Behaviour Research and Therapy 121 (2019) 103463

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Incentive-based extinction of safety behaviors: Positive outcomes competing with aversive outcomes trigger fear-opposite action to prevent protection from fear extinction

Andre Pittig[a,b,*]

[a] Department of Psychology (Biological Psychology, Clinical Psychology, and Psychotherapy), University of Würzburg, Marcusstrasse 9-11, 97070, Würzburg, Germany
[b] Center of Mental Health, University of Würzburg, Würzburg, Germany

ARTICLE INFO

Keywords:
Fear conditioning
Fear extinction
Avoidance
Safety behavior
Anxiety
Positive reinforcement

ABSTRACT

Maladaptive safety behavior maintains fear and anxiety by prohibiting inhibitory learning about the non-occurrence of feared outcomes (known as protection from extinction). Not engaging in safety behavior, however, requires to act opposite to fear-motivated behavioral tendencies. The initiation of such fear-opposite action by positive outcomes, which were in conflict with safety behavior, was tested. Following fear acquisition to a warning signal (CS+), participants acquired safety behavior to prevent the aversive outcome (n = 48). Next, safety behavior also prevented gaining rewards. In a control group (n = 50), neutral outcomes were presented to control for novelty effects of the second outcome. Subsequently, no aversive outcome occurred anymore. Phases with safety behavior were intermitted by phases without safety behavior being available to examine cognitive and physiological indicators of fear and anxiety. Without competing positive outcomes, safety behavior was frequently executed, persisted in absence of the aversive outcome, and prohibited extinction learning. Positive outcomes clearly reduced safety behavior despite equal levels of acquired fear. This enabled fear extinction as soon as the aversive outcome was absent. Importantly, this extinction learning resulted in attenuated fear and anxiety responses when safety behavior became unavailable. Post-hoc findings indicated that the mere anticipation of positive outcomes slightly reduced safety behavior. Thus, competing positive outcomes triggered fear-opposite action that prevented persistent safety behavior and protection from extinction.

In everyday life, humans exhibit a variety of defensive behaviors ranging from automatic reflexes to more deliberate instrumental avoidance and safety behavior (Ledoux & Daw, 2018; Pittig, Treanor, LeBeau, & Craske, 2018). While avoidance entails complete avoidance of threatening stimuli or situations, safety behavior is typically carried out before or during confrontation with threatening stimuli or situations to minimize or prevent a threatening outcome. Oftentimes safety behavior is thus an adaptive precautionary action to prevent harm. However, safety behavior can become maladaptive when it is persistent in the absence of threat and linked to costs and impairments for the individual. In this regard, maladaptive safety behavior is a common characteristic of anxious psychopathology (e.g., Beesdo-Baum et al., 2012; Helbig-Lang & Petermann, 2010; McManus, Sacadura, & Clark, 2008; Salkovskis, Clark, Hackmann, Wells, & Gelder, 1999). For example, individuals with agoraphobia rely on various maladaptive behaviors to prevent a perceived threat of suffocation or fainting when using public transportation (e.g., always being accompanied, intake of unnecessary medication). Experimental psychopathology thus aims to better understand the interaction between safety behavior and the maintenance of fear and anxiety.

In this context, experimental findings demonstrated that safety behaviors preserve threat beliefs and physiological fear responses. This preserving effect has been linked to a mechanism referred to as "protection from extinction" (Lovibond, Davis, & O'Flaherty, 2000; Lovibond, Mitchell, Minard, Brady, & Menzies, 2009). In a seminal study, individuals first learned to associate a warning signal (CS+) with an aversive unconditioned stimulus (US) (Lovibond et al., 2009). Next, participants acquired a safety behavior: Whenever the warning signal was present, participants could prevent the upcoming aversive outcome by pressing a button. Importantly, this safety behavior persisted in the absence of the aversive outcome and thereby protected from fear extinction learning. In other words, continuous safety behavior prohibited

* Department of Psychology (Biological Psychology, Clinical Psychology, and Psychotherapy), University of Würzburg, Marcusstrasse 9-11, 97070, Würzburg, Germany.
E-mail address: andre.pittig@uni-wuerzburg.de.

https://doi.org/10.1016/j.brat.2019.103463
Received 11 March 2019; Received in revised form 27 July 2019; Accepted 19 August 2019
Available online 20 August 2019
0005-7967/

new learning that the warning signal became safe as it was not followed by the aversive outcome anymore. This protection from extinction resulted in persistent threat beliefs and physiological fear responses (Lovibond et al., 2009).

In the clinical context, persistent safety behavior is likewise assumed to play a role in the maintenance of anxiety disorders (e.g., Helbig-Lang & Petermann, 2010; Salkovskis et al., 1999; Wells et al., 1995). This maintaining effect is supposed to be driven by protection from extinction, because individuals engage in safety behaviors when anticipating or being confronted with a feared stimulus. Consequently, a recommendation for exposure-based treatments is to remove all safety behaviors (e.g., Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). Refraining from safety behaviors, however, requires to act against the action tendencies associated with fear. That is, anxious individuals need to perform a fear-opposite action. Facilitating fear-opposite action may thus be beneficial for exposure-based therapy. In this regard, one strategy may be to establish positive outcomes, which are in conflict with performing safety behavior. This operant strategy is here referred to as incentive-based extinction of safety behavior.

Recent experimental studies provided preliminary evidence for the impact of competing positive outcomes on safety and avoidance behavior (for a brief overview see Pittig, et al., 2018). For example, multiple studies demonstrated reduced avoidance of feared or unpleasant stimuli when avoidance was linked to costs or in conflict with gaining real, tangible or virtual, intangible rewards (Aupperle, Sullivan, Melrose, Paulus, & Stein, 2011; Bublatzky, Alpers, & Pittig, 2017; Pittig, Hengen, Bublatzky, & Alpers, 2018; Pittig, Schulz, Craske, & Alpers, 2014; Rattel, Miedl, Blechert, & Wilhelm, 2017; Talmi, Dayan, Kiebel, Frith, & Dolan, 2009). It seems likely, but remains to be tested, that similar effects hold true for safety behaviors. In contrast to these behavioral findings, less is known about the impact of competing positive outcomes on fear and anxiety responses. Some studies demonstrated that an incentive-based reduction of avoidance was accompanied by a reduction of self-reported and physiological indicators of fear when no (more) aversive outcomes occurred (Bublatzky et al., 2017; Pittig, et al., 2018; Rattel et al., 2017). In the presence of aversive outcomes, we recently demonstrated that competing positive outcomes dampen avoidance, but had no effect on fear acquisition (Pittig & Dehler, 2019). Specifically, participants in one group chose to approach an aversive US when this behavior was linked to high rewards. In contrast, a group without competing rewards showed persistent avoidance. Interestingly, the reduction of avoidance acquisition did not modulate the level of acquired fear. Both groups showed comparable fear responses to a warning signal predicting the aversive outcome (Pittig & Dehler, 2019).

These findings suggest that acquired fear is expressed on the cognitive and peripheral-physiological level (US expectancy ratings and SCRs), but does not necessarily express as avoidance or safety behavior when competing positive outcomes are present. This divergence may be used to initiate fear extinction learning: Competing positive outcomes may function as incentives to approach a warning signal, which in turn initiate fear extinction once the aversive outcome is absent. The present study examined this research question. Specifically, the following hypotheses were tested: i) Safety behavior is frequently performed in absence of competing positive outcomes, but less frequent in their presence. ii) A reduction of safety behavior is not associated with a reduction of fear responses as long as the aversive outcome is present. iii) In absence of the aversive outcome, safety behavior is maintained when not in conflict with positive outcomes (i.e., protection from extinction), but competing positive outcomes reduce safety behavior and thereby initiate fear extinction.

So far, investigating the impact of competing positive outcomes focused on the modulation of fear responses. Fear responses are elicited by a specific warning signal (CS+) that is in close proximity to the aversive outcome (i.e., high threat imminence). Typical measures are self-reported expectancy of the aversive outcome or skin conductance responses (SCRs). In contrast to stimulus-specific fear responses, anxiety responses are stimulus-unspecific. Anxiety can be defined as "sustained defensive state towards a future-oriented and not stimulus specific threat" (Lonsdorf et al., 2017). Anxiety responses are thus linked to the broader context and more distal threat (i.e., lower threat imminence). Typical measures are, for example, self-reported anxiety or sustained physiological arousal as measured by skin conductance levels (SCL). Supporting the differentiation of fear and anxiety, proximal and distal threat are linked to different physiological and behavioral response patterns (Löw, Weymar, & Hamm, 2015; Mobbs, 2018; Mobbs et al., 2007, 2009; Wendt, Löw, Weymar, Lotze, & Hamm, 2017). Competing positive outcomes may thus show the same or a different modulation of fear versus anxiety responses. For example, competing positive outcome may not modulate responding under high threat imminence (i.e. fear responses), but under low threat imminence (i.e., anxiety responses). To the best of my knowledge, there is no experimental test on the effect of competing positive outcomes on anxiety responses.

