Available online at www.sciencedirect.com

ScienceDirect

Behavior Therapy 49 (2018) 450–458

Behavior Therapy

www.elsevier.com/locate/bt

# Safety Behavior After Extinction Triggers a Return of Threat Expectancy

Sophie L. van Uijen
Arne Leer
Iris M. Engelhard
Utrecht University

**Safety behavior is involved in the maintenance of anxiety disorders, presumably because it prevents the violation of negative expectancies. Recent research showed that safety behavior is resistant to fear extinction. This fear conditioning study investigated whether safety behavior after fear extinction triggers a return of fear in healthy participants. Participants learned that two stimuli (A and C) were followed by an aversive loud noise ("threat"), and one stimulus (B) was not. Participants then learned to use safety behavior that prevented the loud noise. Next, A and C were no longer followed by the loud noise, which typically led to extinction of threat expectancy. Safety behavior then became available again for C, but not for A and B. All participants used safety behavior on these C trials. In a final test phase, A, B, and C were presented once without the availability to use safety behavior. At each stimulus presentation, participants rated threat expectancy by indicating to what extent they expected that the loud noise would follow. Compared with the last extinction trial, threat expectancy increased for C in the test phase, whereas it did not increase for A and B. Hence, safety behavior after the extinction of classically conditioned fear caused a partial return of fear. The findings suggest that safety behavior may be involved in relapse after exposure-based therapy for anxiety disorders.**

We thank Angelos Krypotos for his advice and helpful comments while writing the manuscript, and Nesrin Tatar for collecting the data.

This study was funded with a Research Talent Grant from the Netherlands Organization for Scientific Research (NWO) awarded to Sophie van Uijen (No. 403-11-406).

Address correspondence to Sophie L. van Uijen, PO Box 80140, 3508 TC Utrecht, the Netherlands; e-mail: s.l.vanuijen@uu.nl.



*Keywords:* safety behavior; exposure; threat expectancy; return of fear; anxiety disorders

COGNITIVE-BEHAVIORAL THERAPY (CBT), in particular, exposure-based therapy, is the treatment of choice for anxiety disorders (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). Patients are exposed to the feared but innocuous stimulus (e.g., going to the supermarket) to learn that the threat they expect (e.g., getting a heart attack) does not follow. In classical, or Pavlovian, conditioning terms, the feared but innocuous stimulus is a conditional stimulus (CS), and the expected threat is an unconditional stimulus (US). Threat expectancy is used as an index of fear in fear conditioning research (Boddez et al., 2013). The decrease in fear (conditional response [CR]) following repeated exposure to the CS is called extinction, a process that is widely used as a model for exposure therapy (e.g., Milad & Quirk, 2012). During extinction learning, the original CS–US relationship is not unlearned, but the CS acquires an additional inhibitory association with the US (i.e., CS - no US). This inhibitory association exists alongside the original excitatory association (Bouton, 2016). Thus, after fear extinction, the CS has an ambiguous status. The context in which the CS is presented determines whether or not it evokes fear (Bouton, 2002), which poses a continuous risk for a return of fear.

Although exposure-based therapy is the treatment of choice, a substantial minority shows insufficient improvement from it (Barlow, Allen, & Choate, 2004) or relapses after initial recovery (e.g., Vervliet, Craske, & Hermans, 2013). Fear conditioning research has indicated several conditions that may trigger a return of fear after extinction, and which are likely involved in relapse (see Bouton, 2002, 2016; Vervliet et al., 2013). Extinguished fears can return with a change in the external context or a person's internal state (which is different from the extinction context; renewal), with the passage of time (spontaneous recovery), and following an unsignaled US (reinstatement). Investigating other conditions that may trigger a return of experimentally conditioned fear can improve our understanding of the processes underlying return of fear in clinical practice.

There are various reasons to hypothesize that safety behaviors may also be involved in the return of fear. Safety behaviors are basically avoidance behaviors to minimize the feared outcome, such as a patient with panic disorder who may carry anxiolytics to feel safe going to the supermarket (e.g., Salkovskis, 1991). They function as safety signals and influence threat expectancy. They are common in individuals with anxiety disorders and can be problematic in treatment, because they can maintain threat expectancy. This was shown in a de novo fear conditioning experiment by Lovibond, Mitchell, Minard, Brady, and Menzies (2009), which entailed a mix of classical and instrumental conditioning. In a Pavlovian (i.e., classical) acquisition phase, participants learned that two neutral stimuli (A and C, which both served as CS+) were followed by shock (US), and one stimulus (B, which served as CS–) was not. Next, in an instrumental conditioning phase, participants learned to use safety behavior during presentation of stimulus A by pressing a button on a response box that effectively canceled the shock. During a subsequent fear extinction phase, stimulus C was no longer followed by shock. The experimental group, but not the control group, was given the opportunity to use safety behavior during C trials. All participants in the experimental group used safety behavior on all C trials. In the following test phase, safety behavior was no longer available for any stimulus. Threat expectancy for C remained high in the experimental group, whereas it had decreased in the control group (Lovibond et al., 2009). Using safety behavior during unreinforced CS presentations thus preserved the subjective threat value of the CS. Because safety behaviors can prevent fear extinction, clinical guidelines recommend to motivate patients to drop all safety behaviors during exposure-based therapy (e.g., Abramowitz, Deacon, & Whiteside, 2011; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Keijsers, van Minnen, & Hoogduin, 2011). However, if individuals continue to use safety behavior after exposure-based therapy, they may again misattribute safety to their own behavior rather than to innocuous properties of the CS.

