Archival Report

Biological Psychiatry: CNNI

Timing Matters: Providing Contingency Instructions to Modify Fear Extinction Memories in Anxiety Disorders

Annalisa Lipp, Christian J. Merz, Beray Macit, Marcella L. Woud, Oliver T. Wolf, Jürgen Margraf, and Armin Zlomuzica

ABSTRACT

BACKGROUND: Cognitive processes play a significant role in fear extinction. Contingency instructions (CIs) about the non-occurrence of an unconditioned stimulus (UCS) can be used to promote fear extinction learning and extinction retrieval. However, previous research has mainly been performed with healthy individuals, and a systematic study on the effects of CIs administered during different phases of the extinction memory acquisition in patients with anxiety disorders (ADs) is lacking.

METHODS: We investigated the effects of CIs administered before and/or after fear extinction training on extinction learning and retrieval processes in 240 participants (120 patients with ADs, 120 healthy control participants) using a 3-day fear conditioning paradigm. Skin conductance responses (SCRs), conditioned stimulus (CS) valence, and UCS expectancy ratings served as readout measures.

RESULTS: CIs before extinction training enhanced extinction learning for both patients with ADs and healthy control participants as indexed by SCRs and affective measures. Similarly, postextinction CI administration led to reduced CS differentiation in SCRs and UCS expectancy during extinction retrieval in healthy control participants and patients. After reinstatement, postextinction CI administration also diminished fear relapse in the original acquisition and novel context. Here, time-dependent effects of CIs seemed to further depend on contextual cues established during initial fear acquisition and extinction.

CONCLUSIONS: Our findings contribute to the existing knowledge on the role of time-dependent administration of CIs in fear extinction and offer novel insights into how cognitive interventions can be integrated to augment existing therapeutic approaches for ADs.

https://doi.org/10.1016/j.bpsc.2026.03.015

Fear extinction, the process by which the association between a conditioned stimulus (CS) and an aversive outcome (the unconditioned stimulus [UCS]) is weakened, was initially thought to be a primarily reflexive and evolutionarily conserved process. However, research has increasingly shown that cognitive processes play a significant role in fear extinction and the reduction of fear responses (1). Understanding these processes is crucial for developing more efficient and sustainable cognitive behavioral treatments for anxiety disorders (ADs) (2–6).

Previous research indicates that instructions about CS-UCS contingencies administered after fear acquisition are sufficient to decrease CS-related fear responding. Furthermore, such contingency instructions (CIs) have been shown to promote subsequent fear extinction learning [e.g., (7–11)] and increase brain activity in areas associated with emotion regulation and safety learning (12). Furthermore, CIs decrease differential responding to CS+ versus CS− across measures, including skin conductance responses (SCRs) (8,10,13–17), fear-potentiated startle, and UCS expectancy (7,17–19). However, CIs have little effect on affective learning, such as valence ratings (11), and the few studies investigating extinction retrieval have reported inconclusive findings (8,10,17).

Although several studies have investigated the effects of CIs on extinction, translation to clinical populations is scarce, despite the potential clinical-mechanistical benefits that such translation could offer. Patients with ADs and individuals with high trait anxiety show accelerated acquisition and/or slower extinction of conditioned fear responses (20–23). To date, only one study by Duits *et al.* (7) investigated the effectiveness of CIs in patients with ADs. The patients exhibited a stronger reduction in CS-modulated startle responses following CIs compared with healthy control participants. However, this study (7) used a within-subjects study design with extinction training immediately after fear acquisition. In the current study, we aimed to extend these findings by using a between-subjects design to investigate the effects of CIs prior to and/or after extinction training on context-dependent extinction learning and retrieval, i.e., fear renewal and reinstatement.

ISSN: 2451-9022

While CIs administered before extinction training facilitate extinction learning, they do not necessarily improve extinction retrieval (8,10). The extinction context may become associated with the CIs, which could contribute to a renewal effect when tested in a different context (24). Repeating CIs directly before extinction retrieval could help decrease conditioned responses toward the CS+ during extinction retrieval (8). The immediate postlearning period is a critical consolidation window during which memories are especially malleable and can be strengthened through intervention (25). Therefore, delivering CIs directly after extinction training may enhance the consolidation of extinction memories, leading to more robust extinction retrieval.

Building on these findings, in the current study, we systematically investigated the timing of CI administration in both healthy control participants and patients with ADs. We used a 3-day paradigm, consistent with recommendations from Lonsdorf *et al.* (26). To this end, we manipulated CI delivery before and/or after extinction training, allowing us to study their time-dependent effects on fear extinction and retrieval and whether these effects differ in the context of clinical anxiety. Consistent with previous findings (22,23), we hypothesized that patients with ADs would show heightened fear acquisition, as indicated by increased conditioned fear responses toward CS− compared with healthy control participants. Additionally and independent of CIs, we hypothesized that healthy control participants would show better extinction learning and retrieval relative to patients with ADs (22,23). Regarding CIs, we hypothesized that CIs provided prior to extinction training would lead to superior extinction learning in healthy control participants and patients with ADs (8,10,13–17). Lastly, we expected that CIs provided after extinction training would improve extinction retrieval and reduce return of fear by diminishing conditioned fear responses toward CS+ (8) and yield even superior effects in combination with CIs delivered prior to extinction training in healthy control participants and patients with ADs.

METHODS AND MATERIALS

Participants

This study was preregistered at ClinicalTrials.gov (NCT05718271). Based on the existing literature closest to our design and population [e.g., (7,10)], we expected a medium effect size ($\eta_p^2 = 0.2$) of CIs on extinction training and retrieval. A sample size of 200 participants was required to detect a significant interaction comprising 1 between-subjects factor (instruction group) and 2 within-subjects factors (CS type and phase) with $1 - \beta \geq 0.95$ power and an α level of 0.05. A sensitivity analysis using the final sample indicated that, with $1 - \beta \geq 0.95$ power and $\alpha = 0.05$, the design including patient group, instruction group, CS type, and context was able to detect small-to-moderate interaction effects of at least $f = 0.116$ ($\eta_p^2 \approx 0.013$).

