Behavioural Brain Research 398 (2021) 112931

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Research report

Thought suppression inhibits the generalization of fear extinction

Augustin C. Hennings [a,b], Sophia A. Bibb [c], Jarrod A. Lewis-Peacock [a,b,c,d,e], Joseph E. Dunsmoor [a,b,e,*]

[a] Institute for Neuroscience, University of Texas at Austin, United States
[b] Center for Learning and Memory, Department of Neuroscience, University of Texas at Austin, United States
[c] Department of Neuroscience, University of Texas at Austin, United States
[d] Department of Psychology, University of Texas at Austin, United States
[e] Department of Psychiatry, Dell Medical School, University of Texas at Austin, United States

ARTICLE INFO

Keywords:
Pavlovian extinction
Directed forgetting
Thought suppression
Protection from extinction

ABSTRACT

A challenge for translating fear extinction research into clinical treatments for stress and anxiety disorders is that extinction learning tends not to generalize beyond the treatment context. This may be because the hippocampus limits the expression of extinction memories. Consequently, downregulating the hippocampus may help to promote the generalization of extinction learning. One nonpharmacological strategy to downregulate hippocampal activity in humans is motivated forgetting, in which a participant deliberately attempts to suppress the encoding and/or retrieval of episodic memories. Here, we evaluated whether this strategy could facilitate extinction generalization by augmenting extinction training with thought suppression. Participants were threat conditioned using two conditioned stimulus (CS) categories paired with an electrical shock. Subsequently, during extinction training, one CS category was accompanied by thought suppression. Participants were tested for extinction generalization 24h later with conceptual variations of the extinguished stimuli. Contrary to our prediction, we found that extinction training paired with thought suppression resulted in enhanced shock expectancy (i.e., worse generalization) relative to standard extinction. We conclude that thought suppression during memory encoding likely acts as an inhibitory cue that blocks the acquisition of extinction memories, and therefore may not be a viable tactic to promote extinction generalization in the treatment of anxiety disorders.

1. Introduction

Learned fear can be stubbornly resistant to change, even after experiences of safety that disconfirm threat expectations. Techniques to maximize safety learning are clinically relevant for optimizing treatment for stress and anxiety disorders [1]. Consequently, there is growing research interest in behavioral and pharmacological strategies that produce persistent safety memories to effectively counteract retrieval and expression of fear memories. This area of research focuses predominately on Pavlovian fear extinction, in which omission of an expected threat diminishes conditioned defensive responses. It is well established that extinction learning produces a new memory but leaves the original conditioned fear memory more or less intact [2]. This secondary extinction memory is transient and often fails to generalize over time, across contexts, and to variations of the feared stimulus not present at the time of extinction training [3]. A number of clever behavioral and pharmacological strategies have been developed to help compensate for the limited nature of extinction training and to promote the generalization of extinction memories [1,4–6]. These strategies are informed by an increased understanding of the neurobehavioral mechanisms of fear and extinction. Here, we investigated a novel approach to modulate long-term extinction memories in humans through top-down suppression of memory encoding processes at the time of extinction training. Our goal was to determine whether a strategy of thought suppression, which putatively downregulates the hippocampus via top-down inhibition from the prefrontal cortex [7], modulates extinction memory formation and enhances extinction memory generalization across time and to conceptual variations of learned threat.

The specificity of extinction memory serves an adaptive function. That is, from an evolutionary standpoint it is better to mistakenly regard harmless stimuli as dangerous than mistakenly treat harmful stimuli as safe, a phenomenon known as anxiety conservation. This “better safe than sorry” approach to threat is supported by mechanisms in the brain that tie extinction memories to the spatiotemporal details present at the

* Corresponding author at: Institute for Neuroscience, University of Texas at Austin, United States.
E-mail address: joseph.dunsmoor@austin.utexas.edu (J.E. Dunsmoor).

https://doi.org/10.1016/j.bbr.2020.112931
Received 16 June 2020; Received in revised form 4 September 2020; Accepted 21 September 2020
Available online 11 October 2020

time of extinction learning, a process largely governed by the hippocampus [3,8]. Emerging evidence from rodent neurophysiology shows that projections from the ventral hippocampus to the medial prefrontal cortex promotes fear expression to an extinguished conditioned stimulus (CS) that is encountered outside the extinction context [9]. Consequently, one possible approach to render extinction more generalizable after learning is to temporarily downregulate hippocampal function to release the specificity of extinction training. Neurobehavioral research in rodents indicates that inactivation of the hippocampus prevents the renewal of extinguished fear when animals are tested outside the extinction context [10]. Administration of the muscarinic cholinergic antagonist scopolamine, which impairs contextual processing in the hippocampus, also promotes extinction generalization across contexts in rodents [11]. However, a recent study translating this pharmacological approach to humans produced mixed results and did not find strong effects of scopolamine on preventing contextual renewal [12].

Another approach to downregulating hippocampal activity is through deliberate attempts to prevent the formation or retrieval of specific memories. One non-pharmacological strategy involves direct suppression of explicit memory encoding processes [13]. Human neuroimaging research shows that instructions to suppress one's thoughts while presented with target memoranda diminishes hippocampal activity through a mechanism of top-down inhibition mediated by the lateral prefrontal cortex [7]. Whether a behavioral strategy that putatively downregulates hippocampal function could serve to release the specificity of extinction training (akin to a temporary lesion or pharmacological inactivation in rodents) is unknown. We modified a multi-day Pavlovian threat conditioning and extinction design by adding an instruction for participants to suppress their thoughts at the moment of extinction memory formation. We tested the effects of thought suppression on extinction retrieval 24h later to variations of the extinguished stimuli as a test of extinction generalization.

There are competing hypotheses for the effect of instructed memory suppression on extinction learning. One hypothesis is in line with rodent studies showing that diminished hippocampal processing releases extinction from specificity, thereby promoting extinction memory generalization [10,11]. If so, we would predict a diminished return of threat when extinction is accompanied by thought suppression as compared to standard extinction. Such a result would be clinically relevant and might suggest a possible enhancement to extinction-based therapy for fear and anxiety.

An alternative hypothesis is that pairing thought suppression with extinction training might serve the same function as an adding an additional inhibitory stimulus that interferes with extinction to the target CS. Such effects are sometimes referred to as *protection from extinction* [14–16] and would be anticipated based on prominent error-correcting associative learning models [17]. That is, learning models of extinction propose that extinction is determined by the surprising omission of the expected unconditioned stimulus (US). This drives the prediction error that diminishes the CS—US association, resulting in a decrease in the conditioned response. If, however, the CS is accompanied by an added stimulus, then the absence of the US is less surprising or could be attributed to the presence of the added stimulus (i.e., conditioned inhibition). If so, then we would predict an *enhanced* return of fear for a CS extinguished during thought suppression as compared to a typical extinction procedure. Such evidence would likewise be informative for extinction-based therapies, as it would support treatments focused on engaging with feared stimuli or situations in order to maximize the discrepancy between predicted and actual outcomes [18].

2. Materials and methods

The goal of the present study was to investigate for the first time whether deliberate thought suppression during extinction training promotes generalization of safety learning. We developed a novel within-subjects Pavlovian threat conditioning task with two CS categories (CS+'s) associated with a mildly aversive shock to the wrist (US), and an unpaired control stimulus (CS-) (Fig. 1A). Following threat acquisition, one CS+ category was presented alone (standard extinction) whereas the other CS+ category was accompanied by an instruction for the subject to momentarily suppress their thoughts (CS+S). Participants returned 24 h later and were presented with novel variations of all three CS categories in a test of extinction generalization. Threat acquisition, extinction, and renewal were measured using trial-by-trial ratings of shock expectancy with a continuous rating bar, as well as skin conductance responses (SCR). We used a category threat conditioning design (see [19]), in which the two CS+'s and CS- were comprised of trial unique (i.e., non-repeating) exemplars from three distinct superordinate semantic categories: animals, tools, and food. This category-conditioning design allowed us to investigate whether extinction to specific exemplars generalizes to novel conceptual variations from the extinguished categories.

