Do safety behaviors preserve threat expectancy?

Sophie L. van Uijen  
Utrecht University, The Netherlands

Edwin S. Dalmaijer  
University of Oxford, UK; University of Cambridge, UK

Marcel A. van den Hout  
Utrecht University, The Netherlands

Iris M. Engelhard  
Utrecht University, The Netherlands

Abstract
Empirical evidence for the detrimental effect of safety behaviors on fear extinction is inconsistent. This fear conditioning study investigated whether the negative effects of safety behavior on extinction learning depend on whether safety behavior precludes the occurrence of threat. In two experiments, participants first underwent fear acquisition. During a subsequent extinction procedure, participants used safety behavior that precluded the occurrence of threat, safety behavior that reduced threat, or no safety behavior. Safety behavior that precluded the occurrence of threat prevented extinction learning in the first and second experiments. Additionally, in the second experiment, safety behavior that reduced threat severity did not prevent extinction for several participants but did prevent it for others. The findings suggest that safety behavior that prevents the possibility of threat prevents extinction, whereas safety behavior that reduces threat does not prevent extinction consistently.

Keywords
Anxiety disorders, exposure, extinction, threat expectancy, safety behavior

Date received: 21 July 2017; accepted: 6 September 2018

Introduction
Clinical guidelines generally recommend to motivate patients to inhibit all safety behaviors during exposure-based therapy (e.g., Abramowitz, Deacon, & Whiteside, 2011; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Keijsers, van Minnen, & Hoogduin, 2011). Safety behaviors are actions aimed at preventing, minimizing, or escaping a feared outcome (e.g., Deacon & Maack, 2008; Salkovskis, 1991). They are common in individuals with anxiety disorders and are functionally related to the expected

Corresponding author:  
Sophie L. van Uijen, Department of Clinical Psychology, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands.  
Email: sophievanuijen@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
threat (Salkovskis, Clark, & Gelder, 1996). For example, a patient with social anxiety disorder may hide her face in social situations, because she fears that others will ridicule and reject her when they see her blush. Safety behaviors can preclude the occurrence of disconfirming experiences, because the patient may misattribute the nonoccurrence of the expected threat to the safety behavior (Salkovskis, 1991). This prevents the violation of negative expectancies (Blakey & Abramowitz, 2016; Craske et al., 2014) and thus prevents extinction. Hiding her face may prevent the patient with social anxiety from learning that others will not ridicule and reject her if they see her blush.

However, whether safety behavior is detrimental to the beneficial effects of exposure is an empirical issue, and overall, the findings from various studies are inconsistent. Several studies found unfavorable effects of exposure with safety behavior compared to exposure without safety behavior (e.g., McManus, Sacadura, & Clark, 2008; Salkovskis, Clark, Hackmann, Wells, & Gelder, 1999; Sloan & Telch, 2002). Other studies found that safety behavior did not reduce the effects of exposure (e.g., Deacon, Sy, Lickel, & Nelson, 2010; Hood, Antony, Koerner, & Monson, 2010; Milosevic & Radomsky, 2008) and even enhanced exposure effects (e.g., Milosevic & Radomsky, 2013; Sy, Dixon, Lickel, Nelson, & Deacon, 2011). In a recent review, Blakey and Abramowitz (2016) concluded that safety behaviors tend to be detrimental to the beneficial effects of exposure but do not always interfere with the therapeutic effects of exposure. A meta-analysis, however, did not find differences in effects between the incorporation and removal of safety behaviors during exposure (Meulders, Van Daele, Volders, & Vlaeyen, 2016).

The crucial question is how to explain this inconsistency. This is not only theoretically relevant but also clinically relevant, because it has obvious consequences for the psychological treatment of anxiety disorders. Rachman, Radomsky, and Shafran (2008) argued that not all safety behaviors necessarily prevent disconfirmatory experiences, and they called for a reconsideration of the categorical rejection of safety behavior during treatment. If the negative effects of safety behavior on extinction learning indeed depend on whether safety behaviors preclude disconfirmatory experiences, then safety behaviors that allow the occurrence of threat should not hamper extinction and may not be detrimental to the beneficial effects of exposure. Preliminary empirical support for this “interference with disconfirmation” hypothesis can be derived from two studies by Milosevic and Radomsky (2008, 2013), in which participants could use protective gear, such as gloves and goggles during exposure to a snake or spider. Exposure with and without the use of these safety behaviors resulted in comparable reductions of a fear of snakes (2008) and spiders (2013). One possible explanation is that this safety behavior did not prevent the corrective learning experience of not getting attacked by the snake or spider and, therefore, did not hinder fear extinction.

