Neurobiology of Learning and Memory 187 (2022) 107572

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Survival of the salient: Aversive learning rescues otherwise forgettable memories via neural reactivation and post-encoding hippocampal connectivity

David Clewett [a,b], Joseph Dunsmoor [c], Shelby L. Bachman [b,d], Elizabeth A. Phelps [e], Lila Davachi [f,g,*]

[a] Department of Psychology, University of California, Los Angeles, United States
[b] Department of Psychology, New York University, United States
[c] Department of Psychology, University of Texas at Austin, United States
[d] Department of Gerontology, University of Southern California, United States
[e] Department of Psychology, Harvard University, United States
[f] Department of Psychology, Columbia University, United States
[g] Nathan Kline Institute, Orangeburg, NY, United States

ARTICLE INFO

Keywords:
Consolidation
Aversive
Hippocampus
Dopamine
Memory
Reactivation

ABSTRACT

The effects of aversive events on memory are complex and go beyond the simple enhancement of threatening information. Negative experiences can also rescue related but otherwise forgettable details encoded close in time. Here, we used functional magnetic resonance imaging (fMRI) in healthy young adults to examine the brain mechanisms that support this retrograde memory effect. In a two-phase incidental encoding paradigm, participants viewed different pictures of tools and animals before and during Pavlovian fear conditioning. During Phase 1, these images were intermixed with neutral scenes, which provided a unique 'context tag' for this specific phase of encoding. A few minutes later, during Phase 2, new pictures from one category were paired with a mild shock (threat-conditioned stimulus; CS+), while pictures from the other category were not shocked. FMRI analyses revealed that, across-participants, individuals who showed aversive learning-related retroactive memory benefits for Phase 1 CS+ items were also more likely to exhibit three brain effects: first, greater spontaneous reinstatement of the Phase 1 context when participants viewed conceptually-related CS+ items in Phase 2; second, greater successful encoding-related VTA/SN and LC activation for Phase 2 CS+ items; and third, learning-dependent increases in post-encoding hippocampal functional coupling with CS+ category-selective cortex. These biases in hippocampal-cortical connectivity also mediated the relationship between VTA/SN aversive encoding effects and across-participant variability in the retroactive memory benefit. Collectively, our findings suggest that both online and offline brain mechanisms may enable threatening events to preserve memories that acquire new significance in the future.

1. Introduction

Aversive information is preferentially processed and remembered (Kensinger et al., 2007; LaBar & Cabeza, 2006; Mather & Sutherland, 2011). However, we don't always know the motivational relevance of information at the moment of encoding. Thus, while some information might seem inconsequential in the moment, an adaptive memory system should retain some seemingly mundane details, at least temporarily, in case this information gains significance in the future. The goal of the present study was to investigate the neural correlates by which an aversive event retroactively prioritizes episodic memory for related events encoded close in time.

Post-encoding arousal influences memory consolidation for recently encoded stimuli. This has been demonstrated with stressors (Andreano & Cahill, 2006; McCullough & Yonelinas, 2013; Sazma, McCullough, et al., 2019), norepinephrine (Southwick et al., 2002), exercise (Nielson et al., 1996), reward (Braun et al., 2018; Murayama & Kitagami, 2014; Patil et al., 2017), and aversive videos (Nielson & Powless, 2007;

* Corresponding author at: Department of Psychology, Columbia University, Schermerhorn Hall #406, 1190 Amsterdam Ave, New York, NY 10027, United States.
E-mail address: ld24@columbia.edu (L. Davachi).

https://doi.org/10.1016/j.nlm.2021.107572
Received 28 May 2021; Received in revised form 17 October 2021; Accepted 26 November 2021
Available online 3 December 2021

Nielson et al., 2005). In some cases, post-encoding arousal enhances memories for some but not all preceding stimuli (Liu et al., 2008; Preuss & Wolf, 2009; Sazma, Shields, et al., 2019; Smeets et al., 2007). Predicting what neutral memories are prioritized by a subsequent aversive event remains a challenge. Recent findings demonstrate that associative learning with aversive or appetitive outcomes selectively and retroactively enhances episodic memory for neutral information conceptually related to the more salient event (Dunsmoor et al., 2015; Hennings et al., 2021; Patil et al., 2017). Conceptual overlap may therefore be key in determining the fate of seemingly mundane details encoded close in time

While the neural processes supporting this memory benefit are unknown, one compelling neurobiological mechanism is 'synaptic tagging' (Frey & Morris 1997) and its behavioral counterpart 'behavioral tagging' (Ballarini et al., 2009; Moncada et al., 2011; Moncada & Viola, 2007; Wang et al., 2010). According to tagging models, a weak learning experience sets a 'learning tag' in activated synapses, creating a weak memory trace that is short-lived. However, if a stronger event is experienced minutes to a few hours later and engages overlapping neural ensembles, this learning tag can be stabilized to create a more enduring long-term memory ((see also Joels et al., 2006). Whether behavioral tagging offers a valid neurobehavioral framework for understanding retroactive episodic memory in humans is unclear. Here, we used a hybrid episodic memory and Pavlovian threat conditioning design during functional neuroimaging (fMRI) to target online and offline mechanisms that relate to the retroactive memory for conceptually-related stimuli.

