

Rewarded Extinction Increases Amygdalar Connectivity and Stabilizes Long-Term Memory Traces in the vmPFC

Nicole E. Keller,¹ Augustin C. Hennings,¹ Emily K. Leiker,² Jarrod A. Lewis-Peacock,^{1,3,4,5} and Joseph E. Dunsmoor^{1,3,5}

¹Institute for Neuroscience, University of Texas at Austin, Austin, Texas 78712, ²Department of Psychiatry School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, ³Center for Learning and Memory, Department of Neuroscience, University of Texas at Austin, Austin, Texas 78712, ⁴Department of Psychology, University of Texas at Austin, Austin, Texas 78712, and ⁵Department of Psychiatry, Dell Medical School, University of Texas at Austin, Austin, Texas 78712

Neurobiological evidence in rodents indicates that threat extinction incorporates reward neurocircuitry. Consequently, incorporating reward associations with an extinction memory may be an effective strategy to persistently attenuate threat responses. Moreover, while there is considerable research on the short-term effects of extinction strategies in humans, the long-term effects of extinction are rarely considered. In a within-subjects fMRI study with both female and male participants, we compared counterconditioning (CC; a form of rewarded-extinction) to standard extinction at recent (24 h) and remote (approximately one month) retrieval tests. Relative to standard extinction, rewarded extinction diminished 24-h relapse of arousal and threat expectancy, and reduced activity in brain regions associated with the appraisal and expression of threat (e.g., thalamus, insula, periaqueductal gray). The retrieval of reward-associated extinction memory was accompanied by functional connectivity between the amygdala and the ventral striatum, whereas the retrieval of standard-extinction memories was associated with connectivity between the amygdala and ventromedial prefrontal cortex (vmPFC). One month later, the retrieval of both standard-extinction and rewarded-extinction was associated with amygdala-vmPFC connectivity. However, only rewarded extinction created a stable memory trace in the vmPFC, identified through overlapping multivariate patterns of fMRI activity from extinction to 24-h and one-month retrieval. These findings provide new evidence that reward may generate a more stable and enduring memory trace of attenuated threat in humans.

Key words: amygdala; fear extinction; fMRI; reward; striatum; vmPFC

Significance Statement

Prevalent treatments for pathologic fear and anxiety are based on the principles of Pavlovian extinction. Unfortunately, extinction forms weak memories that only temporarily inhibit the retrieval of threat associations. Thus, to increase the translational relevance of extinction research, it is critical to investigate whether extinction can be augmented to form a more enduring memory, especially after long intervals. Here, we used a multiday fMRI paradigm in humans to compare the short-term and long-term neurobehavioral effects of aversive-to-appetitive counterconditioning (CC), a form of augmented extinction. Our results provide novel evidence that including an appetitive stimulus during extinction can reduce short-term threat relapse and stabilize the memory trace of extinction in the ventromedial prefrontal cortex (vmPFC), for at least one month after learning.

Introduction

While learning about threats is adaptive, persistent and misattributed fearful responses are characteristic of anxiety disorders. Exposure therapy, based on the principles of Pavlovian extinction, is a widely used treatment for anxiety-related disorders (Abramowitz et al., 2019). Unfortunately, relapse of extinguished behavior is common, and a substantial number of individuals undergoing treatments will drop-out or relapse (Schottenbauer et al., 2008; Markowitz et al., 2015). Notably, even healthy adults tend to show postextinction recovery of learned defensive behavior in new situations, indicating extinction is a fragile form of

Received Jan. 11, 2022; revised May 24, 2022; accepted May 27, 2022.

Author contributions: N.E.K., J.A.L.-P., and J.E.D. designed research; N.E.K., A.C.H., E.K.L., and J.E.D. performed research; N.E.K., A.C.H., E.K.L., J.A.L.-P., and J.E.D. contributed unpublished reagents/analytic tools; N.E.K., A.C.H., E.K.L., J.A.L.-P., and J.E.D. analyzed data; N.E.K. wrote the first draft of the paper; N.E.K., A.C.H., E.K.L., J.A.L.-P., and J.E.D. edited the paper; N.E.K. and J.E.D. wrote the paper.

The authors declare no competing financial interests.

Correspondence should be addressed to Joseph E. Dunsmoor at joseph.dunsmoor@austin.utexas.edu.

<https://doi.org/10.1523/JNEUROSCI.0075-22.2022>

Copyright © 2022 the authors

inhibitory learning bound to the spatiotemporal context in which extinction memories were formed (Bouton, 2002). Several augmented strategies to standard extinction have shown success in promoting relatively short-term (~24 h) retention of extinction memories in humans (Dunsmoor et al., 2015b; Craske et al., 2018). However, evaluating the long-term success (more than one week) of extinction protocols in humans is extremely rare, which limits the clinical translational relevance of extinction research, as symptoms frequently return some time after treatment (Vervliet et al., 2013). Here, we compared the neurobehavioral effects of standard extinction and augmented extinction in healthy adults at recent (24 h) and remote (approximately one month) intervals in the same individuals. Whereas standard extinction involved simply omitting an expected aversive electrical shock, augmented extinction involved replacing the shock with a positive outcome, a paradigm known as aversive-to-appetitive counterconditioning (CC; Dickinson and Pearce, 1977; Keller et al., 2020).

In CC, behavior is modified through a new association with a stimulus of the opposite valence. Research on CC dates to the earliest studies of conditioning in humans (Jones, 1924), and forms the basis for popular treatments for anxiety disorders such as systematic desensitization (Wolpe, 1954, 1968, 1995). Contemporary behavioral research on CC is sparse (Koizumi et al., 2016; van Dis et al., 2019; Keller and Dunsmoor, 2020; Gatzounis et al., 2021), and there are currently no neuroimaging studies directly comparing CC and extinction in humans. It remains unclear whether a reduction of conditioned responses through CC is modulated by similar neural circuitry as standard extinction, and whether the resulting threat attenuation is more enduring over time.

One possibility is that reduced relapse following CC is mediated by augmented activity in networks involved in the formation of extinction memories, specifically activity within and between the ventromedial prefrontal cortex (vmPFC) and amygdala (Hartley and Phelps, 2010; Milad and Quirk, 2012; Giustino and Maren, 2015; Alexandra Kredlow et al., 2022). The presence of a positive stimulus could further engage reward-related regions of the mesostriatal dopamine system shown to be involved in threat extinction (Holtzman-Assif et al., 2010; Raczka et al., 2011; Josselyn and Frankland, 2018; Luo et al., 2018; Kalisch et al., 2019; Salinas-Hernández and Duvarci, 2021). In support of this idea, neurobiological evidence in rats found that rewarded extinction enhanced recruitment of an amygdala-striatal pathway and led to diminished threat relapse at a remote test (Correia et al., 2016). However, other evidence in rodents suggests that CC is less effective than standard extinction at preventing relapse of the original behavior (Holmes et al., 2016). If this were the case, then replacing shock with reward (rather than omitting it) may somehow interfere with processes underlying extinction memory formation and retrieval.

We developed a multiday fMRI protocol to compare the neurobehavioral effects of threat extinction and aversive-to-appetitive CC on threat attenuation at recent and remote time points. The protocol incorporated a within-subjects Pavlovian conditioning design with renewal tests at 24 h and approximately one month later. Based on our prior behavioral findings (Keller and Dunsmoor, 2020), we predicted CC would more effectively attenuate the relapse of conditioned responses. In line with previous research on enhanced extinction (Dunsmoor et al., 2019), we also predicted that CC would more effectively attenuate within-session activity in regions associated with threat appraisal [e.g.,

the insula, thalamus, dorsal anterior cingulate cortex (dACC), brainstem]. Based on prior neurobiological evidence in rats (Correia et al., 2016), we also predicted that amygdala-ventral striatum functional connectivity would be selectively enhanced for stimuli associated with CC versus standard extinction.

To examine the fidelity of extinction and CC memory representations over time, we incorporated multivariate representational similarity analysis (RSA; Kriegeskorte et al., 2008) of encoding-to-retrieval overlap (Ritchey et al., 2013) between extinction learning and 24-h and one-month retrieval. We focused on the vmPFC based on recent fMRI evidence that 24-h extinction retrieval reactivates similar neural activity patterns associated with extinction formation in this region (Hennings et al., 2020, 2021). We predicted that neural similarity in the vmPFC would be enhanced and maintained over time for CC in comparison to standard extinction, indicating a more durable memory trace in a region critical for the encoding, storage, and retrieval of safety memories (Milad and Quirk, 2012; Giustino and Maren, 2015; Tovote et al., 2015).

Materials and Methods

Participants

A sample size of twenty-five healthy participants (15 female; mean age: 23.48 years; SD = 5.51 years, age range 18–36) was a priori based on our prior fMRI research on category threat conditioning and extinction (Hennings et al., 2020). Participants reported no neurologic or psychiatric disorders, and were recruited from the University of Texas at Austin and local Austin community to complete this experiment. Two participants did not return for their third session approximately one month away, therefore twenty-five participants completed the first and second session, and 23 (14 female; mean age 23.69 year; SD = 5.63 years, age range 18–36) completed all three sessions. We collected state individual difference measures [posttraumatic stress disorder (PTSD) checklist for DSM-5 (PCL-5), the Childhood Trauma Questionnaire (CTQ), Beck Anxiety Inventory (BAI)] and trait measures [Intolerance of Uncertainty–Short Form (IUSF), State-Trait Anxiety Inventory (STAI)] of negative affect-related constructs for each participant. All participants provided written informed consent and procedures complied with the Institutional Review Board of University of Texas at Austin (IRB #2017-02-0094).

Task and procedure

The design for this study was based on the behavioral experiment by (Keller and Dunsmoor, 2020). This was a within-subjects functional MRI study that included Pavlovian threat acquisition and extinction/CC on day one, and a renewal test and an episodic memory test 24 h later and approximately one month later (mean length: 26.91 d, SD = 10.26 d, day range: 15–63 d) depending on the participants' availability (Fig. 1A). The episodic memory results are not discussed in this report. Conditioned stimuli (CSs) were pictures of animals, tools, and food on a white background. Pictures of common phobic stimuli (e.g., spiders, snakes, weapons) or highly appealing food items (e.g., pizza), were not used as CSs. The unconditioned stimulus (US) was a 6-ms electrical shock delivered to the index and middle finger of the participant's left hand from the BIOPAC MP160 System with a STM100C module. The task was presented using E-Prime 2.0 and consisted of a trial-unique category conditioning design, meaning that each trial was a different basic-level exemplar with a unique name. For example, there were not two different pictures of a dog. For all phases (Pavlovian threat acquisition, extinction, and the renewal tests), each CS was on the screen for 6 s, followed by a 7–8 s inter-trial interval (ITI) with a fixation cross on a white background. On each trial, subjects rated their expectancy to receive a shock using a two-alternative forced choice scale (2AFC; i.e., yes or no). Trial order was pseudo-randomized so that participants did not see pictures from the same category three trials in a row.