Therefore, this study tested the effect of positive outcomes competing with aversive outcomes on the frequency of safety behaviors and the level of fear and anxiety responses. A single-cue conditioning paradigm was completed by two randomized groups. After passive fear acquisition training, participants learned to perform a safety behavior to prevent the occurrence of the aversive outcome. In the reward group, safety behavior subsequently did not only prevent the aversive outcome, but also prevented gaining rewards. In the no-reward group, a neutral feedback stimulus was presented instead of rewards to control for novelty effects of the second outcome. Level of fear responses (US expectancy ratings and SCRs) were examined during safety behavior phases and repeated test phases, in which safety behavior was unavailable. In addition, anxiety responses (self-reported ratings and SCL) were investigated during inter-trial intervals (ITIs). For this purpose, ITIs during experimental phases in which safety behavior was available versus unavailable were highlighted by different background colors and prompts about (un-)availability of safety behavior.

1. Methods

1.1. Participants

Initially, 100 participants from the student body of the University of Würzburg and the general community were recruited. Participants provided informed consent to procedures approved by the local ethics committee. Exclusion criteria were current or history of psychosis, bipolar disorder, traumatic brain injury, mental retardation, substance dependence, current use of psychotropic medication, any serious medical conditions, and pregnancy. Participants were randomized to two equally sized groups (reward or no-reward group). Two participants of the reward group were excluded for not following instructions. No significant differences between groups were found for age, sex, trait anxiety, current symptoms of anxiety or depression, general risk-taking, or acceptance of unpleasant distress (see Table 1). Groups also did not differ in self-reported consumption of caffeine, nicotine, and alcohol, or amount of physical activity per week, $ts < 1.51$, $ps > .13$; $BF_{01} > 1.72$.

2. Materials and procedures

After electrodes for physiological recording were attached, participants completed questionnaires to assess individual differences that might affect task performance. Questionnaires assessed symptoms of anxiety during the last week (PROMIS Short Form v1.0 - Anxiety 8a; Wahl, Löwe, & Rose, 2011) as well as state and trait anxiety (State-Trait Anxiety Inventory; Spielberger, Gorsuch, Lushene, & Vagg, 1983; anxiety facet of the NEO-PI-R; (Costa & McCrae, 1992). Additionally, questionnaires assessed symptoms of depression (General depression scale; Hautzinger, Bailer, Hofmeister, & Keller, 2012), general risk taking (short-scale risk-taking-1; Beierlein, Kovaleva, Kemper, & Rammstedt, 2014), acceptance of unpleasant or unwanted distress

Table 1
Demographic and questionnaire data.

	Reward (n = 48)		*No-Reward* (n = 50)		t or χ^2	p	d	Bayes factor
Sex = Female (%)	36	(75.0%)	35	(70.0%)	0.31[b]	.580	0.11	BF_{01} = 4.31
Age	25.21	(5.02)	24.48	(3.75)	0.82[a]	.416	0.17	BF_{01} = 3.45
Trait anxiety: NEO-PI-R-N1	15.04	(4.81)	14.78	(5.34)	0.26[a]	.800	0.05	BF_{01} = 4.57
STAI-T	37.44	(9.50)	38.02	(9.61)	−0.30[a]	.765	−0.06	BF_{01} = 4.52
Anxiety symptoms during last week (PROMIS)	14.58	(5.62)	13.94	(5.37)	0.58[a]	.563	0.12	BF_{01} = 4.05
State anxiety (STAI-S)	34.70	(6.65)	36.63	(7.87)	−1.30[a]	.195	−0.26	BF_{01} = 2.21
Depression (ADS-L)	13.24	(9.10)	11.59	(8.42)	0.93[a]	.353	0.19	BF_{01} = 3.20
Risk taking	4.15	(0.95)	4.00	(1.07)	0.71[a]	.477	0.14	BF_{01} = 3.75
Acceptance of unpleasant distress (AS)	29.37	(4.67)	28.78	(4.11)	0.67[a]	.507	0.14	BF_{01} = 3.86
Unpleasantness of last US	62.63	(24.78)	65.68	(23.43)	−0.63[a]	.532	−0.13	BF_{01} = 3.95
Avoidance motivation after task	54.67	(27.91)	86.68	(19.46)	**6.61**[a]	**< .001**	**−1.14**	**BF_{10} > 1000**
Approach motivation after task	79.75	(24.10)	–	–	–	–	–	–

Note. Means (and standard deviations) for the two groups. NEO-PI-R-N1 = anxiety subscale of NEO-PI-R, range = 0–32 (Costa & McCrae, 1992); STAI-S/-T = State-Trait Anxiety Inventory, range = 20–80 (Spielberger et al., 1983); PROMIS Short Form v1.0 - Anxiety 8a (Wahl et al., 2011), range = 8–40; ADS-L = General depression scale, range 0–60 (Hautzinger et al., 2012); Risk taking = Short-scale risk-taking-1, range = 1–7 (Beierlein et al., 2014); AS = Acceptance scale, range = 7–49 (Wolgast, 2014). [a] $t(96)$. [b] $\chi^2(1, 98)$.

(Acceptance scale; Wolgast, 2014), and various sociodemographic data. Next, individual intensity of the US was calibrated. The US was an electrical stimulation to the non-dominant forearm consisting of 125 consecutive 2-ms stimulations delivered through a bar-electrode. US intensity started at a fixed low level (0.2 mA) and was stepwise increased depending on participant's aversiveness ratings until participants indicated that the US was "unpleasant and causing discomfort, but not painful" (anchor at 80/100 on rating scale). Importantly, groups did not differ in actual US intensity, $t(96) = 0.55$, $p = .584$, $d = 0.11$, $CI_{95} = -0.29 - 0.51$; $BF_{01} = 4.38$, or self-reported unpleasantness of the last US delivered in the paradigm (see Table 1). After questionnaires, participants completed the computerized learning paradigm.

2.1. Single-cue fear and safety behavior paradigm

The paradigm consisted of 52 trials subdivided into eight consecutive phases (see Table 2). In each trial, the same geometrical shape was presented for 8s as CS (purple hexagon). Inter-trial-intervals (ITIs) varied from 16 to 19s. Across phases, outcomes following the CS and the availability of safety behavior were manipulated. Before starting the paradigm, participants in the reward group were instructed that i) they may win rewards on some trials, ii) three trials will be randomly selected, and iii) they will be paid the amount of rewards gained in the selected trials. The no-reward group did not receive this instruction.

2.1.1. Fear and safety behavior acquisition to the aversive US

The first four phases were identical for both groups to verify comparable fear and safety behavior acquisition to the aversive US. During *CS habituation* (Phase 1; 4 trials), the CS was presented without any outcomes. CS habituation served as a baseline to contrast fear acquisition in the subsequent phase. During *fear acquisition* training (Phase 2; 8 trials), the CS was followed by the US in every trial (i.e., 100% reinforcement).

For subsequent *safety behavior acquisition* training (Phase 3; 8 trials), participants were instructed that they can "*prevent all outcomes of the upcoming stimulus*" by pressing a safety mouse button (left/right counterbalanced between participants), but that outcomes are not prevented when pressing the other, non-safety mouse button (right/left). At the beginning of each trial, participants thus had to actively decide to prevent vs. not to prevent outcomes of the immediately following CS (i.e., press safety or non-safety button). The CS was presented regardless of the participant's response, i.e., responses reflect safety behavior in terms of US-avoidance, not avoidance behavior in terms of CS-avoidance. In line with the participant's response, the US was omitted when the safety button was pressed, but delivered when the non-safety button was pressed. Thus, safety behavior was operationalized as pressing the safety button. Safety behavior acquisition was followed by the first test phase.

During *Test I* (Phase 4; 2 trials), safety responses were not available anymore, i.e., there was no choice to prevent outcomes of the CS. CS presentations were always followed by a US (i.e., re-acquisition) to not induce extinction learning, which may have biased performance in subsequent phases (i.e., reduced safety behavior may be caused fear extinction rather than competing positive outcomes). The first test thus examined the return of fear responses when acquired safety behavior was not available anymore.

2.1.2. Introducing competing rewards vs. neutral stimuli

In the following *Reward-US* phase (Phase 5; 8 trials), safety behavior was available again. Again, an aversive US was delivered when participants pressed the non-safety button and omitted when they pressed the safety button. However, each CS was now associated with a second outcome, which differed between groups. For the reward group, a fixed reward was obtained when participants did not prevent outcomes ($€_f$, "Gained reward: 0.10€" displayed as green text) and omitted when participants chose to prevent all outcomes ("Missed reward: 0.10€" in red text). These contingencies were used to establish a conflict between competing outcomes: Either, participants decided to prevent all outcomes and were thus safe from the aversive US, but missed the competing reward. Or, participants decided to not prevent consequences and obtained the reward, but also received an aversive US. In this regard, the active choice between safety vs. non-safety button in combination with the instructions to "*prevent all outcomes*" emphasized the

Table 2
Experimental design.