The neural overlap between weak and strong learning is a critical feature of behavioral tagging (Ballarini et al., 2009). To investigate neural overlap in fMRI, we first generated a mental context-tag composed of scene images injected during weak encoding. A multi-voxel pattern analysis was then used to decode spontaneous scene reinstatement patterns during subsequent fear conditioning (e.g., Gershman et al., 2013). We hypothesized that memory for related stimuli would be associated with subsequent reinstatement of the weak learning context at the moment of strong learning (i.e., threat conditioning), thus demonstrating overlap between the weak and strong learning events promotes mnemonic processing. Another critical feature of behavioral tagging is engagement of the noradrenergic and midbrain dopaminergic (DA) systems to trigger release of plasticity-related proteins that stabilize a weak learning tag (Moncada, 2017; Moncada et al., 2011; Moncada & Viola, 2007; Wang et al., 2010). We therefore examined whether activation of catecholaminergic nuclei during threat conditioning promotes memory of previously encoded exemplars. Finally, behavioral tagging is a model for the consolidation of newly formed memories. A substantial literature in humans shows that post-encoding hippocampal-cortical connectivity is associated with later memory for recently encoded categorical stimuli directly associated with a salient outcome (de Voogd et al., 2016; Hermans et al., 2017; Murty et al., 2016; Tambini & Davachi, 2019; Tambini et al., 2010; Tompary et al., 2015). Thus, we examined if post-encoding connectivity between the hippocampus and category-selective cortex also supports retroactive memory for stimuli related to the threat-conditioned category.

2. Methods

2.1. Participants

Twenty-seven healthy young adults were recruited from the New York University Psychology Subject Pool and nearby community to participate in this experiment. All participants provided written informed consent approved by the New York University Institutional Review Board and received monetary compensation for their participation. All eligible individuals were right-handed, had normal or normal-to-corrected vision and hearing, and were not taking psychoactive medications. Nine participants were excluded from data analyses for the following reasons: five participants didn't return for session 2; two people fell asleep during scanning; one participant withdrew from session 1; and one participant had an incidental finding on his anatomical brain scan. Recruiting additional participants was not possible due to planned decommission of the MRI scanner at NYU. In total, data from eighteen participants were analyzed in this study (8 women; $M_{age} = 22$, $SD_{age} = 2.26$).

2.2. Materials

The stimuli consisted of colored photographs used in Dunsmoor et al. (2015) as well as new scene images (Dunsmoor et al., 2015). There were three categories: 120 neutral tool images, 120 neutral animal images, and 240 outdoor scenes. Half of the outdoor scene images were phase scrambled and included in the localizer phase of the experiment. All pictures were originally obtained from the website http://www.lifeonwhite.com and from the internet. The tool and animal images were unique exemplars and had different names from each other.

2.3. Procedure

This study involved two separate sessions that were spaced 24 h apart. In the first session (fMRI), participants' brains were scanned during a scene functional localizer scan, a two-phase incidental encoding task, and three intervening resting-state scans (see Fig. 1). The resting-state scans were collected at the following times: (1) upon entering the scanner (baseline); (2) between Phase 1 and Phase 2 of the incidental encoding task (pre-conditioning); and (3) after Phase 2 of the incidental encoding task (post-conditioning). During the ~6-m resting-state scans, participants were instructed to lay still with their eyes open while viewing a black fixation cross in the middle of the screen. An infrared camera was used to monitor pupil diameter and ensure participants did not fall asleep during these scans.

2.3.1. Phase 1 incidental encoding (preconditioning)

In Phase 1 of the incidental encoding task (preconditioning), participants viewed a series of neutral animal and neutral tool images for 4 s each (Fig. 1). The participants' task was to classify each image as an animal or tool via button press. During the inter-stimulus interval (ISI), a series of four outdoor scene images were presented for 1 s each. Scenes were only presented during this initial encoding phase to create a unique 'mental context tag' for Phase 1 information (Gershman et al., 2013). A total of 30 tools and 30 animals were presented in pseudo-randomized order, such that no more than 3 objects from the same object category appeared in a row.