A second reason to hypothesize that safety behavior may cause a return of fear is that besides signaling safety, it can signal potential threat, even in objectively safe situations. Using a design similar to Lovibond et al. (2009), Engelhard, van Uijen, van Seters, and Velu (2015) found that when safety behavior is used in response to a safety cue (i.e., a CS that has never been paired with shock; CS–), this paradoxically increases threat expectancy to that stimulus when it is subsequently presented without safety behavior. In line with this finding, Deacon and Maack (2008) found that cleaning-related safety behavior increases threat perceptions and contamination anxiety in healthy participants. Likewise, health-related safety behavior increases health anxiety and hypochondriacal beliefs (Olatunji, Etzel, Tomarken, Ciesielski, & Deacon, 2011), and increasing obsessive-compulsive-related checking behavior increases obsession-related threat beliefs (van Uijen & Toffolo, 2015) in healthy participants.

A third reason to hypothesize that safety behavior may cause a return of fear is that Vervliet and Indekeu (2015) recently found that safety behavior is resistant to fear extinction. Comparable to the Lovibond et al. (2009) and Engelhard et al. (2015) studies, participants learned that pressing a button (i.e., safety behavior) during CS+ presentation effectively canceled a subsequent shock. Next, fear extinction was established: threat expectancy and skin conductance responses decreased after repeated unreinforced CS presentations during which safety behavior was unavailable to the participants. However, when safety behavior became available again after fear extinction, participants showed a significant return of safety behavior. Apparently, extinction of classically conditioned fear did not mitigate safety behavior. What is more, threat expectancy to the extinguished CS+ increased when safety behavior was subsequently made unavailable again (Vervliet & Indekeu, 2015).

In summary, previous research has shown (a) that safety behavior can signal safety and prevent fear extinction (Lovibond et al., 2009), (b) that safety behavior can signal threat (Engelhard et al., 2015), and (c) that extinction of classically conditioned fear does not eliminate safety behavior (Vervliet & Indekeu, 2015). However, it remains unclear whether safety behavior after fear extinction has detrimental effects on fear. This is not only theoretically but also clinically relevant (Treanor & Barry, 2017). Using safety behavior after fear extinction logically again causes a misattribution of safety to the behavior instead of to the innocuous CS. Moreover, by signaling threat, safety behaviors may trigger a return of threat expectancy. Retention of safety behavior could then predict relapse. Accordingly, we tested whether using safety behavior after fear extinction triggers a return of fear. We used a within-subjects fear conditioning paradigm adapted from Lovibond et al. (2009) to investigate this hypothesis. Similar to Lovibond et al. (2009), the paradigm was a mix of classical and instrumental conditioning. Pavlovian acquisition took place with two CS+ (A and C) and one CS– (B). However, instead of shock, the US was a loud noise presented through headphones. Next, instrumental conditioning occurred in the safety behavior acquisition phase. Participants learned to use safety behavior by unplugging the headphones during presentation of all three CSs, which canceled the loud noise that would otherwise follow stimuli A and C. Then, during a fear extinction phase, A and C were no longer followed by the US. After the extinction phase, safety behavior was again made available during C trials, but not during A and B trials. We hypothesized that this would increase threat expectancy for C in a subsequent test phase, in which C was presented without safety behavior. More specifically, we hypothesized that, compared with stimuli A and B, threat expectancy for stimulus C would increase from the final extinction phase trial to the test phase. Furthermore, we explored whether safety behavior affected evaluative conditioning in the current paradigm. Evaluative conditioning is the change in the valence of a CS after it has been paired with a US (De Houwer, Thomas, & Baeyens, 2001). A CS itself becomes more negative after it has been paired with an aversive US. Safety behavior may also directly make a stimulus more negative (see Centerbar & Clore, 2006). We therefore compared valence ratings of A, B, and C after the experimental task, and hypothesized that stimulus C would be rated more negative than stimuli A and B.