Additional empirical support for the interference with disconfirmation hypothesis comes from a study by Lovibond, Mitchell, Minard, Brady, and Menzies (2009). In a laboratory fear conditioning experiment, safety behavior that precluded the occurrence of threat prevented the extinction of subjective threat expectancy and skin conductance responses, which are measures of fear in fear conditioning research (Boddez, Baeyens, Luyten, Vansteenwegen, & Hermans, 2013). In a Pavlovian acquisition phase, participants learned that two neutral stimuli (A and C, which both served as conditional stimuli (CS⁺) were followed by shock (unconditional stimulus (US)), and a third neutral stimulus (B, which served as CS⁻) was not. Next, participants learned to use safety behavior during presentation of stimulus A by pressing a button on a response box that effectively cancelled the shock. During a subsequent extinction phase, stimulus C was no longer followed by shock. Participants in the experimental condition, but not those in the control condition, were given the opportunity to use safety behavior during C trials. All participants in the experimental condition used safety behavior on all C trials. In a following test phase, in which safety behavior was no longer available for any stimulus, threat expectancy and skin conductance responses for C remained high in the experimental condition, whereas they had decreased in the control condition (Lovibond, Mitchell, Minard, Brady, & Menzies, 2009). Presumably, participants in the experimental condition misattributed the nonoccurrence of the shock to the safety behavior, which prevented extinction. This suggests that safety behaviors that preclude the occurrence of threat prevent fear extinction. However, neither the study by Lovibond et al. (2009) nor the studies by Milosevic and Radomsky (2008, 2013) are direct, experimental tests of the crucial remaining issue: Does safety behavior that allows the occurrence of threat prevent fear extinction? Answering this question was the aim of the present research.
Experiment 1

In Experiment 1, we aimed to replicate Lovibond et al.’s (2009) finding that safety behavior that precludes the occurrence of threat prevents fear extinction in an adapted version of their fear conditioning paradigm. A first change to the paradigm was that instead of a shock, the US was an aversive loud noise presented through headphones that were connected to the computer with a sound amplifier. This allowed us to operationalize safety behaviors that did and did not prevent the occurrence of threat. The fear conditioning task started with the Pavlovian acquisition phase, during which participants learned the CS–US associations. Stimulus A and C (CS+) were followed by the loud noise, while stimulus B (CS-) was not. Next, during the safety behavior acquisition phase, all participants learned to make a full avoidance response, which was analogous to the safety behavior in the study by Lovibond et al. (2009). However, a second change to the paradigm was that this response was not pressing a button on a response box but unplugging the headphones from the sound amplifier, which prevented participants from hearing the loud noise. The full avoidance response thus precluded the (potential) occurrence of threat. A third change was that all participants also learned to use subtle safety behavior, which entailed taking off the headphones. This did not prevent participants from hearing the loud noise but decreased how loud, and thus how unpleasant, it was for participants. Hence, the subtle safety behavior reduced threat severity but did not preclude the occurrence of threat. In a subsequent extinction phase, the full avoidance response was available during unreinforced presentations of stimulus C for participants in the full avoidance condition but not for participants in the control condition (analogous to Lovibond et al.’s, 2009, experimental condition and control condition, respectively). Note that the effect of subtle safety behavior during unreinforced presentations of stimulus C was investigated in Experiment 2.

A fourth and final change was that we did not measure skin conductance responses, because the full avoidance response and subtle safety behavior required participants to move both hands during CS presentations. Movement causes artifacts in skin conductance responses (Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, 2012). Movement also affects another commonly used psychophysiological measure in fear conditioning studies, namely the fear potentiated startle responses (Blumenthal et al., 2005). Additionally, the auditory probes that are used to induce the startle responses may interfere with the acquisition of contingencies (Lonsdorf et al., 2017). An alternative psychophysiological measure of fear learning is pupillary dilation responses (Leuchs, Schneider, Czisch, & Spoormaker, 2017). Pupil dilations in response to CS+ are a measure of autonomic arousal and covary with skin conductance responses (Bradley, Miccoli, Escrig, & Lang, 2008). They can be used as a measure of associative learning in fear conditioning research (e.g., Visser, Scholte, Beemsterboer, & Kindt, 2013). Hence, instead of skin conductance responses, we included pupil dilation responses as a psychophysiological measure of fear learning in Experiment 1.

We had intended to include a third condition in Experiment 1, in which participants could use subtle safety behavior during the extinction phase: the subtle safety behavior condition. However, we tested several participants in this condition, and they reported different interpretations of the procedure and uncertainty about the experimental task at the debriefing. The cause of these differences was unclear. We, therefore, did not include this condition in Experiment 1, and first investigated whether we could replicate the findings by Lovibond et al. (2009) in the current paradigm. Note that, in Experiment 1, subtle safety behavior was available for all participants during the safety behavior acquisition phase, but none of the participants could use this response during the extinction phase.

In Experiment 1, we investigated the effect of full avoidance versus no avoidance or safety behavior on the extinction of threat expectancy ratings and on a psychophysiological measure of fear learning (cf., Lovibond et al., 2009). We hypothesized that making the full avoidance response during unreinforced CS presentations would prevent extinction. In line with the findings by Lovibond et al. (2009), we hypothesized that threat expectancy and pupil dilation responses for stimulus C in the test phase that followed the extinction phase would be higher, and larger, in the full avoidance condition than in the control condition.

Method

Participants

Participants were 64 student volunteers ($M_{age} = 22.63, SD = 5.12, 41$ women, $23$ men). They gave
written informed consent and received money 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. Participants were randomly allocated to the full avoidance or control condition.

Apparatus, stimuli, and measures

State-Trait Anxiety Inventory. The State-Trait Anxiety Inventory (STAI; Spielberger, Gorusch, & Lushene, 1970) was included to measure state and trait anxiety, because they may affect fear learning (Grillon et al., 2006; Lissek et al., 2005). Each scale contained 20 items, rated on a scale from 1 = not at all to 4 = severely. In this study, Cronbach’s α = .87 and .86 for state and trait anxiety, respectively.