2.3.2. Phase 2 incidental encoding (threat conditioning)

Approximately 6 min after Phase 1 encoding, participants viewed a new set of 30 animal and 30 tool images (Phase 2; Pavlovian threat conditioning phase). Unlike Phase 1 of incidental encoding, however, one of these categories co-terminated with a mild electrical shock on 2/3rds of the trials (20 trials; duration = 200 ms; CS+ category). Each image was presented for 4 s, during which time participants had to indicate whether or not they expected to be shocked on that trial. An 8-s fixation cross centered on a gray background was inserted between each image. The object categories serving as CS+/CS− (i.e., animals or tools) were counterbalanced across participants (shock category sub-groups: $N_{animal} = 7$, $N_{tool} = 11$). To verify successful aversive learning, we performed a repeated-measures analysis of variance (rm-ANOVA) on the number of CS+ and CS− trials that the participants rated they expected a shock to occur. CS Type (CS+, CS−) was modeled as the factor of interest, with Shock Category (animal, tool) as a between-subjects covariate.

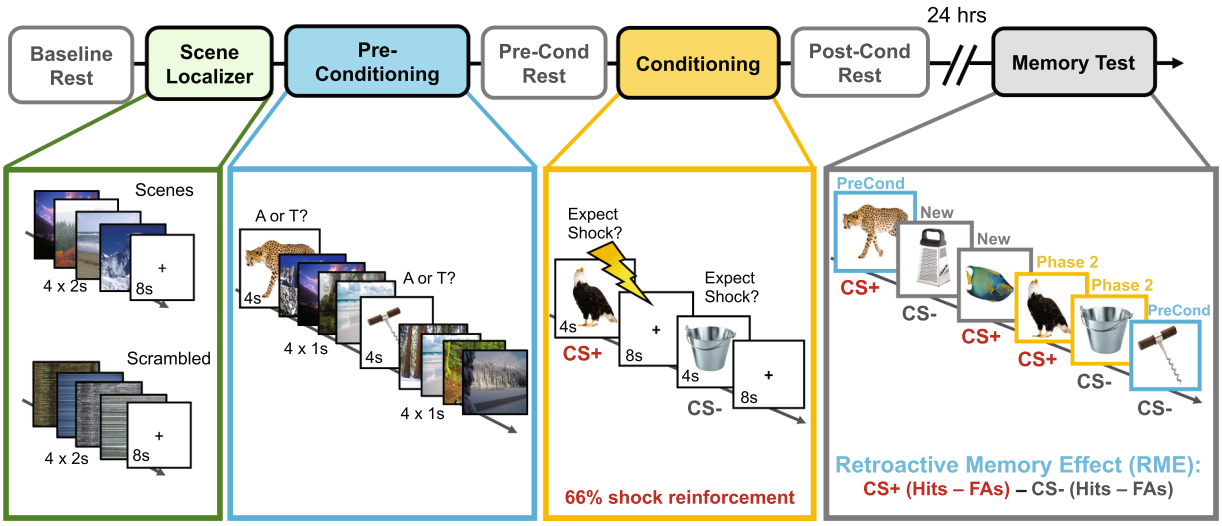

Fig. 1. Overview of experiment design. Prior to the main two-phase incidental encoding task, participants performed a functional localizer task in which they viewed mini-blocks of neutral scenes and scrambled images (green box). Next, participants performed Phase 1 of a two-phase incidental encoding task. During this first phase (preconditioning; blue), participants viewed a series of neutral animal and tool images and had to classify the category of each image via button press. Importantly, four neutral scene images were also inserted in between each tool and animal image to create a unique 'context tag' for Phase 1 information. Approximately 6 min later, participants incidentally encoded novel animal and tool images (Phase 2; gold box). This time, however, a mild electrical shock also co-terminated with 2/3rds of the trials from one visual category (conditioning), thereby making that conceptual information motivationally significant (CS+). When each image appeared, participants rated whether they expected a shock on that trial, which provided a behavioral index of aversive learning. To examine how aversive learning influenced post-encoding hippocampal resting-state functional connectivity, resting-state scans were collected immediately before and after Phase 2 of the encoding task. Participants returned 24 h later for a surprise recognition memory test (gray box). During this memory test, participants viewed all of the animals and tools they had seen during Phase 1 (blue borders) and Phase 2 (gold borders) of the experiment, along with new animal and tool 'lure' items (gray borders). Participants simply indicated whether each item had been seen previously ('old' judgement) or was completely new ('new' judgement). The critical measure in this study was the retroactive memory effect (RME), which was computed by subtracting participants' corrected recognition memory for Phase 1 CS− items from their corrected recognition memory for Phase 1 CS+ items. Higher RME scores index a greater retroactive memory benefit for Phase 1 items that are conceptually-related to the aversive (CS+) category from Phase 2. Lightning bolt indicates shock. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3.3. Delayed recognition test