Group	CS Habi-tuation	Fear Acquisition	Safety Behavior Acquisition	Test I	Reward-US	Test II	Reward-NoUS	Extinction test
Reward	A - (4)	A + (8)	A [+] (8)	A + (2)	A [+, $€_f$] (8)	A + (2)	A [-, $€_i$] (12)	A - (8)
No-Reward	A - (4)	A + (8)	A [+] (8)	A + (2)	A [+, NS_f] (8)	A + (2)	A [-, NS_v] (12)	A - (8)

Note. A = CS (geometric shape), - = no US, + = US, […] = preventing all outcomes possible, $€_{f/i}$ = fixed/increasing reward, $NS_{f/v}$ = fixed/varying neutral stimulus (non-sense letters), number of trials is indicated in parentheses.

competition between performing safety behaviors and obtaining rewards. The non-reward group served as control group to account for mere novelty effects of a second outcome. Instead of competing rewards, a neutral stimulus (NS) was used as second outcome (NS_f, nonsense letters with same color and length as reward feedbacks). US contingencies were identical to the reward group. In both groups, the second outcome was presented after a 2 s fixation cross following the CS (and US). Comparing both groups thus tested whether competing rewards reduce safety behavior in the presence of aversive outcomes.

Test II (Phase 6; 2 trials) was identical to Test I (i.e., no safety behavior available, CS paired with US in each trial), however, rewards and neutral feedback stimuli were continued as in the previous phase. Test II examined whether both groups show comparable fear responses despite differences in safety behavior in the prior phase (i.e., a conceptual replication of Pittig & Dehler, 2019).

2.1.3. Initiating extinction learning

The subsequent *Reward-NoUS* phase (Phase 7; 12 trials) was comparable to the Reward-US phase with two important differences: First, no more USs were delivered in both groups, allowing for fear extinction learning. Second, the competing rewards in the reward group incrementally increased from trial to trial ($€_i$, randomized step size of 0.03€, 0.04€, or 0.05€). Combined, these two design factors aimed to minimize safety behavior and initiate fear extinction learning in the reward group. To control for novelty introduced by varying reward feedbacks, varying neutral stimuli were presented in the no-reward group (NS_V). Participants were not instructed about the changes in contingencies and had to learn from experience.

In the final *extinction test* (Phase 8; 8 trials), the CS was presented in the absence of any outcomes and without safety behavior being available in both groups. For the no-reward group, this test resembles the test phase of Lovibond et al. (2009) to demonstrate protection from extinction. Thus, this final phase examined whether the initiation of fear extinction learning in the reward group during the previous phase resulted in reduced fear responses, i.e., reduced protection from extinction.

2.2. Indicators of fear learning: US expectancy ratings and SCRs

US expectancy ratings and SCRs to the CS were used as cognitive and physiological indicators of fear learning. For US expectancy ratings, participants rated their expectancy of an US occurring after the CS in each trial. Ratings were completed during CS presentation by indicating the expected likelihood on a visual analog scale (0%–100%) via mouse click using the dominant hand to not confound expectancy ratings with SCR analyses (see Pittig & Dehler, 2019).

Skin conductance was recorded with two reusable Ag/AgCl electrodes with electrodermal conducting gel attached to the hypothenar eminence of the non-dominant hand and a constant voltage of 0.5 V using a V-Amp system (Brain Products, Germany; sampling rate = 1000 Hz). Data monitoring, acquisition, and parametrization was conducted with BrainVision Analyzer (Brain Products, Germany). Raw data were filtered with a notch filter (50 Hz) and a 1 Hz FIR lowpass filter to remove high frequency noise. Artifacts (e.g., coughing, excessive movement) were recorded and excluded from analyses. SCRs were obtained with semi-automatic trough-to-peak scoring by calculating the maximum increase in skin conductance in the interval from 1s after CS onset to CS offset in comparison to the corresponding trough (see Boucsein et al., 2012). Specifically, maximum and minimum amplitude during CS presentation was automatically scored with BrainVision Analyzer, visually inspected, and corrected if necessary (e.g., if minimum was scored after maximum). The square root was taken to obtain normal distribution (Dawson, Schell, & Filion, 2007). Three participants ($n = 2$ from the reward group) were excluded from SCR analyses due to technical failure. In addition, one participant in the reward group was excluded due to excessive movements, which biased SCRs in most trials.

2.3. Indicators of anxiety: anxiety ratings and SCL during ITIs

Indicators of anxiety were assessed via self-reported anxiety ratings and SCL during ITIs. Starting with safety behavior acquisition (Phase 3), phases alternated between the availability vs. unavailability of safety behaviors, establishing different contexts. To amplify context differentiation during ITIs, screen backgrounds and texts were manipulated. During CS habituation and fear acquisition training, ITIs consisted of a fixation cross on white background. For all subsequent phases in which safety behavior was available, ITIs consisted of a light-blue background with a central text stating "*Preventing possible*". For phases with safety behavior being unavailable, ITIs consisted of a light-red background with a central text stating "*Preventing NOT possible*". In the middle of each phase (+/− one trial), participants were asked to rate their current level of anxiety ("*How anxious do you feel at the moment?*", visual analog scale, 0–100). In addition, SCL was calculated as the average SCL during the last 4s of each ITI (to omit SCRs to the US in SCL scoring).

2.4. Statistical analysis

Original data of the study are available at https://osf.io/e7ghz/. Main analyses focused on i) the frequency of safety behavior, ii) the level of fear responses (US expectancy, SCRs), and iii) the level of anxiety (anxiety ratings, SCL). For safety behavior, average frequency of safety behavior was calculated for each phase and compared between groups. No group differences were expected during safety behavior acquisition (two-tailed test). Based on Pittig and Dehler (2019), less frequent safety behavior was expected in the presence of competing rewards (i.e., Reward-US and Reward-NoUS phase, one-tailed tests). In addition, change of safety behavior frequency across phases was analyzed by planned 2 × 2 repeated measure ANOVAs with factors Group and Phase.

For fear responses, responses from two consecutive trials were averaged to reduce noise. Analyses were conducted within each phase using pairwise tests or repeated measure ANOVAs (Group x Trial) because different trajectories were expected per phase (e.g., responses decreasing during CS habituation, but increasing during fear acquisition). In addition, changes in fear responses due to the transition between phases were examined by planned 2 × 2 repeated measures ANOVAs (Group x Trial) comparing the last trial of a preceding phase with the first trial of the succeeding phase. As reduced safety behaviors during the Reward-US phase would result in more USs, higher fear responses during this phase were expected in the reward group (one-tailed tests). No differences in the subsequent Test II were expected. In the Reward-NoUS phase, reduced safety behavior was expected, resulting in a decrease of fear responses due to fear extinction in the reward group, whereas no change in fear responses was expected in the no-reward group. Finally, an increase in fear responses was expected during extinction test in both groups, however, with an attenuated increase in the reward group due to prior extinction experience.

For anxiety responses, planned comparisons examined group differences in self-reported anxiety and SCL in each phase. Again, higher anxiety responding was expected in the reward group during the Reward-US phase (one-tailed test) and no differences were expected during Test II. Finally, lower anxiety responses were expected during extinction test due to prior fear extinction experience in the reward group.

For all ANOVAs, Greenhouse-Geisser correction was applied when necessary. Pairwise and follow-up analyses were conducted with *t* tests or non-parametric *U* or *W* tests when assumptions of normal distribution were violated. Analyses were also performed within a Bayesian framework (see Krypotos, Blanken, Arnaudova, Matzke, & Beckers, 2017). The main goal of the analyses was hypothesis testing (i.e., the existence of group differences), for which Bayes factor analyses are recommended (see van Doorn et al., 2019). BF_{10} is reported for

comparing the probability of the data coming from the H1 (e.g., mean difference between groups is not zero) compared to the H0 (e.g., mean difference between groups is zero) and BF_{01} for the reversed comparison. Bayesian and frequentist analyses were performed using corresponding tests in JASP (Version 0.9.2; JASP Team, 2019). Although prior knowledge exists for at least some hypotheses (e.g., Pittig & Dehler, 2019), JASP default priors were used as recommended when prior knowledge is rare or vague (see van Doorn et al., 2019). To test whether the direction of the Bayesian results changed based on the choice of prior distribution, robustness checks as provided within JASP were performed. Tests indicated that the direction of BFs remained the same. In case of multiple factors in Bayesian ANOVAs, BFs refer to analyses of effects (across matched models) in which models including the effect are compared to equivalent models without inclusion of the effect.