2.1. Participants

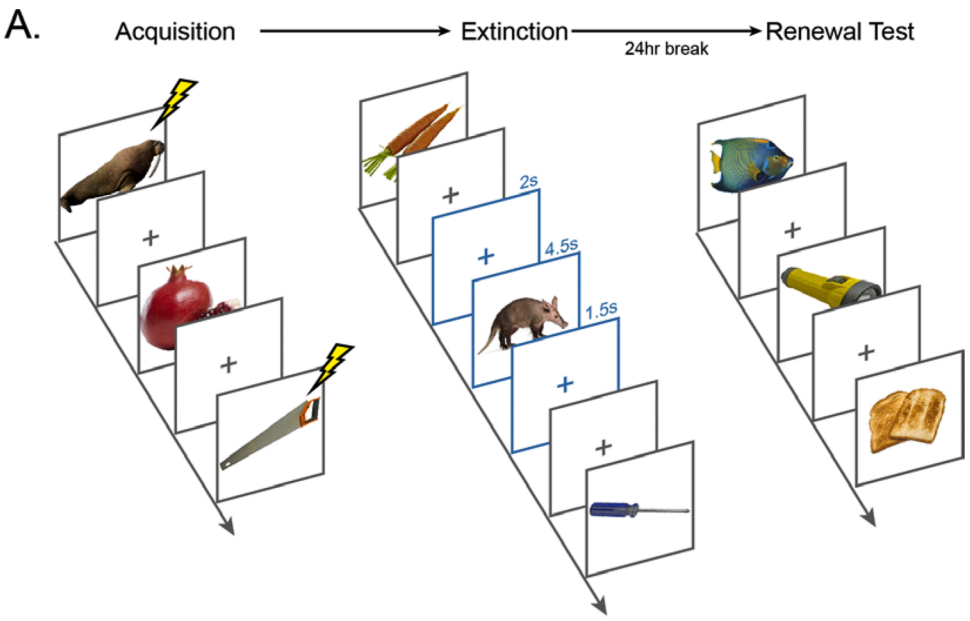
A total of 21 participants were recruited for Experiment 1 (12 Female), and 21 were recruited for Experiment 2 (14 Female). Participants were recruited under the requirements that they be between the ages of 18–45 be able to speak and read English fluently, and self-report no lifetime history of any neurological or psychiatric disorder, as well as not currently taking any psychoactive medication. As detailed below, Experiment 1 and Experiment 2 were nearly identical, with the only exception being that participants in Experiment 1 received a set of instructions regarding a cue to suppress their thoughts during extinction training (Fig. 1B), while participants in Experiment 2 received the same cue but without any prior instructions. All participants provided written informed consent prior to beginning the experiment and were compensated at the rate of $20/hour. All procedures were in compliance with the Institutional Review Board of the University of Texas at Austin (IRB # 2017−02-0094).

2.2. Experimental procedure

The experiment was conducted across two sessions separated by a 24h break. Experimental stimuli consisted of trial unique exemplars from three semantic categories, animals, tools, and food. In this category conditioning procedure, no single basic-level exemplar is repeated, such that participants form an association to an entire semantic category (Fig. 1**A**) (see [19]). For example, there were not pictures of different dogs during threat acquisition, extinction, or the renewal test. The animal and tool categories served as the CS+'s, and the food category always served as the CS-. Common phobic stimuli were excluded from the animals and tools, and food images were filtered to exclude distinctly appetizing images. Stimuli were displayed for 4.5s, and the intertrial interval was randomized to 7.5 or 9.5s. Trial order was pseudorandomized such that no more than 2 of the same CS type were presented consecutively. All participants received the same CS trial order, but the specific basic-level exemplar on each CS trial was randomized for each participant. The first trial of the renewal test on Day 2 was always a CS- to account for initial orienting responses (see also [20,21]). The experiment was built and displayed using PsychoPy in Python [22]. The unconditioned stimulus (US) was a 50-millisecond electric shock delivered to the right wrist using a BIOPAC (Goleta, CA) STM200 module. The shock was calibrated at the beginning of the first session to be "highly annoying and unpleasant, but not painful" in accordance with the IRB and prior research in our laboratory (e.g., [23]).

2.2.1. Threat expectancy and psychophysiology

Throughout the experiment a continuous expectancy bar was used to measure explicit threat expectancy. At the beginning of Day 1, participants were given instructions on how to use the continuous expectancy



B.

Please pay attention to the color of both the fixation cross (+) and the border around the images.
Sometimes right before, during, and briefly after a picture is presented, the color will change to BLUE.
On these trials you are to try to clear the corresponding image from your mind.
Do not think about other images, instead focus on mentally suppressing the current image.
Please do not close your eyes or look away, but try to stop yourself from thinking about the image.

Please still respond whether or not you expect a shock!

Fig. 1. Experimental Procedure. **A. Multi-day associative threat learning and extinction paradigm**. During threat acquisition, two CS + categories (animals and tools) co-terminated with a mild electric shock. Food stimuli served as the CS- and were never paired with shock. Extinction learning followed acquisition. During extinction, one CS + category (animals or tools, counterbalanced across participants) was presented with the suppress cue on every trial (CS + S). The suppress cue was indicated by the color of the image border and fixation cross changing from grey to blue. 24 h later, shock electrodes were re-attached and participants underwent a renewal test using novel images from each CS category. **B. Thought suppression instruction**. The instructions were designed to encourage direct thought suppression, and to discourage thought substitution. These instructions were only given to participants in Experiment 1; participants in Experiment 2 received the cue on CS + S trials during extinction, but they did not receive any instructions regarding the cue.

bar and completed three practice trials with dummy stimuli (colored squares). Participants were told that they would complete a learning experiment, and that their expectancy ratings would not influence the shock contingencies. The right thumb stick of a Logitech F310 gamepad (www.logitech.com) was used to control the expectancy bar in order to answer the question "Do you think that the picture you see will shock you?". The bar displayed a range from 100% No to 100% Yes and was reset to the middle (0%) at the beginning of each trial. During an intertrial interval the previous response was cleared, and the bar could not be moved. For analysis, the range of expectancy ratings were scaled to -1 (100% No) to 1 (100 % Yes).

In addition to explicit threat expectancy, SCR were measured to gauge autonomic arousal. SCR electrodes were placed on the hypothenar eminence of the left palm. Over the course of data collection, two mirror image testing rooms were used, and SCR was measured with either a BIOPAC MP150 or MP160 module (Goleta, CA). Participants always completed both sessions of the experiment in the same room. SCR were scored using previously validated criteria [20]. Specifically, SCR were considered related to a CS if the trough-to-peak deflection occurred within a set time window that extended from 0.5s following CS onset to CS offset (4.5 + 0.5 seconds), lasted between 0.5 and 5.0s, and was greater than 0.02μS. If SCR did not fit these criteria it was scored as zero. SCR were scored using the *Autonomate* script for MATLAB (The MathWorks) [24]. SCR values were square root normalized prior to analysis.