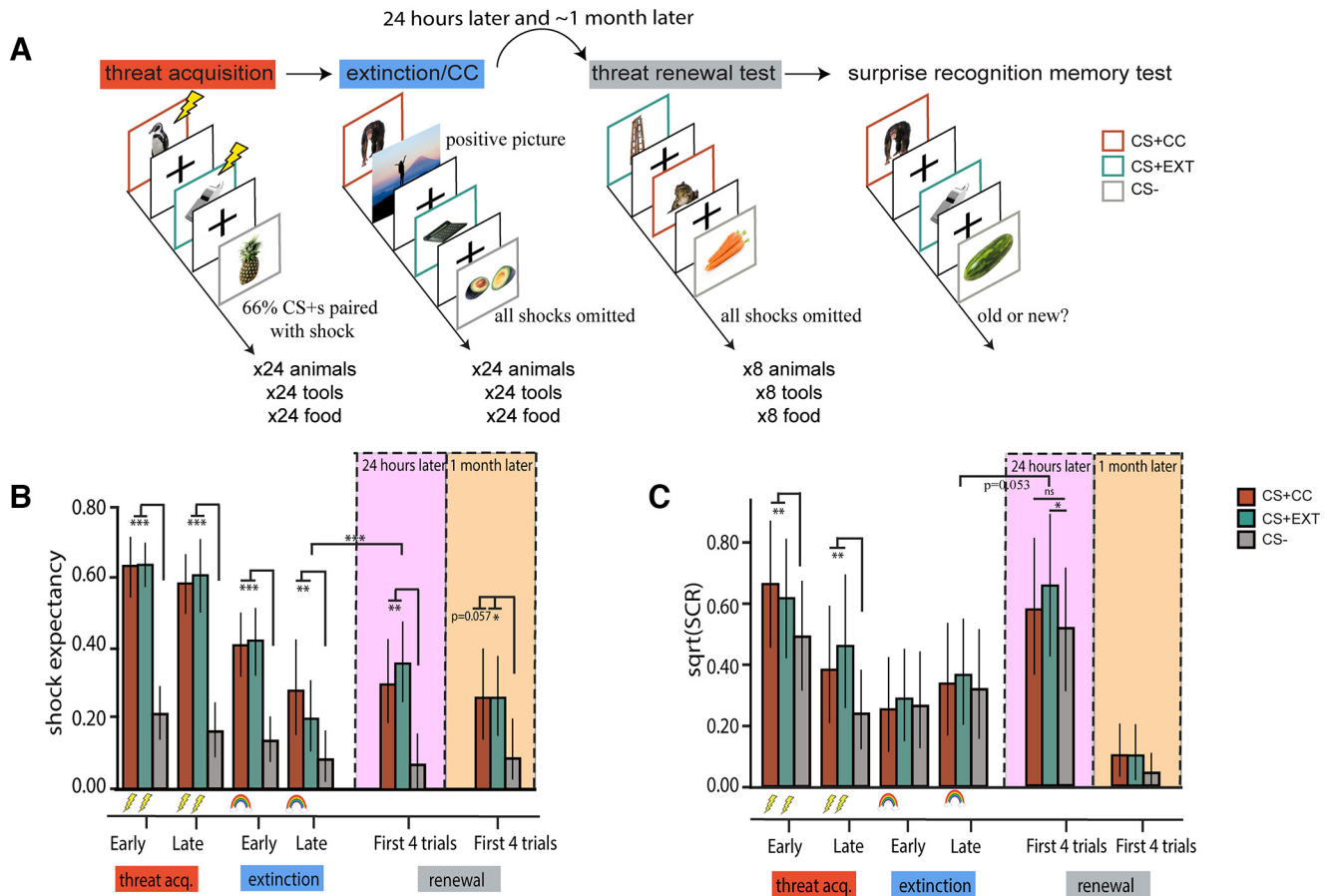


Figure 1. Experimental design and behavioral results. **A**, Participants underwent threat acquisition with category exemplars of animals or tools (CS+s) paired with a shock on a partial reinforcement schedule, and a third category, food (CS-), never paired with shock. Conditioning was followed by the extinction phase, in which the shock was omitted following CS+EXT trials (counterbalanced, tools or animals, in this example tools), and CC, in which the shock was replaced by a positive picture at the end of each CS+CC trial (animals or tools, respectively, in this example animals). Subjects returned 24 h and approximately one month later, and new CSs from the same categories, were presented in the absence of any shocks or positive pictures. **B**, Shock expectancy results confirmed successful acquisition and extinction of threat expectancy. Twenty-four hours after day 1, shock expectancy toward the CS+EXT category significantly increased from the end of extinction to early renewal. Shock expectancy toward the CS+s remained even at the approximately one-month follow-up. **C**, Conditioned SCRs replicated prior findings (Keller and Dunsmoor, 2020), there were no differences between CS+s during acquisition or extinction, but 24 h later, SCRs were higher for the CS+EXT category as compared with the CS- category, and there were no differences between the CS+CC category and the CS- category. One month later, there was no renewal of conditioned SCRs toward the CS+s. Colored borders are for illustrative purposes only. The rainbow, which represents the positive pictures, and the lightning bolt, which represents an electrical shock, depict the outcome following a given CS type. For example, following CS+CC trials during extinction, there is a positive picture. Error bars indicate the 95% confidence interval of the mean; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Day 1

Day 1 included two phases: Pavlovian threat acquisition and extinction. These phases were divided into three separate functional imaging runs: (1) the first half of threat acquisition; (2) the second half of threat acquisition followed without a break by the first half of extinction; and (3) the last half of extinction. Each scanner run occurred consecutively with less than ~1-min break between runs. Before participants entered the scanner, shock electrodes were attached to the index and middle finger on the left hand. The electrical shock was calibrated to be at a level that was deemed “highly annoying and unpleasant, but not painful.” Skin conductance response (SCR) served as a measure of conditioned autonomic arousal and was collected throughout the experiment on each day. SCR electrodes were placed on participants’ left palm and connected to a BIOPAC MP160 System. SCR sampling rate was set to 200 Hz (see below, Psychophysiology analysis). Day 1 included a total of 144 trials across acquisition and extinction. During acquisition, CS stimuli from two categories (CS+, animals and tools, 24 stimuli per category) co-terminated with a shock 66% of the time. Items from the food category (CS-, 24 stimuli) were never paired to a shock, and always served as a within-subjects unpaired control category. Extinction included a total of 24 CS+, 24 CS+EXT, and 24 CS- trials, all unpaired with shock. During extinction, stimuli from one CS+ category (CS+CC, animals or tools, counterbalanced across participants), were not followed by a shock

and were paired with a unique image of a positive valenced picture, depicting vibrant, rich images of visually appealing landscapes/nature scenes that did contain any animals, tools or food. The positive pictures were presented for a duration of 1 s and were pilot rated by a separate group of 19 participants to confirm high valence and low arousal. Stimuli from the other CS+ category (CS+EXT, either tools or animals, respectively), were simply not followed by a shock. Stimuli from the CS- category were not followed by any outcome.

Day 2

Participants returned the next day (~24 h) and underwent a test of threat renewal. The renewal test was followed by a recognition memory test for half the items encoded the previous day including the positive pictures paired with CS+CC during CC (details on the memory test and memory results not reported here). Before participants entered the scanner, shock and SCR electrodes were re-attached. Participants did not receive any new instructions from the day before and were instructed to continue to rate expectancy for receiving shocks on each trial. The renewal test included eight trials each of animals, tools, and food. The CSs were novel category exemplars not shown the previous day. There were no shocks or positive pictures presented during the renewal test on day 2. The first trial on the renewal test was always a discarded CS- trial that was used to capture the initial orienting response.

Behavioral and neural analyses for the renewal test focused a priori on the first four trials (early renewal test) per CS type. Focusing on the first four trials is in line with previous human neuroimaging research on extinction recall (Milad et al., 2009; Schiller et al., 2010; Kroes et al., 2016), as these early trials capture the instance when the possibility for threat is most ambiguous. In the absence of any outcomes, the last portion of renewal most likely reflects processes relevant to further extinction learning rather than extinction memory recall.

One month later

Participants returned for their third and final session approximately one month later. This session followed the same format as day 2, and included four functional imaging runs: a final threat renewal test, a recognition memory test for the rest of the CS exemplars from day 1, and two runs of a perceptual category localizer. Before participants entered the scanner, shock and SCR electrodes were re-attached and the shock was re-calibrated.

Psychophysiology analysis

SCRs were calculated using prior criteria (Keller and Dunsmoor, 2020). SCRs were considered valid to the CS trial if the trough-to-peak deflection of electrodermal activity occurred between 0.5 and 6 s following CS onset and were not $>0.2 \mu\text{S}$. Trials that did not meet these criteria were scored as zero. SCRs were scored by an automated analysis script implemented in MATLAB (Green et al., 2014), and were later visually inspected by research assistants blind to the experimental conditions. SCR data were square-root transformed before statistical analysis to normalize the distributions. Participants were not excluded from the analysis based on any response criteria for SCRs, based on recommendations from the field of human threat conditioning (Lonsdorf et al., 2017). Two-AFC shock expectancy was coded as 1 = expect to receive a shock, 0 = do not expect.

Imaging parameters

Brain images were recorded on a 3T Siemens Vida with 64-channel head coil at the University of Texas at Austin Biomedical Imaging Center. Functional task and localizer data were acquired using T2*-weighted EPI sequences (TR = 1000 ms, TE = 86 ms, FOV = 86×86 mm, 2.5-mm isotropic voxels), with slices oriented parallel to the hippocampal long axis and positioned to provide whole-brain coverage. High-resolution T1-weighted (T1w) anatomic images were obtained using 3D MPRAGE sequences (TR = 2400 ms, TE = 1000 ms, FOV = 208×300 mm, 0.8-mm isotropic voxels) before the EPIs in each session, to aid in co-registration and normalization. Diffusion-weighted images were also acquired but were not examined.

fMRI data preprocessing

MRI data were preprocessed using fMRIPrep 1.5.9 (Esteban et al., 2019) and FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) FEAT 6.00 (FMRI Expert Analysis Tool). Processing in fMRIPrep followed the default steps, with additional options for multiple T1w images per participant (–longitudinal flag) and a framewise displacement threshold of 0.3 mm. T1w images were corrected for intensity nonuniformity and skull-stripped using *N4BiasFieldCorrection* (Tustison et al., 2010) and *BrainExtraction* (both from ANTs 2.2.0; Avants et al., 2008). Segmentation of the skull-stripped T1w images into three tissue classes (CSF, WM, GM) was performed using FSL 5.0.9 *fast* (Y. Zhang et al., 2001), followed by surface reconstruction with FreeSurfer 6.0.1 *recon-all* (Dale et al., 1999). The skull-stripped T1w images were registered using FreeSurfer's *mri_robust_template* to generate a single unbiased T1w-reference map per participant for spatial normalization (Reuter et al., 2010). Spatial normalization to MNI space was performed via non-linear registration (ANTs *Registration*), using skull-stripped versions of both the T1w reference volume and MNI152NLin2009cAsym template (Fonov et al., 2009).