A total of 242 participants were recruited, comprising 122 healthy control participants and 120 participants diagnosed with an AD. Eligibility was determined via a telephone screening assessing gender, age, smoking habits, and current and past mental health disorders, followed by a short diagnostic interview (Mini Diagnostisches Interview für Psychische Störungen) (27). Inclusion criteria included sufficient German language skills, age between 18 and 65 years, current primary diagnosis of an AD as defined by DSM-5 (for patients), or no diagnosis of any mental disorder (for healthy control participants). Exclusion criteria included current psychotherapeutic treatment, diagnosis of a substance use disorder, current or past diagnosis of a psychotic or bipolar disorder, acute suicidality, personality disorder, attention-deficit/hyperactivity disorder, chronic physical or neurological diseases, previous experience with fear conditioning, and for women, being pregnant or breastfeeding. The final sample included 240 participants. Details on participant exclusions for each analysis are provided in the Supplement.

Participants provided written informed consent before the start of the experiment and received monetary compensation or course credit for their participation. All procedures were in accordance with the Declaration of Helsinki and approved by the ethics committee of the Faculty of Psychology at the Ruhr University Bochum (Approval No. 673).

Stimuli

Stimuli and procedures were adopted from Milad *et al.* (28). The CS were presented in different contexts: a picture of an office room (A), a room with a shelf (B), and a conference room (C). Two lamps shining in either blue or yellow served as CS. Electrical stimulation was applied as the UCS via a constant-voltage stimulator (STM200; BIOPAC Systems, Inc.). Two Ag/AgCl electrodes filled with isotonic (0.05 NaCl) electrolyte medium were attached to the middle of the left shin, and stimulus intensity was set individually to be unpleasant but not painful using a gradually increasing rating procedure.

Fear Conditioning Paradigm

The study was conducted in a 3-day fear conditioning paradigm (overview depicted in Figure 1). Participants completed ratings (UCS expectancy and CS valence) at several time points throughout the experiment and were connected to the shock electrodes on all 3 days.

On day 1, participants first completed the Intolerance of Uncertainty (IU) Scale (29), the General Self-Efficacy (GSE) Scale (30), and the trait version of the State-Trait Anxiety Inventory (31) before the conditioning procedures began. After completing the questionnaires, all participants were instructed that at the end of the presentation of a light color, they may or may not receive electrical stimulation. During habituation, the CS were presented twice without a UCS pairing. Immediately afterward, fear acquisition training (in context A) started, in which one stimulus (CS+) was followed by a UCS in 5 of 8 trials, whereas the second stimulus (CS−) was presented 8 times without any UCS pairing (CS allocation counterbalanced). Following fear acquisition training, participants' contingency awareness was checked by letting them choose which CS was paired with the UCS and which was not [similar to the contingency test in Mertens *et al.* (32)].

The next day, extinction training was performed in context B. The CS+ and CS− were each presented 8 times while no UCS was applied. Prior to extinction training, the CI-post and no-CI groups were instructed that their task was to follow the presentation of the images carefully, while the CI-pre and CI-pre+post groups were instructed that their task was to follow