## Method

### PARTICIPANTS

Participants were 45 student volunteers ($M_{age}$ = 22.38, *SD* = 2.38; 36 women, 9 men). They gave written informed consent and received €7 (approximately $8) or course credit for their cooperation. The study was approved by the Ethics Review Board of the Faculty of Social and Behavioural Sciences of Utrecht University.

### APPARATUS AND STIMULUS MATERIALS

The task was programmed in Python (Python Software Foundation, 2012) and presented on a 19-inch monitor. The US was 1-s 100 dB white noise presented through headphones (cf. Leer & Engelhard, 2015) that were connected to the computer with a sound amplifier. CSs were a black square, triangle, and circle.

### EXPERIMENTAL TASK

Each trial consisted of the presentation of a CS for 5 s, followed by a 5-s waiting period during which participants rated threat expectancy, immediately followed by either the US or no US. The intertrial interval was 3 s. A picture of a plug was visible in the upper right corner of the screen during each CS presentation. The color of this plug indicated safety behavior availability. Safety behavior was unavailable to the participant if the plug was gray and available if the plug was green. During safety behavior available trials, participants could unplug the headphones from the sound amplifier, which prevented them from hearing the US. At the end of each safety behavior available trial, an instruction screen told participants to plug the headphones back into the sound amplifier. The design of the study is shown in Table 1. A, B, and C were randomly allocated to the different shapes for each participant. In the Pavlovian acquisition phase, A and C were followed by the US, and B was not. In the safety behavior acquisition phase, safety behavior was available during presentation of A, B, and C. In this phase, A, B, and C were also presented without safety behavior availability to remind participants of the CS–US contingencies. From the extinction phase onward, A and C were no longer followed by the US. In the return of safety behavior phase, safety behavior was again available during C trials, but not for A and B. Finally, in the test phase, A, B, and C were presented once without availability of safety behavior. The experimental task is graphically depicted in Table 2. The order of trial types was randomized within each phase, with the restriction that there were no more than two consecutive trials of the same type. Furthermore, the final three trials of the extinction phase were A, B, and C, and C was always presented last in the test phase.

### MEASURES

*State-Trait Anxiety Inventory (STAI)*

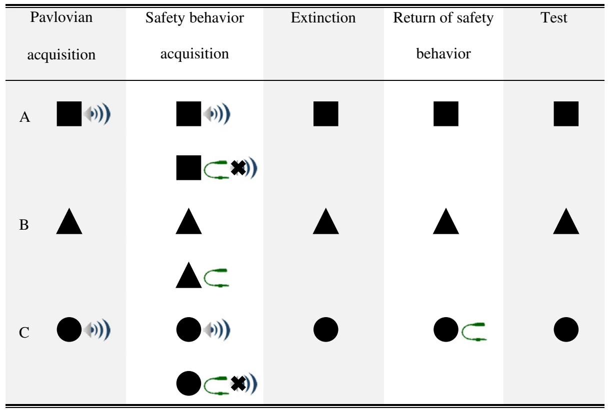
The STAI (Spielberger, Gorusch, & Lushene, 1970) was included to measure state and trait anxiety,

Table 1
Design of the Experimental Task

| Pavlovian acquisition | Safety behavior acquisition | Extinction | Return of safety behavior | Test |
|---|---|---|---|---|
| A+ (3) | A+ (1)<br>A*(+) (2) | A– (7) | A– (6) | A– (1) |
| B– (6) | B– (1)<br>B*– (2) | B– (4) | B– (3) | B– (1) |
| C+ (3) | C+ (1)<br>C*(+) (2) | C– (7) | C*– (6) | C– (1) |

*Note.* A, B, and C refer to visual stimuli (CSs); + and – refer to presence and absence, respectively, of a loud noise (US) following the CSs; * indicates the availability of the safety behavior; (+) indicates that participants heard the loud noise only if they did not use safety behavior; numbers in parentheses indicate the number of trials of each type.

Table 2
Graphical Depiction of the Experimental Task




*Note.* The black square, triangle, and circle are the CS. A, B, and C were randomly allocated to the different shapes for each participant. The picture of a loudspeaker indicates the presence of a loud noise (US) following CS presentation. The picture of the plug indicates the availability of the safety behavior. A cross through the picture of the loudspeaker indicates that participants did not hear the loud noise if they unplugged the headphones (i.e., used safety behavior).

because they may affect fear learning (Grillon et al., 2006; Lissek et al., 2005). Each construct was measured by 20 items, rated on a scale from 1 (*not at all*) to 4 (*severely*). In this study, Cronbach's α = .87 and .94 for state and trait anxiety, respectively.

### *Threat Expectancy*

Immediately after each CS presentation, but before (possible) US presentation, participants rated to what extent they expected the loud noise to follow on a 0 (*certain no loud noise*) to 100 (*certain loud noise*) visual analogue scale (VAS) that was presented on the computer screen.