Experimental task. The experimental task was programmed in Python (Python Software Foundation) using the PyGaze toolbox (Dalmaijer, Mathôt, & van der Stigchel, 2014) and presented on a 21-in ViewSonic P227f CRT monitor (1024 × 768 pixel, 100 Hz) at a distance of 67 cm from the participant. The US was 1 s of 100 dB white noise presented through headphones (cf., Leer & Engelhard, 2015) that were connected to the computer with a sound amplifier. CS were a black (0.24 cd/m²) circle, square, and triangle presented on a light gray (6.41 cd/m²) background. CS were equiluminant and of the same surface area.

Threat expectancy. Immediately after each CS presentation, but before (possible) presentation of the aversive loud noise (US), participants rated to what extent they expected the noise to follow using a 0 = certain no noise to 100 = certain noise visual analog scale (VAS) shown on the computer screen.

Pupil dilations. Pupil size of the left eye was measured during CS presentation with an Eyelink 1000 (SR Research, 2010), a video-based eye tracker sampling at 1000 Hz. Pupil dilation responses to the CS were calculated as the relative growth (proportion) during stimulus presentation from 0 to 5 s after stimulus onset compared to baseline (200 ms before stimulus onset), and the maximum proportional dilation in this window was used for further analysis (cf., Visser et al., 2013).

Contingencies and pleasantness. After the experimental task, participants indicated whether each CS (i.e., the square, triangle, and circle) was followed by the loud noise on a 3-point Likert-type scale with the categories never, sometimes, and always as an additional check of contingency awareness. As a manipulation check, participants rated the (un)pleasantness of the loud noise when they had the headphones on and when they had the headphones off on two 100-mm VAS from extremely unpleasant (left, i.e., 0) to extremely pleasant (right, i.e., 100). Finally, they were asked to describe what happened when they unplugged the headphones and when they took off the headphones to assess whether they had understood the consequences of the full avoidance response and subtle safety behavior.

Procedure

Participants were tested individually in a darkened room with dimmed lights. After the informed consent procedure, participants filled out the STAI. They then 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. They were also instructed and demonstrated how to unplug the headphones (full avoidance) and plug them back into the sound amplifier, and how to take off the headphones (subtle safety behavior) and put them back on. Participants were not told what the consequences of these responses were. Next, they were seated with their head in a chin and forehead rest to restrict head movements. A 9-point calibration procedure was performed for the eye tracker. Participants practiced rating threat expectancy twice, making the full avoidance response twice, and using subtle safety behavior twice. The experimental task then started.

Each trial consisted of the presentation of a CS for 5 s in the middle of the screen, followed by a 5-s waiting period during which participants rated threat expectancy, followed by the US or nothing. The intertrial interval was 3 s. During each CS presentation, a dark gray picture of a plug was visible in the upper right corner of the screen, and a dark gray picture of headphones was visible in the upper left corner of the screen. These pictures indicated availability and unavailability of the full avoidance response and subtle safety behavior. If the picture of the plug turned green, participants could unplug the headphones from the sound amplifier (full avoidance), and if the headphones turned green, participants could take off the headphones (subtle safety behavior). If the full
avoidance response or subtle safety behavior had been available, an instruction screen would appear at the end of the trial to inform participants to plug the headphones back into the sound amplifier or to put the headphones back on.

The design of the study is shown in Table 1. A, B, and C were randomly allocated to the three different shapes for each participant. In the Pavlovian acquisition phase, A and C were followed by the loud noise, while B was not. In the safety behavior acquisition phase, full avoidance and subtle safety behavior were available during presentation of A. A was also presented without full avoidance or subtle safety behavior availability to remind participants that A was still followed by the loud noise if the full avoidance response was not available (cf., Engelhard, van Uijen, van Seters, & Velu, 2015; Lovibond et al., 2009). From the extinction phase onward, C was no longer followed by the loud noise. During the extinction phase, the full avoidance response was available during unreinforced C presentations for participants in the full avoidance condition but not for participants in the control condition. Finally, in the test phase, A, B, and C were presented once without the availability of full avoidance or subtle safety behavior. 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, C was always presented last in the test phase to prevent that the nonoccurrence of the US at this trial would affect the response to stimuli A and B (cf., Lovibond et al., 2009). After finishing the experimental task, participants filled out the questionnaire about contingencies and pleasantness. Next, they were debriefed and given their reward.

**Scoring and analysis**

Pupil data were preprocessed by interpolating blinks via Mathôt’s (2013) method and then dividing the pupil signal from 0 to 5,000 ms after CS onset by the median pupil size during a baseline period of 200 ms before CS onset. The maximum value of the pupil trace in this 0–5,000 ms period was used for further analysis. Fear acquisition had to take place for drawing conclusions about fear extinction (see Londt et al., 2017). Therefore, participants who did not show contingency awareness were excluded from the analyses (Lovibond & Shanks, 2002). Contingency awareness was defined as a higher threat expectancy for A and C than for B on the final trial of the Pavlovian acquisition phase and correctly indicating which CS were followed by the loud noise on the contingencies questionnaire after the experimental task. Analyses were performed on data with and without replaced outliers. Outliers were defined as more than 3 SD from the mean and were replaced with $M + 3 SD$. Replacing outliers did not affect the direction of the main findings. We, therefore, reported the analyses that were performed on the original data. The data were analyzed with mixed analyses of variance ($\alpha = .05$), comparing threat expectancy and pupil dilation responses between stimuli (A vs. B vs. C) and conditions (full avoidance vs. control). Corrected values were reported in case the assumption of sphericity was violated. Planned comparisons were conducted with paired and independent $t$-tests.

