Neuropsychologia 147 (2020) 107573

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Contextual reinstatement promotes extinction generalization in healthy adults but not PTSD

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

[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 Psychiatry, Dell Medical School, University of Texas at Austin, United States
[d] Department of Psychology, University of Texas at Austin, United States

ARTICLE INFO

Keywords:
Fear extinction
Context
Pavlovian conditioning
vmPFC
Amygdala
fMRI
MVPA
PTSD

ABSTRACT

For episodic memories, reinstating the mental context of a past experience improves retrieval of memories formed during that experience. Does context reinstatement serve a similar role for implicit, associative memories such as fear and extinction? Here, we used a fear extinction paradigm to investigate whether the retrieval of extinction (safety) memories is associated with reactivation of the mental context from extinction memory formation. In a two-day Pavlovian conditioning, extinction, and renewal protocol, we collected functional MRI data while healthy adults and adults with PTSD symptoms learned that conditioned stimuli (CSs) signaled threat through association with an electrical shock. Following acquisition, conceptually related exemplars from the CS category no longer signaled threat (i.e., extinction). Critically, during extinction only, task-irrelevant stimuli were presented between each CS trial to serve as "context tags" for subsequent identification of the possible reinstatement of this extinction context during a test of fear renewal the next day. We found that healthy adults exhibited extinction context reinstatement, as measured via multivariate pattern analysis of fMRI data, in the medial temporal lobe that related to behavioral performance, such that greater reinstatement predicted CSs being rated as safe instead of threatening. Moreover, context reinstatement positively correlated with univariate activity in the ventromedial prefrontal cortex and hippocampus, regions which are thought to be important for extinction learning. These relationships were not observed in the PTSD symptom group. These findings provide new evidence of a contextual reinstatement mechanism that helps resolve competition between the retrieval of opposing associative memories of threat and safety in the healthy adult brain that is dysregulated in PTSD.

1. Introduction

An adaptive memory system should be capable of maintaining conflicting memories of related experiences, as well as retrieving the appropriate memory given the current circumstances. How neural competition between conflicting memories is resolved remains an important question. Consider for example how you may feel about seafood after eating a dish containing tainted seabass while on a summer vacation. Nonetheless, you may later consume salmon at a local restaurant and find it delightful with no aversive consequences. This second experience countervails previous learning and creates a new association that "seafood is safe." These two opposing learning events may enter into conflict when there is an opportunity to eat a new seafood dish: does this meal pose a threat or is it harmless? One way the brain resolves ambiguity in these situations is by retrieving past memories of similar situations (Anderson, 1974). Episodic memories, for instance, involve event-specific details embedded in the contextual information present during the time of the original experience (Tulving, 2002). Thinking about a past event can bring its context to mind, and vice versa. Consequently, retrieving contextual details of either the aversive or safe dining experience could bias memory retrieval in favor of either an aversive or a safe association with seafood, thereby inhibiting expression of the alternative association (Bouton, 2002). This same scenario plays out in more extreme emotional situations as well. For example, posttraumatic stress disorder (PTSD) is characterized in part by the inability to inhibit memories of threat in harmless environments

* Corresponding author. 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.neuropsychologia.2020.107573
Received 14 April 2020; Received in revised form 24 July 2020; Accepted 26 July 2020
Available online 29 July 2020

(Liberzon and Abelson, 2016). Here, we investigated whether neural reinstatement of a past context of safety resolves context-dependent emotional memory retrieval during a time of threat ambiguity in healthy adults, and whether this mechanism is disrupted in adults with PTSD symptoms.

To address this question, we leveraged theoretical insights and experimental approaches from two academic traditions that seldom intersect: episodic memory and Pavlovian conditioning. Episodic memories are easier to recall if the context at retrieval matches that from encoding, referred to as the encoding-specificity principle (Tulving and Thomson, 1973) or transfer appropriate processing (Craik and Tulving, 1975). Background spatiotemporal details from the time of episodic memory formation can provide a "mental context," and reinstatement of a mental context can help guide memory retrieval (Howard, 2017). Neuroimaging experiments using multivariate pattern analysis (MVPA) have cleverly incorporated the concept of mental context to covertly decode brain activity related to the retrieval of items that had been encoded in a distinct visual context (Bornstein and Norman, 2017; Gershman et al., 2013; Manning et al., 2016). This method, described in detail in section 2.3, relies on the ability to decode multi-voxel patterns of fMRI data corresponding to natural scene images in scene-selective cortex, the parahippocampal place area (PPA). Whether a mental context framework can be applied to understand context-dependent emotional memory retrieval is unknown.

Combining theoretical models of episodic memory retrieval with an associative fear learning and extinction framework provides a new approach by which to examine how context resolves memory retrieval between two related, but incompatible learning experiences. Experimental studies of conditioned fear and clinical accounts of PTSD demonstrate that extinction memories are contextually specific, and that fear often returns outside the extinction context in a form of fear relapse known as renewal (Maren et al., 2013; Pitman et al., 2012). Extinction generates a secondary memory that can inhibit retrieval and expression of the original fear memory. Extinction also introduces ambiguity to the emotional meaning of a conditioned stimulus (CS), because the stimulus can now signal both the presence or absence of the unconditioned stimulus (US) (Bouton, 2002). The predominant view of extinction retrieval is that of a *competition* between expression of the original

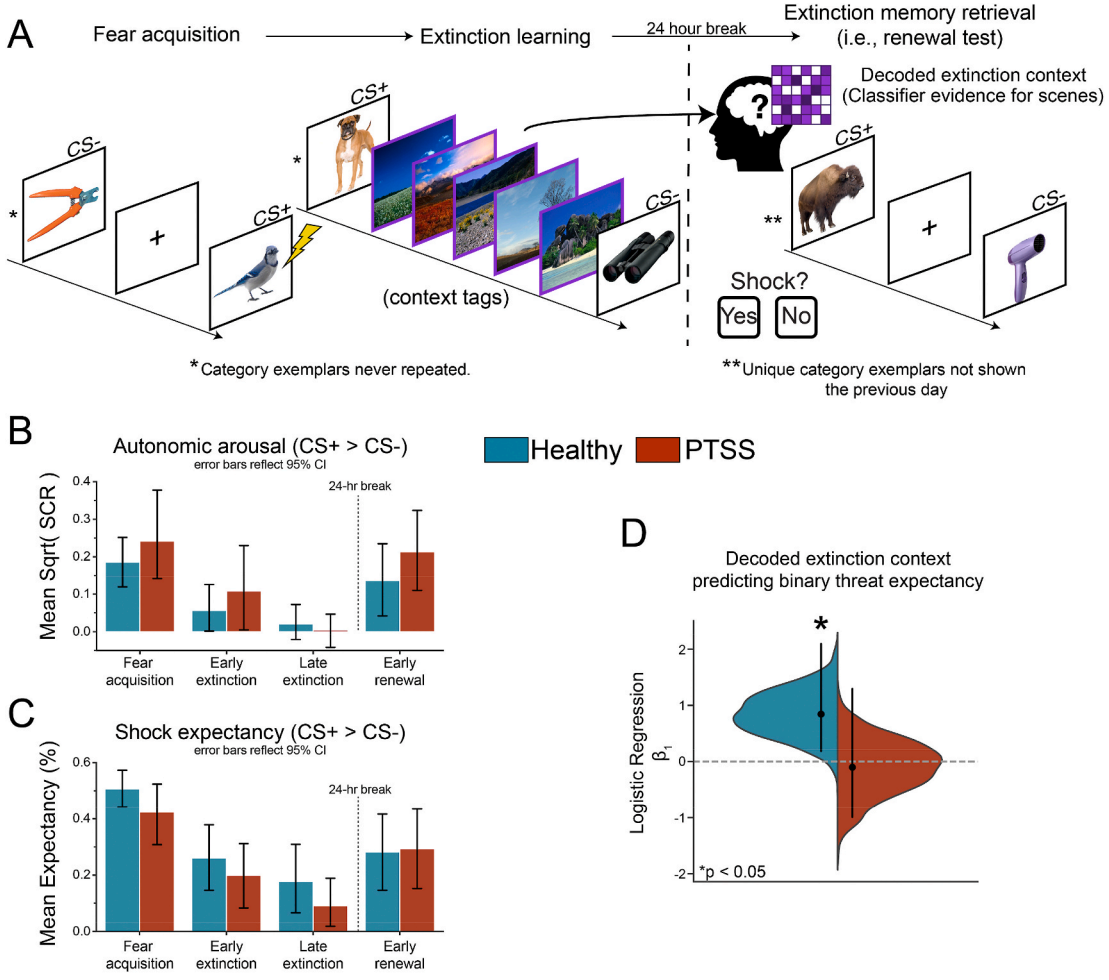

Fig. 1. Extinction mental context decoding A. Experimental Design. During fear acquisition 50% of CS+ co-terminated with a mild electric shock (US). During extinction, no shocks were delivered, and the ITIs were replaced with natural scene context tags. 24 h later, participants were placed back into the scanner and shown novel CS+/- images. MVPA classifier evidence on CS+ trials during the renewal test provided evidence for reinstatement of the mental context associated with scene images from extinction memory formation. B. Mean square root transformed differential SCR (CS+ > CS-) shows successful acquisition, extinction, and renewal for both groups. Error bars reflect 95% bootstrapped confidence intervals (N iteration = 1000). See section 3.1 and Supplementary Fig. 6 & 7 for details. C. Mean differential shock expectancy (CS+ > CS-). For this analysis responses were coded as Expect a Shock = 1, Do not expect = 0, and averaged within each phase, such that higher values indicate more shock expectancy. See section 3.1 and Supplementary Fig. 5 for details C. Decoded extinction mental context related to conscious threat expectancy during early extinction. Subjects responded "Yes" or "No" if they expected a shock on each trial. Error bars represent mean and one-sided 95% bootstrap CIs.

conditioned fear memory and the secondary extinction memory (Bouton, 1993; Maren et al., 2013). The context at the time of retrieval helps resolve this competition and determines which association (i.e., CS is a threat, or CS is safe), and consequent behavior, is most appropriate. Whereas threat associations generalize to novel environments, memory for the countervailing experience of safety is often bound to the context where safety was learned. Computations in and between the ventromedial prefrontal cortex (vmPFC), hippocampus, and amygdala modulate context-dependent extinction memory retrieval (Hermans et al., 2017; Phelps et al., 2004; Quirk and Mueller, 2008; Senn et al., 2014; Tovote et al., 2015).