Functional data from each BOLD run were corrected for field distortion based on a B0-nonuniformity map estimated via AFNI *3dQwarp* (Cox and Hyde, 1997), then co-registered to the corresponding T1w reference using boundary-based registration (Greve and Fischl, 2009) with

6 degrees of freedom (FreeSurfer *bbregister*). Head-motion parameters, including transformation matrices and six rotation and translation parameters, were estimated for each BOLD run before any spatiotemporal filtering (FSL *mcflirt*). Framewise displacement and DVARS were calculated for each functional run using Nipype (Power et al., 2014), and frames exceeding 0.3 mm FD or 1.5 standardized DVARS were annotated as motion outliers. In addition, six principal components of a combined CSF and white matter signal accounting for the most variance were extracted using aCompCor (Behzadi et al., 2007) following highpass filtering (128-s cutoff) with discrete cosine filters. The BOLD runs were then slice-time corrected (AFNI *3dTshift*; Cox, 1996), and resampled onto original native space using custom methodology of fMRIPrep that applies all correction transformations in a single interpolation step. Additional details on the fMRIPrep pipeline may be found in the online documentation (<https://fmriprep.org/en/1.5.9/>).

Following preprocessing in fMRIPrep, we masked the preprocessed BOLD data for each participant with the intersection of the average T1-reference brain mask with the average BOLD reference mask. In final preparation of the MRI data for analysis with FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl; version 6.00), the following prestatistical processing was performed in FSL's FEAT (FMRI Expert Analysis Tool): registration of the T1w-reference map and co-registration of the BOLD reference data to MNI152 space using *FLIRT* with 12 degrees of freedom (Jenkinson and Smith, 2001; Jenkinson et al., 2002), spatial smoothing using a Gaussian kernel of FWHM 5 mm, and grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor.

Confound regressors consisting of the following MRIPrep-derived factors were prepared for functional denoising of individual BOLD runs: six aCompCor components, cosine filters for temporal filtering, 6 rotation and translation parameters and FD and spike regressors to exclude time points with excessive motion (>0.3 mm FD or >1.5 standardized DVARS). MRIQC (Esteban et al., 2019) was used as a preliminary check of data quality. Scan runs were excluded from analysis if $>20\%$ of TRs exceeded a framewise displacement of 0.3 mm. Only a single run (functional run 2, day 1) from one participant was excluded with this threshold.

fMRI analysis

fMRI analysis of the processed data was conducted using FEAT. Individual-level time-series statistical analyses were conducted using FILM with local autocorrelation correction (Woolrich et al., 2001). Separate regressors were specified for the experimental conditions of primary interest (CS+CC, CS+EXT, CS–) in each learning phase (threat acquisition: CS+s > CS–, CS– > CS+s, extinction: CS+CC > CS+EXT, and renewal tests: CS+CC > CS+EXT), by convolving the stimulus function with a double- γ hemodynamic response function (HRF), and adding a temporal derivative. Additional covariates included the electrical shock (following CS+ trials, during acquisition), positive pictures (following CS+CC trials, during CC), and confound regressors derived from fMRIPrep (described above). The higher-level analysis averaged contrasts estimates in each learning phase (acquisition, extinction/CC, and the renewal tests), and was conducted using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 (Beckmann et al., 2003; Woolrich et al., 2004; Woolrich, 2008). Whole-brain Z (Gaussianized T/F) statistic images were thresholded nonparametrically using clusters determined by $Z > 3.1$ and a (corrected) cluster significance threshold of $p = 0.05$ (Worsley, 2001). A left superficial amygdala mask from the Juelich histologic atlas (Amunts et al., 2005; Eickhoff et al., 2005), with a probability threshold of 30%, was used as a prethresholding mask for analysis of 24-h renewal. We then performed a small-volume correction (SVC) within this mask identified at $Z > 3.1$ and cluster corrected at $p = 0.05$ (Worsley, 2001). Anatomical labels in the tables of activation were obtained by converting significant cluster coordinates in MNI space to Talairach space using GingerALE 3.0.2 (<https://www.brainmap.org/>; Laird et al., 2010), and subsequently using Talairach Client (Lancaster et al., 2000).

Region of interest (ROI) selection

A priori ROIs for parameter estimate analysis included brain regions that are reliably characterized in meta-analyses of Pavlovian threat

Table 1. Single group average (paired *t* test) whole-brain contrasts during threat acquisition, identified at $Z > 3.1$ (cluster-corrected $p < 0.05$)

Contrast	Region	MNI coordinates			Size (voxels)	Cluster <i>P</i>	Cluster <i>Z</i>
		<i>x</i>	<i>y</i>	<i>z</i>			
CS+s > CS–	Inferior occipital gyrus, BA 19	–46	–75	2	13676	3.50E-19	7.02
	Inferior temporal gyrus, BA 37	51	–71	2	11861	2.21E-17	7.96
	Cuneus, BA 18	7	–94	20	3206	7.15E-07	5.32
	Precuneus, BA 7	–8	–70	46	2351	1.90E-05	4.58
	Superior frontal gyrus, BA 6	5	8	70	1541	0.000612	4.72
	Posterior cingulate, BA 23	–1	–27	22	952	0.0114	4.56
	Inferior frontal gyrus, BA 47	–33	23	–10	926	0.0132	4.27
CS– > CS+s	Postcentral gyrus, BA 3	–13	–33	74	24383	1.56E-28	6.04
	Clastrum	–35	–9	9	5725	1.99E-10	5.06
	Lingual gyrus, BA 18	20	–82	–8	4205	2.33E-08	5.38
	Clastrum	38	–4	4	3305	5.36E-07	6.07
	Lingual gyrus, BA 8	–17	–94	–9	3235	6.56E-07	5.44
	Medial frontal gyrus, BA 10	–14	50	–1	2516	9.89E-06	4.4
	Superior temporal gyrus, BA 41	58	–17	6	1493	0.000765	4.13
	Lingual gyrus, BA 18	–20	–58	8	1373	0.00135	5.15
	Parahippocampal gyrus, hippocampus	–28	–15	–24	1072	0.00606	4.79
	Posterior cingulate, BA 30	22	–53	8	1019	0.008	3.92
	Postcentral gyrus, BA 3	–44	–25	55	912	0.0188	3.86
	Middle temporal gyrus, BA 37	–59	–60	0	825	0.0297	4.56
	No significant activity	—	—	—	—	—	—
CS+CC > CS+EXT							
CS+EXT > CS+CC							

conditioning and extinction studies, and are involved in threat expression. Specifically insula (MNI R 40,16,–2; L –40,18,–2), dACC (MNI 8,18,42) and thalamus (MNI R 2,–18,6; L –2, 18, –4) coordinates were taken from (Fullana et al., 2016) and periaqueductal gray (PAG) (MNI 1,–29,–12) coordinates from (Linnman et al., 2012). For each of these threat ROIs, a sphere was drawn around peak coordinates with a radius of 10 mm. Parameter estimates for ROIs were extracted using FSL's Featquery tool and input to R Studio for further analyses with paired *t* tests.

The amygdala and nucleus accumbens (NAc) were a priori ROIs for functional connectivity analyses. The amygdala was not identified in the Fullana and coworkers meta-analyses. Therefore, in accordance with human neuroimaging research on the circuitry of amygdala subregions (Roy et al., 2009, 2013; Koch et al., 2016), standardized amygdala ROIs [basolateral amygdala (BLA) and central amygdala (CeM)] were defined using the Juelich histologic atlas (Amunts et al., 2005; Eickhoff et al., 2005) as implemented in FSL. Following (Koch et al., 2016), voxels were included if they had a 50% or higher probability of belonging to the CeM, but because of signal drop-out in the temporal cortex, we used a more stringent threshold of 70% for the BLA. Significantly, in cases of voxel overlap, voxels were assigned to the region for which they had the highest probability of inclusion. Likewise, a standardized NAc mask was derived from the Harvard-Oxford Subcortical Probability Atlas, thresholded at 50%. Standardized masks were then transformed into individual functional space.

The vmPFC, an a priori ROI for functional connectivity and RSA analyses, was defined functionally from the CS– > CS+s contrast during acquisition. A 10-mm sphere was drawn around the coordinates of a significant cluster ($z < 3.1$, cluster corrected $p < 0.05$) corresponding to the medial frontal gyrus (MNI coordinates, –14, 50, –1; Table 1).

Task-based functional connectivity

We used generalized psychophysiological interaction (gPPI) to examine functional connectivity at the 24-h and approximately one-month renewal tests, in two a priori pathways (BLA→NAc and vmPFC→CeM). While biological directionality cannot be inferred from gPPI, we use arrows to illustrate the statistical directionality of gPPI results, i.e., the functional connection between the seeds (BLA and vmPFC) and their respective targets (NAc and CeM). The timeseries for the seeds (BLA and vmPFC) were extracted using FSL's means command and input as regressors in the model. Timeseries extraction took place in the processed functional data from our initial GLM analyses. Interactions between the physiological variable (i.e., the seed's respective timeseries) and each of

the psychological variables (i.e., CS+CC, CS+EXT, and CS–) were computed and included in the design matrix as the variables of interest.

Mean *z* scores of connectivity from target ROIs were extracted using Featquery for each regressor of interest (CS+CC, CS+EXT, and CS–), at both the 24-h and approximately one-month renewal tests. These connectivity means were then input into R studio for further statistical analyses.

RSA

In order to facilitate RSA, LS-S style betaseries were computed for each scanner run (Mumford et al., 2012, 2014). Within each scanner run trial-specific β images were iteratively computed in FEAT using a design matrix which modeled a single trial of interest and all of trials as regressors of no interest based on trial type (e.g., separate CS+CC, CS+EXT, CS– regressors of no interest). FEAT settings were identical as in our univariate analysis, with the exception that no spatial smoothing was applied to respect the boundaries of our a priori ROIs in multivariate analyses. In addition to these trial-specific β estimates, we also generated conventional estimates of average activity for each CS type during each phase (i.e., all CS+CC in one regressor of interest), again without spatial smoothing. For the renewal sessions, separate regressors were used to model the early versus late trials.