**Table 1. Design of the experimental task in Experiment 1.**

| Phase          | Pavlovian acquisition | Safety behavior acquisition | Extinction | Test  |
|----------------|-----------------------|-----------------------------|------------|-------|
| Full avoidance condition | A+ (3) | A*+ (4) | A+ (2) |       |
| B-/C0 (6)      | B+ (1) | A* (1)  | B- (2) |       |
| C+ (3)         | C- (1) | B- (2)  | A+ (1) |       |
| Control condition | A+ (2) | B-/C0 (2) | C- (1) |       |

Note. US = unconditional stimulus. A, B, and C refer to visual stimuli (CS); + and − refer to presence and absence, respectively, of the loud noise (US) following the CS; * indicates the availability of the full avoidance response (i.e., unplugging the headphones); − indicates the availability of subtle safety behavior (i.e., taking off the headphones); (+) indicates that participants only heard the US if they made the full avoidance response (i.e., if they did not unplug the headphones); [+] indicates that the US was less loud for participants if they used subtle safety behavior (i.e., if they took off the headphones); numbers in parentheses indicate the number of trials of each type.
Additionally, safety behavior learning occurred. In the safety behavior acquisition phase, threat expectancy decreased in a linear trend during full avoidance trials (see the A*(+)- trials in Figure 1), \( F(1, 55) = 8.57, p = .005, \eta^2_p = .14 \), which was similar for both conditions, \( F < 1 \). Participants thus learned that the full avoidance response canceled the loud noise. Additionally, threat expectancy increased in a quadratic trend during subtle safety behavior trials (see the A ~[+] trials), \( F(1, 55) = 19.80, p < .001, \eta^2_p = .27 \), which did not differ between conditions, \( F(1, 55) = 1.82, p = .18, \eta^2_p = .03 \). This indicates that participants learned that the subtle safety behavior response did not prevent them from hearing the loud noise. Finally, in the extinction phase, threat expectancy decreased in a quadratic trend over the six C trials, \( F(1, 55) = 22.57, p < .001, \eta^2_p = .29 \). This trend showed a Stimulus \( \times \) Condition interaction effect, because this decrease was steeper in the control condition (see the C− trials in the bottom panel of Figure 1) than in the full avoidance condition (C+− trials in the upper panel) due to the lower starting point in the full avoidance condition (see Figure 1; the first extinction phase C trial is lower in the top panel than in the bottom panel), \( F(1, 55) = 7.09, p = .01, \eta^2_p = .11 \).

In the test phase, threat expectancy showed a main effect of stimulus, \( F(2, 110) = 624.97, p < .001, \eta^2_p = .92 \); condition, \( F(1, 55) = 168.39, p < .001, \eta^2_p = .75 \); and, crucially, a Stimulus \( \times \) Condition interaction, \( F(2, 110) = 159.38, p < .001, \eta^2_p = .74 \). Threat expectancy was higher for A than for B, \( F(1, 55) = 1424.60, p < .001, \eta^2_p = .96 \). This did not differ between conditions, \( F(1, 55) = 1.32, p = .26, \eta^2_p = .02 \), or show a Stimulus \( \times \) Condition interaction, \( F(1, 55) = 1.50, p = .23, \eta^2_p = .03 \). In line with our hypothesis, threat expectancy for C was higher in the full avoidance condition than in the control condition (see Figure 1; C− in the test phase is higher in the top panel than in the bottom panel), \( t(28.90) = 18.30, p < .001, d = 4.77 \). To further decompose the Stimulus \( \times \) Condition interaction, we examined threat expectancy for C compared with the danger stimulus (A) and safety stimulus (B) in the test phase for each condition. In line with our hypothesis, in the control condition, threat expectancy was lower for C than for A, \( t(28) = 13.21, p < .001, d = 3.26 \), and higher for C than for B, \( t(28) = 2.77, p = .01, d = 0.73 \) (see Figure 1, C is lower than A and higher than B in the bottom panel). This indicates that extinction for C occurred in the control condition. In the full avoidance condition,
threat expectancy for C was also higher than for B, \( t(27) = 97.82, p < .001, d = 35.72 \), but it did not differ between A and C, \( t(27) = 1.30, p = .20, d = 0.22 \) (see Figure 1, C is similar to A and higher than B in the top panel). Thus, threat expectancy for C was maintained in the full avoidance group.

**Pupil dilations**

As presented in Figure 2, the conditioning effect observed on threat expectancy was not reflected in pupil dilations. At the end of the Pavlovian acquisition phase, pupil dilation responses did not differ between A, B, and C \( (F < 1) \) between conditions, \( F(1, 54) = 2.97, p = .09, \eta^2_p = .05 \), or show a Stimulus \( \times \) Condition interaction, \( F(2, 108) = 1.70, p = .19, \eta^2_p = .03 \). However, pupil dilation responses increased on full avoidance and subtle safety behavior trials. In the safety behavior acquisition phase, pupil dilation responses were larger on full avoidance trials (see the A* (+) trials in Figure 2), \( F(1, 54) = 14.67, p < .001, \eta^2_p = .21 \), and subtle safety behavior trials (see the A ~ [+]) trials), \( F(1, 55) = 17.69, p < .001, \eta^2_p = .24 \), compared to no avoidance or safety behavior trials (the A+ trial). Additionally, in the extinction phase, pupil dilation responses for C were larger in the full avoidance condition (C*+ trials) than in the control condition (C− trials), \( F(1, 55) = 13.94, p < .001, \eta^2_p = .20 \).