Importantly, while Pavlovian conditioning research typically defines context in terms of the physical environment (i.e., an animal's cage), the conditioning literature accounts for a variety of cues that include spatiotemporal factors and internal bodily or mental states (Bouton et al., 2001). Here, we asked whether reinstated mental context might reveal the context-dependent nature of extinction memory retrieval in humans. We generated a mnemonic signature specific to an episode of extinction learning using MVPA tools that have been previously used to tag and capture the reinstatement of mental context (Bornstein and Norman, 2017; Gershman et al., 2013; Manning et al., 2016). This was combined with a novel Pavlovian conditioning protocol in order to track reinstatement of an extinction context signature during a 24-h test of fear renewal where subjects encountered threat ambiguous CSs that were conceptually related to CSs encoded during both fear acquisition and extinction the previous day (Fig. 1A). We compared results between healthy adults and adults with PTSD symptoms, a disorder characterized by severe dysregulation in contextual processing that might contribute to deficits in the retrieval of extinction memories (Garfinkel et al., 2014; Liberzon and Abelson, 2016).

We hypothesized that neural reinstatement of the extinction context, as measured from scene-specific reactivation in the PPA, would be associated with activity in brain regions which are believed, based on animal models and human neuroimaging, to be crucial for extinction memory retrieval, including the vmPFC and hippocampus. Given that performance after extinction is context-dependent (Bouton, 2004), we also hypothesized that context reinstatement would predict behavioral ratings of threat-ambiguous CSs as safe rather than dangerous. Previous studies have demonstrated deficits for adults with PTSD in contextual processing during extinction learning (Garfinkel et al., 2014; Rougemont-Bücking et al., 2011). Therefore, we hypothesized that individuals with PTSD symptoms would also show deficits in mental context reinstatement. Borrowing from MVPA approaches to probe the fidelity of episodic memory (Ritchey et al., 2013), we also investigated whether retrieval of an extinction memory reactivates similar patterns of neural activity associated with the formation of an extinction memory from the previous day. For this, we employed an MVPA approach from episodic memory research that measures memory fidelity through the overlap in neural activity patterns between encoding and retrieval.

2. Materials and methods

2.1. Participants

Twenty-four healthy adults (Mean age 21 y/o, s.d. 2 y; 15 female, 9 male) were recruited with inclusion criteria of no self-reported history of psychiatric or neurological illness or history of medication for such an illness. In addition, 24 Criterion A trauma-exposed adults reporting current post-traumatic stress symptoms (PTSS, the range of trauma-related symptoms that comprise the PTSD diagnosis) were recruited (Mean age 26 y/o, s.d. 4.7 y; 17 female, 7 male). All participants provided written informed consent in compliance with the IRB at UT Austin (IRB # 2017-02-0094). An additional 3 participants (2 healthy and 1 PTSS) were recruited but did not complete the study. Participants in the PTSS group responded to flyers seeking volunteers with PTSD. We then phone-screened participants to confirm self-reported PTSD diagnosis and the absence of neurological or substance use disorders. Of the 24 participants included in the analysis, 22 self-reported that a primary PTSD diagnosis had been given at some point in the past; two self-reported obsessive-compulsive disorder as their primary diagnosis but indicated significant PTSS related to a Criterion A trauma.

Eligible participants appeared for an in-person screening and completed a dimensional measure of PTSS (PTSD Checklist for DSM-5; PCL-5; M = 26.09, s.e.m = 3.08; Supplementary Fig. 10) (Blevins et al., 2015), as well as a brief assessment of Criterion A trauma type and details (Life Events Checklist; LEC) (Gray et al., 2004) that allowed us to verify that reported PTSS were directly trauma-related. One participant declined to share details of their Criterion A trauma. Participants also completed measures of anxiety (Beck Anxiety Inventory; BAI) (Beck et al., 1988), and depression (Beck Depression Inventory; BDI) symptoms (Beck et al., 1961) (Supplementary Fig. 9). As we did not use a structured diagnostic interview, we refer to this group as PTSS, consistent with modern conceptualizations of PTSD as a dimensional disorder and that those below formal diagnostic cut-offs still have notably pathology (Harpaz-Rotem et al., 2014; Kramer et al., 2016). Given the high co-morbidity between trauma-related symptoms and substance abuse disorder, subjects in the PTSS group were also given a urine toxicology test just prior to going into the MRI. No subjects in the PTSS group tested positive for illicit drugs or benzodiazepines.

2.2. Stimuli

CSs consisted of 168 trial-unique (i.e., non-repeating) pictures of animals (N = 84) and tools (N = 84), obtained from the website www.lifeonwhite.com or publicly available resources on the internet. Each CS used in the experiment across all phases was a different basic-level exemplar (e.g., there were not two different pictures of a chimpanzee). Threatening or typically phobic stimuli were excluded (e.g., spiders, snakes, knives). Stimulus presentation was controlled using E-Prime 3.0. During acquisition, extinction, and the renewal test CSs were presented for 4.5 ± 0.5s (jittered) and followed by a jittered 6 ± 1s intertrial-interval (ITI). The trial order of CSs was pseudorandomized such that no more than 3 images from the same category occurred in a row. We used the same pseudorandomized trial order for every subject, with the exception that the first and second trial during the renewal test was counterbalanced as a CS+ and CS-, in order to control for any non-specific orienting responses during the renewal test (Schiller et al., 2010). The basic-level animal and tool exemplars were randomized throughout the study for each subject.

The US was a 50-ms electrical shock, delivered to the to the index and middle finger of the participant's left hand. Prior to the experiment, the US was calibrated for each participant to a level described as "highly annoying and unpleasant, but not painful," and was controlled using the STMEPM-MRI stimulation system from BIOPAC Systems (Goleta, CA).

2.3. Task and procedures

We developed a novel fMRI task designed to tag and track the encoding and retrieval of an extinction memory by identifying patterns of neural activity associated with the context in which extinction was learned (Fig. 1A). This task was motivated by work on human episodic memory using MVPA to reveal how mental context reinstatement organizes memory retrieval (Bornstein and Norman, 2017; Gershman et al., 2013; Manning et al., 2016). We combined the mental context tagging procedure with a category fear conditioning design in which CSs in each phase consisted of basic-level exemplars of animals and tools (Dunsmoor et al., 2018, 2015a, 2012; Dunsmoor and Kroes, 2019). Animals (or tools) served as CS+ and were reinforced with the shock US, while tools (or animals, respectively) were unpaired control stimuli (CS-); CS+/CS- category assignment was counterbalanced between subjects. Subjects were never instructed about the CS shock contingencies but were told that if they paid attention then they might learn

the association between the pictures and the shock. We measured skin conductance responses (SCRs) throughout the experiment as a measure of autonomic arousal (Supplementary Figures 6 & 7). Electrodes were placed on participants' left palm and connected to a BIOPAC MP100 System (Goleta, CA) SCR sampling rate was set to 200Hz. SCRs were manually scored using previously validated criteria (Dunsmoor et al., 2019). Specifically, the time frame in which trough-to-peak deflection was considered valid extended from 0.5 s following a CS onset to CS offset, lasted between 0.5 and 5.0 s, and with a deflection greater than 0.02 μS. If SCRs did not meet these criteria, they were scored as zero. Prior to analysis, all SCRs were square-root normalized. SCRs from acquisition was analyzed for CS- and unreinforced CS+ trials only (no shock). Due to technical errors, 4 subjects (2 healthy and 2 PTSS) are missing SCR data from extinction learning and are thus excluded from all relevant statistics.