RSA was accomplished using custom Python code. The goal of our analyses was to iteratively compare multivoxel patterns of activity in the vmPFC, between memory encoding in the extinction/CC session, recent renewal, and remote renewal. In order to reduce noise across the multivoxel pattern before estimating pattern similarity, each LS-S β image was weighted (multiplied) by the overall univariate activity estimate of the corresponding CS type and time point (Hennings et al., 2020, 2021; H. Kim et al., 2020; e.g., all images of CS+CC from early 24-h renewal were weighted with the average CS+CC pattern from the same time point). For each CS type, all of the LS-S images were entered into a representational similarity matrix, where each cell represents the Pearson's correlation of the multivoxel patterns of activity between two images in the vmPFC. For each CS type, the correlations were fisher-*z* transformed, and the average similarity was taken for our three comparisons of interest: extinction/CC encoding to recent renewal, extinction/CC encoding to remote renewal, and recent renewal to remote renewal. Average fisher-*z* similarity values were then exported to R studio for statistical analysis.

Analytic plan

All statistical analyses were conducted in the R environment (R Core Team, 2020). Data were analyzed using repeated measures ANOVA,

with the *ez* package (Lawrence, 2016), and included factors for CS Type (CS+CC, CS+EXT, and CS−) and time (e.g., first and second half of phase, or recent and remote renewal phases) where appropriate. Greenhouse–Geisser (GG) correction was applied when sphericity was violated. Main effects or interactions were followed by *post hoc* two-tailed paired *t* tests.

Results

Behavioral results

Threat acquisition and extinction

Analyses of mean shock expectancy and SCRs during the acquisition and extinction phases on day 1 were separated into the first and second half of trials (i.e., early/late; Fig. 1B,C). Shock expectancy was significantly higher for both CS+s in comparison to CS− during both early and late trials of acquisition (all $p < 0.001$; Fig. 1B). A repeated-measures ANOVA of SCR during acquisition revealed a main effect of CS type ($F_{(1.50,36.04)} = 11.462$, $p_{gg} < 0.001$, $\eta^2_G = 0.025$) and a main effect of early/late trials ($F_{(1,24)} = 21.194$, $p < 0.001$, $\eta^2_G = 0.053$), but no interaction ($p_{gg} = 0.071$). *Post hoc* paired *t* tests showed successful acquisition toward both CS+s, as SCRs were significantly higher for CS+CC versus CS− and CS+EXT versus CS− (all $p < 0.01$; Fig. 1C). Importantly, shock expectancy and SCR did not differ between CS+s during acquisition. Thus, participants successfully acquired equivalent expectancy responses and conditioned arousal toward both CS+s.

A repeated measures ANOVA of shock expectancy during extinction revealed a significant main effect of CS Type ($F_{(1.89,45.42)} = 12.810$, $p_{gg} < 0.001$, $\eta^2_G = 0.143$), early/late trials ($F_{(1,24)} = 10.440$, $p = 0.004$, $\eta^2_G = 0.066$) and an interaction of CS type by early/late trials ($F_{(1.57,37.64)} = 5.514$, $p_{gg} = 0.013$, $\eta^2_G = 0.018$). While mean shock expectancy ratings were still significantly higher for CS+s in comparison to CS− during the first half (all $p < 0.001$), and second half (all $p \leq 0.01$) of extinction, there was a significant decrease in shock expectancy for CS+EXT stimuli from the first to the last half of extinction ($t_{(24)} = 5.073$, $p < 0.001$, 95% CI [0.133, 0.314]), but not for CS+CC stimuli ($p = 0.069$; Fig. 1B). A comparison of the decrease in shock expectancy from the first to the second half of extinction between CS+s revealed a significant difference ($t_{(24)} = 2.178$, $p = 0.039$, 95% CI [0.005, 0.181]), indicating that the decrease of shock expectancy to CS+EXT stimuli is larger, as compared with the temporal change that is observed to the CS+CC stimuli.

A repeated-measures ANOVA of SCR means from extinction showed no effect of CS Type ($p_{gg} = 0.471$), no effect of early/late trials ($p = 0.237$), nor an interaction ($p_{gg} = 0.786$), indicating successful diminishment of conditioned SCRs via the absence of shock (Fig. 1C).

Twenty-four-hour threat renewal test

Repeated measures ANOVA of threat expectancy during 24-h renewal revealed a main effect of CS type ($F_{(1.74,41.82)} = 9.80$, $p_{gg} < 0.001$, $\eta^2_G = 0.070$). Mean shock expectancy during early 24-h renewal (first four trials) was higher for both CS+s in comparison to CS− (all $p < 0.01$), and there were no differences between CS+s ($p = 0.387$; Fig. 1B).

Notably, given the limited sensitivity of a 2AFC, we did not expect to see differences between CS+s within sessions. As such, we assessed expectancy during the end of extinction, and compared it to expectancy during the renewal phase. A repeated measures ANOVA with a factor of CS type and phase (last half of extinction and early renewal), revealed a main effect of CS

Type ($F_{(1.73,41.56)} = 11.26$, $p_{gg} < 0.001$, $\eta^2_G = 0.115$), a trend toward a significant main effect of phase ($F_{(1,24)} = 4.04$, $p = 0.056$, $\eta^2_G = 0.010$), but no significant CS type by phase interaction ($p_{gg} = 0.072$). *Post hoc* paired *t* tests revealed that expectancy for CS+EXT significantly increased ($t_{(24)} = 3.894$, $p < 0.001$, 95% CI [0.075, 0.245]) from late extinction to early renewal, but was not different between phases for neither CS+CC ($p = 0.720$) nor CS− stimuli ($p = 0.818$). But, a comparison of the change in shock expectancy from the end of extinction to early renewal between CS+s (CS+EXT vs CS+CC) revealed no significance difference ($p = 0.082$), indicating that the strength of renewal to the CS+EXT is not different from the CS+CC. Nevertheless, at 24 h, participants exhibited renewal of shock expectancy toward items from the category that underwent standard extinction, but not toward items from the control category, nor the CC category.

Repeated-measures ANOVA of SCRs during 24-h renewal revealed a main effect of CS type ($F_{(1.81,43.39)} = 3.732$, $p_{gg} = 0.036$, $\eta^2_G = 0.030$; Fig. 1C). *Post hoc* paired *t* tests revealed greater mean SCRs toward CS+EXT versus CS− ($t_{(24)} = 2.374$, $p = 0.026$, 95% CI [0.018, 0.255]), but no difference between CS+CC versus CS− ($p = 0.186$), nor CS+CC versus CS+EXT ($p = 0.122$). Thus, while SCRs did not differ between CS+s, participants expressed heightened conditioned arousal to items from the CS+EXT category as compared with items from the CS− category, but this difference was eliminated for CS+CC stimuli.

An ANOVA comparing physiological arousal at the end of extinction to early renewal revealed a main effect of CS Type ($F_{(1.90,45.54)} = 4.099$, $p_{gg} = 0.025$, $\eta^2_G = 0.005$), no main effect of phase ($p = 0.062$) and no significant CS type by phase interaction ($p_{gg} = 0.354$). *Post hoc* paired *t* tests revealed that conditioned arousal for CS+EXT stimuli was marginally higher ($t_{(24)} = 2.037$, $p = 0.053$, 95% CI [−0.004, 0.5693]) from late extinction to early renewal, but was not different between phases for neither CS+CC ($p = 0.081$) nor CS− ($p = 0.089$) stimuli.

One-month threat renewal test

Approximately one month later, participants did maintain slightly elevated shock expectancy to each CS+ versus the CS− (Fig. 1B). While a repeated measures ANOVA of mean shock expectancy revealed no significant main effect of CS Type ($p_{gg} = 0.080$), *post hoc* paired *t* tests revealed significantly higher expectancy for CS+EXT in comparison to the CS− ($t_{(24)} = 2.336$, $p = 0.029$, 95% CI [0.0195, 0.328]), a trend toward significantly higher shock expectancy for CS+CC in comparison to the CS− ($t_{(24)} = 2.005$, $p = 0.057$, 95% CI [−0.001, 0.354]), and no differences between CS+s ($p = 1$). Interestingly, autonomic arousal to each CS was exceptionally low (Fig. 1C). A repeated measures ANOVA of mean SCR revealed no main effect of CS type ($p_{gg} = 0.395$). Thus, one month later, participants expressed some retrieval of day 1 CS+ shock contingencies, but did not display heightened physiological arousal toward CS+ items.

Neuroimaging results

Univariate analysis

Extinction. Univariate whole-brain fMRI analysis focused on the extinction and renewal test phases (see Tables 1–Tables 4 for full results from each experimental phases). During extinction, a contrast of CS+CC > CS+EXT revealed significant clusters only in the cuneus and precuneus (Table 2). The inverse contrast (CS+EXT > CS+CC) revealed significant clusters in brain regions traditionally involved in maintaining and expressing

Table 2. Single group average (paired *t* test) whole-brain contrasts during extinction identified at $Z > 3.1$ (cluster-corrected $p < 0.05$)

Contrast	Region	MNI coordinates			Size (voxels)	Cluster P	Cluster Z
		x	y	z			
CS+s > CS–	Superior frontal gyrus, BA 6	–2	25	62	3335	5.36E-07	5.06
	Inferior temporal gyrus, BA 37	52	–70	0	1682	0.000352	4.77
	Precuneus, BA 7	2	–74	50	1216	0.00312	4.87
CS– > CS+s	Declive	19	–76	–10	59754	0	6.99
	Medial frontal gyrus, BA 6	–4	–5	55	39827	1.75E-39	5.79
	Insula, BA 13	36	–26	16	13300	1.20E-18	6.2
	Insula, BA 13	–37	–26	15	6292	4.77E-11	5.11
	Inferior occipital gyrus, BA 17	–13	–94	–7	4245	2.44E-08	5.38
	Precentral gyrus, BA 4	55	–12	37	2726	4.95E-06	4.36
	Precentral gyrus, BA 6	–44	0	30	1786	0.000223	4.62
	Middle temporal gyrus, BA 19	38	–63	19	1306	0.00201	4.47
	Precuneus, BA 7	24	–54	47	1231	0.0029	4.36
	Anterior cingulate, BA 32	8	50	–9	1065	0.00669	4.08
	Precentral gyrus, BA 6	45	3	31	962	0.0115	4.28
	Inferior frontal gyrus, BA 47	28	31	–11	738	0.0396	4.6
	Superior temporal gyrus, BA 22	–51	1	–9	714	0.0454	3.74
	Precuneus, BA 7	4	–65	59	1949	5.10E-05	6.3
	Cuneus, BA 19	3	–85	36	680	0.04	4.56
CS+CC > CS+EXT	Culmen	18	–37	–13	117657	0	6.68
	Postcentral gyrus, BA 2	–36	–24	45	12787	1.95E-19	5.84
	Inferior frontal gyrus, BA 9	47	14	26	6756	1.90E-12	4.4
	Cingulate gyrus, BA 24	–6	4	49	6719	2.12E-12	5.34
	Clastrum	32	24	–1	2292	1.12E-05	5.68
	Precuneus, BA 7	28	–52	45	1722	0.000146	5.19
	Superior temporal gyrus, BA 22	61	–39	14	1371	0.000814	4.48
	Precentral gyrus, BA 4	46	–15	36	993	0.00617	4.15
	Superior temporal gyrus, BA 38	50	9	–19	791	0.0201	5.04
CS+EXT > CS+CC							

threat (Fullana et al., 2016; Table 2; Fig. 2A). To further characterize these fMRI results, we extracted activity associated with each stimulus type (CS+CC, CS+EXT and CS–) from a priori ROIs putatively involved in acquisition and extinction of threat (Fullana et al., 2016, 2018; i.e., dACC, insula, thalamus, and PAG). We focused these ROI analyses on the second half of extinction. This revealed diminished activity to the CS+CC in comparison to the CS+EXT (Fig. 2C), indicating that CC attenuated activity in regions involved in maintaining and expressing threat expectations relative to merely omitting the shock.