2.2.2. Associative threat learning task

Forty-eight stimuli per category were used on Day 1, and an additional 12 stimuli per category were used on Day 2. In order to reduce the influence of event boundaries on dependent measures of fear learning [25], threat acquisition and extinction were split into two runs of equal lengths for a total of 4 runs on Day 1. There was a brief pause between each run that lasted less than one minute. For analysis, these runs are considered early and late acquisition and extinction, respectively. Each run consisted of 36 trials, 12 of each CS type. Threat acquisition occurred over the first two runs, during which 66% of CS+ images, both animals and tools, co-terminated with the US (32 total shocks). Partial CS+US reinforcement (66%) was used to delay extinction [26,27]. Extinction learning immediately followed acquisition, during which US pairings were omitted for both CS+ categories. During extinction learning one CS+ category was presented with the suppress cue 100% of the time. The suppressed CS+ category was counterbalanced across participants and is referred to as the "CS+S". See section **2.3 Thought suppression** for more information on this manipulation. The next day, participants had both sets of electrodes re-attached (shock and SCR) and were told that they would continue the task as before. During this test there were no US presentations or thought suppression cues, and participants were presented with novel, threat-ambiguous stimuli. For these reasons, we consider this a renewal test, as opposed to a test of spontaneous recovery. Consistent with previous literature investigating extinction retrieval, only the early renewal test (4 trials/CS type; 12 trials total) was considered for analysis [20,28]. These early trials are most likely to reflect retrieval of the threat or extinction memory, whereas the later trials are more likely reflect further extinction learning from the renewal test itself.

2.2.3. Recognition memory test

Following the renewal test on Day 2, the electrodes were removed and all participants completed a surprise recognition memory test for stimuli seen on Day 1 (threat acquisition and extinction learning). This phase of the experiment was self-paced. Along with all Day 1 stimuli, 32 novel foils per category were used in the recognition memory test (total of 80 images per category, 40% novel lures). Participants responded on each trial whether the image was "definitely old", "maybe old", "maybe

new", or "definitely new".

2.2.4. Surveys

Following extinction learning on Day 1, participants completed a brief survey designed to evaluate the success of suppression in Experiment 1. Participants in were asked to rate on a scale of 1–10 how annoying or unpleasant they found the shock to be throughout the experiment. They were also asked to estimate the total number of shocks they received (excluding calibration) and if they could state what the "rule" was for the picture-shock pairings. Only participants in Experiment 1 were asked to rate, from 1 to 10, how well they were able to suppress the pictures when cued to during the experiment. Average suppression success rating was 5.86 (s.e.m. 0.44). In addition, they were asked to report the strategy, if any, that they used to suppress pictures during the experiment, as well as what they thought the "rule" was for which pictures they were instructed to suppress. After the recognition memory test on Day 2, all participants completed the Intolerance of Uncertainty (IUS) and the State-trait anxiety inventory (STAI) surveys.

2.3. Thought suppression

At the beginning of Day 1, participants in Experiment 1 were given additional instructions in order to facilitate thought suppression of the CS+S category during extinction. The full instructions are shown in Fig. 1**B**. These instructions were crafted in order to maximize top-down functional suppression of the hippocampus as shown in neuroimaging experiments, and avoid thought substitution, which engages a different functional mechanism for thought removal [7,29]. In order to cue suppression, the color of both the border around each image, as well as the fixation cross immediately preceding and following the picture, changed from grey to blue. On suppress trials, the fixation cross changed to blue 2s before the CS+S picture, the color border remained blue during the CS+S trial, and then the fixation cross remained blue for 1.5s after stimulus offset before changing back to grey. The average time of the intertrial intervals between stimuli did not change (7.5 or 9.5s). The total length of the suppress cue (8s) was selected in order to maximize thought suppression during extinction learning. Participants in Experiment 1 were given three additional practice trials with dummy stimuli at the start of the experiment to familiarize themselves with the thought suppression manipulation. These participants were instructed at the start of the task that the suppress cue could appear at any point during the experiment. For Experiment 2, participants were not given any information about the suppress cue and did not receive suppression practice trials. However, the cue was still presented with all CS+S trials during extinction learning.

2.4. Statistical analysis

Threat expectancy, SCR, and recognition memory for different phases of the experiment were first assessed using repeated measures ANOVAs including within subject factors of phase and condition (CS type). *A priori* comparisons were evaluated with two-tailed paired t-tests. Both threat expectancy and SCR were determined to be non-normally distributed using a Shapiro-Wilk test [30] implemented in the Python package *pingouin* [31]. Accordingly, permutation tests were used to obtain all p-values for these measures. The R package *permuco* [32] was used to permute repeated measures ANOVAs, and permuted difference of means for follow-up comparisons was implemented using custom Python code. The number of permutations was set to 10,000, allowing for a minimum possible p-value of 0.0001. All p-values are reported as two-tailed. For clarity, parametric F and t statistics and corresponding degrees of freedom are reported alongside permutation test p-values. When used, *post-hoc* follow-up comparisons were Bonferroni corrected by multiplying the *post-hoc* p-values by the number of tests.

2.5. Missing data

Due to technical errors, two participants in Experiment 1 are missing SCR from acquisition and extinction, and one subject in Experiment 2 is missing threat expectancy from extinction. Data from these participants are excluded from relevant analyses.

3. Results

3.1. Experiment 1

3.1.1. Acquisition

Threat expectancy results demonstrated successful associative threat learning for the two CS+ categories relative to the CS- (Fig. 2A). There was a main effect of condition $F_{(2,40)} = 58.07$, $P_{perm} = 0.0001$, and main effect of phase (late vs. early acquisition) $F_{(2,\ 40)} = 6.00$, $P_{perm} = 0.007$. There was no significant condition by phase interaction. Planned comparisons during late acquisition show greater expectancy relative to the CS- for both the CS+ ($t_{20} = 6.89$, $P_{perm} = 0.0001$) and CS+S ($t_{20} = 9.76$, $P_{perm} = 0.0001$). There was no difference in expectancy between the CS+ and CS+S ($t_{20} = 1.04$, $P_{perm} = 0.32$).

The same basic pattern of results was observed for SCR (Fig. 2B). There was a main effect of condition ($F_{(2,\ 36)} = 10.91$, $P_{perm} = 0.0006$), but not a significant main effect of phase or interaction of condition by phase. Planned comparisons during late acquisition show greater mean SCR for both the CS+ ($t_{18} = 3.79$, $P_{perm} = 0.0014$) and CS+S ($t_{18} = 3.26$, $P_{perm} = 0.0008$) versus the CS-. There was no difference in SCR between the CS+ and CS+S ($t_{18} = 1.14$, $P_{perm} = 0.27$), indicating equivalent threat learning to both CS+ categories.

3.1.2. Extinction

Extinction followed acquisition, during which one category (CS+S) was accompanied by a cue for participants to suppress their thoughts. Threat expectancy and SCR values confirmed successful extinction. For threat expectancy, there were significant main effects of phase ($F_{(1,\ 20)} = 29.23$, $P_{perm} = 0.0001$) and condition ($F_{(2,\ 40)} = 19.94$, $P_{perm} = 0.0001$), as well as a significant phase by condition interaction ($F_{(2,\ 40)} = 3.62$, $P_{perm} = 0.035$). Planned comparisons during late extinction showed that expectancy for both CS+ and CS+S was still higher than CS- expectancy (CS+ $t_{20} = 3.76$, $P_{perm} = 0.0006$; CS+S $t_{20} = 3.71$, $P_{perm} = 0.0006$). Importantly, post-hoc comparisons of late extinction vs. late acquisition show significantly reduced expectancy for both the CS+ ($t_{20} = -5.75$, Bonferroni corrected $P_{perm} = 0.0002$) and the CS+S ($t_{20} = -5.40$, Bonferroni corrected $P_{perm} = 0.0002$). Notably, there was no difference in threat expectancy between the CS+ and CS+S categories ($t_{20} = 0.78$, $P_{perm} = 0.45$), indicating the thought suppression cue did not influence extinction learning itself.