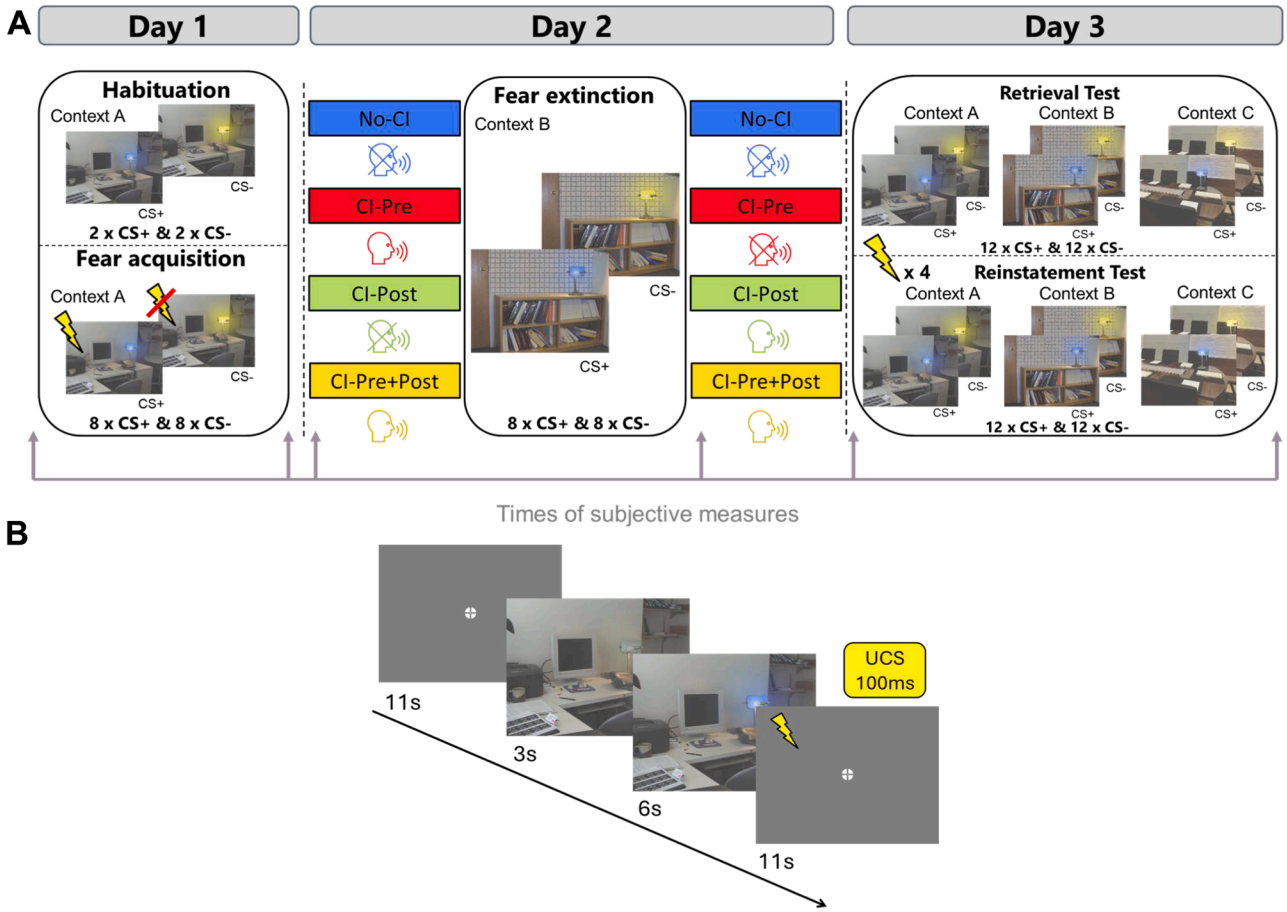

Figure 1. **(A)** Experimental paradigm. The first day started with a habituation phase (in context A) with no electrical stimulation and was followed by fear acquisition training (in context A) in which CS+ was paired with an electrical stimulation 5 of 8 times. The second day included extinction training (in context B) in which no electrical shock was applied. The first group (in blue) received no contingency instructions (CIs) about the absence of the unconditioned stimulus (UCS), the second group (in red) received CIs before extinction training, the third group (in green) received CIs after extinction training, and the fourth group (in yellow) received CIs before and after extinction training. The third day involved an extinction retrieval test in 3 different contexts—A, B, and C; a reinstatement with 4 unannounced electrical stimulations; and a reinstatement test in 3 different contexts—A, B, and C. Participants completed subjective ratings (UCS expectancy and conditioned stimulus [CS] valence) before habituation, after fear acquisition training, before extinction training, after extinction training, before extinction retrieval, and after the reinstatement test. **(B)** Example of a reinforced trial. The trials followed the same structure across all experimental phases. Each trial began with the context being presented for 3 seconds followed by the CS being displayed for 6 seconds in that context; in reinforced trials the UCS was delivered for 100 ms at CS+ offset. The trials were separated by an 11-second fixation cross. Pseudorandomized stimulus orders ensured that no CS appeared more than twice in a row.

the presentation of the images carefully but that there would be no electrical stimulation on that day. After extinction training, all participants were informed that the next session would involve the same task as the previous days. Additionally, the CI-post and CI-pre+post groups were instructed that there would be no electrical stimulation the next day.

The third day included extinction retrieval and reinstatement test. All participants were instructed that their task was to follow the presentation of the images carefully. During extinction retrieval, the CS+ and CS− were each presented 12 times in different contexts (4 presentations in contexts A, B, and C) with no UCS pairing and contexts intermixed in a randomized order. Right after extinction retrieval, 4 electrical stimulations were delivered while participants viewed a black screen. The reinstatement test followed immediately using the same procedure as during extinction retrieval. As a manipulation check, participants from the CI-pre, CI-post, and CI-pre+post groups were asked to indicate whether they had believed the instruction that the electrical stimulation would not occur on the second/third day (yes or no question) [similar to Luck and Lipp (14)].

Conditioning Measures

SCRs were retrieved continuously during all conditioning procedures and were measured with Ag/AgCl electrodes filled with an isotonic (0.05 NaCl) electrolyte gel attached to the hypothenar of the left hand. Data were acquired at 1000 Hz and filtered at 4.5 Hz (MP160 + AMI100D; BIOPAC Systems, Inc.) (software: AcqKnowledge 5.0). Conditioned responses

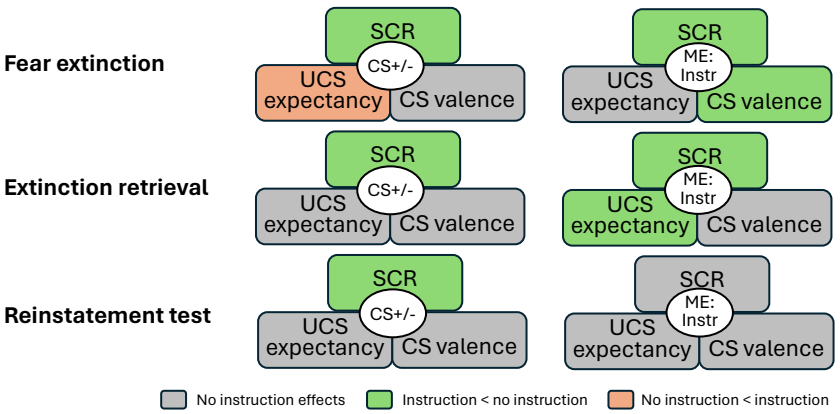

Figure 2. Visualization of results across fear extinction training, extinction retrieval, and reinstatement test. Gray shading indicates no group difference; colored rectangles represent significant instruction (Instr) effects, either as a main effect (ME) or as an interaction effect with conditioned stimulus (CS) type (CS+/CS−). Green denotes lower conditioned responses or smaller CS differentiation in the instructed groups; red denotes lower conditioned responses or greater CS differentiation in the uninstructed groups. SCR, skin conductance response; UCS, unconditioned stimulus.

were analyzed using the MATLAB (version 2023b; The MathWorks, Inc.) EDA-Analysis App (33) and specified as trough-to-peak maximum amplitudes in a time window of 1–6.5 seconds after CS onset. Analyses were conducted with natural log-transformed values (log [1 + SCR in μS]).

CS valence ratings ("How pleasant do you feel when you see this picture?") and UCS expectancy ratings ("Do you think this picture is paired with an electrical stimulation?") were both obtained using visual analog scales presented on the computer screen, without a preceding trial. The scales ranged from 0 (very pleasant/extremely unlikely) to 100 (very unpleasant/extremely likely).

Statistical Analyses

All statistical analyses were carried out in RStudio (34). SCR data were averaged into blocks per phase: habituation (2 trials per CS), acquisition and extinction (block 1: trials 1–4; block 2: trials 5–8), and extinction retrieval and reinstatement test (all 4 trials per context). Fear acquisition was analyzed with mixed analyses of variance (ANOVAs) including patient group (AD vs. healthy control) and instruction group (CI-pre vs. CI-pre+post vs. CI-post vs. no CI) as between-subjects factors and phase (habituation, acquisition blocks 1 and 2) and CS type (CS− vs. CS+) as within-subjects factors. The effects of pre-extinction instructions on extinction were examined using mixed ANOVAs with pooled instruction groups (CI-pre and CI-pre+post vs. CI-post and no CI) and patient group as between-subjects factors and CS type and phase (extinction blocks 1 and 2) as within-subjects factors. For extinction retrieval and the reinstatement test, separate mixed ANOVAs included instruction group and patient group as between-subjects factors and CS type and context (A vs. B vs. C) as within-subjects factors.

Significant ANOVA results were followed by Tukey-adjusted post hoc tests. η_p^2 was calculated as an effect size, and Greenhouse-Geisser correction was applied when sphericity was violated.

RESULTS

An overview of the results concerning the effects of CIs is illustrated in Figure 2. Overall effects of the fear conditioning task (confirming that the task elicited the expected conditioned responses), manipulation check, additional analyses (including only participants who believed the instructions) and detailed results of the subjective ratings are provided in the Supplement.

Sample Characteristics

Patients with ADs and healthy control participants did not differ in age or gender distribution but showed significant differences in IU, trait anxiety, and GSE (see Table 1). No differences emerged across instruction groups.