**Discussion: Experiment 1**

We replicated Lovibond et al. (2009)’s findings for threat expectancy. Presumably, participants in the full avoidance condition misattributed the nonoccurrence of the loud noise during C trials in the extinction phase to the full avoidance response. Hence, threat expectancy for C was maintained in the full avoidance condition, whereas it had decreased in the control condition. In line with our hypothesis and the findings by Lovibond et al. (2009), Experiment 1 showed that safety behavior that prevents the occurrence of threat maintains threat beliefs when safety behavior is no longer available and may, therefore, be detrimental to the beneficial effects of exposure.

Contrary to our hypothesis, pupil dilation responses did not show associative learning effects. Pupil dilations discriminate most strongly between the CS+ and CS− shortly before US onset (Reinhard & Lachnit, 2002), because the peak pupil response occurs in a time window immediately preceding the US (Visser, Kunze, Westhoff, Scholte, & Kindt, 2015). However, the US did not coincide with CS+ offset in the current paradigm, because participants rated threat expectancy for 5 s between CS offset and US onset. Simultaneously measuring pupil dilations during this time window may cause the pupillary responses to be affected by arousal associated with filling out the VAS (Sirois & Brisson, 2014). Additionally, the availability or use of the full avoidance response and subtle safety behavior increased pupil dilation responses. Together, this indicates that pupil dilation responses were not a suitable psychophysiological measure of fear learning in the current paradigm. Pupil dilations were, therefore, not measured in Experiment 2.

In Experiment 1, we did not investigate whether subtle safety behavior that reduces threat severity prevents extinction learning. Participants in the subtle safety behavior condition differed in their interpretation of the experimental task and reported uncertainty about the task at the debriefing. This may be due to the several limitations of Experiment 1. First, in Experiment 1, the instructions in the subtle safety behavior condition may not have been entirely clear to participants. Second, although all participants in Experiment 1 correctly described the consequences of the full avoidance responses and subtle safety behavior after the experimental task, they may not have understood these consequences correctly at the beginning of the experimental task. Third, the availability of the full avoidance response and subtle safety behavior may have functioned as negative occasion setters. An occasion setter is a cue that provides information about whether a CS will be followed by a US. A negative occasion setter inhibits the association...
between a CS and US (Bouton, 2016). At the debriefing, several participants in the subtle safety behavior condition reported that they had reasoned that C was not followed by the loud noise as long as it was combined with the availability of the subtle safety behavior response. Hence, the picture of the green headphones may have functioned as a negative occasion setter.

**Experiment 2**

Experiment 2 entailed the same basic design as Experiment 1 with several improvements to investigate whether safety behavior that allows the occurrence of threat prevents extinction. First, we simplified and clarified the oral and written instructions based on pilot studies. Second, after receiving the oral instructions from the experimenter, participants were asked to explain the experimental task in their own words to check whether they had understood the instructions correctly. Third, participants were told that the full avoidance response prevented a possible subsequent loud noise and that the subtle safety behavior reduced the volume of a possible subsequent loud noise. Fourth, the availability of the full avoidance response and subtle safety behavior was not indicated by a picture of a plug and a picture of headphones that were presented simultaneously with the CS, to prevent that these pictures might function as negative occasion setters. Instead, during full avoidance and subtle safety behavior trials, an instruction screen preceded CS presentation to inform participants that they could unplug or take off the headphones, respectively. Fifth, to further prevent that the full avoidance response and subtle safety behavior may function as negative occasion setters, the full avoidance response and subtle safety behavior were learned during two A and two C trials in the safety behavior acquisition phase, instead of during four A trials in Experiment 1 (see Table 2). Hence, in Experiment 2, stimulus C was not exclusively combined with the full avoidance response and subtle safety behavior on unreinforced presentations but also on reinforced trials.

We hypothesized that we would replicate the findings of Experiment 1 and Lovibond et al. (2009), as evidenced by higher threat expectancy for C in the test phase in the full avoidance condition than in the subtle safety behavior and control condition. Furthermore, we hypothesized that using subtle safety behavior during unreinforced C trials would not prevent extinction learning for C. To be precise, we hypothesized that threat expectancy in the test phase would be lower for C than for A in the subtle safety behavior and control condition.

**Method**

**Participants**

The sample consisted of 64 student volunteers (\(M_{age} = 22.64, SD = 2.89, 44\) women, 20 men). They were randomly allocated to the full avoidance (\(n = 21\)},
subtle safety behavior \((n = 22)\), or control condition \((n = 21)\).

**Apparatus, stimuli, measures, and procedure**

The apparatus, stimuli, and measures were similar to Experiment 1. The design of Experiment 2 is presented in Table 2. The procedure was similar to Experiment 1, except for the changes described under Experiment 2. Participants in the subtle safety behavior condition could use subtle safety behavior (i.e., could take off the headphones) during unreinforced C trials in the extinction phase. Because we did not measure pupil dilations, the eye tracker and chin and forehead rest were not used, and participants were tested in a regularly lighted room. In this study, Cronbach’s \(\alpha\) was .89 for state and trait anxiety.