Day 1 included 3 phases: baseline, fear acquisition, and extinction. During baseline (24 CS+, 24 CS-) no shocks were presented while subjects simply viewed each picture and made a 2-alternative forced choice (2-AFC) rating of which category the object belonged to (animal, tool) using a button box in their right hand (data not included here). During fear acquisition (24 CS+, 24 CS-), 50% of CS+ trials co-terminated with the US. Intermittent reinforcement was used to slightly delay extinction learning, as continuous reinforcement is associated with abrupt extinction (Chan and Harris, 2019). On each trial, subjects rated shock expectancy as a measure of explicit threat expectancy (2-AFC, Yes/No) (Lonsdorf et al., 2017). Fear extinction (24 CS+, 24 CS-) immediately followed acquisition. No shocks were administered, and subjects continued to make shock expectancy ratings. Importantly, while the ITI during baseline and fear acquisition consisted of a crosshair on a white background, the ITI during extinction consisted of a stream of natural scene images. The scene images were displayed for 1 s each with 5, 6 or 7 scenes per ITI. Participants were not given any instructions regarding the scene images. These images served as 'mental context tags' that, when viewed, became bound to the current learning episode, and whose subsequent neural reactivation would signal the reinstatement of the mental context from that episode, as demonstrated in prior studies (Bornstein and Norman, 2017; Gershman et al., 2013; Manning et al., 2016). In accordance with computational models of contextually mediated memory retrieval (Howard and Kahana, 2002; Polyn et al., 2009), scene-related activity should be assimilated into a mental context representation specific to the formation of the extinction memory. During a test of fear renewal the next day, when subjects were not viewing scenes, we decoded the reactivation strength of scene-related information using machine learning approaches as detailed below. When used, early extinction refers to the 1st half (12 CS+ and 12 CS-) of this phase, and late extinction refers to the 2nd half.

On Day 2, subjects underwent a fear renewal test (12 CS+, 12 CS-) during fMRI in which the SCR and shock electrodes were re-attached and subjects continued to rate shock expectancy, but no shocks were ever delivered. Critically, no scene images were presented during fear renewal. Behavioral and neural analyses of the renewal test were focused *a priori* on the first 4 CS+ and 4 CS- trials (early renewal test). Focusing on these early trials is consistent with human neuroimaging studies of extinction retrieval (Dunsmoor et al., 2019; Milad et al., 2009), as these early trials are the most threat ambiguous and most likely reflect attempts at retrieval, whereas later trials more likely reflect continued extinction learning as subjects realize they are not receiving any shocks.

Following the fear renewal test on Day 2 was a recognition memory test (data not included here) and 2 runs of a perceptual MVPA localizer necessary for classifying scene-related neural activity during the fear renewal test, described in detail in section 2.7. After exiting the scanner participants completed various questionnaires (Supplementary Figure 10).

2.4. Functional MRI acquisition

Scanning was completed using the Siemens Skyra 3T Human MRI scanner located at the Biomedical Imaging Center at the University of Texas at Austin. Functional data were acquired with a 32-channel head-coil. Functional image resolution was 3 mm isotropic voxels (TR = 2000 ms, TE = 29 ms; FoV = 228 mm; n(Slices) = 48). A multi-band factor of 2 was used with AC/PC auto alignment. Due to a computer malfunction, 2 subjects had slightly different acquisition parameters on day 1 (TR = 2230 ms, n(Slices) = 66), which were accounted for during pre-processing and analysis. An additional high resolution T1-weighted 3D MPRAGE scan (TR = 1.9s, 1 mm isotropic voxels) was collected on day 1 to aid in registration and region-of-interest definition.

2.5. Image preprocessing

Images were prepared using a combination of FSL (Oxford Centre for Functional MRI of the Brain (FMRIB) Software Library version 5.0.10, ANTs (Advanced Normalization Tools version 2.1.0), FreeSurfer version 6.0.0, and in-house Python scripts (Avants et al., 2011; Fischl, 2012; Jenkinso et al., 2012). DICOM conversion was accomplished using dcm2niix (Li et al., 2016). Skull stripping was accomplished via FreeSurfer recon-all. Functional runs were motion-corrected using FSL mcflirt, denoising of the 6 canonical head motion parameters occurred prior to any analyses. Following motion correction, all functional runs were co-registered to the T1 structural image. In preparation for MVPA, time series data had a linear trend removed, were high-pass filtered (sigma = 128s), and z-scored. Representational similarity and trial unique neural activity analyses required the estimation of LS-S style beta images (Mumford et al., 2012), which were was accomplished using FSL and in-house python scripts. In the LS-S parameter estimation procedure, trial-specific betas are computed iteratively using a design matrix with the trial of interest modeled as one explanatory variable, and all other trials as a single other explanatory variable. When provided, GLM parameter estimates were computed using FSL FEAT (motion parameters applied before modeling) and included pre-whitening, high-pass filters, the canonical double gamma hemodynamic response function, and linear spatial registration to the MNI152 template provided in FSL via FLIRT with 12 degrees of freedom. GLM design matrices included CS+s and CS-s modeled as separate regressors using full trial duration, and where applicable the US was modeled using a finite impulse response function.

2.6. Statistical analysis

A combination of parametric and non-parametric resampling (bootstrapping) analyses were used to statistically evaluate all of the following results. Where mentioned, outliers were defined by data that were beyond [1.5*Inter Quartile Range] in either direction. The type and value of all calculated test statistics and significance values are reported in the text and figure captions. PyCortex was used to display neuroimaging statistical maps (Gao et al., 2015).

2.7. Perceptual localizer

The perceptual category localizer included a stream of category images (1s on, 1s off) including different animals, tools, natural scenes, indoor scenes, and phase-scrambled scenes. Images were shown in blocks of 8, and participants were told to respond if they detected any duplicate images in a block (1 duplicate/block) simply to ensure they were paying attention. Each run (2 total) consisted of 4 blocks of each category, with 16s of rest in-between each block. Each image in the localizer was distinct, and did not repeat from the experimental phases or across localizer runs. For all subjects the localizer had the same

structure and order of image blocks, while the order in which specific images were presented in each block of 8 images was randomized across participants.

2.8. ROI selection

The hippocampus, amygdala, and vmPFC were chosen *a priori* as ROIs based on their known role in extinction learning and retrieval from human and rodent models. Each ROI was defined for each subject anatomically using the relevant FreesSurfer parcellations of the Desikan-Killiany atlas ("mOFC" was used for vmPFC). At the group level, the vmPFC was defined functionally via GLM parameter estimates of CS- > CS+ activity during acquisition. A separate regressor was used to model the US (electric shock) in order to separate the estimates of neural activity resulting from the CS+ and US. A cluster of voxels was selected manually (caudal cut-off applied) after thresholding the whole brain image at $p = 0.001$, uncorrected. In order to facilitate MVPA decoding of the scene images comprising the mental context tags, a group-level parahippocampal place area (PPA) ROI was functionally defined following procedures of Bornstein and Norman (2017). Specifically, GLM parameter estimates of [Scenes > Scrambled Scenes | Objects] were obtained for each subject and threshold at $p = 0.001$, uncorrected. These subject maps were then binarized and stacked. A cluster corresponding to the PPA was selected based on the criteria that it showed activation in 80% of subjects. The group PPA mask was registered to each subject from MNI152 space using FSL FLIRT with 12 degrees of freedom.

2.9. Multivariate pattern analysis

MVPA decoding was accomplished using the Sklearn Logistic Regression classifier in Python (Pedregosa et al., 2012). The classifier was trained to detect natural scene images vs. scrambled scene images in the PPA (Bornstein and Norman, 2017). Prior to analysis of the experimental data, classifier sensitivity was assessed via cross validation of the two localizer runs (mean classifier ROC AUC = 0.91, s.e.m. = 0.03). Due to equipment malfunction, 1 participant in the PTSS group had only 1 localizer run; however, this did not impact decoding. To test for reinstatement of the context-tag scene images, the classifier was trained on scenes vs. scrambled scenes during the localizer, and then used to obtain classifier evidence (probability estimates) for scene-related activity during the renewal test. Classifier evidence for scenes was used to directly operationalize extinction mental context reinstatement. TR-wise data were used for the analysis linking extinction mental context decoding to behavioral shock expectancy (section 3.2), and data were shifted 2 TRs (4 s) to account for the hemodynamic response. LS-S parameter estimates of single trials (Mumford et al., 2012) were used for all other multivariate analyses.

Pattern similarity analysis (Kriegeskorte et al., 2008) was achieved using in-house Python scripts. In order to reduce noise prior to computing a correlation, trial-specific LS-S beta estimates from the early renewal test and extinction learning were weighted (multiplied) by the univariate estimate of the respective CS type from the relative timepoint (e.g. CS+ beta estimates from early renewal were multiplied by univariate estimate of early CS+ activity). For each CS type, a representational similarity matrix was constructed, where each cell is a Pearson correlation between all pairs of images from renewal and extinction. Matrices were Fisher-Z transformed and the mean value taken for the section of the matrix corresponding to the renewal-to-extinction comparisons. Individual subjects' means were then entered into group analyses in order to test for reliability.

2.10. Data and code availability

All deidentified behavioral and neuroimaging data, as well as all custom python analysis code can be found online (https://osf.io/qeg83/).

3. Results

3.1. Behavioral results

Behavioral results showed successful acquisition, extinction, and renewal of SCRs and shock expectancy (Fig. 1B–C). Planned paired two-tailed t-tests of CS+ > CS- differences from fear acquisition to late extinction show successful extinction in healthy adults (SCR: $t(21) = -2.60$, $p = 0.017$; Expectancy: $t(23) = -4.33$, $p = 2.46e\text{-}4$) and individuals with PTSS (SCR: $t(21) = -2.86$, $p = 9.34e\text{-}3$; Expectancy: $t(23) = -3.67$, $p = 1.29e\text{-}3$). Comparisons of CS+ > CS- differences from late extinction to early renewal show significant renewal of SCRs ($t(21) = 3.55$, $p = 2.01e\text{-}3$) and trending renewal of Expectancy ($t(23) = 1.99$, $p = 0.059$) in PTSS individuals, but not in healthy individuals (SCR: $t(21) = 1.40$, $p = 0.18$; Expectancy: $t(23) = 1.26$, $p = 0.22$). Given this report is focused on MVPA approaches to measure neural extinction processes, additional information on the behavioral data are detailed in the Supplementary Results.