Skin Conductance Responses

Fear Acquisition. No differences were found between patients with ADs and healthy control participants ($F_{1,232} = 1.79$, $p = .183$, $\eta_p^2 = 0.01$; all patient group interactions: $ps \geq .066$) (see Figure 3), and no significant effects of random allocation to instruction group were evident ($F_{1,232} = 1.42$, $p = .238$, $\eta_p^2 = 0.02$; all interactions: $ps \geq .066$).

Fear Extinction. During extinction training, participants who received instructions prior to the session exhibited lower overall SCRs compared with uninstructed participants (instruction group × block interaction: $F_{1,235} = 29.75$, $p < .001$, $\eta_p^2 = 0.11$) (see Figure 4), with the effect being most pronounced during the first half of extinction (block 1: $t_{235} = 6.27$, $p < .001$; block 2: $t_{235} = 3.76$, $p < .001$). However, differential responding toward CS was not affected by instructions (instruction group × CS type interaction: $F_{1,235} = 2.94$, $p = .088$, $\eta_p^2 = 0.01$).

Instructed healthy control participants exhibited lower SCRs than instructed patients during the first half of extinction training (instruction group × patient group × block interaction: $F_{1,235} = 5.98$, $p = .015$, $\eta_p^2 = 0.02$; block 1: $t_{235} = -2.44$, $p = .016$), while no differences were found during the second half ($t_{235} = -0.76$, $p = .451$). Among uninstructed participants, no differences were found between patients and healthy control participants (block 1: $t_{235} = -0.77$, $p = .441$; block 2: $t_{235} = -1.18$, $p = .241$).

Patients exhibited larger SCRs toward CS+ compared with the healthy control participants during extinction training

Table 1. Descriptive Statistics for Patients With ADs and Healthy Control Participants

Measure	Patients With ADs, n = 120	Healthy Control Participants, n = 120
Age, Years	25.19 (7.03)	25.28 (6.05)
Gender, Female	75.83%	67.50%
IU	56.27 (13.12)***	39.21 (11.77)***
TA	89.00 (22.08)***	57.14 (21.18)***
GSE	26.33 (4.52)***	31.49 (4.03)***

Values are presented as mean (SD) or %. Sample characteristics (age and questionnaire scores) were compared using analyses of variance with patient group (AD vs. healthy control) and instruction group as between-subjects factors. Gender ratios were tested with a Cochran-Mantel-Haenszel test.

***Significant differences between patients with ADs and healthy control participants, $p < .001$.

AD, anxiety disorder; GSE, General Self-Efficacy; IU, Intolerance of Uncertainty; TA, trait anxiety.

($t_{235} = -2.25$, $p = .025$), while the SCRs toward the CS− were comparable ($t_{235} = -1.25$, $p = .212$; patient group × CS type interaction: $F_{1,235} = 4.61$, $p = .033$, $\eta_p^2 = 0.02$).

A trial-by-trial analysis of SCRs during extinction training revealed an interaction between instruction group, trial, and CS type (see the Supplement). In the uninstructed groups, CS differentiation persisted across most trials and disappeared only in the last trial, whereas in the instructed groups, it was limited to the early trials. These findings suggest improved extinction learning in the instructed participants.

Extinction Retrieval. The CI-pre+post group exhibited lower SCRs at extinction retrieval compared with the no-CI ($t_{231} = 2.95$, $p = .018$) and CI-pre ($t_{231} = 3.50$, $p = .003$) groups (group main effect: $F_{3,231} = 5.28$, $p = .002$, $\eta_p^2 = 0.06$) (see Figure 5). No differences were found between the other instruction groups (all ps ≥ .078).

A significant instruction group × CS type interaction ($F_{3,231} = 4.78$, $p = .003$, $\eta_p^2 = 0.06$) revealed that the CI-post and CI-pre+post groups had no significant CS differentiation (CI-post: $t_{231} = 1.33$, $p = .185$; CI-pre+post: $t_{231} = 1.38$, $p = .169$), while the no-CI and CI-pre groups showed significantly larger SCRs for CS+ compared with CS− (no CI: $t_{231} = 4.84$, $p < .001$; CI-pre: $t_{231} = 5.42$, $p < .001$).

Regarding context effects, larger SCRs were more evident in context A compared with contexts B ($t_{231} = 3.61$, $p = .001$) and C ($t_{231} = 2.82$, $p = .014$; context main effect: $F_{1.79,412.76} = 8.86$, $p < .001$, $\eta_p^2 = 0.04$), with no difference between context B and C ($t_{231} = -1.57$, $p = .260$).

Patients did not differ significantly from healthy control participants (patient group main effect: $F_{1,231} = 0.94$, $p = .333$, $\eta_p^2 < 0.01$), and there were no significant interactions with the patient group (all ps ≥ .139).

Reinstatement Test. During the reinstatement test, a 3-way interaction between instruction group, context, and CS type was detected ($F_{5.99,457.15} = 2.44$, $p = .025$, $\eta_p^2 = 0.03$) (see Figure 6). The no-CI group showed larger SCRs toward CS+ compared with CS− in context A ($t_{229} = 3.22$, $p = .002$), while no CS differentiation was found in contexts B ($t_{229} = 0.82$, $p = .412$) and C ($t_{229} = 0.45$, $p = .650$). In contrast, the CI-pre group exhibited an opposite pattern, with larger SCRs toward the CS+ than the CS− in contexts B ($t_{229} = 3.20$, $p = .002$) and C ($t_{229} = 3.15$, $p = .002$) and no CS differentiation in context A ($t_{229} = 1.37$, $p = .172$).

The CI-post group showed larger SCRs for CS+ compared with CS− in context B ($t_{229} = 2.34$, $p = .020$) and no CS differentiation in context A ($t_{229} = 0.78$, $p = .435$) or C ($t_{229} = 0.38$, $p = .706$). Lastly, the CI-pre+post group showed a pattern similar to the one shown by the no-CI group, with larger SCRs for CS+ than CS− in context A ($t_{229} = 3.34$, $p = .001$) and no CS differentiation in context B ($t_{229} = 0.34$, $p = .733$) or C ($t_{229} = -0.12$, $p = .905$).

Regarding context effects, larger SCRs were evident in context A compared with context C ($t_{229} = 2.65$, $p = .024$; context main effect: $F_{1.94,443.33} = 3.87$, $p = .023$, $\eta_p^2 = 0.02$), while there were no differences between contexts A and B ($t_{229} = 1.46$, $p = .312$) or contexts B and C ($t_{229} = 1.39$, $p = .346$).