Additionally, we compared the outcome of CS+CC (positive picture) to the outcome of CS+EXT (shock omission) during extinction. As expected, a contrast of CS+CC outcome versus CS+EXT outcome revealed activity in the visual cortex for visual scenes, but there were no regions showing significant activation for shock omission alone versus the positive picture.

Twenty-four-hour threat renewal test. Univariate fMRI analysis of the CS+EXT > CS+CC and CS+CC > CS+EXT contrasts did not reveal any significant activity that survived whole-brain correction for multiple comparisons. A more liberal exploratory threshold of $p < 0.001$ (uncorrected) for the CS+CC > CS+EXT contrast revealed a cluster in the left amygdala (MNI –16, –7, –21; 27 voxels, $z = 3.49$, $p_{uncorrected} < 0.001$; cluster corrected at $p < 0.05$ with SVC; Table 3; Fig. 2B). No regions emerged at this liberal threshold for the inverse contrast (CS+EXT > CS+CC).

One-month threat renewal test. No regions emerged at the whole-brain level for the univariate contrasts CS+CC > CS+EXT or CS+EXT > CS+CC at one month, even using a liberal threshold ($p < 0.001$, uncorrected).

Functional connectivity

A BLA → NAc circuit for retrieval of rewarded extinction

To examine the involvement of fMRI derived amygdalar connections, we conducted a gPPI analysis during recent and remote threat renewal tests (Fig. 3A). This analysis was inspired by neurobiological evidence that a BLA to NAc circuit preferentially supports reduced threat relapse of rewarded extinction (Correia et al., 2016). The seed region was an anatomically defined BLA, and the target region was an anatomically defined NAc.

Twenty-four hours following extinction, a repeated measures ANOVA revealed a main effect of CS Type ($F_{(1.93,46.24)} = 5.781$, $p_{gg} = 0.006$, $\eta^2_G = 0.085$). *Post hoc* paired *t* tests revealed that connectivity between the BLA and the NAc, at this recent time point, was enhanced for stimuli from the CS+CC category, in comparison to stimuli from the CS+EXT category ($t_{(24)} = 3.320$, $p = 0.003$, 95% CI [0.217, 0.932]) and the CS– category ($t_{(24)} = 2.631$, $p = 0.015$, 95% CI [0.091, 0.756]). One month after extinction, there were no differences in connectivity between CS types (all $p > 0.3$). But, comparing across renewal test intervals (recent vs remote), *post hoc* paired *t* tests revealed that BLA → NAc connectivity significantly diminished for CS+CC stimuli from the 24-h to the approximately one-month renewal test ($t_{(22)} = -2.087$, $p = 0.048$, 95% CI [–0.990, –0.003]).

A vmPFC → CeM circuit is recruited for CS+ stimuli at a remote renewal test

The medial PFC is considered a critical region that inhibits conditioned defensive responses via projections that inhibit the central nucleus of the amygdala (CeM; Ghashghaei and Barbas, 2002; McDonald et al., 1996). This circuit is considered critical for successful extinction retrieval. We therefore conducted a gPPI during recent and remote threat renewal tests using the vmPFC as the seed region and an anatomically defined region of

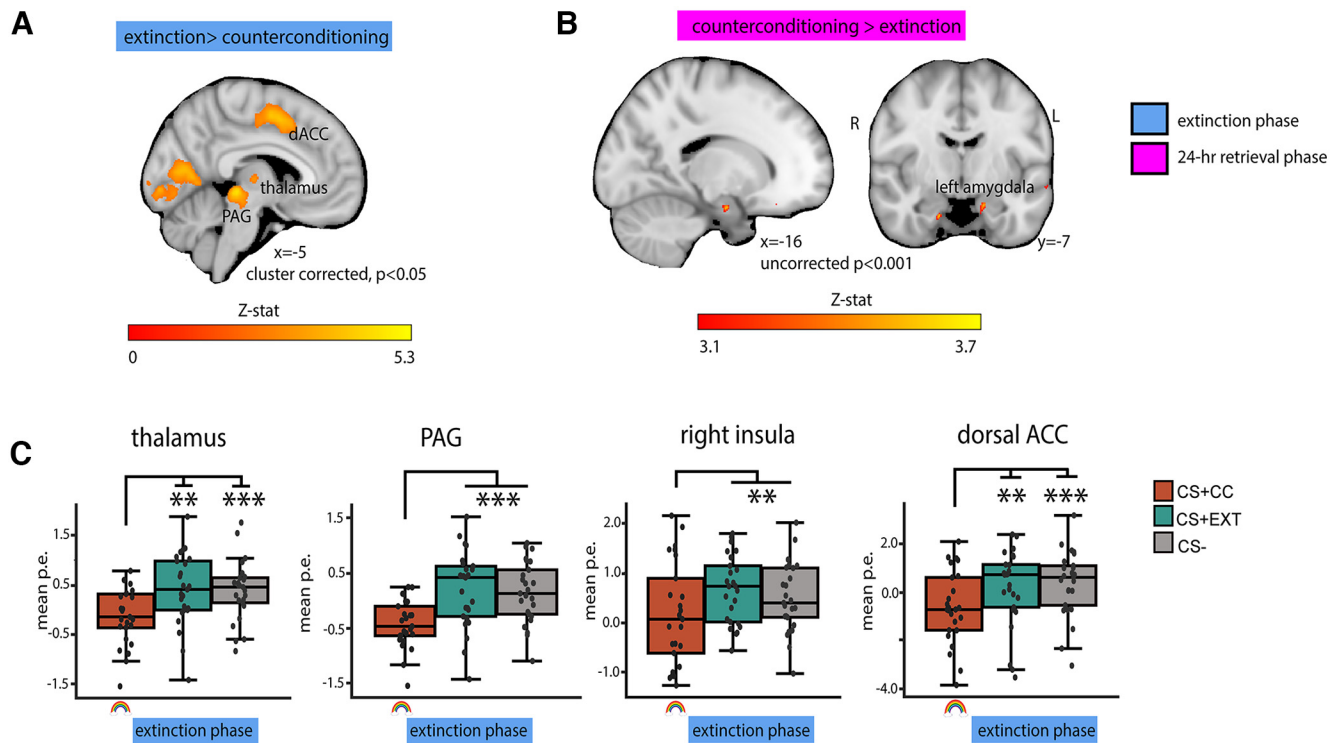


Figure 2. CC was associated with reduced activity in threat ROIs during extinction, and enhanced amygdala activation during 24-h renewal. **A**, A whole-brain contrast of CS+EXT > CS+CC during the extinction phase, identified at $Z > 3.1$, cluster corrected at $p < 0.05$, revealed activity in regions traditionally associated with threat appraisal and expression (e.g., periaqueductal gray, dACC, insula and thalamus). **B**, A whole-brain contrast of CS+CC > CS+EXT during the 24-h renewal phase, identified at $p < 0.001$ uncorrected for multiple comparisons, revealed a cluster in the left amygdala. **C**, Parameter estimates extracted from a priori regions associated with threat (periaqueductal gray, dACC, insula, and thalamus), during the last half of extinction, revealed significantly lower activity for CS+CC stimuli in comparison to both CS+EXT and CS- stimuli. The rainbow, which represents the positive pictures, depict the outcome following CS+CC stimuli; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 3. Single group average (paired t test) whole-brain contrasts during a 24-h threat renewal, identified at $Z > 3.1$ (cluster-corrected $p < 0.05$)

Contrast	Region	MNI coordinates			Size (voxels)	Cluster P	Cluster Z
		x	y	z			
CS+s > CS-	No significant activity	—	—	—	—	—	—
CS- > CS+s	No significant activity	—	—	—	—	—	—
CS+CC > CS+EXT (SVC of left amygdala, cluster corrected $p < 0.05$)	Left amygdala	-16	-7	-21	27	0.0185	3.49
CS+EXT > CS+CC	No significant activity	—	—	—	—	—	—

Table 4. Single group average (paired t test) whole-brain contrasts during approximately one month of fear retrieval, identified at $Z > 3.1$ (cluster-corrected $p < 0.05$)

Contrast	Region	x	y	z	Size (voxels)	Cluster P	Cluster Z
CS+s > CS-	No significant activity	—	—	—	—	—	—
CS- > CS+s	Precentral gyrus, BA 4	48	-12	42	3335	1.90E-08	5.06
	Paracentral lobule, BA 3	21	-32	58	1682	2.58E-08	4.77
	Declive	18	-81	-13	1216	0.01	4.87
CS+CC > CS+EXT	No significant activity	—	—	—	—	—	—
CS+EXT > CS+CC	No significant activity	—	—	—	—	—	—

the CeM as the target region (Fig. 3B). The vmPFC was functionally defined based on a medial frontal gyrus cluster from the CS- > CS+ contrast during acquisition (Table 1), as anatomic labels for the vmPFC are variable across studies of Pavlovian conditioning and extinction.