### *Valence and Pleasantness*

After the experimental task, participants rated the valence of each CS on a 100 mm VAS from *negative* (left) to *positive* (right) with *neutral* in the middle. Participants also rated the (un)pleasantness of the US and the safety behavior on two 100 mm VAS from *extremely unpleasant* (left) to *extremely pleasant* (right).

### PROCEDURE

After the informed consent procedure, participants filled out the STAI. Then they received oral instructions from the experimenter, followed by written instructions on the computer screen. Participants were told that there was a relationship between the CS and the US, and that they should try to discover this relationship. Next they were told that there was a picture of a plug visible during each CS presentation, and that this picture could be gray or green. It was explained that they could unplug the headphones from the sound amplifier when the picture was green. They were shown how to unplug the headphones (i.e., safety behavior) and plug them back into the sound amplifier, but they were not told what the consequence of unplugging the headphones was. At the start of the experimental task, participants practiced rating threat expectancy twice and using safety behavior twice. After finishing the experimental task, participants filled out the questionnaire about valence and pleasantness. Next, they were debriefed, thanked, and rewarded.

### SCORING AND ANALYSIS

Drawing conclusions on the renewal of fear is appropriate only when both fear acquisition and fear extinction took place. We defined Pavlovian awareness as a higher threat expectancy for A and C than for B on the final trial of the Pavlovian acquisition phase, and extinction as threat expectancy ratings below 30 for A and C at the final extinction phase trial (cf. Leer, Engelhard, Dibbets, & van den Hout, 2013). We used Statistical Package for the Social Sciences (SPSS; Version 24; IBM Corp., 2016) for the analyses. Data were analyzed with repeated measures ANOVAs (α = .05), comparing threat expectancy ratings between stimuli (A vs. B vs. C), or stimuli and trials (e.g., final extinction phase trial vs. test phase). Corrected values were reported in case the assumption of sphericity was violated. Planned comparisons were conducted with paired *t*-tests, which were tested one-tailed when we had a specific prediction for the direction of the effect.

## Results

### PARTICIPANTS

Data of 15 participants were excluded from the analyses,[1] because 3 showed no fear acquisition, 11 showed no extinction, and 1 participant thought that the US would be audible from the computer's internal speaker when the headphones were unplugged. This resulted in a final sample of 30 participants ($M_{\text{age}}$ = 21.87, *SD* = 2.18; 25 women, 5 men). All participants used safety behavior on all six C*– trials in the return of safety behavior phase. Excluded participants did not differ from included participants in state ($M_{\text{included}}$ = 31.47, *SD* = 6.24) or trait anxiety ($M_{\text{included}}$ = 35.27, *SD* = 8.99) scores, both *t*s < 1.

### THREAT EXPECTANCY

Figure 1 shows that Pavlovian acquisition occurred. At the end of the Pavlovian acquisition phase,

[1] Results were comparable when all 45 participants were included in the analyses. The change in threat expectancy from the final extinction phase trial to the test phase differed between A, B, and C, $F(2, 88) = 4.08$, $p = .02$, $\eta_p^2 = .09$. Threat expectancy decreased for B, $t(44) = 2.91$, $p = .006$, $d = 0.41$, and did not significantly change for A, $t(44) = 1.89$, $p = .07$. However, it did not significantly change for C either, $t(44) = 1.07$, $p = .15$ (one-tailed).