Patients did not differ significantly from healthy control participants (patient group main effect: $F_{1,229} = 0.23$, $p = .635$,

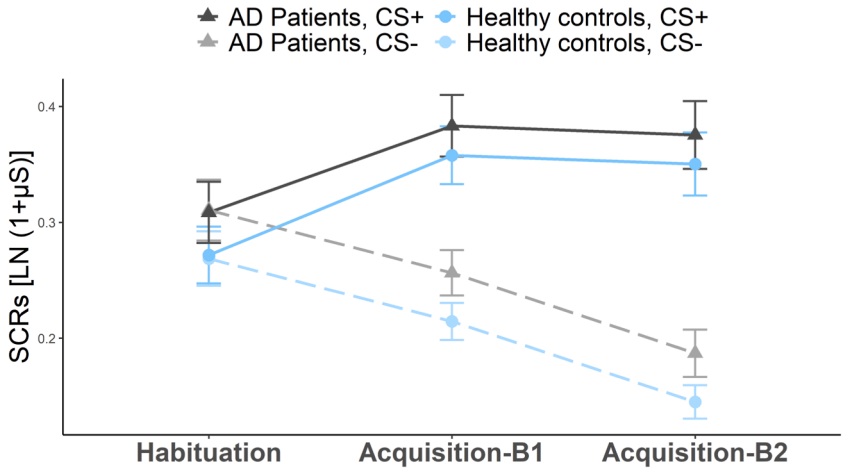

Figure 3. Mean (± SEM) skin conductance responses (SCRs) for CS+ and CS− over the course of habituation (trials 1 and 2) and fear acquisition (block 1: trials 1–4 and block 2: trials 5–8) separately for patients with anxiety disorders (ADs) and healthy control participants. CS, conditioned stimulus.

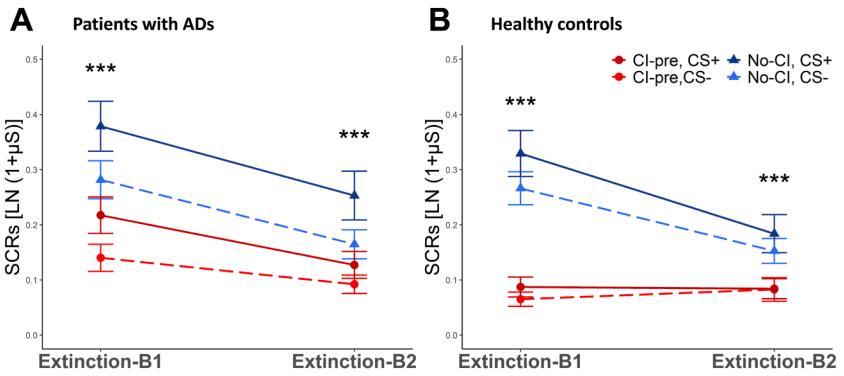

Figure 4. Mean (± SEM) skin conductance responses (SCRs) for CS+ and CS− over the course of extinction (block 1: trials 1–4 and block 2: trials 5–8), depicted separately for patients with anxiety disorders (ADs) **(A)** and healthy control participants **(B)**. Groups receiving contingency instructions (CIs) before extinction training (CI-pre: CI-pre and CI-pre+post in red) and groups receiving no contingency instructions before extinction training (no CI: CI-post and no CI in blue) are contrasted. Significant Tukey-adjusted pairwise instruction group comparisons from the instruction group × block interaction are indicated with ***$p < .001$. CS, conditioned stimulus.

$\eta_p^2 < 0.01$), and no significant interactions with patient group were found (all $ps \geq .095$).

UCS Expectancy

Patients consistently showed higher UCS expectancy than healthy control participants. During extinction training, instructed patients showed elevated UCS expectancy for CS+ compared with uninstructed patients, and at retrieval, the CI-pre+post group exhibited lower UCS expectancy than the CI-pre group. No significant CI effects were observed after the reinstatement test. Further details can be found in the Supplement.

CS Valence Ratings

Across fear acquisition and extinction retrieval, no significant CI effects or patient group differences were observed. During extinction training, instructed participants showed a stronger reduction in negative CS valence than uninstructed participants, although instructions did not affect differential responding. After the reinstatement test, patients rated CS− more negatively than healthy control participants. Further details can be found in the Supplement.

DISCUSSION

In the current study, we investigated the effects of the timing of CI administration on extinction and extinction retrieval in patients with ADs and healthy control participants. During fear acquisition, UCS expectancy (as a purely cognitive measure) for CS− was higher in patients compared with healthy control participants, while SCRs and CS valence did not differ. This is consistent with previous reports of impaired safety learning in ADs (35,36), as well as a recent meta-analysis (23) showing elevated UCS expectancy and affect ratings for CS− but no significant differences in SCRs during fear acquisition in patients with anxiety- or stress-related disorders compared with healthy participants. During extinction training, patients with ADs exhibited larger SCRs toward CS+, higher UCS expectancy for both CS, and no differences in CS valence. The same meta-analysis by Kausche *et al.* (23) revealed somewhat different findings. Patients demonstrated increased affect ratings toward both CS and higher UCS expectancy for CS− compared with healthy control participants, while SCRs did not differ (23). Concerning extinction retrieval in the current study, patients showed higher UCS expectancy compared with healthy control participants and no group differences in SCRs or CS valence, which is consistent with impaired extinction retrieval in patients with posttraumatic stress disorder (37). Elevated conditioned responses by patients with ADs may reflect increased generalization of fear responses to safety cues (36,38) or impaired safety learning (39). These differences may underlie the development and persistence of ADs.

