expectancy for A+ was significantly higher than B- and C-, while B- and C- also differed in the Experimental group (p's<.01) but did not differ within the Control group (p>.05). Trial block two showed that threat expectancy for C- for the Experimental group was only trend level different from the Control group (p=.07) and did not differ from B- (p>.05). Surprisingly, ratings for B- were significantly higher in the Experimental group than in the Control group (*p*=.02). Expectancy for A+ remained significantly higher than both B- and C- (*p*'s<.01). Together, this indicates that the Experiment al group's higher threat expectancy for C- did not persist over time and that expectancies for B- and C- were generally higher throughout the test than in the Control group (Table 2).

SCR analyses did not show any significant main effects or interactions (*F*'s <2.31, *p*'s>.05). Planned comparisons showed that SCRs on A+ trials were not significantly higher than on B- and C- trials (*p*'s>.05). There was also no SCR difference in C- and B- trials between the groups (*p*'s >.05).

Startle EMG data did, however, show a significant main effect of CS ($F_{(1.96, 121.7)}$ =18.07, *p*<.001, η_p^2=.23). Post hoc analyses showed that startle responses were significantly higher in A+ trials than B- and C- indicating fear potentiated startle across both groups (Figure 2 and Table 3).

3.6 Avoidance shift learning vs. Test

To investigate whether the removal of the avoidance cue during Test was responsible for the increase in C- threat expectancy ratings, a 2 (Group) x 3 (CS) x 2 (Phase) mixed ANOVA was conducted using the last trial block during Avoidance shift learning and the first trial block of Test. There was a significant interaction of Group x CS x Phase ($F_{(.1.94, 141.23)}$ =4.74, *p*=.01, η_p^2=.06). Follow up planned comparisons showed that for both groups, ratings for A+ and B- did not significantly differ between the avoidance shift learning and test (*p*'s<.05). C- threat expectancy, however, significantly increased in the Experimental group from avoidance shift learning to test (*p*<.001), but not in the Control group (*p*>.05).

3.7 Avoidance learning vs. Test

To investigate the general effect of the avoidance shift learning phase on shock expectancy ratings for the safety stimuli, a 2 (Group) x 2 (CS) x 2 (Phase) mixed ANOVA comparing the Avoidance learning phase block to the first trial block during Test. There was no significant Group x CS x Phase interaction ($F_{(1, 73)} < 1$, $p > .05$), however the Group x Phase interaction was significant ($F_{(1.31, 95.81)} = 4.49$, $p = .04$, $\eta_p^2 = .06$). Planned comparisons showed that threat expectancy ratings for both B- and C- did not change from Avoidance learning to Test in the Experimental group (*p*'s>.05) but decreased in Test in the Control group (*p*'s<.05). Similar to Engelhard et al.'s (2015) finding, avoidance of C- seemed to not only maintain pre-existing threat for that cue, but also for B-.

3.8 Shift learners vs. non-shift learners

During avoidance shift learning, 14 participants were excluded as they did not meet the learning criteria. Explorative tests yielded no differences in threat expectancy for the CSs or anxiety about receiving shocks (all *t*(49)'s <1, *p*'s>.05).

4. Discussion

The aim of the present study was to replicate Engelhard et al. (2015) with methodological changes consisting of equal numbers of stimulus presentations, the inclusion of SCR and startle EMG measures, ecologically valid stimuli, and an extended test phase. During threat conditioning, C- was successfully conditioned as a safety cue, as threat expectancy, SCR and startle EMG were similar to B-, and both were significantly lower than A+. Furthermore, all measures of B- and C- did not differ during avoidance learning across groups. In line with the original findings, we observed significantly higher threat expectancy for C- in the Experimental group during the first trial of the Test phase compared to both C- and B- in the Control group. Expectancy ratings for A+ and B- did not differ across groups, and there was no difference between B- and C- in the Control group. Higher threat expectancy towards C- in the Experimental group did not persist as ratings for C- did not differ from B- in later test trials and was only marginally higher than C- in the Control group. Threat expectancy for B- was significantly higher between the groups (and remained high for both B- and C- in the Experimental group compared to the Control group). Furthermore, threat expectancy for C- in the Experimental group significantly increased from Shift learning to the Test phase, indicating that the removal of the avoidance cue was responsible for the observed increase in expectancy ratings. In the Experimental group, both SCR and startle measures were undifferentiated for both B- and C- and no group differences were found. Furthermore, comparisons between the safety stimuli pre-and post-avoidance shift phase showed that both B- and C- retained similar levels of threat post-avoidance shift phase for participants in the Experimental group, while there was a significant decrease in threat expectancy for both B- and C- in the Control group.

