**METHODS**

**Participants***.* A total of 48 participants from the community volunteered to complete the two-day functional MRI study. Three additional participants were recruited but did not complete the experiment. Half of the participants (N = 24) were recruited with the criteria that they have no current or past psychiatric or neurological disorders. The remaining participants (N = 24) were recruited after responding to flyers seeking volunteers with PTSD. These participants underwent phone screening and completed additional in-person questionnaires to confirm Criterion A trauma exposure on the PTSD checklist for DSM-5 (Blevins et al., 2015), as well as the absence of other neurological disorders. We refer to this cohort as a post-traumatic stress symptom (PTSS) group as we did not implement a structured diagnostic interview. Additional recruitment details, as well as PCL and other mood survey scores are reported in Hennings et al., 2020. All experimental procedures were approved by the University of Texas at Austin IRB (#2017-02-0094).

**Stimuli.** Conditioned stimuli were images of animals and tools collected from lifeonwhite.com or other publicly available resources on the internet. Critical to the design of the task, each stimulus was a unique exemplar from its category. For example, there were not two different kinds of “dog” used. Typically phobic animals or threatening tools were excluded (e.g., spiders, snakes, knives). Stimulus presentation was controlled using E-Prime 3.0. The unconditioned stimulus (US) was a brief (50ms) electric shock delivered to fingers of the left hand. Prior to entering the scanner, the US was calibrated for each participant to a level described as “highly annoying and unpleasant, but not painful”. A BIOPAC STMEPM-MRI module was used to deliver the US (Goleta, CA).

**Task.** *Associative learning task*.Participants completed an associative learning task in two sessions of about an hour each, roughly 24 hours apart. We note that “fear” is often a misnomer of the emotional construct being studied in research involving human participants. A better term may be “threat conditioning”*,* as it better captures both the actual emotional experience of participants and the acquisition of conditioned responses. Nevertheless, we retain the term “fear*”* to connect the results the broader field of Pavlovian conditioning. For all phases of the associative learning task, images were displayed for 4.5 +/- 0.5s (jittered), and the ITI between trials lasted 6 +/- 0.5s (jittered). The trial order of the CSs was pseudorandomized to ensure no more than 3 CS type were presented in a row. The same pseudorandomized order was used for all subjects, with the exception being that the first trial of the renewal test was counterbalanced to be either a CS+ or CS- across participants. Which phase of the experiment each stimulus was displayed was randomized across participants. Day 1 consisted of pre-conditioning, fear conditioning, and extinction. On Day 1, each phase consisted of 48 trials, 24 animals and 24 tools, for a total of 144 items. During pre-conditioning, participants identified which category each image belonged to (animal or tool). During fear conditioning, 50% of the trials from one category (CS+) co-terminated with the US, for a total of 12 CS+US pairings. Images from the other category were never paired with shock (CS-), and the category of each CS was counterbalanced across participants. Extinction learning followed fear conditioning, during which no shocks were delivered. In order to tag and track the mental context corresponding to extinction, the normal fixation cross displayed during the ITI was replaced with a stream of natural scene images displayed for 1s each (5, 6, or 7 scenes per ITI). Natural scenes have previously been used to successful tag and track the reinstatement of a particular mental context (Bornstein & Norman, 2017; Gershman et al., 2013; Hennings et al., 2020; Manning et al., 2016). The following day, participants had the electrodes reattached prior to entering the scanner for the fear renewal test. The fear renewal test consisted of 12 each of CS+/-, and no shocks were delivered.

During fear conditioning and extinction on Day 1, and the fear renewal test on Day 2, participants responded whether or not they expected a shock on each trial (yes or no). Skin-conductance responses were collected during pre-conditioning, fear condition, extinction, and the fear renewal test. These data are reported in Hennings et al., (2020), and confirm successful acquisition and extinction of discriminatory (CS+ > CS-) expectancy and SCR. There were no differences in behavior between healthy adults and adults with PTSS on Day 1. During the fear renewal test PTSS displayed significant renewal of SCR while healthy individuals did not.

*Recognition memory test*. After completing the fear renewal test on Day 2, participants completed a surprise recognition memory test for the items the had seen the previous day. All 144 old images were included as well as 96 novel foils. The stimuli seen during the fear renewal test were not shown during the recognition memory test. Each image was displayed for 3s, and participants indicated whether each image was old (they had seen it the previous day), or new (never seen before). Participants indicated the confidence of their choice by responding the image was definitely old, maybe old, maybe new, or definitely new. The memory test was split into three fMRI runs of equal length, and trial order was again pseudorandomized to ensure a balance of lures and foils of both CS types and encoding phases across the memory runs. Recognition memory scores are reported in Hennings et al., (preprint). Relevant to the present analyses, we reported no significant difference in recognition memory between healthy and PTSS individuals. Trials during the recognition memory test were removed from analysis if participants failed to make a response (N average per subject).