3.2. Neural reinstatement of the extinction context and behavioral performance

An overview of the study design is shown in Fig. 1A. Important to this design, scene images were only presented during extinction the previous day, and therefore scene-related activity detected during the fear renewal test can be interpreted as reinstatement of the extinction mental context, in keeping with extant work on neural context reinstatement in human episodic memory research (Bornstein and Norman, 2017; Gershman et al., 2013; Manning et al., 2016). Also, the CSs shown at fear renewal were novel exemplars (e.g., new animal pictures) that were conceptually related to CSs encoded during both fear acquisition and extinction the previous day. As a consequence, the stimuli themselves can be considered threat ambiguous. To track neural reinstatement of the extinction context during the renewal test, a machine learning classifier was trained to identify scene-related activity in the PPA using fMRI data from a separate perceptual localizer task. This classifier was then used to estimate reinstatement of the extinction context on threat-ambiguous CS trials during the renewal test by quantifying scene-related neural activity in the PPA.

There was a wide distribution of classifier evidence for scene reinstatement in the PPA in both groups. Notably, the variability in mean classifier evidence was similar between groups (ind. two-tailed $t(46) = -0.94$, $p = 0.35$; Supplementary Fig. 2). If context reinstatement facilitates extinction memory retrieval, then one hypothesis is that it increases the likelihood that subjects perceive threat ambiguous CSs as safe rather than threatening. Accordingly, we assessed the relationship between behavioral threat expectancy ratings (Fig. 1C & Supplementary Fig. 5) and extinction context reinstatement. We used logistic regression to relate reinstated extinction context to behavioral threat expectancy on each CS+ trial of the early renewal test. Responses were recoded as 1 if the subject perceived the stimulus as safe (did not expect a shock) and 0 if a subject perceived the stimulus as threatening (expected a shock). We selected data from the 2 TRs (4 s) associated with the moment preceding the decision (average RTs were below 2s, one sample, two-tailed $t(47) = -12.1$, $p = 5.6e\text{-}16$) in order to better capture the decision phase rather than outcome anticipation. Reaction times did not differ between groups (two-tailed ind. $t(46) = 1.4$, $p = 0.18$). In order to overcome the relatively limited number of trials, we analyzed the fixed effect in each group by combining all trials across all subjects. We evaluated the generalizability of this relationship with a random-effects bootstrap test, randomly resampling whole participants within each group with replacement 1,000 times (Kim et al., 2014). We found that reinstated extinction context reliably predicted subsequent threat expectancy in healthy adults (Fig. 1D, $\beta_1 = 0.84$, one-sided 95% CI =

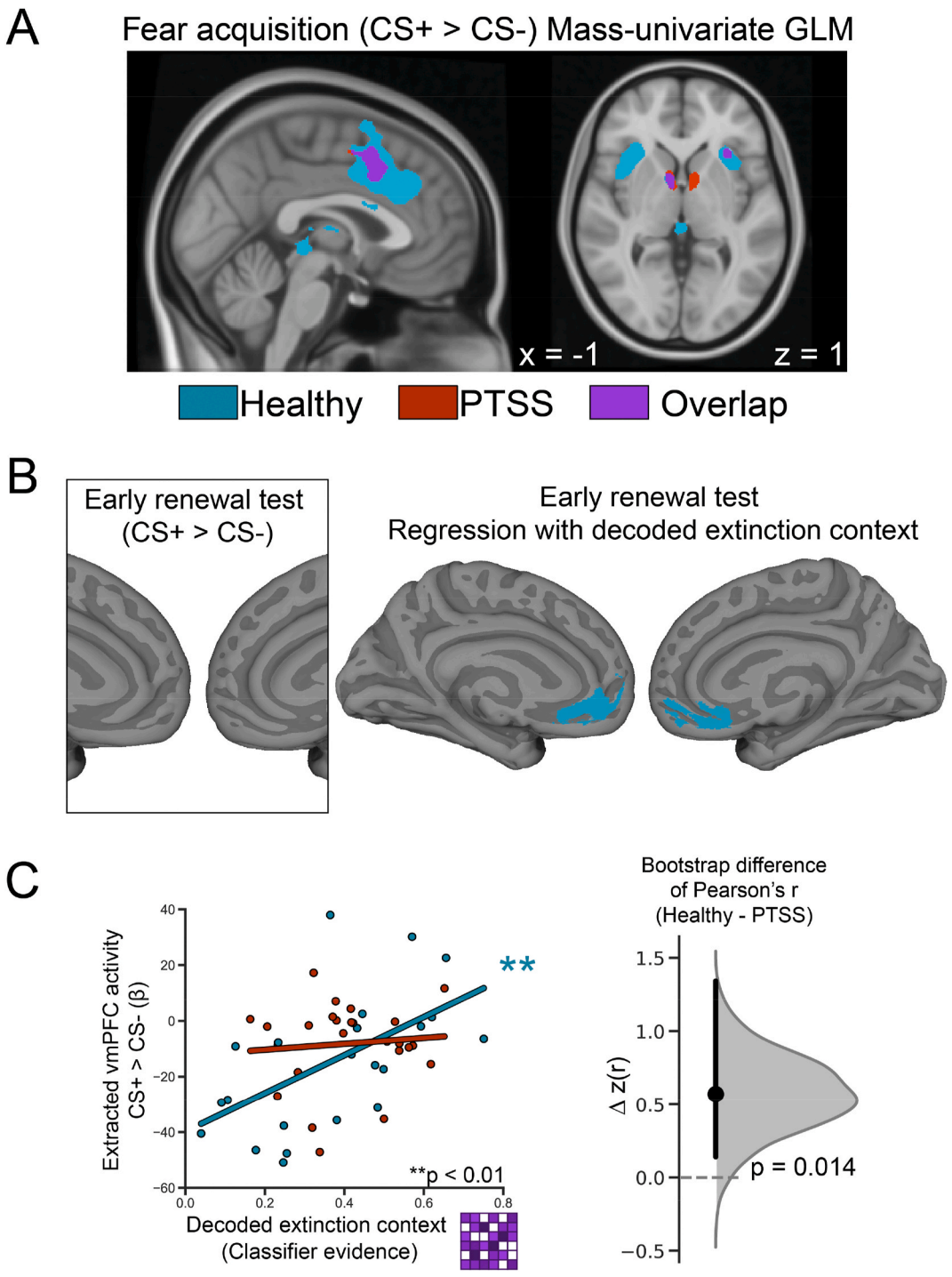

Fig. 2. Traditional GLM analysis and correlations with neural activity in the extinction circuit A. Mass-univariate GLM contrast of CS+ > CS- during fear acquisition reveals activation in the dACC, anterior insula, and thalamus for healthy adults and individuals with PTSS. Statistical maps are displayed at FWE corrected threshold of $p < 0.05$, accomplished via permutation testing (N iteration = 1000). Overlap indicates voxels significant as detected in separate tests after FWE correction. B. Left. Traditional GLM contrast of CS+ > CS- reveals no significant activity for either group during early renewal. Right. MVPA-univariate whole brain regression using vmPFC ROI. 68% of voxels within the group vmPFC mask met the threshold for healthy adults. PTSS group had no significant voxels. Statistical maps are displayed at FWE corrected threshold of $p < 0.05$, accomplished via permutation testing (N iteration = 1000). C. Left During the early renewal test, extracted univariate (CS+ > CS-) vmPFC activity correlated with decoded extinction context, classifier evidence for scenes, in healthy adults) Right. Between groups comparison. Point and bar indicate mean and 95% confidence interval of a one-sided bootstrap comparison (N = 1000 iterations) of fisher z transformed correlation coefficients. A normalized kernel density estimate of the derived bootstrap distribution of differences is shown in grey.

[0.19, 2.19], $p = 0.015$), but not in PTSS (Fig. 1D, $\beta_1 = -0.10$, one-sided 95% CI = [-0.99, 1.29], $p = 0.55$). The observed relationship between reinstated context and threat perception was found to be marginally stronger in healthy adults compared to PTSS (mean difference $\beta_1 = 0.95$, one-sided 95% CI = [-0.10, 2.75], $p = 0.069$). In summary, the degree of mental context reinstatement dissociated healthy adults who perceived the CS+ as safe versus threatening, but this neural measure was unrelated to behavioral ratings of safety or danger in PTSS.