Concerning the effects of CIs, patients and healthy control participants who received CIs prior to extinction training

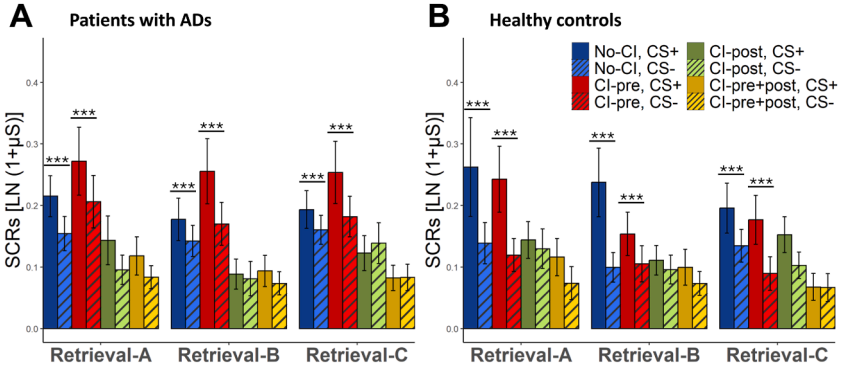

Figure 5. Mean (± SEM) skin conductance responses (SCRs) for CS+ and CS− during extinction retrieval separately for contexts A, B, and C. SCRs are depicted separately for patients with anxiety disorders (ADs) **(A)** and healthy control participants **(B)**. No contingency instructions (CIs) indicates no CIs before extinction training; CI-pre indicates CIs before extinction training; CI-post indicates CIs after extinction training; CI-pre+post indicates CIs before and after extinction training. Significant Tukey-adjusted pairwise conditioned stimulus (CS) type comparisons from the instruction group × CS type interaction effect are indicated with ***$p <$ .001.

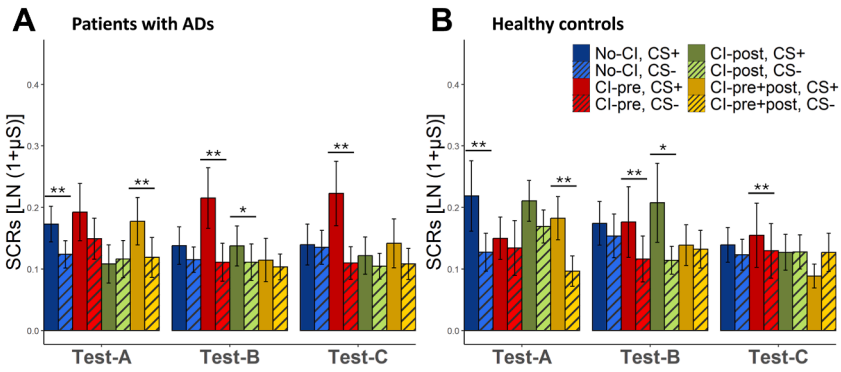

Figure 6. Mean (± SEM) skin conductance responses (SCRs) for CS+ and CS− during the reinstatement test separately for contexts A, B, and C. SCRs are depicted separately for patients with anxiety disorders (ADs) **(A)** and healthy control participants **(B)**. No contingency instructions (CIs) indicates no CIs before extinction training; CI-pre indicates CIs before extinction training; CI-post indicates CIs after extinction training; CI-pre+post indicates CIs before and after extinction training. Significant Tukey-adjusted pairwise comparisons from the instruction group × context × conditioned stimulus (CS) type interaction are indicated with $^{*}p < .05$ and $^{**}p < .01$.

exhibited reduced SCRs and more positive CS valence. However, these effects were not accompanied by greater reductions in UCS expectancy. Furthermore, a trial-by-trial analysis showed that CIs affected CS differentiation, but only during the second half of extinction training, unlike previous studies reporting immediate instruction effects [e.g., (10)], possibly due to the multiday design or the inclusion of patients. These findings suggest that CIs primarily modulated physiological arousal and affective responses, which is partially consistent with prior research highlighting the beneficial effects of CIs in improving extinction learning [e.g., (7–10)]. While CIs are assumed to primarily alter the predictive value of the CS rather than its affective value (11), the current findings show increased positive CS valence. Furthermore, the study extends these results to patients with ADs: While both groups benefited from CIs, healthy control participants showed a steeper reduction of SCRs through CIs than patients during the first half of extinction training. Interestingly, Duits *et al.* (7) reported that CIs led to greater reductions in startle responses among patients with ADs compared with healthy control participants.

Regarding extinction retrieval, participants who received CIs after extinction training, especially in combination with CIs before training, showed altered physiological arousal (absent CS differentiation in SCRs) and reduced threat expectancy. No group differences were found for CS valence. These findings are consistent with studies showing that repeating instructions before extinction retrieval improves retrieval and reduces relapse of fear (8). In the current study, providing CIs only prior to extinction training did not reduce CS differentiation, and conditioning measures were comparable to those of the no-CI group. Similarly, disadvantageous effects of CIs on extinction retrieval after disgust conditioning have been reported, indicated by higher UCS expectancy for CS+ (40). CIs given only before extinction training may not be sufficiently transferred to extinction retrieval. That is, in the absence of a reminder or safety signal before extinction retrieval, participants may revert to expecting a threat. Exploring such mechanisms more systematically, especially across different emotions relevant to anxiety disorders (e.g., disgust), could help clarify conditions under which CIs enhance extinction retrieval (41).

Further analyses of the reinstatement test revealed distinct SCR patterns across instruction groups and contexts. CS differentiation varied by context and instruction group, while no CI effects were observed in UCS expectancy or CS valence. Fear reinstatement was most pronounced in the original acquisition context for the no-CI and CI-pre+post groups, indicating that extinction learning did not sufficiently overwrite the initial fear memory. This is consistent with theoretical models of contextual renewal of fear (24), suggesting that extinction learning is often context dependent, making the return of fear more likely when individuals reencounter the original fear context (42). In contrast, the CI-pre group exhibited a reversed pattern, with fear reinstatement occurring in the extinction and novel context rather than the original acquisition context. The CI-post group demonstrated reduced reinstatement in both the original acquisition and the novel context, suggesting a generalized extinction memory. These findings support the idea that postextinction safety information can enhance fear-inhibitory learning (2) and may reduce fear relapse, particularly by enhancing the transfer of safety learning beyond the original acquisition context.

The current findings highlight the importance of CI timing in modulating fear extinction and its retrieval. While pre-extinction CIs facilitated extinction learning (reflected in greater reductions of physiological arousal and affective responses), this benefit did not translate into better extinction retrieval or reduced fear reinstatement. However, CIs administered after extinction training appeared to enhance extinction memory consolidation and mitigate relapse during the reinstatement test. These effects may be explained by differential engagement of memory systems: Postextinction CIs may support consolidation by reinforcing the nonthreat value of the CS+, thereby preventing return of fear after reinstatement (43,44). In contrast, pre-extinction CI may reduce prediction error and interfere with the full acquisition of extinction learning (2). Furthermore, postextinction CIs may help generalize extinction learning beyond the context in which it was acquired and thereby reduce context specificity. Context changes are a major challenge in exposure therapy (4,42,45,46). Providing patients with targeted, safety-promoting information following exposure may help consolidate fear-inhibitory learning, enhance extinction generalization, and ultimately reduce the likelihood of fear relapse. This could be achieved by providing corrective psychoeducational

information about UCS aversiveness (e.g., that most spiders in Europe are harmless).