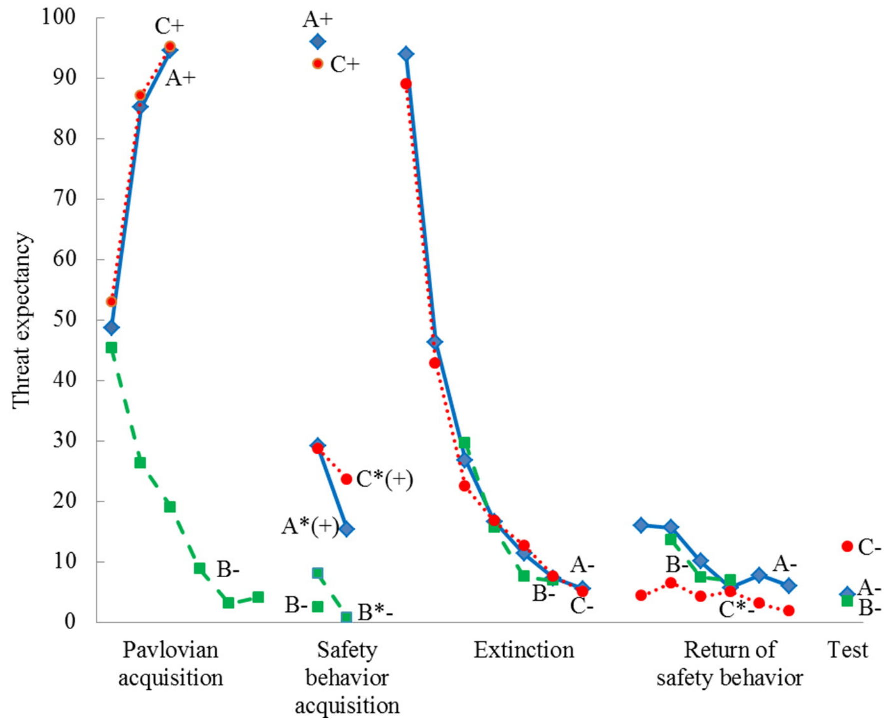



**FIGURE 1** Mean threat expectancy for A, B, and C. See Table 1 for explanation of trial types.

threat expectancy differed between A, B, and C, $F(1.41, 41.01) = 706.30$, $p < .001$, $\eta_p^2 = .96$. Threat expectancy for B was lower than for A, $t(29) = 28.02$, $p < .001$, $d = 8.47$, and for C, $t(29) = 28.62$, $p < .001$, $d = 8.25$. It was comparable for A and C, $t < 1$. The introduction of safety behavior in the safety behavior acquisition phase immediately decreased threat expectancy for A: A+ versus first A*(+), $t(29) = 8.85$, $p < .001$, $d = 2.20$, and for C: C+ versus first C*(+), $t(29) = 8.13$, $p < .001$, $d = 1.94$. In the extinction phase, threat expectancy for A and C declined in a linear trend, $F(1, 29) = 421.36$, $p < .001$, $\eta_p^2 = .93$, with a quadratic component, $F(1, 29) = 106.43$, $p < .001$, $\eta_p^2 = .79$. This trend did not significantly differ between A and C, $F < 1$. Threat expectancy for B initially increased in the extinction phase: safety behavior acquisition phase B– versus first extinction phase B–, $t(29) = 4.28$, $p < .001$, $d = 1.11$. Then it declined in a linear trend, $F(1, 29) = 15.19$, $p = .001$, $\eta_p^2 = .34$. At the end of the extinction phase, there were no significant differences in threat expectancy between A, B, and C, $F < 1$. At the beginning of the return of safety behavior phase, threat expectancy increased for A: final extinction phase A– versus first return of safety behavior phase A–, $t(29) = 2.55$, $p = .02$, $d = 0.60$, and for B: final extinction phase B– versus first return of safety behavior phase B–, $t(29) = 2.13$, $p = .04$, $d = 0.40$, but not for C: final extinction phase C– versus first return of safety behavior phase C*–, $t < 1$. Threat expectancy subsequently declined in a linear trend for A, $F(1, 29) = 9.00$, $p = .005$, $\eta_p^2 = .24$, but not significantly for B, $F(1, 29) = 2.16$, $p = .15$. However, at the end of the return of safety behavior phase there were no significant differences between A, B, and C, $F(2, 58) = 1.31$, $p = .28$. Finally, the change in threat expectancy from the final return of safety behavior trial to the test phase differed between A, B, and C, $F(2, 58) = 5.49$, $p = .007$, $\eta_p^2 = .16$. There were no main effects of trial and stimulus, both $F$s $< 1$. Threat expectancy did not change for A, $t < 1$, and B, $t(29) = 1.02$, $p = .31$, but it increased for C, $t(29) = 2.67$, $p = .01$, $d = 0.70$.

From the final extinction phase trial to the test phase, the change in threat expectancy differed between A, B, and C, $F(1.94, 43.34) = 4.68$, $p = .02$, $\eta_p^2 = .14$. There were no main effects of trial, $F < 1$, and stimulus, $F(1.57, 45.46) = 2.23$, $p = .13$. From the final extinction phase trial to the test phase, threat expectancy did not significantly change for A, $t < 1$, and decreased for B, $t(29) = 2.69$, $p = .01$, $d = 0.33$. Crucially, however, threat expectancy increased for C, $t(29) = 1.89$, $p = .03$, $d = 0.47$ (one-tailed), which was in line with our hypothesis.

### VALENCE AND PLEASANTNESS

Valence ratings after the experimental task differed between A, B, and C, $F(2, 58) = 59.92$, $p < .001$, $\eta_p^2 = .66$. Stimulus B ($M = 78.80$, $SD = 17.08$) was rated more positively than A ($M = 34.07$, $SD = 17.60$),

$t(29) = 8.40$, $p < .001$, $d = 2.54$, and C ($M = 37.80$, $SD = 16.32$), $t(29) = 8.43$, $p < .001$, $d = 2.46$. A and C did not significantly differ, $t < 1$, which suggests that safety behavior did not affect evaluative conditioning.[2] Participants rated the US as unpleasant ($M = 19.53$, $SD = 11.98$), and the safety behavior as pleasant ($M = 69.07$, $SD = 17.81$).