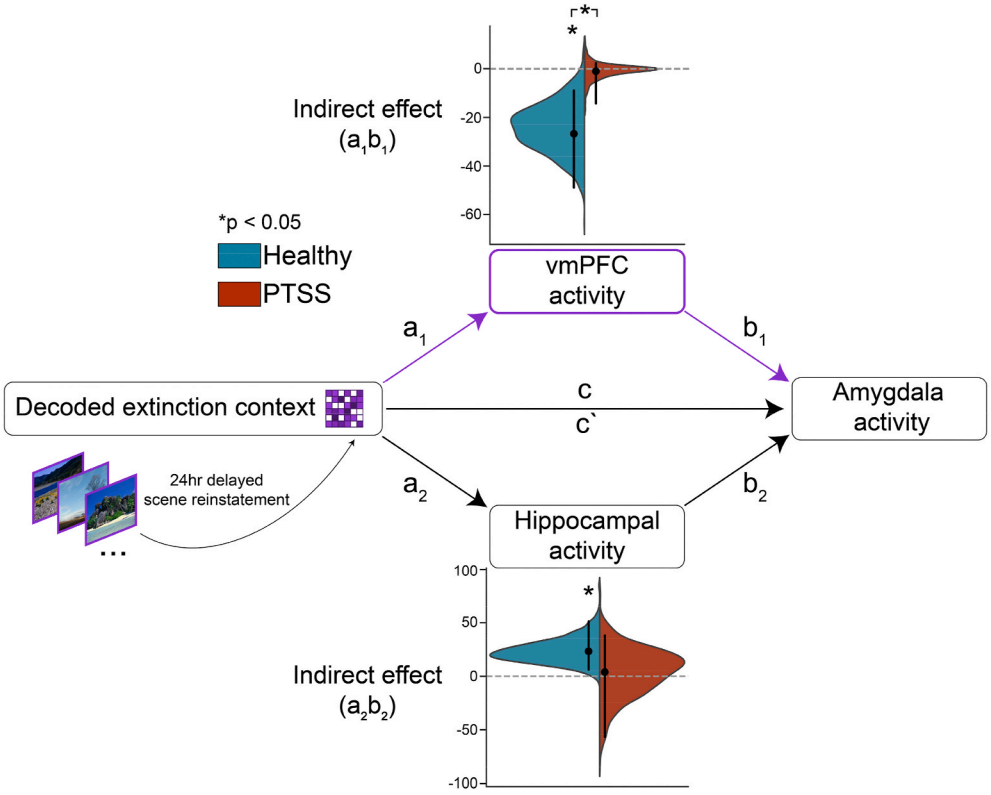

Fig. 3. Reinstatement of extinction context indirectly affects amygdala activity through the vmPFC and hippocampus. Non-causal mediation analysis was performed separately for the healthy and PTSSD groups. Full results for all paths are presented in Supplementary Table 1. No significant total effect (c) or direct (c') effect was observed in either group. We tested if decoded extinction context exerts a significant indirect effect through either the vmPFC or hippocampus. Violin plots display distributions of indirect effects for vmPFC (a_1b_1) and hippocampus (a_2b_2) obtained from bootstrapping (N boots = 1000). Points indicate mean indirect effect, bar indicates bounds of the two-tailed 95% confidence interval. The vmPFC indirect effect significantly differed from zero for healthy adults but not PTSS. The hippocampal indirect effect significantly differed from zero for healthy adults but not PTSS.

3.3. Traditional univariate approach to infer extinction related processes in the human brain

Consistent with mass-univariate neuroimaging analyses of human fear conditioning and extinction (Fullana et al., 2018; Sehlmeyer et al., 2009), the CS+ versus CS- contrast revealed activity in dACC, bilateral insula, and thalamus during fear acquisition (Fig. 2A) that diminished in extinction in both healthy adults and adults with PTSS (Supplementary Fig. 1 & Supplementary Table 2). Also consistent with mass-univariate approaches to measure extinction retrieval (Dunsmoor et al., 2019; Fullana et al., 2018), the CS+ versus CS- contrast failed to reveal activity in the vmPFC (Fig. 2B, left) or hippocampus during extinction retrieval, even at extremely liberal statistical thresholds. The inability to consistently identify robust activity in the vmPFC and hippocampus during extinction learning and recall using fMRI has remained a puzzle, particularly because extensive neurobiological evidence in animal models shows these regions are critical for extinction memory formation and retrieval (Maren et al., 2013; Milad and Quirk, 2012; Quirk and Mueller, 2008; Senn et al., 2014). The lack of vmPFC or hippocampal activity might support an interpretation that these regions are not especially involved in extinction memory retrieval in the human brain. The following analyses shed new light on the relationship between univariate activity associated with extinction memory retrieval and reinstatement of the mental context associated with extinction memory formation.

3.4. Neural reinstatement of the extinction context reveals activity in the vmPFC and hippocampus

We tested whether extinction mental context reinstatement correlated with univariate activity in the canonical extinction neurocircuitry ROIs during the early renewal test. For each subject, average classifier evidence during CS+ trials was correlated with CS+ > CS- univariate activity in the vmPFC, hippocampus, and the amygdala. Healthy adults exhibited a positive correlation between CS+ evoked mental context reinstatement and CS+ > CS- univariate activity in the vmPFC (Fig. 2C, $r = 0.56$, 95% CI = [0.20, 0.79], $p = 0.004$), as well as in the hippocampus ($r = 0.41$, 95% CI = [0.01, 0.70], $p = 0.044$, Supplementary Fig. 3). No significant correlation was observed in the amygdala ($r = 0.07$, 95% CI = [-0.34, 0.46], $p = 0.74$). A whole-brain regression analysis revealed a significant positive correlation between MVPA classifier evidence for scenes and univariate CS+ > CS- parameter estimates in the vmPFC during fear renewal for healthy adults (Fig. 2B, right). These whole brain regression results remained significant following application of a vmPFC mask and permutation testing, indicating a robust relationship.

The PTSS group did not show a relationship between context reinstatement and univariate activity in the vmPFC (Fig. 2B–C, $r = 0.09$, 95% CI = [-0.33, 0.48], $p = 0.68$), hippocampus ($r = 0.04$, 95% CI = [-0.37, 0.44], $p = 0.85$), or amygdala ($r = -0.11$, 95% CI = [-0.49, 0.31], $p = 0.62$) despite similar levels of classifier evidence of context reinstatement in the PPA as healthy adults (Supplementary Fig. 2). The relationship between extinction mental context reinstatement and CS+ specific activity during the renewal test was reliably stronger in healthy adults compared to PTSS in the vmPFC (mean difference = 0.58, 95% CI = [0.17, 1.46], one-tailed $p = 0.014$, Fig. 2C), although not in the hippocampus (mean difference = 0.37, 95% CI = [-0.20, 1.57], one-tailed $p = 0.142$). In sum, the role of the vmPFC and hippocampus in subjects with PTSS appeared to be perturbed, such that activity in this region was insensitive to the strength of extinction context reinstatement.

3.5. Reinstatement of extinction context indirectly affects amygdala activity through the vmPFC and hippocampus

Human neuroimaging studies have also failed to capture extinction related activity in the amygdala (Fullana et al., 2019, 2018), despite the importance of amygdala function for successful extinction retrieval (Marek et al., 2018). We similarly did not observe significant amygdala activity or a relationship between context reinstatement and amygdala activity (Supplementary Figs. 3 & 4). Given the well-defined

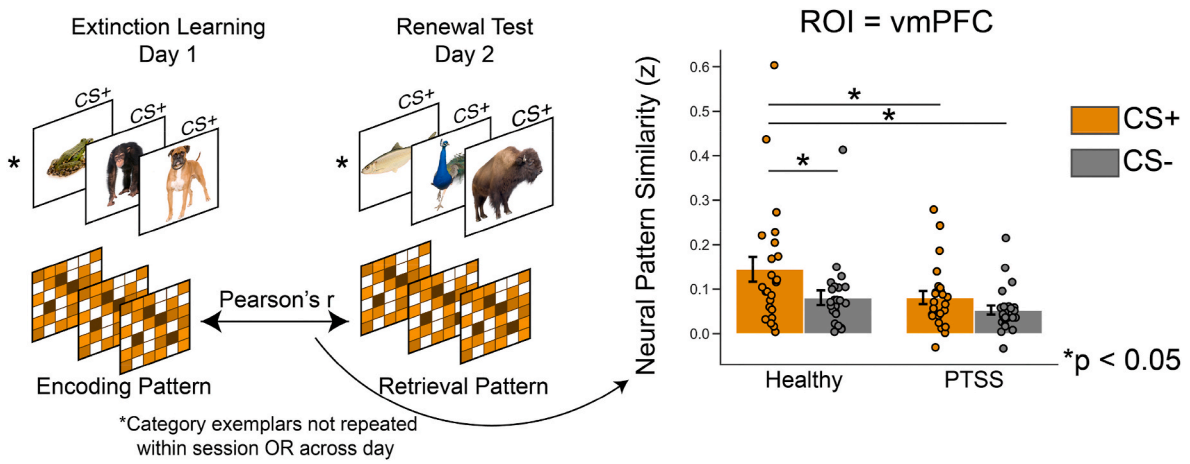

Fig. 4. Extinction encoding-retrieval pattern similarity analysis. Left: CS+/- evoked patterns during extinction learning were correlated with never before seen novel CS+/- evoked patterns from the fear renewal test in the vmPFC. Right: Fischer-Z transformed CS+/- neural patterns for healthy and PTSS groups. Error bars represent ± 1 s.e.m.