Approximately 24 h later, participants returned and were given a surprise recognition memory test while outside of the scanner. During this test, participants viewed all of the object images they had seen in the MRI scanner as well as 60 new animal images and 60 new tool images (lures). A total of 240 images were presented in randomized order. During this self-paced memory test, the participants rated whether each item was 'old' (previously presented) or 'new' (never seen in the scanner) according to their confidence level. There were four options: 'definitely new', 'maybe new', 'maybe old', or 'definitely old'. To examine the effects of aversive learning on recognition memory, a 2 (CS Type: CS+, CS−) × 2 (Encoding Phase: preconditioning, conditioning) mixed ANCOVA was performed on corrected recognition scores (CRS), with Shock Category (tools, animals) as a between-subjects covariate. The CRS values were computed by subtracting participants' hit rates (correctly said 'old') from their false alarm rates (said 'old' but it was new) for each encoding phase and each category, separately.

Our primary behavioral measure of interest was the degree to which participants exhibited a selective and retroactive memory benefit for CS+ exemplars from Phase 1; that is, if participants were shocked on animals during conditioning (Phase 2), they would also show better memory for animals compared to tools from the preconditioning block (Phase 1) of the task. To compute this aversive learning-related memory bias measure, we subtracted participants' corrected recognition performance for CS− exemplars from their performance on CS+ exemplars that were encountered during Phase 1. Henceforth, we will simply refer to this aversive learning-biased retroactive memory effect measure as "RME".

2.4. Skin conductance response (SCR) methods

To index autonomic arousal responses during fear conditioning, SCRs were recorded via MRI-compatible electrodes placed on participants' right wrist and measured with a BIOPAC MP100 System (Goleta, CA). Shocks were delivered to the right wrist using pre-gelled MRI-compatible electrodes connected to a stimulator (Grass Medical Instruments). Upon entering the MRI scanner, the shock electrodes were attached to the right wrist and the shock level was calibrated to be at level deemed "highly annoying but not painful" (e.g., Dunsmoor et al., 2011). Although it was not the focus of this study, we were unable to link SCRs to the appropriate trial labels due to a programming error. Thus, SCRs were not analyzed. Importantly, however, threat conditioning success was validated by trial-by-trial shock expectancy ratings, which are considered a valid measure of human conditioning with strong face- and construct-validity (Boddez et al., 2013).

2.5. fMRI acquisition and analyses

2.5.1. MRI data acquisition

All neuroimaging data were acquired on a 3 T Siemens Allegra scanner located at the Center for Brain Imaging at New York University. The visual stimuli were displayed on a mirror in front of participants' eyes that was attached to a 32-channel matrix head coil. A high-resolution T1-weighted anatomical image (MPRAGE) was acquired to aid with functional image co-registration (slices = 176 axial; TR/TE/TI = 2500 ms/3.93 ms/900 ms; FOV = 256 mm; voxel size = 1 mm^3 isotropic; slice thickness = 1 mm; bandwidth = 130 Hz/Px). Functional images for the three resting-state runs (184 volumes each), scene localizer task (164 volumes), preconditioning run (244 volumes), and conditioning run (364 volumes) were acquired using the same echo-

planar imaging sequence (TR/TE = 2000/15 ms, 34 interleaved slices, FOV = 102 mm; FA = 82°; voxel size = 3 mm^3 isotropic).

2.5.2. Image preprocessing

Image preprocessing was performed using FSL Version 5.0.4 (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The first four volumes of each functional scan were discarded for signal stabilization. Functional volumes were preprocessed by removing non-brain tissue using BET, applying spatial smoothing using a Gaussian kernel of 6 mm full-width-at-half-maximum (FWHM), grand-mean intensity normalization of the 4D data set by a single multiplicative factor, and applying a high-pass temporal filter of 100 s. Additionally, volumes with extreme head motion artifact were regressed from the dataset. Each participant's denoised mean functional volume was co-registered to his/her T1-weighted high-resolution anatomical image using brain-based registration (BBR) with 7 degrees of freedom. Anatomical images were then co-registered to the 2 mm isotropic MNI-152 standard-space brain using an affine registration with 12 degrees of freedom.

2.5.3. Parahippocampal place area region-of-interest (ROI)

Before preconditioning, a functional localizer task was used delineate the parahippocampal place area (PPA), a cortical region in the ventral visual stream that is specialized to process to scene information (Epstein & Kanwisher, 1998). The localizer scan consisted of 40 colored scenes and 40 phase-scrambled scenes. Image presentation was divided into 20 mini-blocks lasting 16-s each. Each mini-block contained either four individual scenes or four individual scrambled images lasting 2 s each. These image quartets were followed by an 8-s fixation cross inter-trial-interval (ITI). Participants were instructed to press a button if one of the images repeated (1-back task). None of the scenes from the localizer task were also used in the preconditioning encoding task.