For SCR, there were significant main effects of both phase ($F_{(1,\ 18)} = 5.92$, $P_{perm} = 0.028$) and condition ($F_{(2,\ 36)} = 3.42$, $P_{perm} = 0.045$), but no significant condition by phase interaction. By late extinction, there was no difference relative to the CS- for the CS+ ($t_{18} = -1.67$, $P_{perm} = 0.11$). However, SCR for the suppressed category (CS+S) was still higher relative to the CS- ($t_{18} = 2.33$, $P_{perm} = 0.03$). Again, post-hoc comparisons of late extinction vs. late acquisition show reduced SCR for both the CS+ ($t_{18} = -4.32$, Bonferroni corrected $P_{perm} = 0.0006$) and the CS+S ($t_{18} = -4.01$, Bonferroni corrected $P_{perm} = 0.0006$). In contrast to threat expectancy results, SCR for the suppressed CS+S was significantly greater than for CS+ ($t_{18} = 2.49$, $P_{perm} = 0.025$).

3.1.3. Renewal test

The following day, participants had SCR and shock electrodes reattached and completed a renewal test with novel category exemplars. No new instructions were given on Day 2, and participants were simply told that the experiment would resume. No thought suppression cues or shocks were presented during renewal test. Results of threat expectancy revealed a main effect of condition ($F_{(2,\ 40)} = 28.11$, $P_{perm} = 0.0001$).

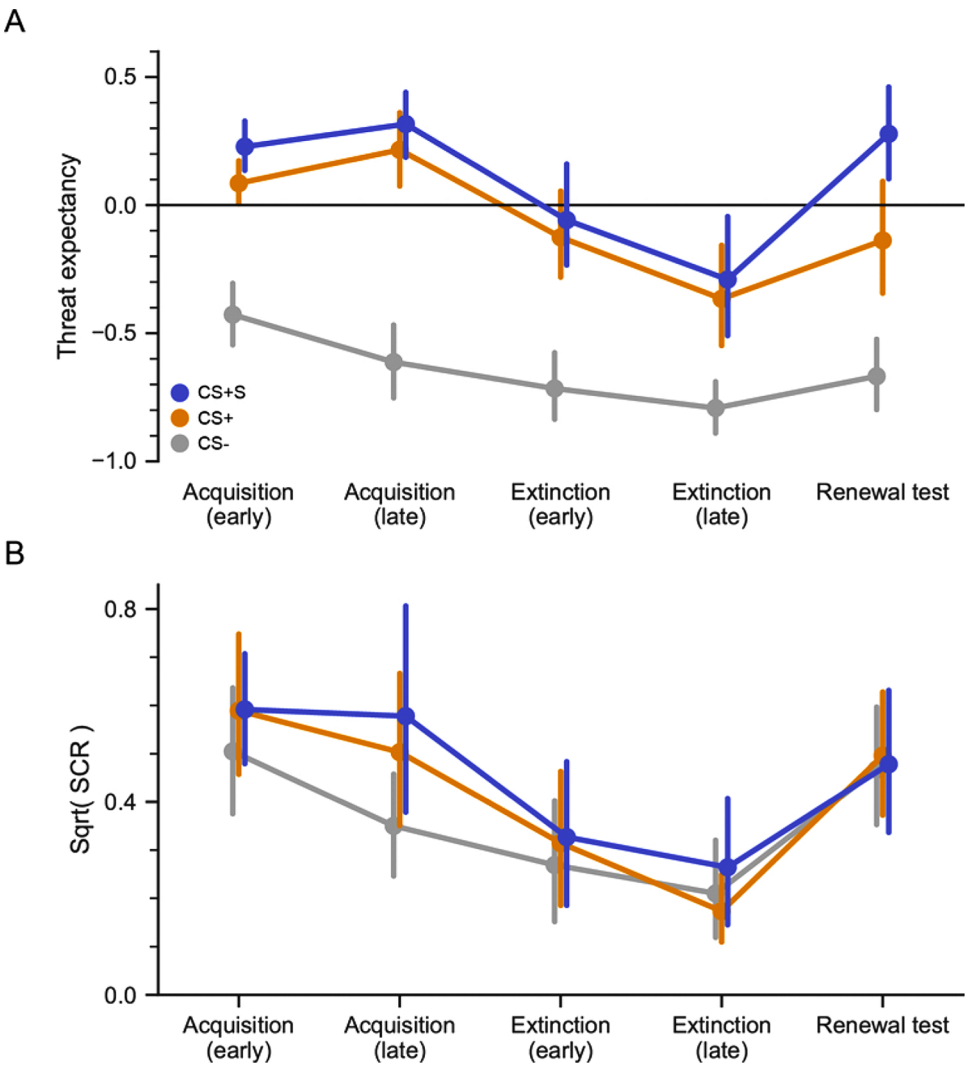


Fig. 2. Experiment 1 behavioral results. **A. Threat expectancy.** Expectancy for receiving an aversive electrical shock to the wrist was measured using a continuous expectancy bar during all phases of the experiment. The bar ranged from -1 (100% Do not expect a shock) to 1 (100% Expect a shock). Results confirm successful acquisition and extinction of both CS + categories. During renewal, expectancy for receiving the shock was significantly higher on trials from the CS+S than the CS+ and CS-. **B. Skin conductance responses.** Results confirm successful acquisition and subsequent extinction for both CS+ categories relative to the CS-. However, there was a general increase in arousal for all three CS categories during the renewal test. Points and error bars correspond to mean and 95% bootstrapped confidence intervals, respectively.

Planned comparisons showed that there was significant renewal of threat expectancy relative to the CS- for both the CS+ ($t(20) = 5.18$, $P_{perm} = 0.0001$) and the CS+S ($t_{20} = 7.20$, $P_{perm} = 0.0001$). Critical to our main hypothesis, we tested the difference in renewal of threat expectancy between the previously suppressed and extinguished CS+S, and the extinguished CS+ (without thought suppression). During the renewal test, threat expectancy was significantly higher for the CS+S compared to the CS+ (Fig. 3; $t_{20} = 2.92$, $P_{perm} = 0.0068$). These results indicate that thought suppression during extinction learning did not produce greater extinction generalization, and instead resulted in greater renewal of shock expectancy than standard extinction.

SCRs showed a generalized increase in arousal for each CS condition, and there was no significant main effect of condition ($F_{(2, 40)} = 0.19$, $P_{perm} = 0.83$).

3.1.4. Recognition memory test

Following the renewal test, participants completed a recognition memory test for all items seen during acquisition and extinction with the addition of novel foils. Previous work has shown that thought suppression results in decreased recognition memory for previously suppressed stimuli [29,33]. But notably, memory studies of this type traditionally include both a "forget" or "suppress" cue as well as an explicit "remember" or "view" cue. We did not include a cue to remember a particular CS category in the present study. High confidence corrected recognition (hits – false alarms) was scored for each CS type and encoding phase, acquisition and extinction. A repeated measures ANOVA revealed only a main effect of phase ($F_{(1, 20)} = 24.00$, $P = 8.68e\text{-}5$). No significant main effect of condition or phase by condition interaction was observed. A post-hoc comparison showed that overall, participants remembered fewer items from extinction compared to acquisition ($t_{20} = -3.77$, $P = 0.0012$). This difference in recognition memory between extinction and acquisition is in-line with previous experiments utilizing a similar category conditioning paradigm [23,25,34]. That we did not find an additional effect of thought suppression on recognition memory is perhaps owed to the absence of an explicit "remember" or "view" cue accompanying another CS category.