The present study confirmed that, following avoidance availability and its subsequent removal, perceived low-level threat is not only maintained for the conditioned safety stimulus, but also spreads to other, unambiguously safe, stimuli (Engelhard et al., 2015). A prior learning history that includes the opportunity to engage in avoidance thus not only maintains the threat-value of previously safe stimuli but also increases generalized threat perception. Such findings indicate that the (over-)judicious use of safety behaviors in exposure-based treatment for excessive threat-avoidance should be closely monitored (Blakey & Abramowitz, 2016; Meulders, Van Daele, Volders, & Vlaeyen, 2016).

There are three potential explanations for how avoidance may have maintained the threat value of the safety cues during the test phase. First, avoidance responding in the presence of C- in the Experimental group may have blocked participants from learning that C- was never followed by the US. During the avoidance shift phase, participants may have attributed the absence of shock to avoidance, effectively providing protection from extinction by blocking safety learning towards C- (Lovibond et al., 2009). This maintenance of pre-existing threat perception assumes that threat perception towards C- was sufficiently established prior to Avoidance shift learning to warrant avoidance. Although threat conditioning was successful, it is possible that residual, perceived threat value remained high enough to warrant avoidance responding. While protection from extinction may explain Experimental participants' higher threat expectancy towards C- , it does not explain the maintenance of threat expectancy in B- in the latter trials of the Test phase since B- was presented equally and without any avoidance availability in both groups.

Second, performing the avoidance response in the presence of C- may have led to retroactive threat inference. As Engelhard et al. (2015) suggested, performing safety behaviors in a safe context can lead to increased threat perception or 'behavior as information' effects (Gangemi et al., 2012; van den Hout et al., 2014). That there would be any avoidance behavior at all despite learning that C- was paired with the absence of shock, was perhaps not unexpected. Even when explicitly stated that no shock would occur during CS-, avoidance still occurred (Lommen, Engelhard & van den Hout, 2010; Xia, Dymond, Lloyd & Vervliet, 2017). Avoidance here may have led to a form of cognitive dissonance which participants sought to reduce via threat expectancy (Uijen et al., 2017). Similar to protection of extinction, however, this explanation also cannot account for the maintained threat expectancy for B- in Experimental participants.

Finally, generalization from perceptual similarity of our CS+ may have led to a reduced spread of the excitatory properties of CS+ to both CS-. Haddad, Pritchett, Lissek and Lau (2012) have shown that threat perception may unintentionally generalize from a face CS+ to another face CS- based on perceptual similarity. It is unclear, however, if the generalization seen in

Haddad et al. can only be attributed to perceptual similarity; both CS-s were categorically dissimilar as well, which may have also facilitated category/non-category related fear generalization (Dymond, Dunsmoor, Vervliet, Roche & Hermans, 2015). In the present study, it is unlikely that perceptual generalization alone influenced threat perception as threat perception of B- and C- remained low in Control participants. Therefore, the maintenance of threat perception towards B- and C- may have been the result of a combination of perceptual generalization and either protection from extinction or behavior as information effects.

Future research should investigate the effect of methodological factors on the rate of avoidance shift learning. For example, avoidance shift learning may be less effective given a more perceptually dissimilar CS- or when the safety context is made explicitly clear by removing shock electrodes (Gillan et al., 2014). Avoidance may serve various functions, and it follows that accurately determining which function it serves has important implications for therapeutic success (Treanor & Barry, 2017). Furthermore, in both the present study and Engelhard et al. (2015), there was a high proportion of participants (27% and 28%, respectively) who did not successfully acquire avoidance shift learning. Further research is needed to reduce attrition, perhaps through the addition of a response-produced safety signal (Crawford & Masterson, 1978). Avoidance shift learning may also be influenced by individual differences. While explorative tests yielded no differences between participants who did or did not shift their avoidance behavior to C-, this may be partly attributed to low and unequal sample sizes. Given that deficiency in safety learning is a hallmark trait for anxious individuals (Craske, Treanor, Conway, Zbozinek & Vervliet, 2014; Duits et al., 2015), the awareness that shifting avoidance behavior is necessary may be an indicator of anxiety resilience. Finally, it has been suggested that safety behaviors can increase the perceived threat level of safety cues (Engelhard et al., 2015; van den Hout et al., 2014; van Uijen & Toffolo, 2015). The present study, as well as Engelhard et al., instead found that low-level threat was maintained. To date, the increase in threat level following use of safety behaviors is limited to clinical research (Gangemi et al., 2012; van den Hout et al., 2014; van Uijen & Toffolo, 2015). Further laboratory-based treatment research is needed to test the commonly held clinical assumption that safety behaviors increase threat value.