When evaluating these findings, certain methodological aspects warrant attention. Our experimental setting provided high certainty of stimulus outcomes for some participants, whereas others remained uncertain due to the absence of instructions. However, in real-life situations, feared outcomes often carry inherent uncertainty in which complete safety can rarely be guaranteed. While such control is necessary for experimental rigor, it limits ecological validity (47). The complexity of the within-subjects design with 3 retrieval contexts might have limited accurate estimation of true context effects, which should be tested in a large sample with a between-subjects design during retrieval in different contexts. Our findings are still highly relevant because modifying CS-UCS associations is central to behavioral treatments for ADs. While we examined only short-term effects of CIs, extending our findings to long-term outcomes is especially relevant for clinical applications, given the risk of return of fear or spontaneous recovery (4). The focus on short-term effects in this study is a necessary first step and proof of principle for establishing efficacy under controlled conditions, while future studies could focus on evaluating the durability of CI effects. Lastly, while the use of multiple outcome measures is valuable to capture different aspects of fear learning (26), there is an increased risk of obtaining spurious findings that may complicate interpretation of findings. Therefore, future studies could use multivariate analyses to directly compare the differences between measures.

Conclusions

The current study underscores the importance of CI as an effective tool for enhancing fear extinction processes. CIs improved extinction learning in healthy control participants and patients with ADs. When CIs were provided after extinction training, participants showed the strongest extinction retrieval and reinstatement performance. They exhibited the least return of fear in the original fear acquisition and novel context. These findings suggest that CIs strengthen extinction learning and retrieval and may enhance the generalization of safety learning, which is an important goal for improving cognitive behavioral treatment outcomes beyond the therapy setting.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) (Project No. 316803389–SFB 1280; project title: The impact of modification of stimulus-outcome contingencies on extinction and exposure in anxiety disorders). AL and BM are funded via CRC 1280 and CJM, MLW, OTW, JM, and AZ are principal investigators within CRC 1280. MLW is also supported by the Emmy Noether Program of DFG (Project title: Mechanisms of interpretation biases in anxiety – Bridging experimental investigation and clinical translation; WO 2018/3-1; Project No. 442163275). The funders were not involved in the study design; data collection, analysis, or interpretation; writing the article; or deciding to submit the article for publication.

A previous version of this article was published as a preprint on PsyArXiv: https://osf.io/preprints/psyarxiv/kma2e_v4.

The authors report no biomedical financial interests or potential conflicts of interest.

ClinicalTrials.gov: Effects of Verbal Instructions on Fear Extinction and Extinction Retrieval; https://clinicaltrials.gov/study/NCT05718271; NCT05718271.

ARTICLE INFORMATION

From the Department of Behavioral and Clinical Neuroscience, Ruhr University Bochum, Bochum, Germany (AL, AZ); Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany (AL, BM, MLW, JM, AZ); Department of Cognitive Psychology, Ruhr University Bochum, Bochum, Germany (CJM, OTW); and Department of Clinical Psychology and Experimental Psychopathology, and Georg-Elias-Mueller-Institute of Psychology, University of Göttingen, Göttingen, Germany (MLW).

Address correspondence to Annalisa Lipp, M.Sc., at annalisa.lipp@rub.de.

Received Sep 11, 2025; revised Mar 17, 2026; accepted Mar 23, 2026.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsc.2026.03.015.

REFERENCES

1. Hofmann SG (2008): Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. Clin Psychol Rev 28:199–210.
2. Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B (2014): Maximizing exposure therapy: An inhibitory learning approach. Behav Res Ther 58:10–23.
3. Margraf J, Zlomuzica A (2015): Changing the future, not the past: A translational paradigm shift in treating anxiety. EMBO Rep 16:259–260.
4. Vervliet B, Craske MG, Hermans D (2013): Fear extinction and relapse: State of the art. Annu Rev Clin Psychol 9:215–248.
5. Zlomuzica A, Schneider S, Konrad C, Merz CJ, Wolf OT, Raeder F, Margraf J (2020): Clinical implications of fear extinction in anxiety disorders. Neuroforum 26:143–149.
6. Zlomuzica A, Kodzaga I, Piovesan K, Lipp A (2025): Treating anxiety comorbidity: Lessons from exposure generalization studies. Behav Brain Res 481:115409.
7. Duits P, Richter J, Baas JMP, Engelhard IM, Limberg-Thiesen A, Heitland I, *et al.* (2017): Enhancing effects of contingency instructions on fear acquisition and extinction in anxiety disorders. J Abnorm Psychol 126:378–391.
8. Javanbakht A, Duval ER, Cisneros ME, Taylor SF, Kessler D, Liberzon I (2017): Instructed fear learning, extinction, and recall: Additive effects of cognitive information on emotional learning of fear. Cogn Emot 31:980–987.
9. Mertens G, Boddez Y, Sevenster D, Engelhard IM, De Houwer J (2018): A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. Biol Psychol 137:49–64.
10. Wendt J, Hufenbach MC, König J, Hamm AO (2020): Effects of verbal instructions and physical threat removal prior to extinction training on the return of conditioned fear. Sci Rep 10:1202.
11. Luck CC, Lipp OV (2016): Instructed extinction in human fear conditioning: History, recent developments, and future directions. Aust J Psychol 68:209–227.
12. Javanbakht A, Grasser LR, Madaboosi S, Chowdury A, Liberzon I, Diwadkar VA (2021): The neurocircuitry underlying additive effects of safety instruction on extinction learning. Front Behav Neurosci 14:576247.
13. Luck CC, Patterson RR, Lipp OV (2020): "Prepared" fear or sociocultural learning? Fear conditioned to guns, snakes, and spiders is eliminated by instructed extinction in a within-participant differential fear conditioning paradigm. Psychophysiology 57:e13516.
14. Luck CC, Lipp OV (2015): A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. Behav Res Ther 66:18–31.
15. Luck CC, Lipp OV (2015): To remove or not to remove? Removal of the unconditional stimulus electrode does not mediate instructed extinction effects. Psychophysiology 52:1248–1256.