3.2. Experiment 2

In Experiment 1, we found that thought suppression during extinction learning did not increase extinction generalization, but instead resulted in an increase in shock expectancy relative to standard extinction. One possibility is that thought suppression during extinction training served the same function as an inhibitory cue, which acted to thwart extinction learning by preventing the omission of the US to diminish associative value of the CS, i.e., protection from extinction [14–16]. However, by our design, there are two possible routes by which thought suppression protected the CS from extinction: the cognitive process of thought suppression or the mere presence of the perceptual cue on those trials. That is, the thought suppression cue (blue border and

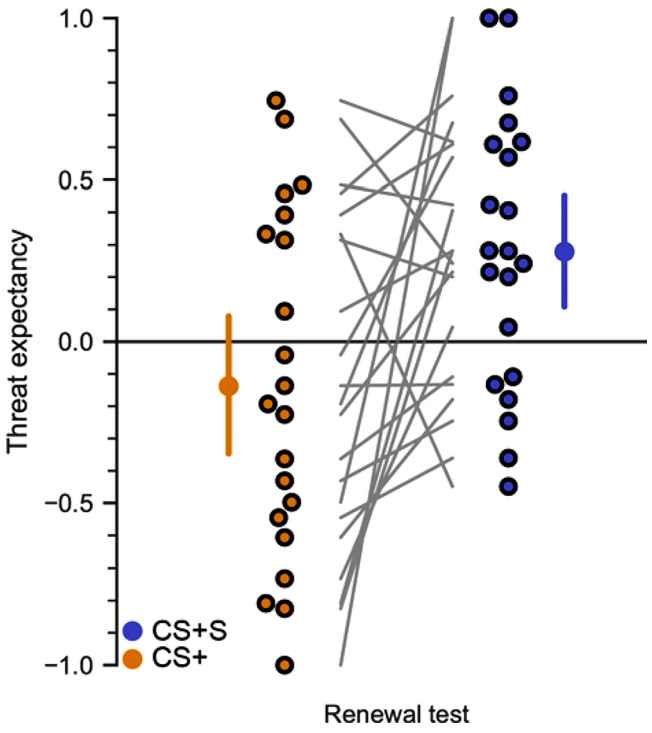


Fig. 3. Experiment 1 renewal test. Thought suppression increased renewal of threat expectancy during the 24h delayed renewal test (CS+S vs. CS+ t_{20} = 2.92, P_{perm} = 0.0068). Mean values and 95% confidence intervals are also shown in Fig. 2**A**. Individual participants are shown as points, with connecting lines for each subject between the two conditions.

fixation cross) could itself serve as an inhibitory cue, irrespective of the instructions and cognitive operation of thought suppression. To address this possibility, we conducted a second experiment that was identical to Experiment 1 in design, but participants were not given instructions regarding the colored border during extinction trials. Thus, the change in the colored border and fixation cross on CS+S trials during extinction served only as an added feature to CS+S trials. If the cue itself was sufficient to protect the CS+S category from extinction, then we would predict a replication of Experiment 1 of increased renewal of threat expectancy to the CS+S compared to the CS+ . However, if the cognitive operation of thought suppression is the relevant feature that protects the CS+S from extinction, then there should not be a difference in threat expectancy between the CS+S and CS+ during the renewal test. The post-experiment surveys confirmed that there was no difference in either STAI or IUS scores between participants in Experiments 1 and 2. Additionally, participants in both experiments did not differ either in the perceived number of received shocks on day 1 or in retrospective shock intensity.

3.2.1. Acquisition

Threat expectancy confirmed successful acquisition, and results were similar to Experiment 1. There was a main effect of condition ($F_{(2, 40)}$ = 48.48, P_{perm} = 0.0001). There was no main effect of phase, nor significant condition by phase interaction. Planned comparisons during late acquisition show greater expectancy for both the CS+ (t_{20} = 6.26, P_{perm} = 0.0003) and the CS+S (t_{20} = 6.33, P_{perm} = 0.0001) relative to the CS-. There was no difference in threat expectancy between the CS+ and CS+S (t_{20} = 0.33, P_{perm} = 0.75).

SCR results also confirmed successful acquisition. There was a significant main effect of condition ($F_{(2, 40)}$ = 11.60, P_{perm} = 0.0001) and phase ($F_{(1, 20)}$ = 4.91, P_{perm} = 0.040). There was no condition by phase interaction. Planned comparisons show greater SCR relative to the CS- for both CS+ (t_{20} = 3.63, P_{perm} = 0.0009) and the CS+S (t_{20} = 3.05, P_{perm} = 0.0057). There was no difference in SCR between the CS+ and CS+S (t_{20} = 0.87, P_{perm} = 0.40).

3.2.2. Extinction

As in Experiment 1, extinction immediately followed acquisition. The cue (colored border and fixation cross) was presented with the CS+S exemplars, but participants had not previously seen the cue nor received any instructions regarding it. Threat expectancy confirmed successful extinction. There were main effects of both condition ($F_{(2, 38)}$ = 11.97, P_{perm} = 0.0001) and phase ($F_{(1, 19)}$ = 49.01, P_{perm} = 0.0001), as well as a significant condition by phase interaction ($F_{(2, 38)}$ = 5.64, P_{perm} = 0.0077). As in Experiment 1, planned comparisons during late extinction showed that threat expectancy for CS+ was still higher than CS- expectancy (t_{19} = 3.56, P_{perm} = 0.0003). However, unlike in Experiment 1, there was no difference between CS+S and CS- expectancy (t_{19} = 1.60, P_{perm} = 0.12). We conducted post-hoc comparisons of CS+ and CS+S expectancy from late acquisition to late extinction, and found that it was reduced in late extinction for both CS+'s (CS+ t_{19} = -10.85, Bonferroni corrected P_{perm} = 0.0002; CS+S t_{19} = -9.09, Bonferroni corrected P_{perm} = 0.0002). There was no difference in threat expectancy between the CS+ and CS+S (t_{19} = -1.37, P_{perm} = 0.19).

SCR again confirmed successful extinction. There was a significant main effect of condition ($F_{(2, 40)}$ = 3.94, P_{perm} = 0.026). No significant main effect of phase nor interaction of condition by phase was observed. Planned comparisons showed no difference between either the CS+ (t_{20} = 1.28, P_{perm} = 0.23), or the CS+S (t_{20} = 1.20, P_{perm} = 0.28) and the CS-. There was no difference in SCR between the CS+ and CS+S (t_{20} = 0.36, P_{perm} = 0.81).

3.2.3. Renewal test

As in Experiment 1, participants completed a renewal test on the following day with novel category exemplars. Again, no thought suppression cues or shocks were presented during this test. Threat expectancy showed a main effect of condition ($F_{(2, 40)}$ = 18.49, P_{perm} = 0.0001). There was significant renewal of threat expectancy relative to the CS- for both the CS+ (t_{20} = 6.05, P_{perm} = 0.0001) and the CS+S (t_{20} = 4.30, P_{perm} = 0.0002). Critical to our hypothesis, there was no difference in threat expectancy between the CS+ and CS+S (Fig. 5; t_{20} = -0.49, P_{perm} = 0.63). These results demonstrate that the presentation of the suppression cue alone, without the engagement of thought suppression, was insufficient to produce the significant renewal of threat expectancy for CS+S that was observed in Experiment 1.

Also, as in Experiment 1, there was no main effect of condition for SCR ($F_{(1, 20)}$ = 0.60, P_{perm} = 0.55), and no follow-up comparisons were considered.