A general linear model (GLM) was fit to each participant's localizer functional data to localize and delineate the left PPA and right PPA. The GLM included separate square wave-form regressors for the scene and scrambled image mini-blocks that were convolved with a double-gamma HRF. Whole-brain statistical parametric maps were calculated for the scene > scramble contrast using a one-sample *t*-test. To correct for multiple comparisons, Z-statistic images were thresholded using clusters determined by $Z > 2.3$ and a corrected cluster significance threshold of $P = .05$ (Worsley, 2001). The final four mini-blocks were excluded from data analyses due to a computer error. The functionally-defined ROI masks for the left/right PPA regions were defined as 6 mm spheres centered upon peak voxels in the parahippocampal gyrus within the scene > scrambled whole-brain contrast map. These spheres were then merged with the same uncorrected statistical maps thresholded at $Z = 2.57$ to retain gray matter voxels that were selective for processing scene information. Using this approach, we were able to define the PPA for all participants (mean peak MNI coordinates across participants: Left PPA [-27–50 −10]; Right PPA [28–47 −11]; see Fig. 3).

2.5.4. Regions-of-interest definitions

Target anatomical ROIs were defined for the ventral tegmental area/substantia nigra (VTA/SN) and locus coeruleus (LC), as well as animal/tool category-selective sensory cortex and left/right hippocampus. The VTA/SN anatomical mask was derived from an existing probabilistic atlas (Murty et al., 2014) and thresholded at 50% probability. A standard-space LC anatomical mask was derived from a separate study that used neuromelanin-sensitive weighted MRI to identify LC neurons in the pontine tegmentum (Keren et al., 2009).

Animal and tool-selective cortical ROIs were defined as 4 mm spheres centered upon peak voxel coordinates reported in a previous fMRI study (Dunsmoor et al., 2014). The animal-selective cortical ROIs included areas of right inferior occipital gyrus and right lateral fusiform gyrus, whereas the tool-selective cortical ROIs included areas of left middle occipital gyrus and left medial fusiform gyrus. Participant-specific left and right hippocampal anatomical ROIs were extracted using the FIRST tool in FSL. These masks were then thresholded at 25% probability and binarized.

2.5.5. Neural context reinstatement analysis

One of our primary goals was to determine if strong, aversive events modulate the reactivation of prior mental contexts, and whether the degree of neural context reactivation relates to the selective consolidation of conceptually related exemplars to the aversive stimuli. To test these ideas, we inserted neutral scene images between the object images in Phase 1 of the incidental encoding task (preconditioning). Because scene images were only presented during this phase of incidental encoding, they provided a unique 'context tag' for Phase 1 mental representations (Gershman et al., 2013). Thus, we interpret any evidence of scene information during Phase 2 as neural reinstatement of the Phase 1 mental context.

To measure the amount of scene reinstatement during aversive learning, we first trained a multivoxel pattern classifier to discriminate scenes versus scrambled scene images. Specifically, an L2-regularized multinomial logistic regression classifier was trained on multivoxel patterns of PPA BOLD signal (betas) during the scene localizer (see Fig. 3a). An eightfold cross-validation procedure verified that the pattern classifier was highly accurate at discriminating multivoxel patterns of PPA activation between processing scene images versus scrambled scene images (mean accuracy = 96% +/ 0.021%).

To determine if the Phase 1 mental context was reinstated during new aversive learning, we used a Least Squares All (LSA) approach to analyze the conditioning-phase data and acquire trial-by-trial estimates of multivoxel activity in the PPA. In this approach, BOLD signal for each of the conditioning-phase trials was estimated simultaneously in a single voxel-wise GLM (Rissman et al., 2004). Separate trial regressors were created by modeling the onset time for each animal and tool image with a duration of 4 s and convolving these regressors with a double-gamma hemodynamic response function (HRF). The six motion parameters, extreme head motion outliers, and shock deliveries (1-s stick function) were modeled as nuisance regressors in the GLM.

For each participant, individual parameter estimates were extracted separately from the left and right PPA voxels for each of the 60 conditioning-phase trials. The pattern classifier was then tested on these parameter estimates to estimate the amount of scene evidence (discrete values ranging between 0 and 1) during each conditioning-phase trial. To examine whether aversive learning biased the degree of Phase 1 neural context reinstatement, the classifier estimates of scene evidence were sorted by CS trial type (CS+ or CS−).