3. Results

3.1. Safety behavior and indicators of fear learning

Frequency of safety behavior and level of fear responses across the different phases are shown in Fig. 1.

3.1.1. Fear and safety behavior acquisition to the aversive US

As the first four phases were identical for both groups, main results are summarized (for details see supplemental material). Both groups showed successful acquisition of fear as indicated by increasing US expectancy ratings and SCRs during fear acquisition training and compared to CS habituation. Both groups also acquired safety behavior. Unexpectedly, average safety behavior was slightly less frequent in the reward ($M = 0.67$, $SD = 0.30$) compared to the no-reward group ($M = 0.83$, $SD = 0.26$) during safety behavior acquisition, $U = 794$, $p = .003$, $r = 0.34$, $BF_{10} = 8.11$. During the safety behavior acquisition phase, US expectancy ratings and SCRs significantly decreased. When safety behaviors became unavailable in Test I, both US expectancy and SCRs significantly increased. Importantly, there were no group differences for US expectancy ratings and SCRs in these phases (see Fig. 1 and supplemental material). In sum, the reward and no-reward group both acquired fear and safety behavior, with no differences in fear responses at the end.

Reward anticipation explains reduced safety behavior acquisition, but does not limit subsequent *findings.* The slightly reduced safety behavior acquisition in the reward group was unexpected. It may relate to the instruction that rewards can be gained based on decisions *before* safety behavior acquisition. This early instruction may have motivated exploration whether or not rewards can be obtained by not performing safety behavior. This explanation was tested by post-hoc recruitment of a third group (n = 39). The *Delayed reward instruction* group did not differ in sociodemographic or questionnaire data from the other groups (see supplemental material) and completed the same procedures as the reward group. However, instructions about rewards were delivered before the Reward-US phase, i.e., after safety behavior acquisition. As a result, this group showed no difference in safety behavior acquisition ($M = 0.81$, $SD = 0.21$) compared to the no-reward group, $U = 1111$, $p = .237$, $r = 0.14$, $BF_{01} = 3.30$, but significantly more frequent safety behavior compared to the reward group, $U = 699$, $p = .001$, $r = 0.25$, $BF_{10} = 2.41$. Importantly, the third group showed no differences in the frequency of safety behavior compared to the reward group in the remaining phases, $Us > 757$, $ps > .09$, $r < 0.19$, $BF_{01} > 2.25$. Moreover, the course of fear responses fully matched the reward group (see Fig. 1). Thus, the third group replicated the main findings in the reward group, i.e., slightly reduced safety behavior acquisition cannot account for the results of the following phases.

3.1.2. Introducing competing rewards vs. neutral stimuli

Table 3 summarizes main significant results for safety behavior, US expectancy, and SCRs are for all subsequent phases.

3.1.2.1. Reward-US phase (Trial 23–30). Safety behavior. Within the phase, average safety behavior was clearly less frequent in the reward ($M = 0.33$, $SD = 0.28$) compared to the no-reward group ($M = 0.91$, $SD = 0.18$), $U = 157$, $p < .001$, $r = 0.87$, $BF_{10} > 1000$. Regarding

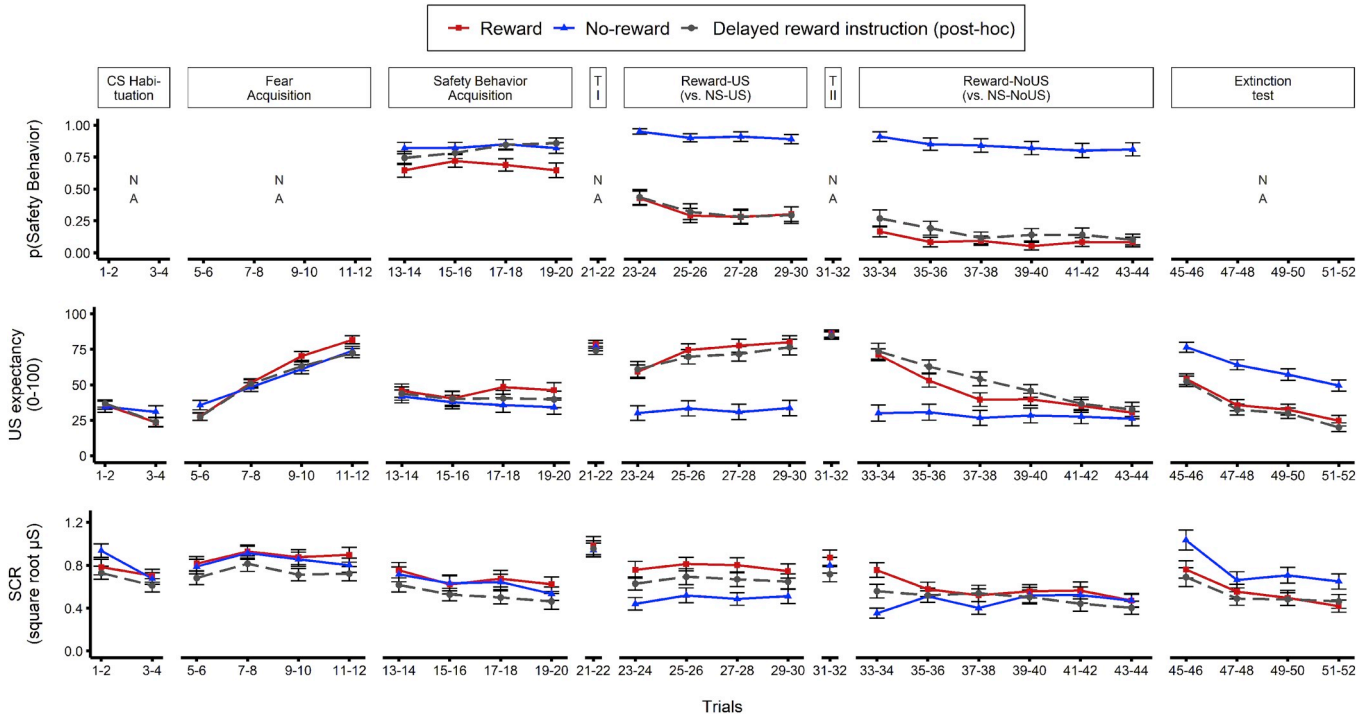

Fig. 1. Frequency of safety behavior (top), US expectancy ratings (middle), and SCRs (bottom) across phases (with SEM), averaged for two consecutive trials. T I/II = Test I/II. NA = Safety behavior not available during the corresponding phase.

Table 3
Summary of main statistical results for each phase.

Phase and measure	Transition analyses (Previous to indicated phase)	Within phase analyses (Trials with the phase)
Reward-US		
Safety behavior	Reward ↓ No-reward ↑	Reward < No-reward
US expectancy	Reward ↓ No-reward ↓↓	Reward ↑ > No-reward ↔
SCR	Reward ↓ No-reward ↓↓	Reward ↔ > No-reward ↔
Test II		
US expectancy	Reward ↔ No-reward ↑	Reward = No-reward
SCR	Reward ↑ No-reward ↑	Reward = No-reward
Reward-NoUS		
Safety behavior	Reward ↓ No-reward ↔	Reward < No-reward
US expectancy	Reward ↓ No-reward ↓↓	Reward ↓ No-reward ↔ Start: Reward > No-reward End: Reward = Reward
SCR	Reward ↔ No-reward ↓	Reward ↓ No-reward ↔ Start: Reward > No-reward End: Reward = Reward
Extinction test		
US expectancy	Reward ↑ No-reward ↑↑	Reward ↓ < No-reward ↓
SCR	Reward ↑ No-reward ↑↑	Reward ↓ < No-reward ↓

Note. Reward/No-reward = Results for the reward/no-reward group.
Within group effects: ↑ or ↓ = significant increase or decrease within group; ↔ = no change; Reward ↓ vs. No-reward ↓↓ indicates stronger decrease in no-reward compared to reward group.
Between group effects: Reward > or < No-reward = significantly larger or smaller in reward compared to no-reward group. " = " indicates no group differences.
"Reward ↑ > No-reward ↔" = Significant increase in reward group, no change in no-reward group and significantly larger responses in the reward group throughout the phase.

phase transition, the reward group showed a significant reduction in safety behavior compared to the previous phase, whereas the no-reward group showed a significant increase, Group x Phase interaction: $F(1, 96) = 71.05$, $p < .001$, $\eta^2 = 0.368$; $BF_{10} > 1000$; follow-up tests reward group: $W = 903.5$, $p < .001$, $r = 0.54$, $BF_{10} > 1000$, follow-up tests no-reward group: $W = 15$, $p < .001$, $r = 0.98$, $BF_{10} = 20.95$.