3.2.4. Recognition memory test

Following the renewal test, participants completed a surprise recognition memory test. Considering high-confidence corrected recognition, there was only a significant main effect of phase ($F_{(1, 20)}$ = 49.27, P = 0.0001). No significant main effect of condition or condition by phase interaction was observed. A post-hoc comparison showed that memory was overall lower for items encoded during extinction compared to acquisition (t_{20} = -7.02, P = 8.27e-7).

4. Discussion

Laboratory techniques to alleviate negative associations and expectations are of increasing interest as potential avenues to innovate and improve psychotherapy for disorders of fear, anxiety, and stress. As extinction is characterized by its transience and specificity, strategies to strengthen extinction and promote its generalization are of particular interest. Based on rodent studies showing that hippocampal down-regulation helps prevent fear renewal [10,11], we examined whether a behavioral instruction associated with of top-down hippocampal

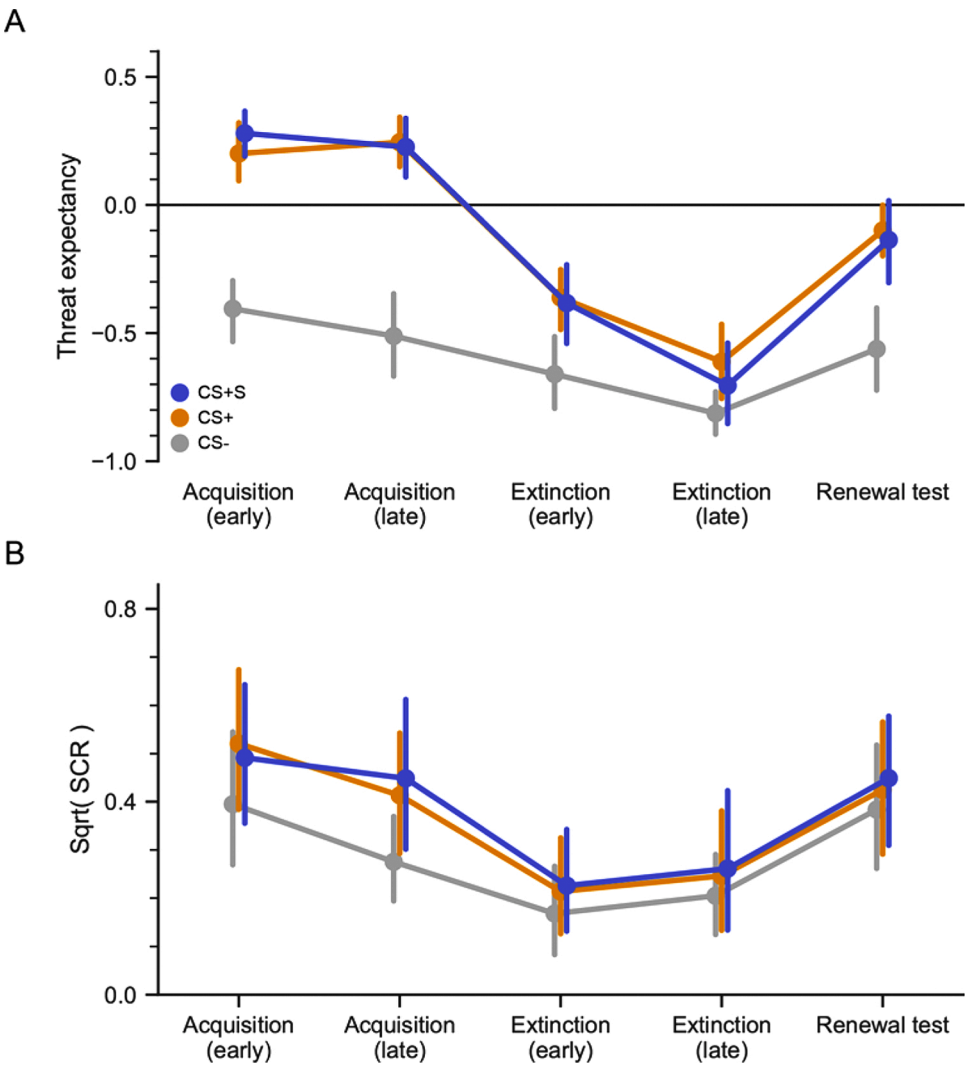


Fig. 4. Experiment 2 behavioral results. Experimental design was nearly identical to Experiment 1, however participants were not instructed to engage in thought suppression during extinction learning. The blue border and fixation cross still accompanied presentations of the CS+S as in Experiment 1. **A. Threat expectancy.** Continuous threat expectancy confirmed successful threat acquisition, extinction, and renewal. No differences were observed between the CS + and CS + S at any time point. **B. Skin conductance responses.** SCR confirmed successful acquisition and extinction, and a generalized increase in arousal to all CS types at renewal test. Points and error bars correspond to means and 95 % bootstrapped confidence intervals.

modulation likewise could serve to promote extinction generalization in humans. Contrary to promoting extinction generalization, an instruction to suppress one's thoughts resulted in a greater return of threat expectancy as compared to standard extinction. A follow-up experiment supported the inference that thought suppression itself served as an inhibitor during extinction learning that "protected" the CS from extinction. These findings have implications for therapies based on principles of extinction, predominately exposure therapy.

One possible explanation for elevated renewal to the suppressed category is a substantive shift in cognitive states (or task demands) between extinction and renewal test for the CS+S category. That is, in Experiment 1 participants were asked to engage in a particular cognitive operation (thought suppression) only during the extinction phase. This may have resulted in hyper-specific learning that the CS+S category is safe only in conjunction with the engagement of thought suppression. Consequently, participants may have reverted to the original CS-US association at the time of renewal test, when there was no instruction to suppress thoughts. Another possible explanation is that having the CS+S be the target of both thought suppression and extinction learning prevented later retrieval of the extinction association. In this case, the thought suppression may have inhibited encoding of both contextual features and stimulus features, such that the novel CS+S stimuli presented during the renewal test did not trigger retrieval of the extinction association. In either case, that threat expectancy was not different between CS+ categories in Experiment 2 (which lacked any thought suppression) suggests that the instruction itself, and not the colored cue, served as salient contextual information during extinction learning.

Although several contextual features differed between extinction learning and the renewal test (presence of the suppression cue, novel category exemplars, change in temporal context), it could be that the observed effect is a type of selective spontaneous recovery. Future work could test whether more salient shifts in context between acquisition, extinction, and test lead to different effects of thought suppression on extinction (i.e. "ABB" vs. "ABA" designs). It is also unclear whether thought suppression would lead to increased renewal of threat appraisal in a design with delayed as opposed to immediate extinction. Previous work has demonstrated that delayed extinction leads to less spontaneous recovery and renewal of fear [35]. The key difference is that in delayed extinction, the threat association has been consolidated. However, the goal of thought suppression during extinction would still be to impair encoding of contextual information during the new learning of the extinction association. It could be then that thought suppression would have a similar outcome as in immediate extinction.