**Results**

**Participants**

Five participants did not show contingency awareness using the predefined criteria (see Experiment 1). Analyses were performed on the data of the remaining 59 participants \((M_{\text{age}} = 22.56, SD = 2.96, 42\) women, 17 men, full avoidance condition \(n = 19\), subtle safety behavior condition \(n = 21\), control condition \(n = 19\)). There were no significant differences between conditions in age, \(F(2, 56) = 1.53, p = .23, \eta_p^2 = .05\), and trait anxiety scores, \(F(2, 56) = 1.49, p = .23, \eta_p^2 = .05\). There was a trend for state anxiety scores to differ between conditions, \(F(2, 56) = 3.10, p = .053, \eta_p^2 = .10\). State anxiety scores were higher in the control condition than in the subtle safety behavior condition, \(t(38) = 2.38, p = .02, d = 0.75\), but did not significantly differ between the full avoidance and subtle safety behavior condition, \(t(38) = 1.68, p = .10, d = 0.53\), and between the full avoidance and control condition \((t < 1)\). Results were similar when analyses were performed with state anxiety scores included as a covariate, which suggests that the differences between conditions in state anxiety did not affect the results. We, therefore, reported the analyses without state anxiety as a covariate. Participants rated the loud noise as unpleasant when they had the headphones on \((M = 12.51, SD = 12.16)\) and as neutral when they had the headphones off \((M = 53.56, SD = 16.33)\).

**Threat expectancy**

As presented in Figure 3, Pavlovian acquisition occurred for all three conditions. At the end of the Pavlovian acquisition phase, threat expectancy was higher for A and C than for B, \(F(1.44, 80.56) = 2718.18, p < .001, \eta_p^2 = .98\). There was no main effect of condition or a Stimulus × Condition interaction, both \(F$s < 1. Safety behavior learning also occurred. In the safety behavior acquisition phase, threat expectancy for A and C was lower on full avoidance trials (see the A*(+) and C*(+) trials in Figure 3) than on no avoidance or safety behavior trials (A+ and C+ trials in the safety behavior acquisition phase), \(F(1, 58) = 508.45, p < .001, \eta_p^2 = .90\), and subtle safety behavior trials (A ~[+] and C ~[+] trials), \(F(1, 58) = 424.31, p < .001, \eta_p^2 = .88\). In the extinction phase, threat expectancy decreased over
the six C trials in all conditions, which was indicated by a quadratic trend for stimulus, $F(1, 56) = 53.55, p < .001, \eta^2_p = .49$. This decrease was steeper in the control condition (see the C− trials in the bottom panel of Figure 3) and subtle safety behavior condition (C~− trials in the middle panel) than in the full avoidance condition (C+− trials in the top panel), $F(2, 56) = 8.98, p = .01, \eta^2_p = .24$, which was due to the lower starting point in the full avoidance condition (see Figure 3; the first extinction phase C trial is lower in the top panel than in the middle and bottom panel), $F(2, 56) = 139.34, p < .001, \eta^2_p = .83$.

In the test phase, threat expectancy showed a main effect of stimulus, $F(1.12, 62.77) = 437.53, p < .001, \eta^2_p = .68$, and a Stimulus × Condition interaction, $F(2, 112) = 47.20, p < .001, \eta^2_p = .63$. Threat expectancy was higher for A than for B, $F(1, 56) = 7603.51, p < .001, \eta^2_p = .99$. This did not differ between conditions or show a Stimulus × Condition interaction, both $F$s < 1. Thus, the experimental manipulation of full avoidance, subtle safety behavior, or no avoidance or safety behavior during the extinction phase had not differentially influenced threat expectancy ratings for danger stimulus A and safety stimulus B. The experimental manipulation had, however, caused differences between conditions in threat expectancy ratings for C in the test phase. In line with our hypothesis, threat expectancy for C was higher in the full avoidance condition than in the control condition, $t(20.53) = 66.53, p < .001, d = 21.59$, and in the subtle safety behavior condition, $t(20.05) = 5.36, p < .001, d = 1.61$ (see Figure 3, C− in the test phase is higher in the top panel than in the middle and bottom panel). This indicates that full avoidance maintained threat expectancy for C. Additionally, threat expectancy for C was higher in the subtle safety behavior condition than in the control condition, $t(20.74) = 3.94, p = .001, d = 1.19$ (see Figure 3, C− in the test phase is higher in the middle panel than in the bottom panel). This suggests that not avoiding or using subtle safety behavior during unreinforced C trials (i.e., during the extinction phase) resulted in a larger reduction of threat expectancy than when subtle safety behavior was used during unreinforced C trials.

Next, we examined threat expectancy for C compared with A and B in the test phase for the subtle safety behavior and control condition to investigate whether using subtle safety behavior during unreinforced C trials had prevented extinction learning. In line with our hypothesis, threat expectancy was higher for A than for C in the control condition, $t(18) = 44.07, p < .001, d = 16.72$, and subtle safety behavior condition, $t(20) = 5.09, p < .001, d = 1.62$ (see Figure 3, C− is lower than A− in the test phase in the middle and bottom panel). This indicates that extinction learning occurred in the subtle safety behavior and control condition. However, in the control condition, threat expectancy did not significantly differ between B and C, $t < 1$, whereas threat expectancy was higher for C than for B in the subtle safety behavior condition, $t(20) = 3.73, p = .001, d = 1.19$ (see Figure 3, C− is similar to B− in the bottom panel, but C− is higher than B− in the middle panel). This again indicates that, on average, the control condition resulted in a larger reduction of threat expectancy for stimulus C than the subtle safety behavior condition.