Twenty-four hours following extinction, *post hoc* paired t tests revealed that connectivity between the vmPFC→CeM was heightened for CS+EXT stimuli versus CS+CC ($t_{(24)} = 2.999$,

$p = 0.006$, 95% CI [0.140, 0.755]). One month following extinction, *post hoc* paired t tests revealed that connectivity between CS+CC and CS+EXT stimuli no longer differed ($p = 0.466$). At this remote time point, CS+CC stimuli ($t_{(22)} = 2.250$, $p = 0.035$, 95% CI [0.027, 0.661]) showed stronger vmPFC→CeM connectivity than the CS- stimuli, but there were no differences between CS+EXT and CS- stimuli ($p = 0.065$). Finally, there was a significant main effect of CS Type ($F_{(1,75,38.49)} = 5.93$, $p_{gg} = 0.008$, $\eta^2_G = 0.043$), and renewal test interval ($F_{(1,22)} = 5.24$, $p = 0.032$, $\eta^2_G = 0.047$), but no significant CS type by renewal interval interaction ($p_{gg} = 0.328$). *Post hoc* paired t tests revealed that vmPFC→CeM connectivity for CS+CC significantly increased from the 24-h to the one-month renewal test ($t_{(22)} = 3.370$, $p = 0.003$, 95% CI [0.239, 1.005]).

Multivariate RSA

Pattern similarity between extinction/CC memory encoding retrieval

To assess the fidelity of the extinction and CC memory traces over time, we used RSA (Kriegeskorte et al., 2008) to compare

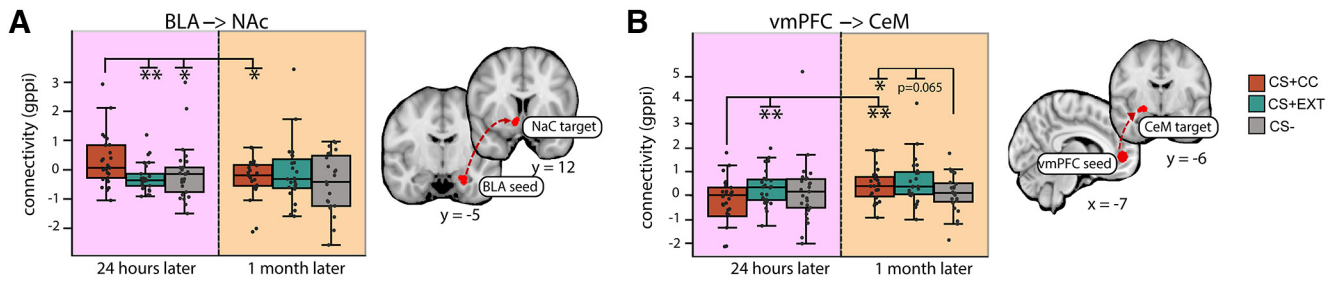


Figure 3. Functional connectivity during recent and remote renewal tests in two a priori pathways. **A**, Functional connectivity using the BLA as a seed region and the NAc as a target region, was enhanced for CS+CC stimuli during 24-h renewal, but was not different between CS types at approximately one-month renewal. **B**, Functional connectivity using the vmPFC as a seed region and the CeM as a target region, was enhanced for CS+EXT, in comparison to CS+CC stimuli during 24-h renewal. At approximately one month, connectivity for CS+CC stimuli significantly increased and was not different from CS+EXT stimuli. At this remote time point, both CS+s were associated with a greater functional vmPFC→CeM connection than CS- stimuli; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

patterns of fMRI activity during extinction/CC and 24-h and one-month renewal tests. We focused this analysis on the vmPFC, as this region is associated with successful extinction recall in humans (Phelps et al., 2004; Milad et al., 2007). Voxel-wise patterns of activity elicited by CS+CC, CS+EXT, and CS- stimuli, were correlated with the pattern of activity elicited by novel stimuli from the same categories at the renewal test 24 h (extinction → 24-h renewal), approximately one month later (extinction → one-month renewal), and across renewal sessions (24-h renewal → one-month renewal). Notably, one innovation to the category-conditioning design (Dunsmoor et al., 2014; Hennings et al., 2020) is that participants are exposed to new category exemplars composing each CS category. Thus, pattern similarity cannot be driven simply by perceptual overlap of CSs, as different basic level items are presented at each phase.

A CC memory trace is stable in the vmPFC from encoding to recent and remote renewal tests

A repeated measures ANOVA on pattern similarity from encoding to recent renewal (extinction→24-h renewal), revealed a main effect of CS Type ($F_{(1.75,41.88)} = 4.20$, $p_{gg} = 0.026$, $\eta^2_G = 0.081$; Fig. 4A). *Post hoc* paired *t* tests revealed that at 24-h similarity from encoding to retrieval in the vmPFC was selectively enhanced for CC stimuli in comparison to CS+EXT stimuli ($t_{(24)} = 2.169$, $p = 0.040$, 95% CI [0.003, 0.110]) and CS- stimuli ($t_{(24)} = 2.491$, $p = 0.020$, 95% CI [0.001, 0.105]). At approximately one month (extinction → one-month renewal), neural similarity for CS+CC stimuli was enhanced in comparison to CS- stimuli ($t_{(22)} = 2.147$, $p = 0.043$, 95% CI [0.002, 0.093]; Fig. 4B). Notably, memory traces from the extinction phase on day 1 did not significantly change from recent to remote renewal, as a repeated measures ANOVA with factors of CS type and renewal phase (extinction → 24-h renewal and extinction → one-month renewal) revealed no main effect of phase ($p_{gg} = 0.286$), a significant main effect of CS Type ($F_{(1.61,35.35)} = 5.73$, $p_{gg} = 0.011$, $\eta^2_G = 0.077$), but no CS type by phase interaction ($p_{gg} = 0.702$; Fig. 4A,B). Thus, both at recent and remote timepoints, the CC memory trace was stable in the vmPFC.

Similarity patterns in the vmPFC across recent and remote renewal are enhanced for CC stimuli

A repeated measures ANOVA of pattern similarity from recent to remote renewal (24-h renewal session → one-month renewal session) revealed no main effect of CS Type ($p_{gg} = 0.083$). *Post hoc* paired *t* tests revealed that across renewal phases, similarity was marginally enhanced for CS+CC in comparison to CS-

stimuli ($t_{(22)} = 2.047$, $p = 0.052$, 95% CI [-0.001, 0.155]), but not in comparison to CS+EXT stimuli ($p = 0.887$; Fig. 4C).

Discussion

As extinction is a transient form of inhibitory learning, there is interest in optimized strategies that more effectively inhibit relapse of extinguished threat. CC may be more effective than standard extinction (Keller et al., 2020), but the neurobehavioral mechanisms of CC in humans have remained unclear. Further, to our knowledge, the long-term neurobehavioral effects of threat attenuation strategies (more than one week) have remained unexamined in humans. Here, we found that, in comparison to standard extinction, rewarded extinction using CC attenuated activity in regions associated with threat appraisal and expression and reduced 24-h conditioned responses. Twenty-four-hour renewal was accompanied by enhanced functional connectivity between the BLA and NAc for stimuli from the CC category, and connectivity between the vmPFC and CeM for stimuli from the standard extinction category. One-month renewal was associated with reduced conditioned responses and accompanied by connectivity between vmPFC and CeM for both extinction strategies. RSA showed that memory traces of CC are stable in the vmPFC across recent and remote time points.

An overarching question about CC is whether it should simply be considered another form of extinction or whether it operates through different neural mechanisms (Keller et al., 2020). Whole-brain univariate analyses did not reveal the vmPFC nor the NAc, two major ROIs, to be strongly, nor differentially, activated during extinction or renewal, between CS+CC and CS+EXT items (Tables 2 and 3, respectively). Nevertheless, we found that in comparison to standard extinction, CC attenuated activity in regions associated with threat appraisal and expression (insula, thalamus, dACC, PAG), suggesting that providing a positive experience during extinction may facilitate safety learning. Notably, this finding is consistent with a recent fMRI study in which a shock was replaced with a neutral outcome (a tone; Dunsmoor et al., 2019). As previously suggested, replacing shock with a nonaversive stimulus might reduce ambiguity and uncertainty otherwise generated when a shock is merely omitted (Dunsmoor et al., 2015a). Future research should consider the role individual differences play in neural activity of extinction versus CC. For example, do individual differences in activity during aversive learning predict subsequent responding during safety learning? This approach could be applied to psychiatric populations with anxiety-related disorders or PTSD, where one possibility is that neural responses during acquisition may

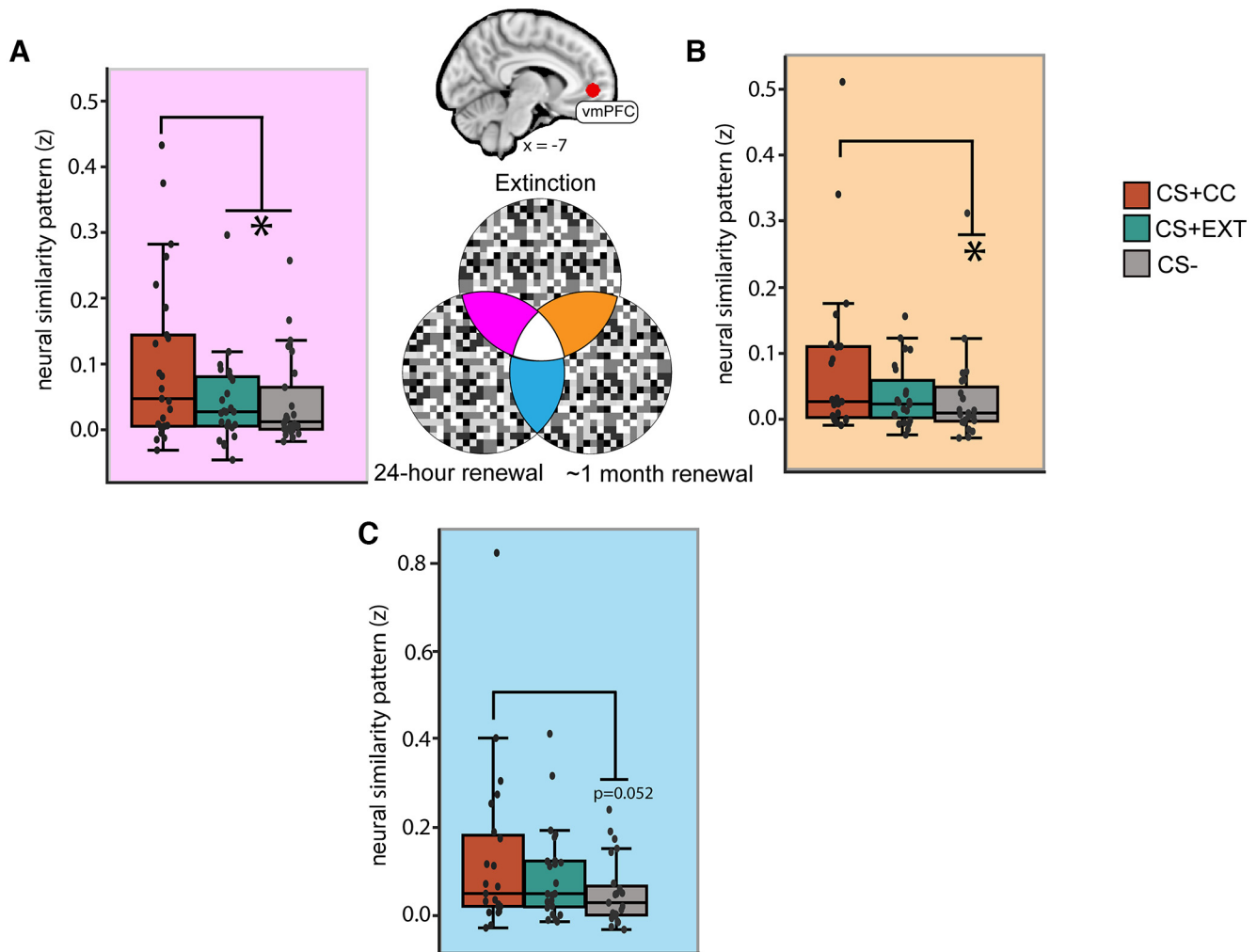


Figure 4. Stimuli that underwent CC were associated with a heightened pattern of similarity in the vmPFC. **A**, Pattern similarity from encoding to recent renewal (extinction→24-h retrieval) in the vmPFC was enhanced for the CS+CC category in comparison to both the CS+EXT and CS- categories. **B**, Pattern similarity from encoding to remote renewal (extinction→ approximately one-month renewal) in the vmPFC was enhanced for the CS+CC category in comparison to the CS- category. **C**, Pattern similarity from recent to remote renewal (extinction→24-h renewal) in the vmPFC was marginally enhanced for the CS+CC category in comparison to the CS- category; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