The current experiment is based on neuroimaging research showing that thought suppression downregulates hippocampal activity [7], however we cannot verify that this top-down regulation is at play in this behavioral study without neuroimaging data. That is, it is possible our observed effects were a general effect of task instructions, rather than thought suppression specifically. Because of our lack of a trial-by-trial readout of hippocampal activity, we also cannot verify with certainty

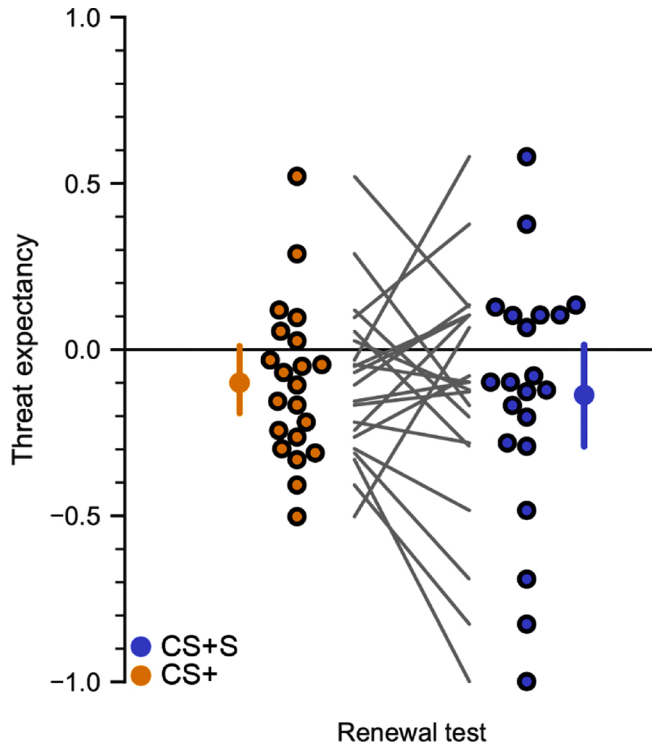


Fig. 5. Experiment 2 renewal test. The visual thought suppression cue, not paired with instructions to engage in thought suppression, does not increase renewal of threat expectancy during a 24h delayed test (CS+S vs. CS+ (t_{20} = -0.49, P_{perm} = 0.63). Mean values and 95% confidence intervals are replicated from Fig. 4. Individual participants are shown as points, with connecting lines for each subject between the two conditions.

how successful participants were at suppressing their thoughts on CS+S trials, or whether the suppression was in fact limited to these trials. The vast majority of item-method directed forgetting studies contrast a to-be-forgotten (TBF) condition with a to-be-remembered (TBR) condition and find that memory for TBF items is worse compared to TBR items. However, studies have shown that when memory for TBF items is either no different than, or better than memory for items not paired with an instruction (uncued) [36–38]. In no case is memory for TBF items poorer than memory for uncued items. These results suggest that the difference in memory between TBR and TBF items observed in item-method directed forgetting studies may be the result of increased processing or active rehearsal of TBR items, in contrast to disrupted encoding of TBF items [36–38]. This could explain why we did not observe differences in episodic memory between the CS conditions in the present study. In the current design, one category was paired with a salient thought suppression cue during extinction learning, and the other two CS conditions were uncued. A future variation of this design could incorporate remember cues into extinction learning, as well as suppression cues.

We can also speculate on whether a different type of thought suppression would produce the same putative outcome, or instead lead to enhanced extinction generalization. One alternative to the item-method directed forgetting paradigm is the think/no-think paradigm (TNT) [39]. TNT designs induce suppression by instructing participants to suppress the retrieval of a learned cue-target association when presented with the cue. Not only does TNT suppression reduce episodic memory for suppressed targets, it can also reduce the influence of suppressed information for future decisions [40,41]. For this reason, TNT suppression may also be more clinically relevant than thought suppression during encoding. Intrusive thoughts are a common symptom in many anxiety disorders, and can be triggered by secondary associations [42]. The ability to suppress these thoughts may be adaptive, and a recent neuroimaging study observed deficits in functional connectivity during TNT suppression in individuals with post-traumatic stress disorder (PTSD) compared to both trauma exposed and healthy controls [43]. In the latter study, individuals that were trauma exposed but did not go on to develop PTSD had stronger connectivity between prefrontal control regions and posterior memory regions during top-down suppression compared to trauma exposed individuals with PTSD. Thus, disruptions in the ability to suppress thoughts may contribute to the development and maintenance of PTSD. If this is the case, patients may benefit from additional suppression training in the context of therapy. Before this idea can be translated into treatment, research should investigate whether TNT suppression combined with extinction learning results in a different outcome than the one we observed with thought suppression in Experiment 1. It could be that TNT is more potent than item-method thought suppression. However, as we report here, introducing an additional cognitive operation during extinction learning can result in a greater return of fear later on, and thus it is possible that even a TNT manipulation could also lead to an increased renewal of threat appraisal.

Finally, the limitation presented by the SCR results should be considered. Specifically, SCR results did not dissociate renewal between any of the CS categories, but were instead enhanced for all CS types, even the unpaired CS-. One possibility is that the use of novel category exemplars during the test generated an unexpected degree of non-associative orienting to the images that was reflected in arousal but not in cognitive threat appraisal. Another account of the divergence between threat expectancy and SCR may be that these behavioral measures are assaying different psychological constructs. A current theory of negative affective processing in humans suggests that there is a dissociation between the cognitive experiences of emotion and physiological reactivity ([44]; however see [45]). This idea is supported by recent findings that the neural representations of fear and physiological reactivity (i.e. SCR) are related, but ultimately separable [46]. It could be the case that top-down thought suppression does not influence the cognitive and physiological systems equally. Future work should explore how top-down inhibitory control can be implemented alongside extinction learning in a way to impact both cognitive and physiological outcomes.

5. Conclusion

We investigated whether thought suppression during extinction learning could facilitate extinction generalization. In contrast to our primary hypothesis, we found that extinction with thought suppression had a protective effect, such that there was a significant renewal of threat appraisal for the suppressed category during a delayed associative memory test relative to standard extinction. In a second experiment, we confirmed that a suppression cue alone was insufficient to replicate this effect, suggesting that the cognitive engagement of thought suppression was necessary to interfere with extinction learning. Future work should investigate whether other forms of suppression, such as retrieval suppression, could better facilitate the generalization of safety learning.

Data availability

All de-identified behavioral data available upon request to J.E. Dunsmoor (joseph.dunsmoor@austin.utexas.edu).

Author contributions

A.C.H, J.A.L.-P., and J.E.D. conceived of and designed the experiment; A.C.H. and S.A.B. recruited participants and performed the experiments; A.C.H. and S.A.B. analyzed data; A.C.H, J.A.L.-P., and J.E.D. wrote the manuscript.

Funding

This work was supported by NIH R00MH106719 to J.E.D.