US expectancy. Regarding phase transition, US expectancy ratings significantly decreased from test (Trial 21–22) to beginning of the Reward-US phase (Trial 23–24). This decrease was less pronounced in the reward group, resulting in higher US expectancy in the reward compared to no-reward group in Trial 23–24, Group x Trial interaction: $F(1, 96) = 12.75$, $p < .001$, $\eta^2 = 0.070$; $BF_{10} = 92.56$; follow-up tests, $U = 1759$, $p > .001$, $r = 0.47$, $BF_{10} = 531.20$.

Within the phase, US expectancy ratings were higher in the reward compared to the no-reward group in all trials, $Us > 1909$, $ps < .001$, $rs > 0.59$, $BF_{10} > 1000$. In addition, US expectancy ratings significantly increase in the reward group, but not in the no-reward group, Group x Trial interaction: $F(2.42, 96) = 4.53$, $p = .008$, $\eta^2 = 0.042$; $BF_{10} = 6.64$; follow-up test reward group: $F(2.30, 47) = 7.34$, $p < .001$, $\eta^2 = 0.135$; $BF_{10} = 172.40$; follow-up test no-reward group: $F(2.05, 49) = 0.99$, $p = .379$, $\eta^2 = 0.020$; $BF_{01} = 12.20$.

Skin conductance responses. SCRs showed a similar pattern of results. Regarding phase transition, SCRs significantly decreased from test to the beginning of the Reward-US phase. This decrease was less pronounced in the reward group, resulting in higher SCRs in the reward compared to no-reward group in Trial 23–24, Group x Trial interaction: $F(1, 92) = 8.14$, $p = .005$, $\eta^2 = 0.052$; $BF_{10} = 7.57$, follow-up tests, $t(92) = 3.29$, $p = .001$, $d = 0.68$, $BF_{10} = 43.89$.

Within the phase, SCRs were consistently higher in the reward compared to the non-reward group, main effect Group: $F(1, 92) = 14.34$, $p < .001$, $\eta^2 = 0.135$; $BF_{10} = 94.15$. There was no significant main effect of Trial or interaction effect, $Fs < 0.79$, $ps > .50$, $\eta^2 < 0.01$; $BFs_{01} > 23.26$.

3.1.2.2. Test II (Trial 31–32). US expectancy. Importantly, within Test II, there were no differences between groups, $U = 1333$, $p = .345$, $r = 0.11$, $BF_{01} = 3.30$. Regarding phase transition, US expectancy ratings did not change from the end of the Reward-US phase (Trial 29–30) to Test II in the reward group, but strongly increased in the no-reward group, Group x Trial interaction: $F(1, 96) = 30.34$, $p < .001$, $\eta^2 = 0.173$; $BF_{10} > 1000$; follow-up test reward group: $W = 394$, $p = .641$, $r = 0.33$, $BF_{01} = 3.50$; follow-up test no-reward group: $W = 90.50$, $p < .001$, $r = 0.86$, $BF_{10} > 1000$.

Skin conductance responses. Importantly, there were no differences between groups within Test II, $t(92) = 0.69$, $p = .493$, $d = 0.14$, $BF_{01} = 3.74$. SCRs analyses showed a similar pattern of results. Regarding phase transition, SCRs significantly increased from the end of the Reward-US phase to the second test, main effect Trial: $F(1, 92) = 14.44$, $p < .001$, $\eta^2 = 0.133$; $BF_{10} = 95.30$.

3.1.2.3. Summary Reward-US phase and Test II. Summarized, safety behavior was clearly less frequent in the reward group during the Reward-US phase. In the presence of the US, this reduction of safety behavior was associated with higher US expectancy ratings and SCRs in the reward compared to the no-reward group. However, when safety behavior became unavailable, there were no group differences in US expectancy and SCRs (i.e., during Test II).

3.1.3. Initiating extinction learning

3.1.3.1. Reward-NoUS phase (Trial 33–44). Safety behavior. Within the phase, safety behavior was almost absent in the reward group ($M = 0.09$, $SD = 0.22$) and clearly less frequent compared to the no-reward group ($M = 0.84$, $SD = 0.32$), $U = 162.5$, $p < .001$, $r = 0.87$, $BF_{10} > 1000$. Regarding phase transition, the reward group again showed a reduction in safety behavior compared to the Reward-US phase, whereas the no-reward group showed no change, Group x Phase interaction: $F(1, 96) = 11.91$, $p < .001$, $\eta^2 = 0.078$; $BF_{10} = 30.81$; follow-up tests reward group: $W = 849.5$, $p < .001$, $r = 0.45$, $BF_{10} > 1000$, $BF_{10} > 1000$; follow-up tests no-reward group: $W = 117.5$, $p = .054$, $r = 0.82$, $BF_{10} = 2.06$.

US expectancy. Regarding phase transition, US expectancy ratings significantly decreased from Test II to the beginning of the Reward-NoUS phase (Trial 33–34). This decrease was more pronounced in the no-reward group, Group x Trial interaction, $F(1, 96) = 29.50$, $p < .001$, $\eta^2 = 0.137$; $BF_{10} > 1000$; follow-up test reward group: $W = 779.5$, $p < .001$, $r = 0.33$, $BF_{10} = 58.76$; follow-up test no-reward group: $W = 115$, $p < .001$, $r = 0.81$, $BF_{10} > 1000$. Moreover, while there were no significant group differences in Test II (see above), the reward group showed higher US expectancy at the beginning of the Reward-NoUS phase, $U = 1859$, $p < .001$, $r = 0.55$, $BF_{10} = 198.14$.

Within the phase, US expectancies significantly decreased in the reward group, whereas no change was found in the no-reward group, Group x Trial: $F(2.62, 96) = 17.26$, $p < .001$, $\eta^2 = 0.124$; $BF_{10} > 1000$; follow-up test reward group: $F(2.58, 47) = 30.14$, $p < .001$, $\eta^2 = 0.391$, $BF_{10} > 1000$; follow-up test no-reward group: $F(2.37, 49) = 1.01$, $p = .377$, $\eta^2 = 0.020$, $BF_{01} = 29.41$. Moreover, the reward group showed clearly higher US expectancy at the beginning of the Reward-NoUS phase (Trial 33–34 and 35–36), $Us > 1636$, $ps < .003$, $rs > 0.36$, $BFs_{10} > 16.65$, and slightly higher ratings in the middle of the phase (Trial 37–38 and 39-40-20), $Us > 1524$, $ps < .02$, $rs > 0.27$, $BFs_{10} > 1.83$. Importantly, these group differences vanished at the end, $U = 1411$, $p = .129$, $r = 0.18$, $BF_{01} = 2.13$.

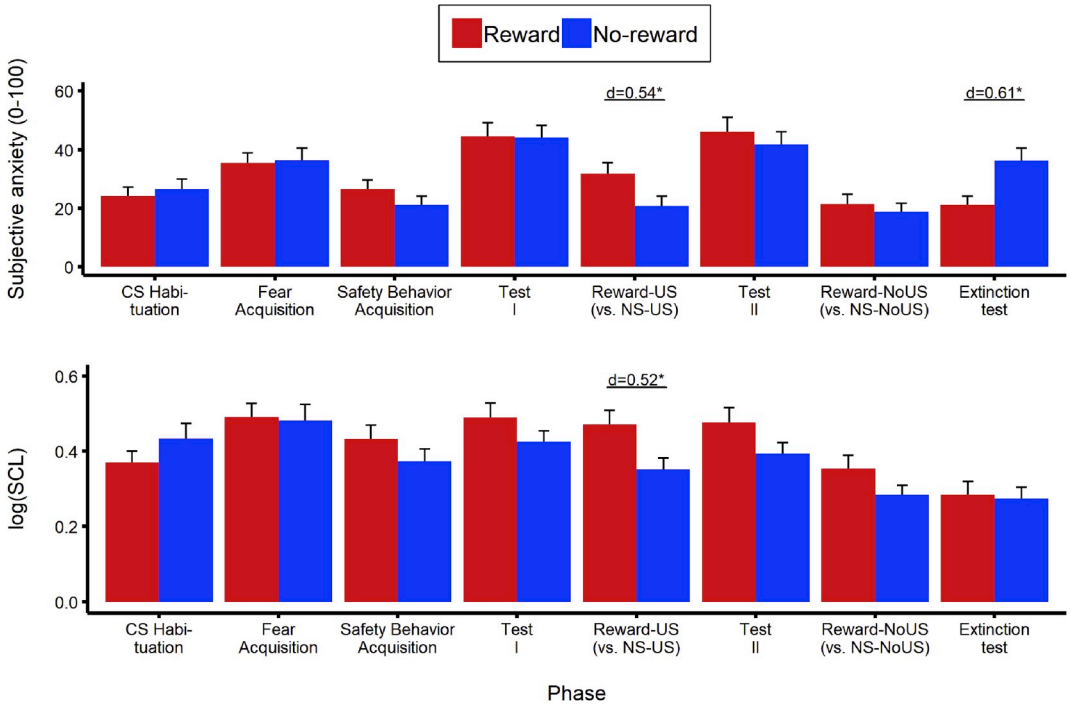

Fig. 2. Self-reported anxiety (top) and SCL (bottom) across phases (with SEM). *ds* indicate Cohen's d effect size for group differences with $p < .05$.