## Discussion

We investigated whether extinguished fears can return as a result of safety behavior. Our main findings are twofold. First, in line with Vervliet and Indekeu's (2015) findings, safety behavior did not diminish after the extinction of classically conditioned fear. Second, subsequent removal of safety behavior availability increased threat expectancy compared with the last extinction trial. Previous studies showed that safety behavior maintains preexisting threat expectancy (Engelhard et al., 2015; Lovibond et al., 2009). The current findings and those of Vervliet and Indekeu (2015) add that safety behavior after fear extinction causes a partial return of threat expectancy. There are several potential explanations for our findings.

First, Vervliet and Indekeu (2015) argued that residual threat expectancy may have motivated individuals to use safety behavior after fear extinction. In their study, threat expectancy and skin conductance responses remained higher for the CS+ than for the CS– at the end of the extinction phase. In the current study, participants who showed insufficient fear extinction were excluded. Although this does not eliminate the possibility that there was residual threat expectancy, at the end of the extinction phase threat expectancy was low (see Figure 1), and comparable for A, B, and C. We therefore consider it unlikely that residual threat expectancy explains the current findings.

Second, the *availability* of safety behavior may have signaled threat and activated alarm mechanisms (Sloan & Telch, 2002). People engage in safety behavior to avert a feared outcome, and therefore safety behavior is meaningfully linked to perceived threat (Salkovskis, Clark, & Gelder, 1996). In this study, participants learned to use safety behavior to prevent the expected US in the safety behavior acquisition phase, and thereby, safety behavior became meaningfully linked to the threat. Hence, the availability of safety behavior may have increased threat expectancy, and motivated safety behavior use after fear extinction.

Third, participants may have inferred threat from their *use* of safety behavior. Gangemi, Mancini, and van den Hout (2012), and van den Hout et al. (2014, 2017) found that individuals may infer threat from safety behavior, "I avoid, so there must be danger" (akin to "emotional reasoning"; see Engelhard & Arntz, 2005). Although "behavior as information" effects are mainly found in anxious individuals and not in healthy control participants, using safety behavior in an objectively safe situation may have induced cognitive dissonance. Participants may have reduced this dissonance by bringing their threat expectancy in line with their behavior (Festinger & Carlsmith, 1959).

Fourth, participants may have perceived different contexts within the experimental task based on CS–US contingencies and safety behavior availability and unavailability, which may explain the findings (see Bouton, 2002, 2016; Vervliet et al., 2013). In the Pavlovian acquisition phase and the safety behavior acquisition phase, stimuli A and C were always followed by the US, and when safety behavior was available it was effective in preventing the US. Participants may have perceived these two phases as a similar context (context A; signaling [potential] danger). Participants may have perceived the extinction phase as a different context (context B; signaling safety), because A and C were no longer followed by the US, and safety behavior was not available. The reintroduction of safety behavior availability for C in the return of safety behavior phase may have given participants the impression that they had returned to context A. This may explain why threat expectancy for A and B initially increased in the return of safety behavior phase. Our data do not show whether threat expectancy increased for C, because all participants used safety behavior on all C*– trials. However, this safety behavior use may have been motivated by increased threat expectancy for C. Threat expectancy for A and B reextinguished during the return of safety behavior phase. However, using safety behavior during C*– trials may have prevented this for C (i.e., protection from reextinction), which may explain the relatively high threat expectancy for C in the test phase. To recapitulate, reintroducing safety behavior availability after fear extinction may have triggered a return of threat expectancy, which was protected from reextinction for C by safety behavior use.

It remains unclear whether the return of threat expectancy was caused by the *availability* or *use* of safety behavior, or by both. To investigate the direct threat-signaling properties of safety behavior availability, future studies could assess whether safety

[2] Additionally, after the experimental task participants rated the safety of each CS on a 100 mm VAS from *dangerous* (left) to *safe* (right) with *uncertain* in the middle. The results for dangerous ratings showed a similar pattern to valence ratings: they differed between A, B, and C, $F(2, 58) = 85.33$, $p < .001$, $\eta_p^2 = .75$. B ($M = 89.67$, $SD = 12.43$) was rated safer than A ($M = 42.30$, $SD = 23.54$), $t(29) = 10.05$, $p < .001$, $d = 2.52$, and C ($M = 44.03$, $SD = 20.07$), $t(29) = 11.18$, $p < .001$, $d = 2.73$. A and C did not significantly differ, $t < 1$.

behavior availability (i.e., the green plug) alone (i.e., without simultaneous CS presentation) increases threat expectancy. If that is the case, then the inclusion of extinction trials in which the green plug is present (but the use of safety behavior is prevented) might extinguish the potential threat-signaling properties of safety behavior availability. However, Vervliet and Indekeu (2015) found that using a visual cue to indicate safety behavior (un)availability and verbally instructing participants about the (un)availability of the safety behavior response resulted in comparable returns of safety behavior use after fear extinction. This suggests that safety behavior's resistance to fear extinction cannot be attributed to the threat-signaling properties of a cue that indicated the availability of safety behavior (i.e., the green plug).