neurocircuitry of extinction retrieval, we hypothesized that reinstated context might have an *indirect* effect on amygdala activity through upstream inputs. In order to evaluate possible indirect effects, we used a non-causal mediation analysis within a bootstrap framework. Neural activity in the amygdala was entered as the outcome, reinstated mental context as the predictor, and neural activity in vmPFC and hippocampus were both used as parallel mediators (Hayes and Rockwood, 2017; Vallat, 2018) (Fig. 3). The correlations between the mediators and the predictor and outcome met necessary assumptions (Supplementary Table 1 & Supplementary Fig. 3). In healthy adults, reinstated context interacted with both the vmPFC and hippocampus resulting in significant intervening effects on amygdala activity (path $a_1b_1 = -26.7$, 95% CI = [-48.8, −8.95], p = 0.02; $a_2b_2 = 23.3$, 95% CI = [6.17, 51.3], p = 0.01). There were no significant intervening effects in the PTSS group (path $a_1b_1 = -1.06$, 95% CI = [-14.3, 2.30], p = 0.79, $a_2b_2 = 3.96$, 95% CI = [-56.6, 38.2], p = 0.78). In order to evaluate the relative strength of these indirect effects in healthy adults relative to PTSS, we computed the differences between the obtained bootstrap distributions of indirect effects. The indirect effect of reinstated mental context through the vmPFC was reliably stronger in healthy adults compared to PTSS (mean difference = −23.4, one-tailed 95% CI = [-62.7, −6.00], p = 0.014). However, the indirect hippocampal effect was not significantly stronger for healthy adults (mean difference = 20.0, one-tailed 95% CI = [-18.4, 130], p = 0.22). Thus, this model posits that reinstated extinction context in the PPA is signaled in the vmPFC and hippocampus, and that information is relayed to the amygdala in the healthy adult brain. This is consistent with current understanding of the directionality of the extinction circuit from rodent neurophysiology (Marek et al., 2018; Quirk and Mueller, 2008).

3.6. Neural similarity between extinction memory encoding and retrieval

A complementary approach to assess successful memory retrieval is to quantify the overlap in neural activity between encoding and retrieval (Rugg et al., 2008). In episodic memory research, the match between activity at encoding and retrieval predicts memory performance (Ritchey et al., 2013; Tompary and Davachi, 2017). We adapted a measure of encoding-retrieval overlap to estimate the fidelity of extinction memory retrieval in the vmPFC, based on the importance placed on this region for encoding and retrieving extinction memories. A pattern similarity analysis (Kriegeskorte et al., 2008) was used to probe similarity of CS evoked neural patterns of activity in the vmPFC between extinction learning and retrieval. In order to identify extinction-learning specific patterns, we also calculated pattern similarity for trials from the CS- category across extinction learning and renewal test for comparison (Fig. 4). Notably, subjects were not engaged in explicit memory retrieval; rather, fear renewal tests constitute a more subtle probe of memory retrieval between competing associative memories (fear vs. safety). Thus, successful extinction memory retrieval can be construed within the framework of cortical reinstatement or transfer-appropriate processing by quantifying neural similarity between extinction learning and extinction retrieval. The use of *different items* across days also mitigates the potential role of mere perceptual overlap between stimuli used at encoding and retrieval. A CS condition by group mixed ANOVA revealed significant main effects of group ($F(1,46) = 4.22$, p = 0.045) and condition ($F(1,46) = 10.4$, p = 0.002). Planned t-tests revealed healthy adults display significantly greater encoding-retrieval similarity of CS+ extinction memories in the vmPFC compared to PTSS participants (two-tailed ind. $t(46) = 2.01$, p = 0.05, without outliers $t(32.9) = 2.84$, p = 0.007). Healthy adults also showed a selectively enhanced similarity of evoked patterns of neural activity in the vmPFC for CS+ images over CS- (two-tailed paired $t(23) = 2.74$, p = 0.01), whereas PTSS participants did not (two-tailed paired $t(23) = 1.71$, p = 0.09). Enhanced pattern similarity to the CS+ versus CS- is important, because it indicates that pattern similarity is selective to the extinguished cue, and not generally more activity in healthy adults across days irrespective of stimulus type. Healthy adults also evinced higher overall neural pattern similarity in the hippocampus (CS by group mixed ANOVA main effect of group; $F(1,46) = 4.12$, p = 0.04; Supplementary Fig. 8). Healthy adults also displayed selectively higher similarity for CS+ patterns in the PPA (two-tailed ind. $t(46) = 2.12$, p = 0.04, Supplementary Fig. 8) as compared to PTSS participants, however there was no effect of CS condition nor any interactions. In sum, these results complement the earlier extinction context reinstatement findings by showing that healthy adults reinstate neural patterns associated with extinction learning as compared to PTSS participants, and this perhaps helps to promote successful retrieval of extinction memories outside the extinction context.

4. Discussion

We present new evidence that neural reinstatement of the extinction context at a time of threat ambiguity is associated with neural and behavioral correlates of extinction memory retrieval in healthy adults, but not in adults with PTSS. Understanding the context-dependent nature of emotional memory retrieval helps advance our understanding of affective disorders, for which dysregulated contextual processing may be a core feature. In healthy adults, a mnemonic signature of extinction

context reinstatement positively correlated with activity in the vmPFC and hippocampus. Neurobiological research shows these regions are critical for encoding and retrieving extinction memories, but evidence from human fMRI studies utilizing mass-univariate approaches have so far only weakly supported translation of these findings from rodents to humans. Despite a well-recognized role for the vmPFC and hippocampus in extinction processes (Quirk and Mueller, 2008; Senn et al., 2014), failures to observe vmPFC and hippocampus activity in human fear extinction are common (Fullana et al., 2018). Here, we used MVPA approaches borrowed from neuroimaging research on human episodic memory to reveal activity in these regions by accounting for the strength of context reinstatement. Reinstatement of encoding patterns associated with the spatiotemporal context in which extinction memories are formed may help balance retrieval of extinction memories against the renewal of fear.

The inability to detect vmPFC activity using univariate contrasts is likely related to certain methodological, analytical, and technical factors related to univariate neuroimaging analyses (Fullana et al., 2019; Morriss et al., 2018). For instance, the CS- is a poor baseline condition to compare to an extinguished CS+, because the CS- has already acquired inhibitory properties as an unpaired stimulus during acquisition. Also, extinction learning and retrieval are dynamic and temporally sensitive processes and many studies do not include time as a factor - BOLD activity in the vmPFC and medial temporal lobe is elevated at rest and typically "deactivates" during task (Harrison et al., 2017, 2011). A similar issue has plagued the detection of amygdala activity during fear acquisition in human neuroimaging (Fullana et al., 2016; Mechias et al., 2010), despite overwhelming evidence for its role in conditioned fear learning in other mammals. One approach for reliably detecting amygdala activity in fear conditioning has been MVPA (Bach et al., 2011; Dunsmoor et al., 2014; Staib and Bach, 2018). Likewise, we show here that MVPA approaches are better suited for detecting vmPFC and hippocampal involvement in extinction retrieval processes that is otherwise undetectable using conventional mass-univariate analytical approaches. A non-causal mediation analysis further revealed that mental context reinstatement has an indirect effect on amygdala activity through the vmPFC and hippocampus. The role of these three structures in learning and retrieving extinction is well known from animal models, but translation to human neuroimaging has remained a challenge. Taken together, this study provides new support for the use of MVPA to investigate extinction memory processes in human neuroimaging research.

As expected, the PTSS group did not show a relationship between reinstated extinction context and neural or behavioral correlates of extinction retrieval. This is new evidence that individuals with PTSS do not utilize mental context to resolve threat ambiguity after extinction, further supporting the idea that contextual processing deficits are a pathogenic marker at the core of the disease (Garfinkel et al., 2014; Liberzon and Abelson, 2016). This observed abnormal vmPFC activity in the PTSS group is generally consistent with neuroimaging data that show dysfunction in both the structure and function of the hippocampus and vmPFC in PTSD (Pitman et al., 2012), as well as prior work showing diminished vmPFC involvement in PTSD during extinction recall (Milad et al., 2009). It is worth noting that a nearly equal number of healthy adults and adults with PTSS perceived the CS+ as safe during fear renewal. In healthy adults, these behavioral ratings were associated with increased mental context reinstatement, supporting the idea that performance after extinction is context-dependent. Yet no relationship between context reinstatement and behavioral threat expectancy was observed in PTSS. This raises the question of whether PTSS subjects who successfully extinguish a fear memory do so through a different, perhaps compensatory, mechanism that does not involve retrieving the memory of extinction. Associative fear learning involves multiple independent components (emotional, temporal, conceptual) that can be either implicit or explicit in nature (Delamater, 2012, 2004; Dunsmoor et al., 2015b), and extinction may affect only some elements of the CS-US association while leaving other elements intact. As such, extinction training might effectively diminish the CS-US contingency ("the shock is less likely to occur now") in PTSS without extinguishing other associative elements. Another possibility is that some subjects were playing the odds that the CS+ would not shock them, given that the shock was probabilistic on Day 1, and this strategy would not rely on retrieving a memory of safety, per se. In such a case, performance at test should be minimally related to extinction learning. Given the nature of the behavioral task, it is difficult to discern in detail what neurocognitive processes were involved in subjects' determination that the CS+ was either safe or threatening. There could also be heterogeneity in the PTSS population that contributes to more or less effective extinction learning and retrieval that is obscured at the group level. Overall, the observed relationship between context reinstatement and feelings of safety in healthy adults, and the absence of such a finding in PTSD, warrants further investigation.