predict the effectiveness of one treatment over another (CC vs standard extinction).

At 24-h and one-month renewal tests, there was a surprising lack of differentiation in whole-brain fMRI activity between the retrieval of CC and standard extinction memories. A more liberal statistical threshold did reveal greater activity for CC in the left amygdala at 24-h renewal. On one hand this finding may seem counterintuitive, given that the amygdala is critical for threat learning and expression (Phelps and LeDoux, 2005) and conditioned responses were slightly more attenuated by CC. However, the amygdala also responds to rewarding stimuli (J. Kim et al., 2016; Beyeler et al., 2018; X. Zhang and Li, 2018; X. Zhang et al., 2020), and the BLA contains neural populations that code for extinction memory (Herry et al., 2010) and neurons that respond to reward overlap with those involved in extinction (X. Zhang et al., 2020). Thus, it is possible the amygdala plays an important role in retrieving reward-associations connected with the memory of CC.

We used a functional connectivity analysis to further assess the neural differences between CC and standard extinction in two a priori pathways: BLA→NAc and vmPFC→CeM (seed to target). At 24-h, functional connectivity between the vmPFC and CeM was enhanced for standard extinction in comparison to

CC; in contrast, functional connectivity between the BLA and the NAc was enhanced for CC in comparison to standard extinction. Our findings can be interpreted in the well-explored neurocircuitry of threat extinction in rodents. Previous research in rodents has shown that infralimbic (rodent homolog of vmPFC) projections to the BLA excite GABAergic intercalated cells that inhibit CeM neurons thereby inhibiting conditioned responses (Amano et al., 2010; Pape and Pare, 2010; Strobel et al., 2015). Moreover, a BLA to NAc circuit has been identified during rewarded-extinction in rats, and is associated with reduced threat relapse (Correia et al., 2016). Further evidence for the role of the BLA-to-NAc circuit comes from recent studies on rescuing behavioral deficits induced by chronic stress (Dieterich et al., 2021; Sun et al., 2021). Collectively, the present results help extend rodent neurobiological findings to humans and indicate that separate patterns of connectivity dissociate CC from standard extinction. Interestingly, connectivity between vmPFC and CeM was observed at one month for both CS types, suggesting that over longer periods of time, extinction recruits medial prefrontal inhibition of the amygdala regardless of the particular threat inhibition strategy. It is worth noting that the 24-h renewal test served as another standard extinction session, as positive outcomes were not included at test. Thus, the memory

of CC at the one-month test comprised a mix of CC (from day 1) and standard extinction (from day 2) that may be reflected in the switch in connectivity from BLA→NAc to vmPFC→CeM over time.

A multivariate RSA was used to further interrogate the fidelity of CC and standard extinction memories. The reactivation of neural activity patterns from extinction were enhanced by CC in the vmPFC both 24-h and one month later. It is notable that the vmPFC showed neural reactivation patterns for CC, as functional connectivity analyses indicated a vmPFC→amygdala connection was selectively enhanced 24-h following standard extinction but not CC. However, neurobiological evidence shows that activation of the BLA→NAc circuit by rewarded extinction increases activity in the IL to prevent threat relapse (Correia et al., 2016). Thus, CC may likewise enhance involvement of the vmPFC for storing long-term memory traces of safety.

The results from the one-month retrieval test were intriguing for several reasons. First, although shock expectancy returned slightly, autonomic arousal was remarkably low. This might indicate that both threat attenuation strategies were successful over the long-term. It is notable that functional connectivity between the vmPFC and the CeM was evident for both CS+ categories at one month (albeit only at a marginal level for CS+EXT), suggesting this is a mechanism for successfully reducing conditioned responses over long durations in humans. It is also important to note that participants were all reportedly free of psychopathology, and thus memory of laboratory conditioned threat might simply weaken over long durations in the healthy brain. This calls for future studies comparing the return of threat over longer intervals in patients with anxiety disorders, particularly PTSD. Threat conditioning is a popular model for PTSD (Mahan and Ressler, 2012), but immediate dysregulated responses to a CS may better reflect Acute Stress Disorder, which refers to the stress symptoms that arise in the first month after a traumatic event (Bryant, 2019). A key criteria in a PTSD diagnosis is the persistence of symptoms at least one month following the trauma (American Psychiatric Association, 2013). Importantly, acute stress disorder can develop when PTSD does not, and vice versa (Bryant, 2010). More research is warranted on the long-term endurance of different extinction strategies in clinical populations who display extinction retrieval deficits.

A limitation of the present study concerns the broad definition of “reward” for the outcomes used to replace shocks in CC. Simply put, were the pictures actually rewarding? More generally, by what operational definition should “reward” be applied? It is worth noting that the pictures used in this study were rated highly in positive valence by a separate group of participants. CC paradigms have employed a wide variety of appetitive outcomes (see Table 1; Keller et al., 2020), as well as different methodology for the subject to obtain the reward (e.g., passively delivered vs an instrumental behavior; Thomas et al., 2012). From a purely neural perspective, extinction does recruit reward-responsive dopaminergic systems (McNally et al., 2011; Kalisch et al., 2019; Salinas-Hernández and Duvarci, 2021). Further, the mere absence of an expected shock could be construed as a psychological reward (or at least a relief). It is therefore possible that facilitating extinction through any number of strategies simply promotes engagement of a threat-inhibition process that overlaps with reward-responsive neurocircuitry. One way future research could evaluate whether there is a unique effect of “reward,” would be to compare outcomes that vary in reward intensity, such as comparing positive pictures to primary reinforcers, like food or juice, or to compare passive delivery

versus instrumental responses (Thomas et al., 2012). Future design implementations could also include reinforcing multiple cues with different valence outcomes, such as CSs that are either always paired with negative or positive outcomes. As counter-CSs are first followed by negative and then positive outcomes, such study designs could allow for direct comparisons on how opposite valences can differentially affect neurobehavioral processes.

Insofar as Pavlovian extinction serves as a theoretical foundation for exposure therapy, and symptoms frequently return following treatment (Vervliet et al., 2013), examining the neurobehavioral endurance of different threat attenuation strategies is important. These results provide new evidence that the presence of a rewarding stimulus during extinction may boost threat attenuation through an amygdala-striatal pathway, and stabilize memory representations in the vmPFC over long time intervals. These results extend neurobiological findings on the overlap between reward and threat extinction from rodents to healthy humans. While neuroimaging research comparing these strategies in clinical populations is warranted, this type of research could serve as a foundation for translational efforts that result in a paradigm shift for exposure therapy.

References

- Abramowitz JS, Deacon BJ, Whiteside SP (2019) Exposure therapy for anxiety: principles and practice. New York: Guilford Publications.
- Alexandra Kredlow M, Fenster RJ, Laurent ES, Ressler KJ, Phelps EA (2022) Prefrontal cortex, amygdala, and threat processing: implications for PTSD. *Neuropsychopharmacology* 47:247–259.
- Amano T, Unal CT, Paré D (2010) Synaptic correlates of fear extinction in the amygdala. *Nat Neurosci* 13:489–494.
- American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, Ed 5. Arlington, VA.
- Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah N, Habel U, Schneider F, Zilles K (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl)* 210:343–352.
- Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* 12:26–41.
- Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in FMRI. *Neuroimage* 20:1052–1063.
- Behzadi Y, Restom K, Liu J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37:90–101.
- Beyeler A, Chang C-J, Silvestre M, Lévêque C, Namburi P, Wildes CP, Tye KM (2018) Organization of valence-encoding and projection-defined neurons in the basolateral amygdala. *Cell Rep* 22:905–918.
- Bouton ME (2002) Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry* 52:976–986.
- Bryant RA (2010) Acute stress disorder as a predictor of posttraumatic stress disorder: A systematic review. *J Clin Psychiatry* 71:381.
- Bryant RA (2019) Post-traumatic stress disorder: a state-of-the-art review of evidence and challenges. *World Psychiatry* 18:259–269.
- Correia SS, McGrath AG, Lee A, Graybiel AM, Goossens KA (2016) Amygdala-ventral striatum circuit activation decreases long-term fear. *Elife* 5:e12669.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Cox RW, Hyde JS (1997) Software tools for analysis and visualization of fMRI data. *NMR Biomed* 10:171–178.
- Craske MG, Hermans D, Vervliet B (2018) State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Phil Trans R Soc B* 373:20170025.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Dickinson A, Pearce JM (1977) Inhibitory interactions between appetitive and aversive stimuli. *Psychol Bull* 84:690–711.