Because accumulated evidence from Pavlovian conditioning studies suggest that amygdala signatures of threat acquisition are more robust during earlier versus later phases of threat conditioning (Buchel et al., 1998; Dunsmoor et al., 2014; LaBar et al., 1998), we split the Phase 2 trials evenly into early and late conditioning-phase bins to see whether reinstatement of the prior context was more robust earlier on during aversive learning. A 2 (Hemisphere: left, right) × 2 (Conditioning Phase: early, late) × 2 (CS Type: CS+, CS−) mixed ANCOVA was performed on scene evidence values to examine the effects of aversive learning and timing on neural context reinstatement. Shock Category was modeled as a between-subjects covariate. Our logic for splitting the aversive learning phase was that, insofar as scene context representations are reactivated in PPA during aversive learning, arousal should up-regulate those activation patterns even further in the same manner as actually viewing scene information. Because neural signatures of aversive learning and arousal tend to be more evident earlier compared to later than conditioning (as indexed by amygdala activation patterns; (LaBar et al., 1998), we expected PPA reinstatement patterns to be enhanced during times when the modulatory effects of arousal were expected to be greatest. Indeed, a supplementary GLM analysis of whole-brain activation patterns during the aversive learning phase verified that right amygdala activation was greater during the early compared to late phase of conditioning (see Supplementary Materials).

To test our main hypothesis that online memory reactivation relates to the retroactive memory effect, we then performed partial Pearson's correlations between participants' RME scores and the amount of scene evidence output by the classifier for CS+ and CS− trials during conditioning, while also controlling for Shock Category. Importantly, we first conducted Shapiro-Wilk's tests on the independent and dependent variables as well as Breusch-Pagan tests to verify that all variables used in these linear regressions were normally distributed and the correlations were homoscedastic. All of the reported regression analyses met the statistical assumptions for a Pearson's correlation coefficient analysis.

2.5.6. Aversive memory encoding GLM analysis

In addition to Phase 1 context reinstatement, we explored whether the level of engagement of neuromodulatory systems during Phase 2 threat conditioning was also related to the selective retroactive memory benefit. We reasoned that, insofar as aversive events amplify encoding processes, these enhancements may also selectively strengthen the ongoing storage of conceptually-related representations from Phase 1. To test this possibility, we first performed a subsequent memory GLM analysis for the CS+ and CS− items from Phase 2 to dissociate the specific effects of aversive stimuli on new encoding processes (see Fig. 4, left panel). We then correlated individual differences in aversive encoding-related BOLD signal during Phase 2 to individual differences in the behavioral retroactive memory effect for Phase 1 items.

In this Phase 2 subsequent memory GLM, we modeled separate event-related regressors for the CS+ and CS− exemplars with durations of 4-s each. Each task regressor was then convolved with a dual-gamma canonical hemodynamic response function. Next, we sorted the 60 conditioning-phase images by CS trial-type (CS+/CS−) and by subsequent memory status (remembered: Hit; forgotten: Miss). Participant-level GLMs were constructed using 4 task regressors: (1) CS+ Hit, (2) CS+ Miss, (3) CS− Hit, and (4) CS− Miss. Additional nuisance regressors for the 6 motion parameters and extreme head movement outliers were also included in these models.

The resulting whole-brain contrast images were analyzed in higher-level mixed-effects analysis using FMRIB's local analysis of mixed effects (FLAME 1 (Beckmann et al., 2003). A single group average for each of the contrasts-of-interest was calculated using a one-sample *t*-test. To correct for multiple comparisons, Z-statistic images were thresholded using clusters determined by $Z > 2.3$ and a corrected cluster significance threshold of $P = .05$ (Worsley, 2001).

To test whether neuromodulatory activity was related to aversive memory enhancements for Phase 2 items, we extracted parameter estimates for each of our four regressors-of-interest from the VTA/SN and LC. These brainstem ROI values were then submitted to separate 2 (CS Type: CS+, CS−) × 2 (Memory: hit, miss) mixed ANCOVA's with Shock Category (shocked on animals or shocked on tools) as a between-subjects covariate to examine whether activation of mesolimbic dopaminergic nuclei (VTA/SN) and noradrenergic nuclei (LC) promote the encoding of aversive information. Because there is substantial evidence implicating neuromodulators in aversive memory enhancements (e.g., McGaugh, 2013; Shohamy & Adcock, 2010), we had strong directional hypotheses and chose to use one-tailed t-tests for these analyses

Our primary goal was to examine if neuromodulatory effects during aversive learning were also associated with the preferential consolidation of Phase 1 CS+ exemplars. To this end, we first computed separate aversive memory enhancement scores for VTA/SN and LC encoding-related BOLD signal using the following formula: [CS+ (Hit > Miss)] > [CS− (Hit > Miss)]. These brain-related Phase 2 aversive encoding scores were then linearly correlated with Phase 1 RME scores using Pearson's correlation coefficient analyses. Importantly, participants were only included in the conditioning-phase-related analyses if they had trials for all four memory/CS type bins. One participant had no trials for CS− Miss and was therefore excluded in these brainstem-related analyses (remaining N = 17).