Scene images were used as context tags because these stimuli robustly activate the PPA. Future work could advance the utility of the context reinstatement protocol for fear extinction research by using multisensory naturalistic contexts. This technique might also have application to test the efficacy of certain clinical treatments for fear and anxiety disorders by measuring neural evidence for context reinstatement associated with the treatment setting. Another advancement of these findings would be an attempt to *enhance* the neural representation of the extinction context using closed-loop neurofeedback, with the goal of biasing memory retrieval toward extinction and away from fear when extinguished stimuli are encountered in novel environments. Such a technique might have clinical utility for psychiatric disorders characterized by inability to retrieve adaptive memories to sustain behavioral change.

Neural pattern similarity analysis also provided a new tool for assaying the strength of extinction memory representations over time. Reactivation of encoding-related activity during memory retrieval is widely thought to modulate the strength of memory (Nyberg et al., 2000; Rugg et al., 2008). We found that patterns of neural activity at fear renewal overlapped with patterns from extinction learning, which might reflect a possible mechanism by which healthy adults retrieve the memory trace of extinction (Ritchey et al., 2013). Importantly, the items presented during fear renewal were threat ambiguous, and did not perceptually match the items encoded during extinction the previous day (i.e., the items in both phases were from the same category but were different exemplars). It is therefore noteworthy that encoding-retrieval neural similarity was selectively enhanced for CS+ as compared to CS- items in healthy adults, as neural similarity could not be driven purely by perceptual overlap. Using MVPA to evaluate the strength of encoding-retrieval overlap may be a valuable new approach to quantify the strength of extinction memory representations over time in human neuroimaging. Future work should look to see whether the strength of neural similarity diminishes at remote memory tests, and whether neural similarity can be enhanced via optimized extinction learning strategies (Craske et al., 2014). One possibility is that enhancing extinction learning in PTSD might rescue extinction memory representations coded in the vmPFC over time.

A limitation of the current study is that PTSSs were self-reported via the PCL, and not assessed with a structured clinical interview. This likely resulted in a lower minimum level of symptom severity than seen in formally diagnosed PTSD groups, with an average PCL score that aligns with a "sub-threshold" description (Supplementary Fig. 10). That said, the broader range and normal distribution of the PCL scores in our study is consistent with a dimensional conceptualization of PTSD, which is a more quantitatively parsimonious fit with symptom data than a binary classification (Harpaz-Rotem et al., 2014). However, a sample of participants with more severe clinical symptoms could serve to further disambiguate the observed extinction recall deficits we describe here. Another limitation of this study is the absence of a full clinical assessment for our healthy adult group, whereas all participants in our PTSS

group reported negative outcomes and current symptoms following exposure to trauma. Nevertheless, our results provide evidence that individuals with self-diagnosed PTSS do not engage the same neural mechanisms as healthy individuals without trauma-related symptoms.

5. Conclusion

We synthesized theoretical approaches from neurobiological models of extinction from animal studies with technical approaches from the cognitive neuroscience of episodic memory. Through this combination of approaches we found that a multivariate neural signature of contextual reactivation, as well as representational overlap of extinction memories from encoding to retrieval, reveals the role of key extinction neurocircuitry in the healthy adult brain and dysfunction in PTSD. These findings also highlight new approaches to investigate the context-dependent nature of fear extinction memory in the human brain, consequently helping bridge the substantial translational divide between fine-scale molecular imaging of activity-dependent neural tagging in animal neuroscience (Josselyn and Tonegawa, 2020) and human neuroimaging. That the link between neural reinstatement and extinction memory retrieval is compromised in PTSD suggests a potential target for a disorder characterized by dysregulated contextual processing and extinction retrieval deficits.

Author contributions

Augustin C. Hennings conceived of and designed the experiment, performed the fMRI experiments, analyzed data, and wrote the manuscript. Mason McClay recruited fMRI participants and performed the fMRI experiments. Jarrod A. Lewis-Peacock conceived of and designed the experiment and wrote the manuscript. Joseph E. Dunsmoor conceived of and designed the experiment and wrote the manuscript.

Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Joseph E. Dunsmoor (joseph.dunsmoor@austin.utexas.edu).

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

We thank Michael Drew, Greg Fonzo, and Suzannah Creech for helpful discussions and comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2020.107573.

Funding

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

References

Anderson, J.R., 1974. Retrieval of propositional information from long-term memory. Cognit. Psychol. 6, 451–474. https://doi.org/10.1016/0010-0285(74)90021-8.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025.

Bach, D.R., Weiskopf, N., Dolan, R.J., 2011. A stable sparse fear memory trace in human amygdala. J. Neurosci. 31, 9383–9389. https://doi.org/10.1523/JNEUROSCI.1524-11.2011.

Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. J. Consult. Clin. Psychol. 56, 893–897. https://doi.org/10.1037/0022-006X.56.6.893.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatr. 4, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004.

Blevins, C.A., Weathers, F.W., Davis, M.T., Witte, T.K., Domino, J.L., 2015. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J. Trauma Stress 28, 489–498. https://doi.org/10.1002/jts.22059.

Bornstein, A.M., Norman, K.A., 2017. Reinstated episodic context guides sampling-based decisions for reward. Nat. Neurosci. 20, 997–1003. https://doi.org/10.1038/nn.4573.

Bouton, M.E., 2004. Context and behavioral processes in extinction. Learn. Mem. 11, 485–494. https://doi.org/10.1101/lm.78804.

Bouton, M.E., 2002. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. Biol. Psychiatr. https://doi.org/10.1016/S0006-3223(02)01546-9.

Bouton, M.E., 1993. Context, time, and memory retrieval in the interference paradigms of pavlovian learning. Psychol. Bull. 114, 80–99. https://doi.org/10.1037/0033-2909.114.1.80.

Bouton, M.E., Mineka, S., Barlow, D.H., 2001. A modern learning theory perspective on the etiology of panic disorder. Psychol. Rev. 108, 4–32. https://doi.org/10.1037/0033-295X.108.1.4.

Chan, C.K.J., Harris, J.A., 2019. The partial reinforcement extinction effect: the proportion of trials reinforced during conditioning predicts the number of trials to extinction. J Exp Psychol Anim Learn Cogn 45, 43–58. https://doi.org/10.1037/xan0000190.

Craik, F.I., Tulving, E., 1975. Depth of processing and the retention of words in episodic memory. J. Exp. Psychol. Gen. 104, 268–294. https://doi.org/10.1037/0096-3445.104.3.268.

Craske, M.G., Treanor, M., Conway, C.C., Zbozinek, T., Vervliet, B., 2014. Maximizing exposure therapy: an inhibitory learning approach. Behav. Res. Ther. 58, 10–23. https://doi.org/10.1016/j.brat.2014.04.006.

Delamater, A.R., 2012. On the nature of CS and US representations in Pavlovian learning. Learn. Behav. 40, 1–23. https://doi.org/10.3758/s13420-011-0036-4.

Delamater, A.R., 2004. Experimental extinction in Pavlovian conditioning: behavioural and neuroscience perspectives. Q. J. Exp. Psychol. Sect. B Comp. Physiol. Psychol. https://doi.org/10.1080/02724990344000097.

Dunsmoor, J.E., Davachi, L., Phelps, E.A., Murty, V.P., 2015a. Emotional learning selectively and retroactively strengthens memories for related events. Nature 520, 345–348.

Dunsmoor, J.E., Kragel, P.A., Martin, A., La Bar, K.S., 2014. Aversive learning modulates cortical representations of object categories. Cerebr. Cortex 24, 2859–2872. https://doi.org/10.1093/cercor/bht138.

Dunsmoor, J.E., Kroes, M.C., 2019. Episodic memory and Pavlovian conditioning: ships passing in the night. Curr Opin Behav Sci 26, 32–39. https://doi.org/10.1016/J.COBEHA.2018.09.019.

Dunsmoor, J.E., Kroes, M.C.W., Li, J., Daw, N.D., Simpson, H.B., Phelps, E.A., 2019. Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. J. Neurosci. 2713–2718. https://doi.org/10.1523/JNEUROSCI.2713-18.2019.

Dunsmoor, J.E., Kroes, M.C.W., Moscatelli, C.M., Evans, M.D., Davachi, L., Phelps, E.A., 2018. Event segmentation protects emotional memories from competing experiences encoded close in time. Nat Hum Behav 2, 291–299. https://doi.org/10.1038/s41562-018-0317-4.

Dunsmoor, J.E., Martin, A., LaBar, K.S., 2012. Role of conceptual knowledge in learning and retention of conditioned fear. Biol. Psychol. 89, 300–305. https://doi.org/10.1016/j.biopsycho.2011.11.002.

Dunsmoor, J.E., Niv, Y., Daw, N., Phelps, E.A., 2015b. Rethinking extinction. Neuron 88, 47–63. https://doi.org/10.1016/j.neuron.2015.09.028.

Fischl, B., 2012. FreeSurfer. *Neuroimage* 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

Fullana, M.A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., Harrison, B.J., 2019. Amygdala where art thou? Neurosci. Biobehav. Rev. 102, 430–431. https://doi.org/10.1016/j.neubiorev.2018.06.003.

Fullana, M.A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., Harrison, B.J., 2018. Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants. Neurosci. Biobehav. Rev. 88, 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002.

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatr. 21, 500–508. https://doi.org/10.1038/mp.2015.88.

Gao, J.S., Huth, A.G., Lescroart, M.D., Gallant, J.L., 2015. Pycortex: an interactive surface visualizer for fMRI. Front. Neuroinf. 9, 23. https://doi.org/10.3389/fninf.2015.00023.