A limitation was the absence of significant SCR or startle responding during the test phase, which may be related to habituation. Startle responding is however less influenced by cognitive processes (Hamm & Weike, 2005) and the significant results for A+ during the Test indicates that the maintained threat value of B- and C- in the Experimental group was not motivated by implicit, automatic fear.

In conclusion, the present study confirms that perceived threat value increases following the removal of low-cost avoidance for a known safety stimulus, which may maintain fear and thereby impair safety learning.

5. References

Barlow, D. H., Raffa, S. D., & Cohen, E. M. (2002). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In P.E. Nathan & J.M. Gorman (Eds.), *A guide to treatments that work* (pp. 301-335). Oxford University Press, New York.

Blakey, S. M., & Abramowitz, J. S. (2016). The effects of safety behaviors during exposure therapy for anxiety: Critical analysis from an inhibitory learning perspective. *Clinical Psychology Review*, *49*, 1-15. doi:10.1016/j.cpr.2016.07.002

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*, 1-15. doi: 10.1111/j.1469-8986.2005.00271.x.

Boddez, Y., Baeyens, F., Hermans, D., & Beckers, T. (2014). A learning theory approach to anxiety disorders: Human fear conditioning and the added value of complex acquisition procedures. In P. Emmelkamp, & T. Ehring (Eds.*), International Handbook of Anxiety Disorders: Theory, Research and Practice* (pp. 85-104). Wiley-Blackwell.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, *11*, 485–494.

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

Crawford, M., & Masterson, F. (1978). Components of the flight response can reinforce bar-press avoidance learning. *Journal of Experimental Psychology*, *4*, 144–151.

Dawson, M. E., & Schell, A. M. (1985). Information processing and human autonomic classical conditioning. *Advances in Psychophysiology*, *1*, 89–165

De Houwer, J., Crombez, G., & Baeyens, F. (2005). Avoidance behavior can function as a negative occasion setter. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 101-106.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M., & Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depression & Anxiety, 32, 239-253.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. *Behavior Therapy*, *46*, 561-582. doi: 10.1016/j.beth.2014.10.001

Engelhard, I. M., van Uijen, S. L., van Seters, N., & Velu, N. (2015). The effects of safety behavior directed towards a safety cue on perceptions of threat. *Behavior Therapy, 46,* 604-610. doi: 10.1016/j.beth.2014.12.006

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28,* 1-11.

Gangemi, A., Mancini, F., & van den Hout, M. (2012). Behavior as information: "If I avoid, then there must be a danger". *Journal of Behavior Therapy and Experimental Psychiatry, 43,*1032-1038. doi:10.1016/j.jbtep.2012.04.005

Gigerenzer, G. (2006). Out of the frying pan into the fire: Behavioral reactions to terrorist attacks. *Risk Analysis, 26*, 347-351. doi:10.1111/j.1539-6924.2006.00753.x

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A-M., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W., (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry, 75*, 631–638. (doi:10.1016/j.biopsych.2013.02.002)

Haddad, A. D., Pritchett, D., Lissek, S., & Lau, J. Y. (2012). Trait anxiety and fear responses to safety cues: stimulus generalization or sensitization? *Journal of Psychopathology and Behavioral Assessment, 34(3),* 323-331. doi:10.1007/s10862-012-9284-7

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology, 57,* 5-14. doi:10.1016/j.ijpsycho.2005.01.006

Handley, R. V., Salkovskis, P. M., Scragg, P., & Ehlers, A. (2009). Clinically significant avoidance of public transport following the London bombings: Travel phobia or subthreshold posttraumatic stress disorder? *Journal of Anxiety Disorders, 23,* 1170-1176. doi:10.1016/j.janxdis.2009.07.023

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

Klumpers, F., Morgan, B., Terburg, D., Stein, D. J., & van Honk, J. (2015). Impaired acquisition of classically conditioned fear-potentiated startle reflexes in humans with focal bilateral basolateral amygdala damage. *Social Cognitive and Affective Neuroscience*, *10*, 1161-1168. doi:10.1093/scan/nsu164.