2.5.7. Aversive learning-dependent changes in post-encoding hippocampal functional connectivity

Our next important question was whether hippocampal-cortical functional connectivity was related to the selective retroactive memory benefit. Previous fMRI work has shown that, following reward or aversive learning, the hippocampus becomes more functionally coupled with regions that process the information associated with those motivationally-significant events (de Voogd et al., 2016; Murty et al., 2016). Inspired by these studies, we predicted that aversive learning may also bias subsequent hippocampal connectivity with CS+ visual cortical regions, and that this functional coupling may help preserve memory for the now-salient items encountered in Phase 1.

Here, we performed hippocampal seed-based functional connectivity analyses. We first extracted the mean hippocampal BOLD timeseries from each participant's three resting-state scans (see Fig. 1 for timings). These hippocampal activity timeseries were then modeled as regressors of interest in separate whole-brain GLM's. Additional nuisance regressors for the 6 motion parameters, extreme head movement outliers, and both white matter and cerebrospinal signals were also included in these models. For the latter two signals, we used FSL FAST to acquire probabilistic white matter and CSF voxel-wise masks for each participant. These masks were thresholded at 30% tissue-type probability and then binarized prior to extracting nuisance BOLD signal.

To quantify experience-dependent changes in hippocampal functional connectivity, we subtracted hippocampal-ROI connectivity estimates from the pre-conditioning resting-state scan from connectivity estimates from the post-conditioning resting-state scan. Finally, we performed Pearson's partial correlations to test our hypothesis that increased hippocampal connectivity with neuromodulatory nuclei and CS+ cortex after aversive learning would relate to RME scores across participants. Because we did not have explicit predictions about hippocampal laterality effects, we collapsed the ROI results across brain hemispheres.

2.5.8. Mediation analysis

In the final analysis, we examined if aversive learning-related changes in hippocampal functional connectivity accounts for the relationship between VTA/SN aversive encoding processes and the retroactive memory effect. For this, we used the mediation package in R. The significance of this mediation model was tested using nonparametric bootstrapping with 1000 iterations. We report on the average causal mediation effect (ACME).

3. Results

3.1. Shock expectancy ratings and recognition memory

During the conditioning phase, participants were significantly more likely to indicate that they expected a shock on CS+ compared with CS− trials, verifying successful fear acquisition at the category level, $F(1,16) = 159.51$, $p < .001$, $\eta_p^2 = 0.91$ (Fig. 2a).

Delayed recognition memory performance is displayed in Fig. 2b-e. The results revealed no significant difference in false alarm rates between CS+ and CS− category lure items, $F(1,16) = 0.14$, $p = .71$, $\eta_p^2 = 0.009$ (Fig. 2b). A 2 (Encoding Phase: preconditioning, conditioning) × 2 (CS Type: CS+, CS−) ANCOVA with Shock Category as a between-subjects covariate revealed a marginally significant main effect of Encoding Phase, $F(1,16) = 4.34$, $p = .054$, $\eta_p^2 = 0.21$, on corrected recognition scores, with memory performance being better for items encoded during the conditioning compared to the preconditioning phase of encoding (Fig. 2d). In addition, we observed a significant encoding phase-by-type interaction effect on corrected recognition scores, $F(1,16) = 5.29$, $p = .035$, $\eta_p^2 = 0.25$, such that participants were significantly better at remembering CS+ compared with CS− items during the conditioning phase (Phase 2) compared with the preconditioning phase (Phase 1). However, there was no significant main effect of CS Type on