Skin conductance responses. Again, SCR analyses yielded a similar pattern. Regarding phase transition, SCRs in the no-reward group decreased from Test II to the beginning of the Reward-NoUS phase, but did not change in the reward group, Group x Trial interaction: $F(1, 92) = 14.03$, $p < .001$, $\eta^2 = 0.096$; $BF_{10} = 76.48$; follow-up test reward group: $F(1, 45) = 2.73$, $p = .106$, $\eta^2 = 0.057$; but $BF_{01} = 1.06$; follow-up test no-reward group: $F(1, 47) = 13.73$, $p < .001$, $\eta^2 = 0.226$; $BF_{10} > 1000$.

Within the phase, SCRs decreased in the reward group, but not in the no-reward group, Group x Trial interaction: $F(5, 92) = 5.73$, $p < .001$, $\eta^2 = 0.058$; $BF_{10} = 432.82$; follow-up test reward group: $F(5, 45) = 5.95$, $p < .001$, $\eta^2 = 0.117$, $BF_{10} = 479.00$; follow-up test no-reward group: $F(4.13, 47) = 2.25$, $p = .063$, $\eta^2 = 0.046$, $BF_{01} = 2.57$. Importantly, the reward group showed higher SCRs at the beginning of the Reward-NoUS phase, $t(92) = 4.90$, $p < .001$, $d = 1.01$, $BF_{10} > 1000$. Importantly, there were no group differences in the rest of the trials, $ts < 1.41$, $p > .16$, $d < 0.29$, $BF_{01} > 1.93$, due to the reduction in SCR observed in the reward group.

3.1.3.2. Extinction test (Trial 45–52). US expectancy. Regarding phase transition, US expectancy ratings significantly increased from the end of the Reward-NoUS phase to the beginning of extinction test. Importantly, this increase was less pronounced in the reward group, resulting in significantly lower US expectancy ratings in the reward group at the beginning of extinction test, Group x Trial interaction: $F(1, 96) = 13.06$, $p < .001$, $\eta^2 = 0.063$; $BF_{10} = 75.58$; follow-up test for Trial 45–46, $U = 595$, $p < .001$, $r = 0.50$, $BF_{10} = 53.78$.

Within the phase, there was a significant reduction in US expectancy ratings, main effect Trial: $F(1.88, 96) = 72.30$, $p < .001$, $\eta^2 = 0.428$; $BF_{10} > 1000$. Importantly, the reward compared to the no-reward group showed lower US expectancy ratings across extinction test, main effect Group: $F(1, 96) = 27.38$, $p < .001$, $\eta^2 = 0.222$; $BF_{10} > 1000$. There was no significant Group x Trial interaction, $F(1.88, 96) = 0.47$, $p = .466$, $\eta^2 = 0.004$; $BF_{01} = 14.08$.

Skin conductance responses. SCRs showed a similar pattern. Regarding phase transition, SCRs significantly increased from the Reward-NoUS phase to the first extinction test trial. Again, this increase was less pronounced in the reward group, resulting in significantly lower SCRs in the reward group at the beginning of extinction test, Group x Trial interaction: $F(1, 92) = 5.32$, $p = .023$, $\eta^2 = 0.036$; $BF_{10} = 2.27$; follow-up test for Trial 45–46, $t(92) = 2.22$, $p = .029$, $d = 0.46$, but $BF_{10} = 1.84$.

Within the phase, SCRs decreased across extinction test, main effect Trial: $F(2.60, 92) = 26.02$, $p < .001$, $\eta^2 = 0.218$; $BF_{10} > 1000$. Again, the reward group showed lower SCRs compared to the no-reward group, main effect Group: $F(1, 92) = 5.20$, $p = .025$, $\eta^2 = 0.053$; $BF_{10} = 2.75$. There was no significant Group x Trial interaction, $F(2.60, 92) = 1.24$, $p = .296$, $\eta^2 = 0.010$; $BF_{01} = 8.55$.

3.1.3.3. Summary Reward-NoUS phase and extinction test. Summarized, safety behavior was almost absent in the reward group and clearly less frequent compared to the no-reward group during the Reward-NoUS phase. In the absence of the US, this reduction of safety behavior was associated with higher US expectancy and SCRs at the beginning of the phase. However, US expectancy and SCRs decreased in the reward group, indicating successful fear extinction learning. No group differences were found at the end of the phase. When safety behavior became unavailable (i.e., during extinction test), a strong increase in US expectancy and SCRs was found in the no-reward group, indicating protection from extinction. Most importantly, this increase was significantly reduced in the reward group.

3.2. Indicators of anxiety

Anxiety ratings and SCL during the different phases are shown in Fig. 2. On average, subjective anxiety, $W = 295.5$, $p < .001$, $r = 0.88$, $BF_{10} > 1000$, as well as SCL, $t(93) = 4.58$, $p < .001$, $d = 0.47$, $BF_{10} > 1000$, were higher in phases with unavailable compared to available safety behavior.

3.2.1. Fear and avoidance acquisition to the aversive US

There were no differences between groups during CS habituation, fear acquisition, safety behavior acquisition, or Test I for anxiety ratings, $Us < 1376.5$, $ps > .20$, $rs < 0.15$, $BFs_{01} > 2.45$, or SCL, $ts < 1.35$, $ps > .18$, $ds < 0.28$, $BFs_{01} > 2.49$.

3.2.2. Introducing competing rewards vs. neutral stimuli

In line with CS-specific responses, the reward group showed higher anxiety ratings, $U = 1515$, $p = .012$, $r = 0.26$, $BF_{10} = 3.26$, as well as elevated SCL during the Reward-US phase, $t(92) = 2.49$, $p = .007$, $d = 0.52$, $BF_{10} = 6.38$. During Test II, there were again no differences between groups in anxiety ratings, $U = 1264$, $p = .652$, $r = 0.05$, $BF_{01} = 3.99$, or SCL, $t(92) < 1.69$, $p = .094$, $d = 0.35$, $BF_{01} = 1.31$.

3.2.3. Initiating extinction learning

During the Reward-NoUS phase, there were no differences between groups in anxiety ratings, $U = 1247$, $p = .739$, $r = 0.04$, $BF_{01} = 4.43$, or SCL, $t(92) = 1.63$, $p = .106$, $d = 0.34$, $BF_{01} = 1.44$. Importantly, the reward group showed lower anxiety ratings during extinction test, $U = 854.5$, $p = .012$, $r = 0.29$, $BF_{10} = 5.85$, but no differences in SCL were found, $t(92) = 0.20$, $p = .845$, $d = 0.04$, $BF_{01} = 4.54$.

4. Discussion

The present study examined the impact of positive outcomes competing with aversive outcomes on the frequency of safety behavior and the level of fear and anxiety responses. Main findings demonstrate a clear reduction of safety behaviors by competing positive outcomes. Reduced safety behavior resulted in higher fear and anxiety responses when the aversive outcome still occurred (i.e., in the Reward-US phase). However, there were no group differences in fear or anxiety responses when safety behavior became unavailable (Test II). This suggests differences in expression of safety behavior despite comparable fear acquisition. When the aversive outcome did not occur anymore, safety behavior persisted in the absence of competing positive outcomes (i.e., in the no-reward group during Reward-NoUS phase). This persistence resulted in protection from extinction as indicated by a strong increase in fear and anxiety responses when safety behavior became unavailable again. In contrast, successful fear extinction was initiated in the presence of competing positive outcomes. Importantly, this fear extinction learning resulted in lower fear and anxiety responses when safety behavior became unavailable. Thus, operant incentive-based reduction of safety behavior prohibited protection from extinction.