A closer look at the data, however, showed that threat expectancy ratings for C varied substantially between participants within the subtle safety behavior condition (SD = 47.53) and showed a dichotomous distribution (see Figure 4). This was not the case in the full avoidance (SD = 1.61) and control condition (SD = 6.14). In the subtle safety behavior condition, threat expectancy for C had decreased for approximately half of the participants (n = 11), whereas for approximately the other half of the participants, it had persisted (n = 9) or had become ambiguous (n = 1). State and trait anxiety scores did not differ between participants within the subtle safety behavior condition whose threat expectancy had, and had not, decreased, $ts < 1$.

**General discussion**

The aim of the current study was to investigate whether the negative effects of safety behavior on fear extinction depend on whether safety behavior precludes the occurrence of threat. In Experiments 1 and
2, we replicated Lovibond et al.’s (2009) finding that safety behavior that precludes the occurrence of threat prevents fear extinction. In line with our hypothesis, threat expectancy ratings for C in the test phase were higher in the full avoidance condition than in the subtle safety behavior and control condition. This is presented in Figure 1 for Experiment 1 and in Figure 3 for Experiment 2 by the red dots (C−) in the test phase that are higher in the full avoidance condition (upper panel of Figures 1 and 3) than in the subtle safety behavior (middle panel of Figure 3) and control condition (bottom panel of Figures 1 and 3). Furthermore, in Experiment 2, it seemed that safety behavior that reduced threat, but did not preclude its occurrence (subtle safety behavior), did not completely prevent extinction. Threat expectancy was lower for C than for the danger stimulus A in the subtle safety behavior and control condition, which was in line with our hypothesis. This is depicted in Figure 3 by the red dot (C−) that is lower than the blue dot (A+) in the middle panel for the subtle safety behavior condition and in the bottom panel for the control condition. However, a closer look at the data showed that extinction occurred for approximately half of the participants in the subtle safety behavior condition but did not occur for the other half. At the debriefing, participants whose threat expectancy had decreased described that they had learned that C was no longer followed by the loud noise, irrespective of whether they could use subtle safety behavior. Participants whose threat expectancy had not decreased explained that they had learned that C was followed by the loud noise, except on trials when they could use subtle safety behavior.

The current findings replicate the results of Lovibond et al. (2009). Together they indicate that, generally, safety behaviors aimed at preventing the occurrence of threat prevent fear extinction and may thus be detrimental to the beneficial effects of exposure. Safety behavior aimed at minimizing the severity of threat (“subtle safety behavior”), however, may allow extinction learning. Nevertheless, extinction learning did not occur for several participants who used subtle safety behavior during an extinction procedure in Experiment 2. There are various explanations for this. Two parsimonious explanations come from the inhibitory learning model of fear extinction (Craske et al., 2014). The inhibitory learning model posits that fear extinction involves new learning of an inhibitory association between the CS and the US (CS–no US), which then exists alongside the excitatory CS–US association (Bouton, 2002, 2004, 2016; Craske et al., 2014). A first potential explanation is that the availability of subtle safety behavior functioned as a negative occasion setter (Bouton, 2016) and prevented inhibitory learning, despite the changes we had made to the experimental task in Experiment 2 to prevent this. Participants may have perceived the instruction screen that preceded the CS on full avoidance and subtle safety behavior trials or taking off the headphones itself as a cue that inhibited the CS–US association, preventing them from learning the inhibitory CS–US association.

A second potential explanation is that participants who did not show extinction in the test phase may have perceived different contexts in the experimental task, which caused contextual renewal of threat expectancy for C in the test phase. Extinction learning is context dependent (Bouton, 2004), which means that the inhibitory association is dominant in the context in which extinction learning occurred. A change in the external context or in a person’s internal state after extinction can cause a return of fear for the CS (contextual renewal; Bouton, 2002, 2004, 2016; Vervliet, Craske, & Hermans, 2013). Extinction learning occurred in the “headphones off” context for participants in the subtle safety behavior condition, whereas threat expectancy in the test phase was assessed in the “headphones on” context. The test phase context, which was similar to the acquisition context, was related to the excitatory CS–US association and may have led to the return of threat expectancy for C. In this case, subtle safety behavior did not prevent extinction learning, but extinction learning was specific to the “headphones off” context.

It is unclear what caused the between-subjects differences in extinction learning within the subtle safety behavior condition. The data were obtained from a group of participants that was small (n = 21 for the subtle safety behavior condition) and specific (i.e., undergraduate students who were approximately 23 years old), and yet they showed maximal variation in threat expectancy ratings for C in the test phase. This variation was not caused by between-subjects differences within the subtle safety behavior condition in state or trait anxiety. It may be caused by an individual difference factor that was unmeasured in the current study and that impacts fear conditioning processes (see Lonsdorf et al., 2017, for a list of individual difference factors). Another possibility is that participants differed in beliefs they held regarding the conditioning task (e.g., about their ability to cope with the loud noise) that may have influenced their response to the experimental task. However, the between-subjects
differences within the subtle safety behavior condition were an unexpected finding and should be interpreted with caution. Replication of the current findings in a larger and more heterogeneous sample and on additional outcome measures is warranted.