Garfinkel, S.N., Abelson, J.L., King, A.P., Sripada, R.K., Wang, X., Gaines, L.M., Liberzon, I., 2014. Impaired contextual modulation of memories in PTSD: an fMRI and psychophysiological study of extinction retention and fear renewal. J. Neurosci. 34, 13435–13443. https://doi.org/10.1523/JNEUROSCI.4287-13.2014.

Gershman, S.J., Schapiro, A.C., Hupbach, A., Norman, K.A., 2013. Neural context reinstatement predicts memory misattribution. J. Neurosci. 33, 8590–8595. https://doi.org/10.1523/JNEUROSCI.0096-13.2013.

Gray, M.J., Litz, B.T., Hsu, J.L., Lombardo, T.W., 2004. Psychometric properties of the life events checklist. Assessment 11, 330–341. https://doi.org/10.1177/1073191104269954.

Harrison, B.J., Fullana, M.A., Via, E., Soriano-Mas, C., Vervliet, B., Martínez-Zalacaín, I., Pujol, J., Davey, C.G., Kircher, T., Straube, B., Cardoner, N., 2017. Human ventromedial prefrontal cortex and the positive affective processing of safety signals. Neuroimage 152, 12–18. https://doi.org/10.1016/j.neuroimage.2017.02.080.

Harrison, B.J., Pujol, J., Contreras-Rodríguez, O., Soriano-Mas, C., López-Solà, M., Deus, J., Ortiz, H., Blanco-Hinojo, L., Alonso, P., Hernández-Ribas, R., Cardoner, N., Menchón, J.M., 2011. Task-Induced deactivation from rest extends beyond the default mode brain network. PloS One 6, e22964. https://doi.org/10.1371/journal.pone.0022964.

Hayes, A.F., Rockwood, N.J., 2017. Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation. Behav. Res. Ther. 98, 39–57. https://doi.org/10.1016/j.brat.2016.11.001.

Hermans, E.J., Kanen, J.W., Tambini, A., Fernández, G., Davachi, L., Phelps, E.A., 2017. Persistence of amygdala-hippocampal connectivity and multi-voxel correlation structures during awake rest after fear learning predicts long-term expression of fear. Cerebr. Cortex 27, 3028–3041. https://doi.org/10.1093/cercor/bhw145.

Howard, M.W., 2017. Temporal and spatial context in the mind and brain. Curr Opin Behav Sci. https://doi.org/10.1016/j.cobeha.2017.05.022.

Howard, M.W., Kahana, M.J., 2002. A distributed representation of temporal context. J. Math. Psychol. 46, 269–299.

Jenkinso, M., Beckmann, C., Behrens, T., Woolrich, M., Smith, S., 2012. FSL. *Neuroimage* 62, 782–790.

Josselyn, S.A., Tonegawa, S., 2020. Memory engrams: recalling the past and imagining the future. Science 80. https://doi.org/10.1126/science.aaw4325.

Kim, G., Lewis-Peacock, J.A., Norman, K.A., Turk-Browne, N.B., 2014. Pruning of memories by context-based prediction error. Proc. Natl. Acad. Sci. U. S. A. 111, 8997–9002. https://doi.org/10.1073/pnas.1319438111.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 1–28. https://doi.org/10.3389/neuro.06.004.2008.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods 264, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001.

Liberzon, I., Abelson, J.L., 2016. Context processing and the neurobiology of post-traumatic stress disorder. Neuron. https://doi.org/10.1016/j.neuron.2016.09.039.

Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas, J.M.P., Merz, C.J., 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. https://doi.org/10.1016/j.neubiorev.2017.02.026.

Manning, J.R., Hulbert, J.C., Williams, J., Piloto, L., Sahakyan, L., Norman, K.A., 2016. A neural signature of contextually mediated intentional forgetting. Psychon. Bull. Rev. 23, 1534–1542. https://doi.org/10.3758/s13423-016-1024-7.

Marek, R., Jin, J., Goode, T.D., Giustino, T.F., Wang, Q., Acca, G.M., Holehonnur, R., Ploski, J.E., Fitzgerald, P.J., Lynagh, T., Lynch, J.W., Maren, S., Sah, P., 2018. Hippocampus-driven feed-forward inhibition of the prefrontal cortex mediates relapse of extinguished fear. Nat. Neurosci. 21, 384–392. https://doi.org/10.1038/s41593-018-0073-9.

Maren, S., Phan, K.L., Liberzon, I., 2013. The contextual brain: implications for fear conditioning, extinction and psychopathology. Nat. Rev. Neurosci. 14, 417–428. https://doi.org/10.1038/nrn3492.

Mechias, M.L., Etkin, A., Kalisch, R., 2010. A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. Neuroimage 49, 1760–1768. https://doi.org/10.1016/j.neuroimage.2009.09.040.

Milad, M.R., Pitman, R.K., Ellis, C.B., Gold, A.L., Shin, L.M., Lasko, N.B., Zeidan, M.A., Handwerger, K., Orr, S.P., Rauch, S.L., 2009. Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biol. Psychiatr. 66, 1075–1082. https://doi.org/10.1016/j.biopsych.2009.06.026.

Milad, M.R., Quirk, G.J., 2012. Fear extinction as a model for translational neuroscience: ten years of progress. Annu. Rev. Psychol. 63, 129–151. https://doi.org/10.1146/annurev.psych.121208.131631.

Morriss, J., Hoare, S., van Reekum, C.M., 2018. It's time: a commentary on fear extinction in the human brain using fMRI. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2018.06.025.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59, 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076.

Nyberg, L., Habib, R., McIntosh, A.R., Tulving, E., 2000. Reactivation of encoding-related brain activity during memory retrieval. Proc. Natl. Acad. Sci. U. S. A. 97, 11120–11124. https://doi.org/10.1073/pnas.97.20.11120.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Phelps, E.A., Delgado, M.R., Nearing, K.I., Ledoux, J.E., 2004. Extinction learning in humans: role of the amygdala and vmPFC. Neuron 43, 897–905. https://doi.org/10.1016/j.neuron.2004.08.042.

Pitman, R.K., Rasmusson, A.M., Koenen, K.C., Shin, L.M., Orr, S.P., Gilbertson, M.W., Milad, M.R., Liberzon, I., 2012. Biological studies of post-traumatic stress disorder. Nat. Rev. Neurosci. https://doi.org/10.1038/nrn3339.

Polyn, S.M., Norman, K.A., Kahana, M.J., 2009. A context maintenance and retrieval model of organizational processes in free recall. Psychol. Rev. 116, 129–156.

Quirk, G.J., Mueller, D., 2008. Neural mechanisms of extinction learning and retrieval. Neuropsychopharmacology. https://doi.org/10.1038/sj.npp.1301555.

Ritchey, M., Wing, E.A., LaBar, K.S., Cabeza, R., 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cerebr. Cortex 23, 2818–2828. https://doi.org/10.1093/cercor/bhs258.

Rougemont-Bücking, A., Linnman, C., Zeffiro, T.A., Zeidan, M.A., Lebron-Milad, K., Rodriguez-Romaguera, J., Rauch, S.L., Pitman, R.K., Milad, M.R., 2011. Altered processing of contextual information during fear extinction in PTSD: an fMRI study. CNS Neurosci. Ther. 17, 227–236. https://doi.org/10.1111/j.1755-5949.2010.00152.x.

Rugg, M.D., Johnson, J.D., Park, H., Uncapher, M.R., 2008. Chapter 21 Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspectiveProgress in Brain Research, pp. 339–352. https://doi.org/10.1016/S0079-6123(07)00021-0.

Schiller, D., Monfils, M.H., Raio, C.M., Johnson, D.C., LeDoux, J.E., Phelps, E.A., 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. Nature 463, 49–53. https://doi.org/10.1038/nature08637.

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., Konrad, C., 2009. Human fear conditioning and extinction in neuroimaging: a systematic review. PloS One. https://doi.org/10.1371/journal.pone.0005865.

Senn, V., Wolff, S.B.E., Herry, C., Grenier, F., Ehrlich, I., Gründemann, J., Fadok, J.P., Müller, C., Letzkus, J.J., Lüthi, A., 2014. Long-range connectivity defines behavioral specificity of amygdala neurons. Neuron 81, 428–437. https://doi.org/10.1016/j.neuron.2013.11.006.

Staib, M., Bach, D.R., 2018. Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. Neuroimage 166, 276–284. https://doi.org/10.1016/j.neuroimage.2017.11.009.

Tompary, A., Davachi, L., 2017. Consolidation promotes the emergence of representational overlap in the Hippocampus and medial prefrontal cortex. Neuron 96, 228–241. https://doi.org/10.1016/j.neuron.2017.09.005 e5.

Tovote, P., Fadok, J.P., Lüthi, A., 2015. Neuronal circuits for fear and anxiety. Nat. Rev. Neurosci. https://doi.org/10.1038/nrn3945.

Tulving, E., 2002. Episodic memory: from mind to brain. Annu. Rev. Psychol. https://doi.org/10.1146/annurev.psych.53.100901.135114.

Tulving, E., Thomson, D.M., 1973. Encoding specificity and retrieval processes in episodic memory. Psychol. Rev. 80, 352–373. https://doi.org/10.1037/h0020071.

Vallat, R., 2018. Pingouin: statistics in Python. J Open Source Softw 3, 1026.