4.1. Incentive-based extinction of safety behavior and anticipation effects

On the behavioral level, competing positive outcomes clearly reduced safety behavior, evident during the presence and absence of the aversive US (i.e., Reward-US and Reward-NoUS phase). In this regard, participants tolerated higher fear and anxiety responses and even the aversive USs in favor of gaining positive outcomes. This clear reduction of safety behavior is in line with previous findings demonstrating incentive-based reduction of avoidance (Aupperle et al., 2011; Bublatzky et al., 2017; Pittig, et al., 2018; Pittig et al., 2014; Talmi et al., 2009). Recently, Dymond (2019) reviewed strategies for overcoming avoidance, which can be transferred to safety behavior. One major operant strategy includes the mere removal of the aversive outcome. This strategy is assumed to reduce avoidance by making it unnecessary (Dymond, 2019). However, low-cost avoidance has been found to persist after mere removal of the aversive outcome (Pittig & Dehler, 2019; Vervliet & Indekeu, 2015). In the present study, low-cost safety behavior likewise persisted in the absence of the US (i.e., in the no-reward group). In contrast, safety behavior was nearly eliminated when in conflict with obtaining positive outcomes. These findings highlight that reducing safety behavior is more likely when alternative outcomes motivate behavioral change. Incentive-based extinction of safety behavior thus represents a promising operant strategy to prevent the persistence of low-cost safety behavior.

Unexpectedly, the frequency of safety behavior during initial acquisition was already lower in the reward compared to no-reward group, i.e., even before positive outcomes were presented. The major conclusion regarding incentive-based extinction of safety behaviors, however, is verified by the further reduction in subsequent phases. During initial safety behavior acquisition, group differences were small and participants in the reward group performed the safety response in approx. 70% of the trials. When rewards were introduced (Reward-US phase), the frequency of safety responses significantly decreased in the reward group, but increased in the no-reward group. A similar interaction was found for the transition from Reward-US to Reward-NoUS phase. These interaction effects cannot be explained by a mere starting difference during safety behavior acquisition.

Most importantly, the differences in initial safety behavior could be explained by the timing of instructions about rewards. Participants in the reward group were instructed about the opportunity to gain rewards before starting the paradigm. This instruction motivated participants to explore whether or not rewards may be obtained by not performing safety behavior during initial acquisition training. In support, the early reduction of safety behavior vanished when reward instructions were given after acquisition training (as tested by a post-hoc group with delayed reward instructions). Importantly, competing positive outcomes again clearly reduced subsequent safety behavior in the group with delayed reward instructions and the course of fear responses matched the reward group. Thus, incentive-based extinction of safety behavior was replicated in a group that did not show reduced safety behavior during initial acquisition. Summarized, the mere anticipation of positive outcomes motivated occasional non-execution of safety behavior to explore outcome contingencies. The actual occurrence of positive outcomes, however, reduced safety behaviors to a much greater extent. Thus, while mere anticipation of positive outcomes may induce some reduction of safety behavior, actual outcomes had a stronger impact.

4.2. Divergence between safety behavior and fear and anxiety responses

The group differences in safety behavior modulated fear and anxiety responding. US expectancy ratings and SCRs showed similar patterns of results. As expected, both groups showed successful fear acquisition, which was accompanied by elevated anxiety responses. Along with the acquisition of safety behavior, fear and anxiety responses decreased, but increased again in both groups when safety behavior became unavailable (i.e., during Test I). These findings indicate comparable fear acquisition in both groups. Thus, the subsequent reduction of safety behavior in the reward group occurred despite comparable fear acquisition. Reduced safety behavior was indeed associated with elevated fear responses during the Reward-US phase. The present results thus conceptually replicate our recent finding that competing positive outcomes facilitate a divergence between cognitive and physiological expressions of acquired fear and its behavioral expressions in avoidance or safety behavior (Pittig & Dehler, 2019). Moreover, these findings are extended to context-specific anxiety response. The reduction in safety behavior occurred despite elevated anxiety responses. Thus, competing positive outcomes did neither reduce proximal fear responses nor more distal anxiety responses. This supports the notion that, in the presence of aversive outcomes, competing positive outcomes do not directly modulate acquired fear and anxiety responses, but rather function as additional factor influencing action selection. In other words, while competing positive outcomes seem to keep original fear and anxiety acquisition intact, action selection is less strongly guided by acquired fear.

4.3. Preventing protection from extinction

Most importantly, the divergence between fear responses and safety behavior influenced subsequent fear reduction in the absence of the aversive outcome. In the no-reward group, safety behavior persisted and thereby prohibited inhibitory learning that no more aversive outcomes occurred. Consequently, fear responses strongly increased when

safety behavior became unavailable (i.e., during extinction test). Thus, protection from extinction due to persistent safety behavior was replicated in the absence of competing positive outcomes (see Lovibond et al., 2009). In contrast, safety behavior was almost absent when in conflict with positive outcomes. Initially, this reduction of safety behavior again resulted in elevated US expectancy and SCRs (i.e., during the Reward-NoUS phase). Absence of safety behavior, however, enabled fear extinction learning as indicated by a gradual decrease in US expectancy and SCRs. As a result, no group differences in fear responding at the end of the phase were found. When safety behavior became unavailable, fear responses somewhat increased, presumably because of renewal effects as no previous extinction learning had occurred when safety behavior was unavailable. Most importantly, this increase was attenuated in the reward group. Similar effects were found for self-reported anxiety. These group differences indicate that fear extinction experience from the previous phase carried over to the context with unavailable safety behavior. Thus, incentive-based extinction of safety behavior enabled fear extinction learning, thereby prevented the protection from extinction, which resulted in reduced fear and anxiety responses even when safety behavior became unavailable.

It could be argued that habituation might be an alternative explanation for group differences during the last two phases. In typical fear conditioning studies, decreasing SCRs to the CS + are oftentimes observed even under ongoing pairing with the US, presumably due to habituation. Because of reduced safety behavior in the Reward-US phase, participants in the reward group experienced the US significantly more often. If habituation would cause subsequent group differences, group difference should be strongest immediately following the more frequent US experience, i.e. during Test II. However, both groups showed equal SCR magnitudes (an US expectancy) during this test phase. Thus, mere habituation effects cannot account for reduced SCR and SCL in the last two phases.

4.4. Implications and limitations

The present findings have theoretical and clinical implications. Theoretically, the results highlight the distinction between fear learning and its behavioral expression. While fear learning can be a major determinant of action selection when other influencing factors are limited (e.g., in the no-reward group), the present results indicate that alternative factors may easily override the impact of acquired fear on action selection. In addition, the present results highlight the motivational-volitional antecedents of fear extinction learning. Traditionally, fear extinction is studied during passive observation, i.e., participants passively observe a CS not being followed by a US anymore. In real-life as well as in exposure therapy, individuals, however, need to take action and decide whether or not to confront a feared stimulus before new learning can occur. Exposure in real-life and in treatment thus represents a two-step process: First, individuals need to refrain from or reduce avoidance and safety behaviors. Subsequently, extinction learning can occur during exposure. The present paradigm resembles this process by demonstrating how safety behavior can be reduced to initiate fear extinction. In this regard, the present study supports the impact of competing positive outcomes and also found preliminary evidence for an effect of mere anticipation of positive outcomes. The paradigm may thus be suited to explore and compare different strategies of behavior modification and further examine the effects of anticipation. The effects of positive outcome anticipation represent a promising future direction, especially, because willingness to confront feared stimuli is typically motivated by anticipation of long-term positive consequences during CBT and value-guided treatments for anxiety disorders (e.g., Acceptance-Commitment therapy; Hayes, 2004; Hayes, Luoma, Bond, Masuda, & Lillis, 2006). However, the present study is limited to healthy individuals, thus, further research in clinical samples and translation to clinical (analog) interventions is necessary (see Richter, Pittig, Hollandt, & Lueken, 2017).

So far, the effects of competing positive outcomes were interpreted in terms of instrumental learning, i.e., how positive outcomes changed behavioral responses and how this in turn modulates fear and anxiety responses. Counterconditioning may be another mechanism involved in the modulation of fear and anxiety responses. Counterconditioning refers to pairing a learned CS+ with a US of opposite valence compared to the US involved in original learning (De Houwer, 2011). For example, a single aversive US is replaced with a single appetitive US. This procedure differs from present procedures as reward outcomes were linked to non-performance of safety behavior, not CS presentation (the CS also preceded missed-reward feedback). In this regard, the CS had no predictive value for reward occurrence. Whereas a CS-US association was established during fear acquisition training (i.e., before instrumental learning), no acquisition or counterconditioning training for a CS-reward association occurred. Moreover, rewards did not replace the aversive US, but were presented successively (in the Reward-US phase). Importantly, the aversive US immediately followed the CS, whereas the reward feedback was presented after a short inter-stimulus interval (1.5 s) in both the Reward-US and Reward-NoUS phase. In the Reward-NoUS phase, reward outcomes were presented in US omission. However, again, rewards did not replace the US, but were already established as outcome of the instrumental behavior. It thus seems likely that the CS was associated with the aversive US or its omission.