References

[1] M.G. Craske, D. Hermans, B. Vervliet, State-of-the-art and future directions for extinction as a translational model for fear and anxiety, Philos. Trans. R. Soc. B Biol. Sci. 373 (2018), 20170025.
[2] M.E. Bouton, E.W. Moody, Memory processes in classical conditioning. Neuroscience and Biobehavioral Reviews, 2004, pp. 663–674.
[3] S. Maren, K.L. Phan, I. Liberzon, The contextual brain: implications for fear conditioning, extinction and psychopathology, Nat. Rev. Neurosci. 14 (2013) 417–428.
[4] J.E. Dunsmoor, Y. Niv, N. Daw, E.A. Phelps, Rethinking extinction, Neuron 88 (2015) 47–63.
[5] P.J. Fitzgerald, J.R. Seemann, S. Maren, Can fear extinction be enhanced? A review of pharmacological and behavioral findings, Brain Res. Bull. 105 (2014) 46–60.
[6] N. Singewald, C. Schmuckermair, N. Whittle, A. Holmes, K.J. Ressler, Pharmacology of cognitive enhancers for exposure-based therapy of fear, anxiety and trauma-related disorders, Pharmacol. Ther. 149 (2015) 150–190.
[7] M.C. Anderson, S. Hanslmayr, Neural mechanisms of motivated forgetting, Trends Cogn. Sci. 18 (2014) 279–292.
[8] I. Izquierdo, C.R.G. Furini, J.C. Myskiw, Fear memory, Physiol. Rev. 96 (2016) 695–750.
[9] R. Marek, J. Jin, T.D. Goode, T.F. Giustino, Q. Wang, G.M. Acca, R. Holehonnur, J. E. Ploski, P.J. Fitzgerald, T. Lynagh, et al., Hippocampus-driven feed-forward inhibition of the prefrontal cortex mediates relapse of extinguished fear, Nat. Neurosci. 21 (2018) 384–392.
[10] J.A. Hobin, J. Ji, S. Maren, Ventral hippocampal muscimol disrupts context-specific fear memory retrieval after extinction in rats, Hippocampus 16 (2006) 174–182.
[11] M. Zelikowsky, T.A. Hast, R.Z. Bennett, M. Merjanian, N.A. Nocera, R. Ponnusamy, M.S. Fanselow, Cholinergic blockade frees fear extinction from its contextual dependency, Biol. Psychiatry 73 (2013) 345–352.
[12] M.G. Craske, M. Fanselow, M. Treanor, A. Bystritksy, Cholinergic modulation of exposure disrupts hippocampal processes and augments extinction: proof-of-Concept study with social anxiety disorder, Biol. Psychiatry 86 (2019) 703–711.
[13] E.L. Bjork, R.A. Bjork, M.C. Anderson, Varieties of goal-directed forgetting, in: J. M. Golding, C.M. MacLeod (Eds.), Intentional Forgetting: Interdisciplinary Approaches, Erlbaum, Hillsdale, NJ, 1998, pp. 103–137.
[14] P.F. Lovibond, N.R. Davis, A.S. O'Flaherty, Protection from extinction in human fear conditioning, Behav. Res. Ther. 38 (2000) 967–983.
[15] P.F. Lovibond, C.J. Mitchell, E. Minard, A. Brady, R.G. Menzies, Safety behaviours preserve threat beliefs: protection from extinction of human fear conditioning by an avoidance response, Behav. Res. Ther. 47 (2009) 716–720.
[16] R.A. Rescorla, Protection from extinction, Learn. Behav. 31 (2003) 124–132.
[17] R.A. Rescorla, A.R. Wagner, A theory of classical conditioning: variations in the effectiveness of reinforcement and nonreinforcement, Class. Cond. II Curr. Res. Theory 21 (1972) 64–99.
[18] M.G. Craske, M. Treanor, C.C. Conway, T. Zbozinek, B. Vervliet, Maximizing exposure therapy: an inhibitory learning approach, Behav. Res. Ther. 58 (2014) 10–23.
[19] J.E. Dunsmoor, M.C. Kroes, Episodic memory and Pavlovian conditioning: ships passing in the night, Curr. Opin. Behav. Sci. 26 (2019) 32–39.
[20] J.E. Dunsmoor, M.C.W. Kroes, J. Li, N.D. Daw, H.B. Simpson, E.A. Phelps, Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction, J. Neurosci. (2019) 2713–2718.
[21] D. Schiller, M.H. Monfils, C.M. Raio, D.C. Johnson, J.E. LeDoux, E.A. Phelps, Preventing the return of fear in humans using reconsolidation update mechanisms, Nature 463 (2010) 49–53.
[22] J. Peirce, J.R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, J.K. Lindeløv, PsychoPy2: experiments in behavior made easy, Behav. Res. Methods 51 (2019) 195–203.
[23] N.E. Keller, J.E. Dunsmoor, The effects of aversive-to-appetitive counterconditioning on implicit and explicit fear memory, Learn. Mem. 27 (2020) 12–19.
[24] S.R. Green, P.A. Kragel, M.E. Fecteau, K.S. LaBar, Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis, Int. J. Psychophysiol. 91 (2014) 186–193.
[25] J.E. Dunsmoor, M.C.W. Kroes, C.M. Moscatelli, M.D. Evans, L. Davachi, E. A. Phelps, Event segmentation protects emotional memories from competing experiences encoded close in time, Nat. Hum. Behav. 2 (2018) 291–299.
[26] A.K. Grady, K.H. Bowen, A.T. Hyde, S.K. Totsch, D.C. Knight, Effect of continuous and partial reinforcement on the acquisition and extinction of human conditioned fear, Behav. Neurosci. 130 (2016) 36–43.
[27] L.C. Humphreys, The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions, J. Exp. Psychol. 25 (1939) 141–158.
[28] M.R. Milad, R.K. Pitman, C.B. Ellis, A.L. Gold, L.M. Shin, N.B. Lasko, M.A. Zeidan, K. Handwerger, S.P. Orr, S.L. Rauch, Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder, Biol. Psychiatry 66 (2009) 1075–1082.
[29] R.G. Benoit, M.C. Anderson, Opposing mechanisms support the voluntary forgetting of unwanted memories, Neuron 76 (2012) 450–460.
[30] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (Complete samples), Biometrika 52 (1965) 591.
[31] R. Vallat, Pingouin: statistics in Python, J. Open Source Softw. 3 (2018) 1026.
[32] J. Frossard, O. Renaud, permuco: Permutation Tests for Regression, Repeated Measures) ANOVA/ANCOVA and Comparison of Signals, 2019.
[33] T.H. Wang, K. Placek, J.A. Lewis-Peacock, More is less: increased processing of unwanted memories facilitates forgetting, J. Neurosci. 39 (2019) 3551–3560.
[34] J.E. Dunsmoor, L. Davachi, E.A. Phelps, V.P. Murty, Emotional learning selectively and retroactively strengthens memories for related events, Nature 520 (2015) 345–348.
[35] N.C. Huff, J.A. Hernandez, N.Q. Blanding, K.S. LaBar, Delayed extinction attenuates conditioned fear renewal and spontaneous recovery in humans, Behav. Neurosci. 123 (2009) 834–843.
[36] H. Gao, M. Qi, Q. Zhang, Forgetting cues are ineffective in promoting forgetting in the item-method directed forgetting paradigm, Int. J. Psychophysiol. 144 (2019) 25–33.
[37] S. Schindler, J. Kissler, Too hard to forget? ERPs to remember, forget, and uninformative cues in the encoding phase of item-method directed forgetting, Psychophysiology 55 (2018).
[38] B. Zwissler, S. Schindler, H. Fischer, C. Plewnia, J.M. Kissler, "Forget me (not)?" - remembering forget-items versus un-cued items in directed forgetting, Front. Psychol. 6 (2015) 1741.
[39] M.C. Anderson, C. Green, Suppressing unwanted memories by executive control, Nature 410 (2001) 366–369.
[40] P. Gagnepain, R.N. Henson, M.C. Anderson, Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition, Proc. Natl. Acad. Sci. U. S. A. 111 (2014).
[41] Y. Wang, A. Luppi, J. Fawcett, M.C. Anderson, Reconsidering unconscious persistence: Suppressing unwanted memories reduces their indirect expression in later thoughts, Cognition 187 (2019) 78–94.
[42] S. Mineka, R. Zinbarg, A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was, Am. Psychol. 61 (2006) 10–26.
[43] A. Mary, J. Dayan, G. Leone, C. Postel, F. Fraisse, C. Malle, T. Vallée, C. Klein-Peschanski, F. Viader, V. de la Sayette, et al., Resilience after trauma: the role of memory suppression, Science 367 (2020) eaay8477.
[44] J.E. LeDoux, D.S. Pine, Using neuroscience to help understand fear and anxiety: a two-system framework, Am. J. Psychiatry 173 (2016) 1083–1093.
[45] M.S. Fanselow, Z.T. Pennington, A return to the psychiatric dark ages with a two-system framework for fear, Behav. Res. Ther. 100 (2018) 24–29.
[46] V. Taschereau-Dumouchel, M. Kawato, H. Lau, Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates, Mol. Psychiatry (2019) 1–13.