After the experimental task, A and C (the two CS+) were rated comparably negative, which indicates that safety behavior did not affect evaluative conditioning. The finding that stimulus B (the CS–) was still rated more positively than A and C after the experimental task suggests that evaluative conditioning was resistant to fear extinction, which is in line with previous findings (e.g., Engelhard, Leer, Lange, & Olatunji, 2014). Negative valence is related to behavioral avoidance tendencies (Rinck & Becker, 2007), and activates avoidance behavior (Krieglmeyer, Deutsch, De Houwer, & De Raedt, 2010). Future research should investigate whether negative valence motivates safety behavior use after fear extinction, and whether changing negative valence (e.g., via counterconditioning; Engelhard et al., 2014) can enhance the extinction of safety behavior and can prevent the return of threat expectancy.

Operant conditioning research indicates ways in which instrumentally conditioned behavior may be extinguished (Bouton, 2016). First, behavior decreases when it is associated with negative consequences. In a recent study by Rattel, Miedl, Blechert, and Wilhelm (2017), avoidance decreased when it was associated with cost for participants (i.e., they had to take a lengthy detour to avoid a potential shock). Avoidance and safety behaviors are often costly for patients with anxiety disorders. Among other things, these behaviors can make patients miss out on the positive consequences of approach behavior and impede achieving aspired goals. Future studies may investigate whether safety behavior in the current paradigm can be extinguished by introducing negative consequences for the behavior, for example, by presenting an aversive stimulus (e.g., mild electrical stimulation) when participants use safety behavior after fear extinction. In addition to increasing the negative consequences of safety behavior, rewarding approach behavior decreases avoidance (Bublatzky, Alpers, & Pittig, 2017; Pittig, Brand, Pawlikowski, & Alpers, 2014). Hence, introducing a reward for not using safety behavior after fear extinction may extinguish safety behavior in the current paradigm. Finally, safety behavior may extinguish when the perceived positive consequences of the safety behavior are removed. For example, in an instrumental extinction phase, participants may learn that safety behavior no longer prevents them from hearing the aversive loud noise, which may reduce their use of safety behavior. Investigating the extinction of instrumentally conditioned safety behavior can provide insights for the improvement of the long-term effects of exposure-based therapy.

Furthermore, future studies may include more extinction trials to increase extinction learning and fully diminish threat expectancy. This could also reduce the exclusion rate for insufficient extinction. Note that previous studies used similar exclusion procedures, which resulted in comparable exclusion rates. For instance, Leer et al. (2013) excluded 29 out of 109 participants (i.e., 27%) based on acquisition and extinction, and Dibbets, Poort, and Arntz (2012) excluded 22 out of 70 participants (i.e., 31%). Similarly, we excluded 14 out of 45 participants (i.e., 31%) based on acquisition and extinction criteria, and 1 additional participant who had misinterpreted the experimental task (i.e., a total of 33%). Finally, we cannot eliminate the possibility that participants were engaged in reasoning that was specific to the laboratory setting. Although the fear conditioning model has proven to be a robust and valid preclinical laboratory model for the return of fear (Milad & Quirk, 2012; Vervliet et al., 2013), research that investigates whether safety behavior is resistant to fear extinction and whether safety behavior is involved in the return of fear in a clinical setting is needed.

In sum, the current findings imply that safety behavior is resistant to extinction of classically conditioned fear, and can trigger a return of fear. This suggests that safety behavior may be involved in relapse. The current data were collected in healthy participants using a fear conditioning paradigm. Future studies should investigate whether safety behavior is resistant to fear extinction during exposure-based therapy in clinical samples, and whether safety behavior after fear extinction causes a return of clinical fear. Investigating how the extinction of safety behavior can be improved and the return of fear can be diminished might provide ways to enhance the long-term effects of exposure-based therapy.

## Conflict of Interest Statement

The authors declare there were no conflicts of interest.

## References


Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. (2011). *Exposure therapy for anxiety: Principles and practice*. New York, NY: Guilford Press.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, *35*, 205–230. http://dx.doi.org/10.1016/S0005-7894(04)80036-4

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

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*, 976–986. http://dx.doi.org/10.1016/j.bbr.2011.03.031

Bouton, M. E. (2016). *Learning and behavior: A contemporary synthesis* (2nd ed.). Sunderland, MA: Sinauer.