Krypotos, A., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience, 9,* 189. doi:10.3389/fnbeh.2015.00189

LeDoux, J. E., Moscarello, J., Sears, R., & Campese, V. (2016). The birth, death and resurrection of avoidance: A reconceptualization of a troubled paradigm. *Molecular Psychiatry*, *22*, 24-36. doi:10.1038/mp.2016.166

Lommen, M. J., Engelhard, I. M., & van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? *Personality and Individual Differences*, *49*, 1001-1006. doi:10.1016/j.paid.2010.08.012

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas J. M. P., & 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 & Biobehavioral Reviews, 77*, 247-285. doi:10.1016/j.neurobiorev.2017.02.026

Lovibond, P. F., Hanna, S. K., Siddle, D. A. T., & Bond, N.W. (1994). Electrodermal and subjective reactions to fear-relevant stimuli under threat of shock. *Australian Journal of Psychology, 46,* 73–80. doi:10.1080/00049539408259476

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. doi:10.1016/j.brat.2009.04.013

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). *OpenSesame*: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44,* 314-324. doi:10.3758/s13428-011- 0168-7.

Meulders, A., Van Daele, T., Volders, S., & Vlaeyen, J. W. S. (2016). The use of safety-seeking behavior in exposure-based treatment for fear and anxiety: Benefit or burden? A meta-analytic review. *Clinical Psychology Review, 45*, 144-156. doi: 10.1016/j.cpr.2016.02.002

Prager, F., Asay, G. R., Lee, B., & Winterfeldt, D. V. (2011). Exploring reductions in London Underground passenger journeys following the July 2005 bombings. *Risk Analysis*, *31*, 773-786. doi:10.1111/j.1539-6924.2010.01555.x

Rachman, S., Radomsky, A. S., & Shafran, R. (2008). Safety behaviour: A reconsideration. *Behaviour Research and Therapy*, *46*, 163-173. doi:10.1016/j.brat.2007.11.008

Renard, T. (2016, September 9). Fear not: A critical perspective on the terrorist threat in Europe. *Egmont Institute.* Retrieved from http://www.egmontinstitute.be/fear-not-a-critical-perspective-on-the-terrorist-threat-in-europe/

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, NY: Appleton-Century-Crofts.

Rubin, G. J., Brewin, C. R., Greenberg, N., Hughes, J. H., Simpson, J., & Wessely, S. (2007). Enduring consequences of terrorism: 7-month follow-up survey of reactions to the bombings in London on 7 July 2005. *British Journal of Psychiatry, 190*, 350-356. doi:10.1192/bjp.bp.106.029785

Shaver, M. (2015, November 23). You're more likely to be fatally crushed by furniture than killed by a terrorist. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2015/11/23/youre-more-likely-to-be-fatally-crushed-by-furniture-than-killed-by-a-terrorist/?utm_term=.836039a5f30e

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A. ... & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242-249. doi: 10.1016/j.psychres.2008.05.006

Treanor, M., & Barry, T. J. (2017). Treatment of avoidance behavior as an adjunct to exposure therapy: Insights from modern learning theory. *Behaviour Research and Therapy, 96,* 30-36. doi:10.1016/j.brat.2017.04.009

Van den Hout, M., Gangemi, A., Mancini, F., Engelhard, I. M., Rijkeboer, M. M., van Dam, M., & Klugkist, I. (2014). Behavior as information about threat in anxiety disorders: A comparison of patients with anxiety disorders and non-anxious controls. *Journal of Behavior Therapy and Experimental Psychiatry, 45*, 489-495. doi:10.1016/j.jbtep.2014.07.002

van de Veer, E., de Lange, A. M., van der Haar, E., & Karremans, J. (2012). Feelings of safety: Ironic consequences of police patrolling. *Journal of Applied Social Psychology, 42*, 3114-3125. doi:10.1111/j.1559-1816.2012.00967.x.

Van Uijen, S. L., Leer, A., & Engelhard, I. M. (2017). Safety behavior after extinction triggers a return of threat expectancy. *Behavior Therapy.* doi:10.1016/j.beth.2017.08.005

van Uijen, S. L., & Toffolo, M. B. (2015). Safety behavior increases obsession-related cognitions about the severity of threat. *Behavior Therapy*, *46*, 521-531. doi:10.1016/j.beth.2015.04.001

Vervliet, B., & Indekeu, E. (2015). Low-cost avoidance behaviors are resistant to fear extinction in humans. *Frontiers in Behavioral Neuroscience, 9*, 351. doi:10.3389/fnbeh.2015.00351

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: application to models of anxiety and depression. *Psychological Medicine, 43*, 2241-2244. doi: 10.1017/S0033291712002267