16. Rowles ME, Lipp OV, Mallan KM (2012): On the resistance to extinction of fear conditioned to angry faces. Psychophysiology 49:375–380.
17. Sevenster D, Beckers T, Kindt M (2012): Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. Psychophysiology 49:1426–1435.
18. Morriss J, Van Reekum CM (2019): I feel safe when i know: Contingency instruction promotes threat extinction in high intolerance of uncertainty individuals. Behav Res Ther 116:111–118.
19. Scheveneels S, Boddez Y, De Ceulaer T, Hermans D (2019): Ruining the surprise: The effect of safety information before extinction on return of fear. J Behav Ther Exp Psychiatry 63:73–78.
20. Caulfield MD, McAuley JD, Servatius RJ (2013): Facilitated acquisition of eyeblink conditioning in those vulnerable to anxiety disorders. Front Hum Neurosci 7:348.
21. Mosig C, Merz CJ, Mohr C, Adolph D, Wolf OT, Schneider S, *et al.* (2014): Enhanced discriminative fear learning of phobia-irrelevant stimuli in spider-fearful individuals. Front Behav Neurosci 8:328.
22. Duits P, Cath DC, Lissek S, Hox JJ, Hamm AO, Engelhard IM, *et al.* (2015): Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress Anxiety 32:239–253.
23. Kausche FM, Carsten HP, Sobania KM, Riesel A (2025): Fear and safety learning in anxiety- and stress-related disorders: An updated meta-analysis. Neurosci Biobehav Rev 169:105983.
24. Bouton ME (2002): Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. Biol Psychiatry 52:976–986.
25. Robertson EM (2012): New insights in human memory interference and consolidation. Curr Biol 22:R66–R71.
26. Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, 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. Neurosci Biobehav Rev 77:247–285.
27. Margraf J, Cwik JC, Pflug V, Schneider S (2017): Structured clinical interviews for mental disorders across the lifespan: Psychometric quality and further developments of the DIPS Open Access interviews [[Strukturierte klinische Interviews zur Erfassung psychischer Störungen über die Lebensspanne: Gütekriterien und Weiterentwicklungen der DIPS-Verfahren.]]. Z Klin Psychol Psychother 46:176–186.
28. Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007): Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biol Psychiatry 62:446–454.
29. Gerlach AL, Andor T, Patzelt J (2008): Die Bedeutung von Unsicherheitsintoleranz für die Generalisierte Angststörung Modellüberlegungen und Entwicklung einer Deutschen Version der Unsicherheitsintoleranz-Skala. Z Klin Psychol Psychother 37:190–199.
30. Schwarzer R, Jerusalem M (1999). In: Skalen zur Erfassung von Lehrer-und schülermerkmalen. Dokumentation der Psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen, 144. Berlin: Freie Universität Berlin. Available at: http://userpage.fu-berlin.de/%7Ehealth/self/skalendoku_selbstwirksame_schulen.pdf. Accessed May 5, 2026.
31. Grimm J (2009): State-Trait-Anxiety Inventory Nach Spielberger. Deutsche Lang - Und Kurzversion. Methodenforum Universidad Wien MF Work Pap. Available at: https://empcom.univie.ac.at/fileadmin/user_upload/p_empcom/pdfs/Grimm2009_StateTraitAngst_MFWorkPaper2009-02.pdf. Accessed May 5, 2026.
32. Mertens G, van Dis EAM, Krypotos A-M, Engelhard IM (2021): Does an unconditioned stimulus memory devaluation procedure decrease disgust memories and conditioned disgust? Results of two laboratory studies. J Anxiety Disord 82:102447.
33. Otto T, Wolf OT, Merz CJ (2023): EDA-Analysis App, version 5.13. Zenodo. Available at: https://doi.org/10.5281/zenodo.8407428. Accessed May 5, 2026.
34. Posit team (2022): RStudio: Integrated Development Environment for R, version 2022.12.0.353. Boston, MA: Posit Software, PBC. Available at: http://www.posit.co/. Accessed May 5, 2026.
35. Lau JYF, Lissek S, Nelson EE, Lee Y, Roberson-Nay R, Poeth K, *et al.* (2008): Fear conditioning in adolescents with anxiety disorders: Results from a novel experimental paradigm. J Am Acad Child Adolesc Psychiatry 47:94–102.
36. Lissek S, Kaczkurkin AN, Rabin S, Geraci M, Pine DS, Grillon C (2014): Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. Biol Psychiatry 75:909–915.
37. Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, *et al.* (2009): Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biol Psychiatry 66:1075–1082.
38. Cooper SE, Van Dis EAM, Hagenaars MA, Krypotos A-M, Nemeroff CB, Lissek S, *et al.* (2022): A meta-analysis of conditioned fear generalization in anxiety-related disorders. Neuropsychopharmacology 47:1652–1661.
39. De Kleine RA, Hutschemaekers MHM, Hendriks GJ, Kampman M, Papalini S, Van Minnen A, Vervliet B (2023): Impaired action-safety learning and excessive relief during avoidance in patients with anxiety disorders. J Anxiety Disord 96:102698.
40. Lipp A, Macit B, Woud ML, Dere E, Zlomuzica A (2023): Conscious knowledge of CS-UCS contingency information affects extinction retrieval of conditioned disgust responses: Findings from an online de novo disgust conditioning task. Int J Clin Health Psychol 23:100368.
41. Lipp A, Zlomuzica A (2025): Differences in conditioning using unconditioned stimuli evoking fear, disgust or both emotions simultaneously. Behav Res Ther 192:104805.
42. Kodzaga I, Heistermann J, Zlomuzica A (2025): Generalization of exposure therapy: Systematic review and recommendations for future research. Behav Res Ther 190:104751.
43. Raeder F, Woud ML, Schneider S, Totzeck C, Adolph D, Margraf J, Zlomuzica A (2019): Reactivation and evaluation of mastery experiences promotes exposure benefit in height phobia. Cogn Ther Res 43:948–958.
44. Zbozinek TD, Hermans D, Prenoveau JM, Liao B, Craske MG (2015): Post-extinction conditional stimulus valence predicts reinstatement fear: Relevance for long-term outcomes of exposure therapy. Cogn Emot 29:654–667.
45. Kodzaga I, Dere E, Zlomuzica A (2023): Generalization of beneficial exposure effects to untreated stimuli from another fear category. Transl Psychiatry 13:401.
46. Preusser F, Margraf J, Zlomuzica A (2017): Generalization of extinguished fear to untreated fear stimuli after exposure. Neuropsychopharmacology 42:2545–2552.
47. Zambetti PR, Schuessler BP, Lecamp BE, Shin A, Kim EJ, Kim JJ (2022): Ecological analysis of Pavlovian fear conditioning in rats. Commun Biol 5:830.