Bublatzky, F., Alpers, G. W., & Pittig, A. (2017). From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behaviour Research and Therapy*, *96*, 47–56. http://dx.doi.org/10.1016/j.brat.2017.01.003

Centerbar, D. B., & Clore, G. L. (2006). Do approach–avoidance actions create attitudes? *Psychological Science*, *17*, 22–29. http://dx.doi.org/10.1111/j.1467-9280.2005.01660.x

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. http://dx.doi.org/10.1016/j.brat.2014.04.006

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

Deacon, B., & Maack, D. J. (2008). The effects of safety behaviors on the fear of contamination: An experimental investigation. *Behaviour Research and Therapy*, *46*, 537–547. http://dx.doi.org/10.1016/j.brat.2008.01.010

Dibbets, P., Poort, H., & Arntz, A. (2012). Adding imagery rescripting during extinction leads to less ABA renewal. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 614–624. http://dx.doi.org/10.1016/j.jbtep.2011.08.006

Engelhard, I. M., & Arntz, A. (2005). The fallacy of ex-consequentia reasoning and the persistence of PTSD. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 35–42. http://dx.doi.org/10.1016/j.jbtep.2004.11.004

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

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. http://dx.doi.org/10.1016/j.beth.2014.12.006

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal Social Psychology*, *58*, 203–210. http://dx.doi.org/10.1037/h0041593.

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. http://dx.doi.org/10.1016/j.jbtep.2012.04.005

Grillon, C., Pine, D. S., Baas, J. M., Lawley, M., Ellis, V., & Charney, D. S. (2006). Cortisol and DHEA-S are associated with startle potentiation during aversive conditioning in humans. *Psychopharmacology*, *186*, 434–441. http://dx.doi.org/10.1007/s00213-005-0124-2

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*, 427–440. http://dx.doi.org/10.1007/s10608-012-9476-1

IBM Corp. (2016). IBM SPSS Statistics for Windows, version 24.0 [Computer software]. Armonk, NY: Author.

Keijsers, G. P. J., van Minnen, A., & Hoogduin, C. A. L. (2011). *Protocollaire behandelingen voor volwassenen met psychische klachten*. Amsterdam, Netherlands: Boom.

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved: Valence activates approach–avoidance behavior independently of evaluation and approach–avoidance intentions. *Psychological Science*, *21*, 607–613. http://dx.doi.org/10.1177/0956797610365131

Leer, A., & Engelhard, I. M. (2015). Countering fear renewal: Changes in the UCS representation generalize across contexts. *Behavior Therapy*, *46*, 272–282. http://dx.doi.org/10.1016/j.beth.2014.09.012

Leer, A., Engelhard, I. M., Dibbets, P., & van den Hout, M. A. (2013). Dual-tasking attenuates the return of fear after extinction. *Journal of Experimental Psychopathology*, *4*, 325–340. http://dx.doi.org/10.5127/jep.029412

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy*, *43*, 1391–1424. http://dx.doi.org/10.1016/j.brat.2004.10.007

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. http://dx.doi.org/10.1016/j.brat.2009.04.013

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

Olatunji, B. O., Etzel, E. N., Tomarken, A. J., Ciesielski, B. G., & Deacon, B. (2011). The effects of safety behaviors on health anxiety: An experimental investigation. *Behaviour Research and Therapy*, *49*, 719–728. http://dx.doi.org/10.1016/j.brat.2011.07.008

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

Python Software Foundation. (2012). Python Programming Language, version 2.7.3. https://www.python.org

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

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 105–120. http://dx.doi.org/10.1016/j.jbtep.2006.10.001

Salkovskis, P. M. (1991). The importance of behaviour in the maintenance of anxiety and panic: A cognitive account. *Behavioural Psychotherapy*, *19*, 6–19. http://dx.doi.org/10.1017/S0141347300011472

Salkovskis, P. M., Clark, D. M., & Gelder, M. G. (1996). Cognition–behaviour links in the persistence of panic. *Behaviour Research and Therapy*, *34*, 453–458. http://dx.doi.org/10.1016/0005-7967(95)00083-6

Sloan, T., & Telch, M. J. (2002). The effects of safety-seeking behavior and guided threat reappraisal on fear reduction during exposure: An experimental investigation. *Behaviour Research and Therapy*, *40*, 235–251. http://dx.doi.org/10.1016/S0005-7967(01)00007-9

Spielberger, C. D., Gorsuch, R. L., & Lushene (1970). *STAI manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.jbtep.2014.07.002

Van den Hout, M., Gangemi, A., Mancini, F., Engelhard, I. M., Rijkeboer, M. M., van Dam, M., & Klugkist, I. (2017). Erratum to "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]. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 307. http://dx.doi.org/10.1016/j.jbtep.2016.05.004

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

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

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

RECEIVED: May 4, 2017
ACCEPTED: August 9, 2017
AVAILABLE ONLINE: 18 August 2017