- Dieterich A, Floeder J, Stech K, Lee J, Srivastava P, Barker DJ, Samuels BA (2021) Activation of basolateral amygdala to nucleus accumbens projection neurons attenuates chronic corticosterone-induced behavioral deficits in male mice. *Front Behav Neurosci* 15:17.
- Dunsmoor JE, Kragel PA, Martin A, LaBar KS (2014) Aversive Learning Modulates Cortical Representations of Object Categories. *Cerebral Cortex* 24:2859–2872.
- Dunsmoor JE, Campese VD, Ceceli AO, LeDoux JE, Phelps EA (2015a) Novelty-facilitated extinction: providing a novel outcome in place of an expected threat diminishes recovery of defensive responses. *Biol Psychiatry* 78:203–209.
- Dunsmoor JE, Niv Y, Daw N, Phelps EA (2015b) Rethinking extinction. *Neuron* 88:47–63.
- Dunsmoor JE, Kroes MCW, Li J, Daw ND, Simpson HB, Phelps EA (2019) Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. *J Neurosci* 39:3264–3276.
- Eickhoff SB, Stephan KE, Mohlbeg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ (2019) fMRIPrep: a robust pre-processing pipeline for functional MRI. *Nat Methods* 16:111–116.
- Fonov VS, Evans AC, McKinstry RC, Alml C, Collins D (2009) Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* 47:S102.
- Fullana M, Harrison B, Soriano-Mas C, Vervliet B, Cardoner N, Ávila-Parcet A, Radua J (2016) Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21:500–508.
- Fullana MA, Albajes-Eizaguirre A, Soriano-Mas C, Vervliet B, Cardoner N, Benet O, Radua J, Harrison BJ (2018) Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants. *Neurosci Biobehav Rev* 88:16–25.
- Gatzounis R, De Bruyn S, Van de Velde L, Meulders A (2021) No differences in return of pain-related fear after extinction and counterconditioning. *Emotion*. Advance online publication. Retrieved Jun 17, 2021. doi: 10.1037/emo0000960.
- Ghashghaei HT, Barbas H (2002) Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* 115:1261–1279.
- Giustino TF, Maren S (2015) The role of the medial prefrontal cortex in the conditioning and extinction of fear. *Front Behav Neurosci* 9:298.
- Green SR, Kragel PA, Fecteau ME, LaBar KS (2014) Development and validation of an unsupervised scoring system (autonomate) for skin conductance response analysis. *Int J Psychophysiol* 91:186–193.
- Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48:63–72.
- Hartley CA, Phelps EA (2010) Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology* 35:136–146.
- Hennings AC, McClay M, Lewis-Peacock JA, Dunsmoor JE (2020) Contextual reinstatement promotes extinction generalization in healthy adults but not PTSD. *Neuropsychologia* 147:107573.
- Hennings AC, Lewis-Peacock JA, Dunsmoor JE (2021) Emotional learning retroactively enhances item memory but distorts source attribution. *Learn Mem* 28:178–186.
- Herry C, Ferraguti F, Singewald N, Letzkus JJ, Ehrlich I, Lüthi A (2010) Neuronal circuits of fear extinction. *Eur J Neurosci* 31:599–612.
- Holmes NM, Leung HT, Westbrook RF (2016) Counterconditioned fear responses exhibit greater renewal than extinguished fear responses. *Learning & Memory* 23:141–150.
- Holtzman-Assif O, Laurent V, Westbrook RF (2010) Blockade of dopamine activity in the nucleus accumbens impairs learning extinction of conditioned fear. *Learn Mem* 17:71–75.
- Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–156.
- Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.
- Jones MC (1924) A laboratory study of fear: the case of Peter. *J Gen Psychol* 31:308–315.
- Josselyn SA, Frankland PW (2018) Fear extinction requires reward. *Cell* 175:639–640.
- Kalisch R, Gerlicher AM, Duvarci S (2019) A dopaminergic basis for fear extinction. *Trends Cogn Sci* 23:274–277.
- Keller NE, Dunsmoor JE (2020) The effects of aversive-to-appetitive counterconditioning on implicit and explicit fear memory. *Learn Mem* 27:12–19.
- Keller NE, Hennings AC, Dunsmoor JE (2020) Behavioral and neural processes in counterconditioning: past and future directions. *Behav Res Ther* 125:103532.
- Kim J, Pignatelli M, Xu S, Itohara S, Tonegawa S (2016) Antagonistic negative and positive neurons of the basolateral amygdala. *Nat Neurosci* 19:1636–1646.
- Kim H, Smolker HR, Smith LL, Banich MT, Lewis-Peacock JA (2020) Changes to information in working memory depend on distinct removal operations. *Nat Commun* 11:1–14.
- Koch SB, van Zuiden M, Nawijn L, Frijling JL, Veltman DJ, Olff M (2016) Intranasal oxytocin normalizes amygdala functional connectivity in post-traumatic stress disorder. *Neuropsychopharmacology* 41:2041–2051.
- Koizumi A, Amano K, Cortese A, Shibata K, Yoshida W, Seymour B, Kawato M, Lau H (2016) Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nat Hum Behav* 1:0006.
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—Connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Kroes MC, Tona KD, den Ouden HE, Vogel S, van Wingen GA, Fernández G (2016) How administration of the beta-blocker propranolol before extinction can prevent the return of fear. *Neuropsychopharmacology* 41:1569–1578.
- Laird AR, Robinson JL, McMillan KM, Tordesillas-Gutiérrez D, Moran ST, Gonzales SM, Ray KL, Franklin C, Glahn DC, Fox PT, Lancaster JL (2010) Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: validation of the Lancaster transform. *Neuroimage* 51:677–683.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120–131.
- Lawrence MA (2016) ez: easy analysis and visualization of factorial experiments. R package version 4.
- Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, Heitland I, Hermann A, Kuhn M, Kruse O (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.
- Linnman C, Moulton EA, Barmettler G, Becerra L, Borsook D (2012) Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage* 60:505–522.
- Luo R, Uematsu A, Weitemier A, Aquili L, Koivumaa J, McHugh TJ, Johansen JP (2018) A dopaminergic switch for fear to safety transitions. *Nat Commun* 9:2483.
- Mahan AL, Ressler KJ (2012) Fear conditioning, synaptic plasticity and the amygdala: implications for posttraumatic stress disorder. *Trends Neurosci* 35:24–35.
- Markowitz JC, Petkova E, Neria Y, Van Meter PE, Zhao Y, Hembree E, Lovell K, Biyanova T, Marshall RD (2015) Is exposure necessary? A randomized clinical trial of interpersonal psychotherapy for PTSD. *Am J Psychiatry* 172:430–440.
- McDonald A, Mascagni F, Guo L (1996) Projections of the medial and lateral prefrontal cortices to the amygdala: a *Phaseolus vulgaris* leucoagglutinin study in the rat. *Neuroscience* 71:55–75.
- McNally GP, Johansen JP, Blair HT (2011) Placing prediction into the fear circuit. *Trends Neurosci* 34:283–292.
- Milad MR, Quirk GJ (2012) Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol* 63:129–151.
- 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.
- Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, Zeidan MA, Handwerker K, Orr SP, Rauch SL (2009) Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol Psychiatry* 66:1075–1082.

- Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59:2636–2643.
- Mumford JA, Davis T, Poldrack RA (2014) The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage* 103:130–138.
- Pape HC, Pare D (2010) Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol Rev* 90:419–463.
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43:897–905.
- Phelps EA, LeDoux JE (2005) Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* 48:175–187.
- Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84:320–341.
- Raczka K, Mechias M, Gartmann N, Reif A, Deckert J, Pessiglione M, Kalisch R (2011) Empirical support for an involvement of the mesostriatal dopamine system in human fear extinction. *Transl Psychiatry* 1:e12.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.
- Reuter M, Rosas HD, Fischl B (2010) Highly accurate inverse consistent registration: a robust approach. *Neuroimage* 53:1181–1196.
- Ritchey M, Wing EA, LaBar KS, Cabeza R (2013) Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb Cortex* 23:2818–2828.
- Roy AK, Shehzad Z, Margulies DS, Kelly AC, Uddin LQ, Gotimer K, Biswal BB, Castellanos FX, Milham MP (2009) Functional connectivity of the human amygdala using resting state fMRI. *Neuroimage* 45:614–626.
- Roy AK, Fudge JL, Kelly C, Perry JSA, Daniele T, Carlisi C, Benson B, Castellanos FX, Milham MP, Pine DS, Ernst M (2013) Intrinsic functional connectivity of amygdala-based networks in adolescent generalized anxiety disorder. *J Am Acad Child Adolesc Psychiatry* 52:290–299.
- Salinas-Hernández XI, Duvarci S (2021) Dopamine in fear extinction. *Front Synaptic Neurosci* 13:10.
- Schiller D, Monfils MH, Raio CM, Johnson DC, LeDoux JE, Phelps EA (2010) Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* 463:49–53.
- Schottenbauer MA, Glass CR, Arnkoff DB, Tendick V, Gray SH (2008) Nonresponse and dropout rates in outcome studies on PTSD: review and methodological considerations. *Psychiatry* 71:134–168.
- Strobel C, Marek R, Gooch HM, Sullivan RK, Sah P (2015) Prefrontal and auditory input to intercalated neurons of the amygdala. *Cell Rep* 10:1435–1442.
- Sun L, You J, Sun F, Cui M, Wang J, Wang W, Wang D, Liu D, Xu Z, Qiu C, Liu B, Yan H (2021) Reactivating a positive feedback loop VTA-BLA-NAc circuit associated with positive experience ameliorates the attenuated reward sensitivity induced by chronic stress. *Neurobiol Stress* 15:100370.
- Thomas BL, Cutler M, Novak C (2012) A modified counterconditioning procedure prevents the renewal of conditioned fear in rats. *Learn Motiv* 43:24–34.
- Tovote P, Fadok JP, Lüthi A (2015) Neuronal circuits for fear and anxiety. *Nat Rev Neurosci* 16:317–331.
- Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 29:1310–1320.
- van Dis EA, Hagenaaars MA, Bockting CL, Engelhard IM (2019) Reducing negative stimulus valence does not attenuate the return of fear: two counterconditioning experiments. *Behav Res Ther* 120:103416.
- Vervliet B, Craske MG, Hermans D (2013) Fear extinction and relapse: state of the art. *Annu Rev Clin Psychol* 9:215–248.
- Wolpe J (1954) Reciprocal inhibition as the main basis of psychotherapeutic effects. *AMA Arch Neurol Psychiatry* 72:205–226.
- Wolpe J (1968) Psychotherapy by reciprocal inhibition. *Cond Reflex* 3:234–240.
- Wolpe J (1995) Reciprocal inhibition: Major agent of behavior change. In *Theories of behavior therapy: Exploring behavior change* (O'Donohue WT, Krasner L, eds), pp 23–57. American Psychological Association.
- Woolrich M (2008) Robust group analysis using outlier inference. *Neuroimage* 41:286–301.
- Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* 14:1370–1386.
- Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* 21:1732–1747.
- Worsley KJ (2001) Statistical analysis of activation images. In: *Functional MRI: an introduction to methods*, pp 251–270. Oxford; New York: Oxford University Press.
- Zhang X, Li B (2018) Population coding of valence in the basolateral amygdala. *Nat Commun* 9:5195.
- Zhang X, Kim J, Tonegawa S (2020) Amygdala reward neurons form and store fear extinction memory. *Neuron* 105:1077–1093.
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57.