Still, it is possible that some degree of CS-reward association was established during the Reward-US phase. If so, this association should be absent in the no-reward group, which should express in different CS responding between groups during the subsequent test (i.e., Test II). However, there were no differences in any indicator of fear or anxiety responding. It remains possible that these indicators were not suitable to detect subtle counterconditioning effects. Counterconditioning is assumed to specifically reduce negative valence of an aversive conditioned CS+ (De Houwer, 2011; Engelhard, Leer, Lange, & Olatunji, 2014). Incorporating measures of valence in future research may help to pinpoint a potential contribution of counterconditioning. However, even if a weak contribution of counterconditioning cannot ultimately be ruled out, it could not explain the initial reduction of safety behavior. Rewards only occurred when no safety behavior was performed. Thus, fear-opposite action had to occur before any new association that reduced fear responding could be established.

5. Conclusion

In sum, the present study provided comprehensive evidence for an incentive-based extinction of safety behavior. Without positive outcomes, safety behavior was frequently executed and persisted in the absence of the aversive outcome, which resulted in protection from extinction. Positive outcomes competing with safety behavior strongly reduced the frequency of safety behavior despite successful fear acquisition. This incentive-based extinction of safety behavior enabled fear extinction learning as soon as the aversive outcome did not occur anymore. Ultimately, incentive-based reduction of safety behavior prevented protection from fear extinction.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) grant PI1269/2–1 - 389569971 to AP. The author has no conflicts of interest regarding this manuscript. The author would like to thank Eda Kir and Julian Faβe for their help with data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2019.103463.

References

Aupperle, R. L., Sullivan, S., Melrose, A. J., Paulus, M. P., & Stein, M. B. (2011). A reverse translational approach to quantify approach-avoidance conflict in humans. *Behavioural Brain Research, 225*(2), 455–463. https://doi.org/10.1016/j.bbr.2011.08.003.

Beesdo-Baum, K., Jenjahn, E., Höfler, M., Lueken, U., Becker, E. S., & Hoyer, J. (2012). Avoidance, safety behavior, and reassurance seeking in generalized anxiety disorder. *Depression and Anxiety, 29*(11), 948–957. https://doi.org/10.1002/da.21955.

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). *Eine Single-Item-Skala zur Erfassung von Risikobereitschaft: Die Kurzskala Risikobereitschaft-1 (R-1)*. GESIS-Working Papers.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., et al. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x.

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. https://doi.org/10.1016/j.brat.2017.01.003.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual: Revised NEO Personality Inventory (NEO PI-R) and NEO Five-factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

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

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.). *Handbook of psychophysiology* (pp. 159–181). (3rd ed.). Cambridge: Cambridge University Press.

De Houwer, J. (2011). Evaluative conditioning: A review of functional knowledge and mental process theories. *Associative learning and conditioning theory: Human and non-human applications*https://doi.org/10.1093/acprof:oso/9780199735969.003.0130.

Dymond, S. (2019). Overcoming avoidance in anxiety disorders: The contributions of Pavlovian and operant avoidance extinction methods. *Neuroscience & Biobehavioral Reviews, 98*(December 2018), 61–70. https://doi.org/10.1016/j.neubiorev.2019.01.007.

Engelhard, I. M., Leer, A., Lange, E., & Olatunji, B. O. (2014). Shaking that icky feeling: Effects of extinction and counterconditioning on disgust-related evaluative learning. *Behavior Therapy, 45*(5), 708–719. https://doi.org/10.1016/j.beth.2014.04.003.

Hautzinger, M., Bailer, M., Hofmeister, D., & Keller, F. (2012). Allgemeine depressionsskala (ADS). *Psychiatrische Praxis, 39*(6), 302–304.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy, 35*, 639–665.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006.

Helbig-Lang, S., & Petermann, F. (2010). Tolerate or eliminate? A systematic review on the effects of safety behavior across anxiety disorders. *Clinical Psychology: Science and Practice, 17*(3), 218–233. https://doi.org/10.1111/j.1468-2850.2010.01213.x.

JASP Team (2019). *JASP*. ([Computer Software]).

Krypotos, A.-M., Blanken, T., Arnaudova, I., Matzke, D., & Beckers, T. (2017). A primer on Bayesian analysis for experimental psychopathologists. *Journal of Experimental Psychopathology, 8*(2), 140–157. https://doi.org/10.5127/jep.057316.

Ledoux, J. E., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience, 19*(5), 269–282. https://doi.org/10.1038/nrn.2018.22.

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

Lovibond, P. F., Davis, N. R., & O'Flaherty, a S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research and Therapy, 38*(10), 967–983.

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*(8), 716–720. https://doi.org/10.1016/j.brat.2009.04.013.

Löw, A., Weymar, M., & Hamm, A. O. (2015). When threat is near, get out of here: Dynamics of defensive behavior during freezing and active avoidance. *Psychological Science, 26*(11), 1706–1716. https://doi.org/10.1177/0956797615597332.

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*(2), 147–161. https://doi.org/10.1016/j.jbtep.2006.12.002.

Mobbs, D. (2018). The ethological deconstruction of fear(s). *Current Opinion in Behavioral Sciences, 24*, 32–37. https://doi.org/10.1016/j.cobeha.2018.02.008.

Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., et al. (2009). From threat to fear: The neural organization of defensive fear systems in humans. *Journal of Neuroscience, 29*(39), 12236–12243. https://doi.org/10.1523/JNEUROSCI.2378-09.2009.

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., et al. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science, 317*(5841), 1079–1083. https://doi.org/10.1126/science.1144298.

Pittig, A., & Dehler, J. (2019). Same fear responses, less avoidance: Rewards competing with aversive outcomes do not buffer fear acquisition, but attenuate avoidance to accelerate subsequent fear extinction. *Behaviour Research and Therapy, 112*, 1–11. November 2018 https://doi.org/10.1016/j.brat.2018.11.003.

Pittig, A., Hengen, K., Bublatzky, F., & Alpers, G. W. (2018). Social and monetary incentives counteract fear-driven avoidance: Evidence from approach-avoidance decisions. *Journal of Behavior Therapy and Experimental Psychiatry, 60*, 69–77. https://doi.org/10.1016/j.jbtep.2018.04.002.

Pittig, A., Schulz, A. R., Craske, M. G., & Alpers, G. W. (2014). Acquisition of behavioral avoidance: Task-irrelevant conditioned stimuli trigger costly decisions. *Journal of Abnormal Psychology, 123*(2), 314–329. https://doi.org/10.1037/a0036136.

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience & Biobehavioral Reviews, 88*(February), 117–140. https://doi.org/10.1016/j.neubiorev.2018.03.015.

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. https://doi.org/10.1016/j.brat.2016.12.010.

Richter, J., Pittig, A., Hollandt, M., & Lueken, U. (2017). Bridging the gaps between basic science and cognitive-behavioral treatments for anxiety disorders in routine care: Current status and future demands. *Zeitschrift Fur Psychologie/Journal of Psychology, 225*(3), 252–267. https://doi.org/10.1027/2151-2604/a000309.

Salkovskis, P. M., Clark, D. M., Hackmann, A., Wells, A., & Gelder, M. G. (1999). An experimental investigation of the role of safety-seeking behaviours in the maintenance of panic disorder with agoraphobia. *Behaviour Research and Therapy, 37*(6), 559–574. https://doi.org/10.1016/S0005-7967(98)00153-3.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., & Vagg, P. R. (1983). *Manual for the state-trait anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologist Press.

Talmi, D., Dayan, P., Kiebel, S. J., Frith, C. D., & Dolan, R. J. (2009). How humans integrate the prospects of pain and reward during choice. *Journal of Neuroscience, 29*(46), 14617–14626. https://doi.org/10.1523/JNEUROSCI.2026-09.2009.

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., et al. (2019). *The JASP Guidelines for conducting and reporting a Bayesian Analysis*. PsyArXivhttps://doi.org/10.31234/OSF.IO/YQXFR.

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

Wahl, I., Löwe, B., & Rose, M. (2011). Das Patient-Reported Outcomes Measurement Information System (PROMIS): Übersetzung der Item-Banken für Depressivität und Angst ins Deutsche. *Klinische Diagnostik Und Evaluation, 4*, 236–261.

Wells, A., Clark, D. M., Salkovskis, P. M., Ludgate, J., Hackmann, A., & Gelder, M. (1995). Social Phobia : The role of in-situation safety behaviors in maintaining anxiety and negative beliefs method subjects. *Behavior Therapy, 26*(1), 153–161. https://doi.org/10.1016/S0005-7894(05)80088-7.

Wendt, J., Löw, A., Weymar, M., Lotze, M., & Hamm, A. O. (2017). Active avoidance and attentive freezing in the face of approaching threat. *NeuroImage, 158*(June), 196–204. https://doi.org/10.1016/j.neuroimage.2017.06.054.

Wolgast, M. (2014). What does the acceptance and action questionnaire (AAQ-II) really measure? *Behavior Therapy, 45*(6), 831–839. https://doi.org/10.1016/j.beth.2014.07.002.