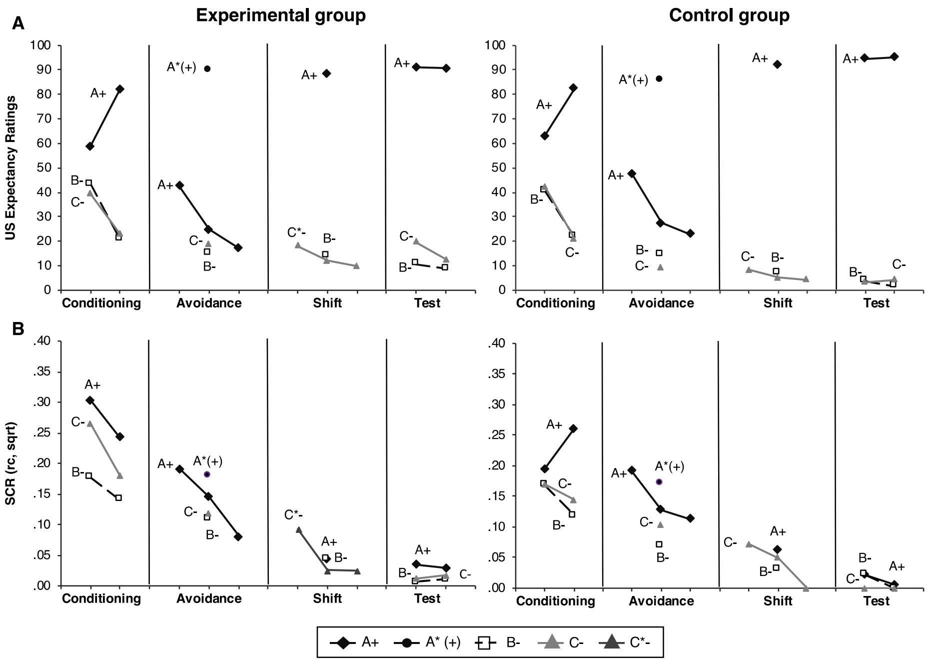
Volders, S., Meulders, A., de Peuter, S., Vervliet, B., & Vlaeyen, J. (2012). Safety behavior can hamper the extinction of fear-related pain: An experimental investigation in healthy participants. *Behaviour Research and Therapy, 50*, 735-746. doi:10.1016/j.brat.2012.06.004

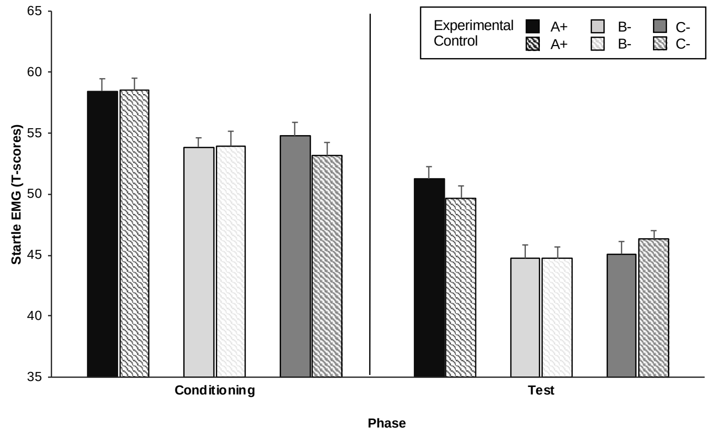
Xia, W., Dymond, S., Lloyd, K., & Vervliet, B. (2017). Partial reinforcement of avoidance and resistance to extinction in humans. *Behaviour Research and Therapy, 96*, 79-86. doi:10.1016/j.brat.2017.04.002

Figure captions

Figure 1. Mean shock expectancy ratings and SCR results during A+, B- and C- for both Experimental and Control groups. (A): Mean shock expectancy ratings during all phases of the experiment. (B): Mean SCR during all phases of the experiment. SCR are reported in μV, range-corrected and square root transformed.

Figure 2. Mean startle EMG data during threat conditioning and test phases. Startle EMG is reported in T-scores. *Note:* Error bars represent SEM.


A
Experimental group
Control group
US Expectancy Ratings
Conditioning
Avoidance
Shift
Test
A+
B-
C-
A*(+)
C*-
B
SCR (rc, sqrt)
A+ A* (+) B- C- C*-


Experimental A+ B- C-
Control A+ B- C-
Startle EMG (T-scores)
65
60
55
50
45
40
35
Conditioning
Test
Phase