*Perceptual localizer*. A perceptual localizer followed the recognition memory test to facilitate MVPA decoding of the mental context tags. Categories of images (animals, tools, indoor scenes, outdoor scenes, and phase-scrambled scenes) were shown in a stream (1s on, 1s off) in blocks of 8 images. Participants completed an N-back duplicate image detection task to facilitate attention to the images (1 duplicate was included per block). Each of the two localizer runs consisted of 4 blocks of each category, with 16s of rest in-between each block. The images used in the perceptual localizer were unique and not repeated from any other phase of the experiment. The block structure was the same for all participants while the presentation of stimuli was randomized across participants.

**Functional MRI acquisition.**

Neuroimaging was accomplished 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, with 3mm isotropic resolution (TR = 2000ms; TE = 29ms; FoV = 228; 48 slices). A multi-band factor of 2 was used with automatic AC/PC alignment. As discussed in Hennings et al., (2020), due to a computer malfunction, 2 subjects had slightly different acquisition parameters on Day 1 (TR = 2230ms; 66 slices), which were accounted for during preprocessing and analysis. An T1-weighted 3d MPRAGE scan (TR = 1900ms; 1mm isotropic resolution) was collected on Day 1 to aid in functional image registration and region of interest definition.

**Image preprocessing**

Functional MRI data were processed using *fMRIprep* (v1.5.4)*,* an open source software suite designed to increase reproducibility and develop common best practices for image processing. The following boilerplate has been included unchanged, as recommended by the package maintainers.

*Anatomical data preprocessing.* The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al., 2009).

*Functional data preprocessing.* For each of the 9 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Susceptibility distortion correction (SDC) was omitted as no field maps were collected. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox & Hyde, 1997). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by (Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri\_vol2surf (FreeSurfer).

**Region of interest selection**

The dACC, vmPFC, hippocampus, and amygdala were selected *a priori* to test for the presence of encoding specificity of fear and extinction memories. Prefrontal ROIs were based on peak coordinates previously reported in literature. Specifically, dACC coordinates (MNI 1, 21, 27) were taken from Milad et al., 2007 in which a univariate contrast of CS+ > CS- during fear conditioning was used. vmPFC coordinates (MNI -4, 34, -6) were taken from an fMRI meta-analysis of extinction recall, using a univariate contrast of extinguished CS+ > unextinguished CS+ (Fullana et al., 2018, cluster labeled subgenual cingulate cortex in orignal paper). For each ROI, a sphere was drawn around the coordinates with a radius of 10mm, and was then restricted to grey matter using a grey matter probability mask with a threshold of 50%. The masks were then warped to subject space to achieve native functional resolution (3mm3) for multivariate analyses. Registration was accomplished using flirt using 12 degrees of freedom and nearest neighbor interpolation for each binary mask (FSL 5.0.9).

The hippocampus and amygdala were masked and segmented into subfields using Freesurfer’s segmentHA\_T1 on the preprocessed T1w anatomical images from recon-all (Freesurfer 7.0). The hippocampus was segmented into head (anterior), body, and tail (posterior) subfields along the long axis. Anterior to posterior segmentation was used based on previous research which suggests different roles in fear and extinction for the homologous structures in the rodent hippocampus [Cite]. The amygdala was segmented into the basolateral (BLA), and central nucleus (CeM) subfields based on their functional roles in associative learning and memory expression [good cite]. The anatomical segmentations were registered to functional space using mri\_label2vol and binary masks created using fslmaths.

A functionally defined parahippocampal place area (PPA) mask was created to facilitate decoding of the mental context tag learned during extinction (Bornstein & Norman, 2017; Hennings et al., 2020). See Hennings et al., (2020) for a full description of the localizer task used. In brief, a GLM was used to estimate the contrast Scenes > Scrambled Scenes and Objects. Each subject’s estimate was threshold at P=0.001 uncorrected, binarized, and stacked. A cluster was corresponding to the PPA was selected using the criteria that at least 95% of subjects show activation. The mask was registered to functional resolution using flirt with 12 degrees of freedom and nearest neighbor interpolation for each subject.

**Multivariate pattern analysis**

*Neural decoding.* Extinction mental context tags were decoded using the same procedure as in Hennings et al., (2020), implemented in Python. In brief, a logistic regression classifier (Sklearn, (Pedregosa et al., 2012) was trained to discriminate scene images vs. scrambled images [check]…

After preprocessing with *fMRIprep*, we computed a LS-S style betaseries to facilitate the encoding-retrieval similarity analysis (Mumford et al., 2012, 2014).

**Statistical analyses.** All statistical analyses were accomplished using linear mixed effects models using the *afex* package in R with maximum likelihood estimation. Model specifications are discussed in the **Results**, but in general all experimental variables were modeled as fixed effects and subject was modeled as a random intercept. Significance of the main effects and interactions of the fixed effects was evaluated using Chi-square tests, comparing the log-likelihoods of a model with and without the term of interest. Planned and *post-hoc* contrasts were accomplished using the *emmeans* package in R to test the estimated marginal means and trends of the various models. Asymptotic degrees of freedom were used, as in general the number of observations in each model was quite large (>10,000). Parametric 95% confidence intervals are reported along with FDR corrected P-values where appropriate.