Extending these findings to additional, clinically relevant outcome measures is needed to ensure that the current results generalize to fear and anxiety. In the current study, participants were specifically instructed to rate their expectancy of the loud noise (i.e., the US), because threat expectancy (or US expectancy) is a measure of fear in fear conditioning research (Boddez et al., 2013). Unexpectedly, however, participants gave high threat expectancy ratings on subtle safety behavior trials in the safety behavior acquisition phase (i.e., trials in which they took the headphones off: A~[+] and C~[+] trials), which suggests that they may have rated their expectancy of noise in general. Taking off the headphones (subtle safety behavior) decreased the volume of the loud noise to such an extent that it was considered a neutral stimulus. This quieter noise may not be considered a US, and it is, therefore, unclear whether subtle safety behavior reduced or removed threat.

Finally, future research should investigate the role of individual differences in the negative effects of safety behavior on extinction learning. Safety behaviors that reduce threat severity, but do not prevent the occurrence of threat, may allow extinction learning for some, but not all, individuals. In a recent fear conditioning study, patients with anxiety disorders more often showed impaired extinction learning than healthy control participants (Duits et al., 2017). Furthermore, impaired extinction learning predicted poorer treatment outcomes. Identifying individual characteristics that predict the negative effects of safety behaviors on fear extinction can provide insights for the development of personalized treatment and may improve treatment outcomes.

In conclusion, safety behavior that precludes the occurrence of threat prevents extinction learning and may, therefore, be detrimental to the beneficial effects of exposure. Furthermore, in Experiment 2, we found that safety behavior that reduces threat, but does not prevent its occurrence, may allow extinction learning, at least for some individuals. This finding supports the proposition for the reconsideration of the categorical rejection of safety behavior during treatment made by Rachman et al. (2008). However, for several participants, safety behavior that reduced threat severity prevented extinction learning. The negative effects of safety behavior on extinction learning may not only depend on whether safety behavior aims to preclude the occurrence of threat. Future research is needed to investigate which safety behaviors should be eliminated during exposure-based therapy and which safety behaviors may be incorporated into treatment, and for whom.

Authors' note
The research was carried out at Utrecht University, The Netherlands.

Acknowledgements
The authors would like to thank Vera Bouwman and Rinske van den Heuvel for their help with the data collection, Stefan van der Stigchel for providing us with the lab and equipment to collect the data for Experiment 1, Angelos Krypotos for extracting the data, and Kevin van Schie for his helpful input while writing the manuscript.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a Research Talent Grant from the Netherlands Organization for Scientific Research awarded to Sophie van Uijen (no 403-11-406).

References
Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. (2011). Exposure therapy for anxiety: Principles and practice. New York, NY: Guilford Press.
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., Luyten, L., Vansteenwegen, D., Hermans, D., & Beekers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. Journal of Behavior Therapy and Experimental Psychiatry, 44, 201–206. doi: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. doi:10.1016/j.bbr.2011.03.031

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, 11, 485–494. doi:10.1101/lm.78804

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

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45, 602–607. doi:10.1111/j.1469-8986.2008.00654.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. doi:10.1016/j.brat.2014.04.006

Dalmaijer, E. S., Mathoût, S., & van der Stigchel, S. (2014). *PyGaze*: An open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments. *Behavior Research Methods*, 46, 913–921. doi:10.3758/s13428-013-0422-2

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. doi:10.1016/j.brat.2008.01.010

Deacon, B. J., Sy, J. T., Lickel, J. J., & Nelson, E. A. (2010). Does the judicious use of safety behaviors improve the efficacy and acceptability of exposure therapy for claustrophobic fear? *Journal of Behavior Therapy and Experimental Psychiatry*, 41, 71–80. doi:10.1016/j.jbtep.2009.10.004

Duits, P., Baas, J. M. P., Engelhard, I. M., Richter, J., Huisman – van Dijk, H. M., Limberg, A., … Cath, D. C. (2017). Latent fear conditioning trajectories in anxiety disorders and their predictive value on treatment outcome. Retrieved from https://dspace.library.uu.nl/handle/1874/339826

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

Eyelink 1000 [Apparatus and software]. Mississauga, Canada: SR Research.

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. doi:10.1007/s00213-005-0124-2

Hood, H. K., Antony, M. M., Koerner, N., & Monson, C. M. (2010). Effects of safety behaviors on fear reduction during exposure. *Behaviour Research and Therapy, 48*, 1161–1169. doi:10.1016/j.brat.2010.08.006

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

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

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017). Neural correlates of pupil dilation during human fear learning. *NeuroImage*, 147, 186–197. doi:10.1016/j.neuroimage.2016.11.072

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. doi:10.1016/j.brat.2004.10.007

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

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

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 3–26. doi:10.1037/0097-7403.28.1.3

Mathôt, S. (2013). Pupil size reconstruction examples. *Figshare*. [Software]. Retrieved from doi:10.6084/m9.figshare.688002

McManus, F., Sacadura, C., & Clark, D. M. (2008). Why social anxiety persists: An experimental investigation of the role of safety behaviours as a maintaining factor. *Journal of Behavior Therapy and Experimental Psychiatry*, 39, 147–161. doi:10.1016/j.jbtep.2006.12.002

Meulders, A., Van Daele, T., Volders, S., & Vlaeyen, J. W. (2016). The use of safety-seeking behavior in exposure-based treatments 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

Milosevic, I., & Radomsky, A. S. (2008). Safety behaviour does not necessarily interfere with exposure therapy.