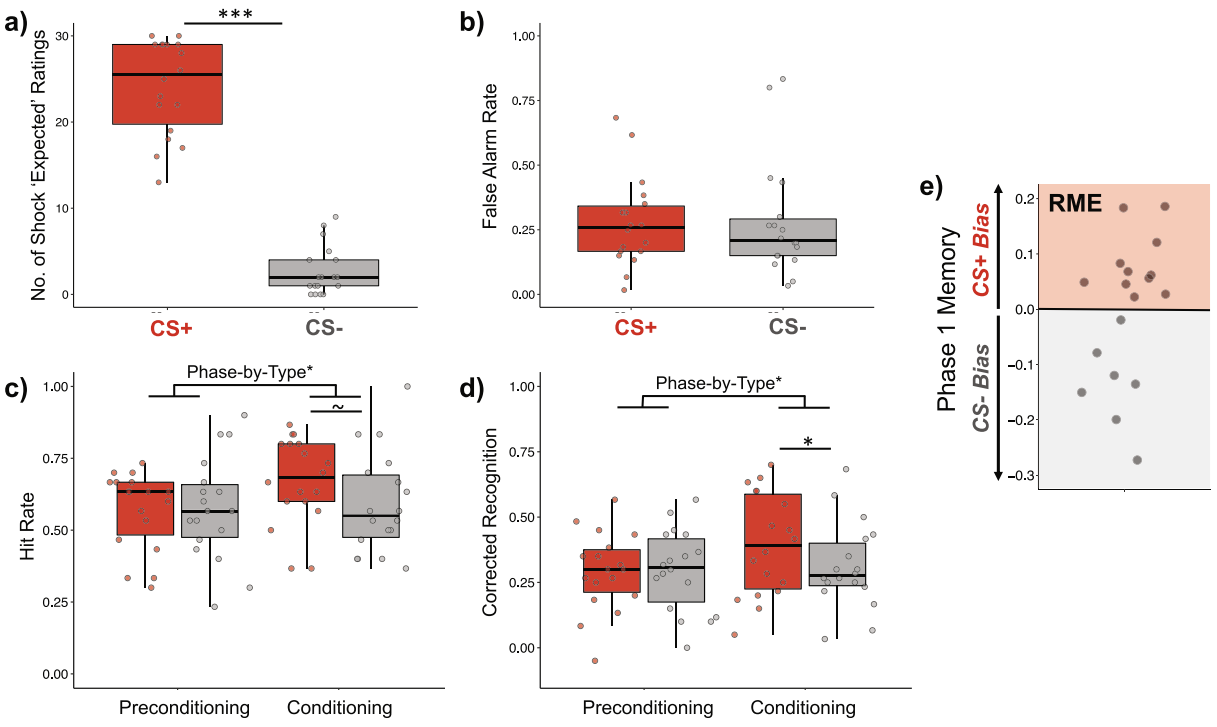

Fig. 2. Aversive learning enhanced memory for CS+ items during conditioning, but had highly variable effects on memory biases for Phase 1 items. (a) Shock expectancy ratings for CS+ (red bar) and CS− (gray bar) trials verified that participants were able to learn which visual category was paired with shock. Values reflect the number of trials that participants indicated they expected to be shocked, broken down by CS Type (out of 30 trials each). Delayed recognition memory test results for (b) false alarm rates and (c) hit rates broken down by CS Type. (c) Corrected recognition scores were computed by subtracting participants' false alarm rates from their hit rates for each CS Type, separately. After a 24-hr delay, participants showed an aversive learning-related memory enhancement for CS+ (red bars) compared to CS− items (gray bars) from the conditioning phase of the experiment (Phase 2). However, aversive learning did not retroactively bias memory in favor of conceptually-related items from the preconditioning phase (Phase 1). For plots a-d, colored boxplots represent 25th–75th percentiles of the data, the center line the median, and the error bars the s.e.m. Overlaid dots represent individual participants. (c) Across-participant variability in the aversive learning-related selective and retroactive memory effect (RME) for Phase 1 items. A subtraction score between corrected recognition rates for CS+ minus CS− items from Phase 1 is plotted on the y-axis. Values below zero represent a memory bias towards remembering items from the CS− category (gray box). Values above zero represent a memory bias towards remembering items from the CS+ category (red box), an effect that is termed "RME". ~ p = .050; *p < .05; ***p < .001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

corrected recognition scores, $F(1,16) = 2.00$, $p = .18$, $\eta_p^2 = 0.11$.

Separate follow-up planned repeated-measures ANCOVA's on the two encoding phases revealed that corrected recognition performance was significantly better for CS+ compared with CS− category items from the conditioning phase, $F(1,16) = 4.83$, $p = .043$, $\eta_p^2 = 0.23$ (Fig. 2d). However, memory performance at the group level did not significantly differ between CS+ and CS− exemplars from the preconditioning encoding phase, $F(1,16) = 0.031$, $p = .86$, $\eta_p^2 = 0.002$. For completeness, raw hit rates are also displayed in Fig. 2c. These followed the same pattern and statistical significance as the corrected recognition scores, our main measure of interest, with one slight exception: the p-value for the main effect of CS Type for Phase 2 on hit rates was exactly $p = .05$ (Fig. 2c).

As shown in Fig. 2e, there was substantial variability in memory performance across individuals as a function of CS type. In the subsequent fMRI analyses, we leveraged this variability in memory performance to examine how individual differences in RME were related to various online and offline brain measures of neural reactivation and consolidation.

3.2. Neural context reinstatement analysis

To determine if Phase 1 neural context representations were reactivated during aversive learning, we measured the amount of scene classifier evidence that was present when participants viewed CS+ and CS− exemplars during conditioning (Phase 2). A 2 (Hemisphere: left, right) × 2 (Conditioning Phase: early, late) × 2 (CS Type: CS+, CS−) repeated-measures ANCOVA with Shock Category as a covariate revealed significantly greater scene evidence during the first compared with the second half of fear conditioning, $F(1,16) = 5.81$, $p = .028$, $\eta_p^2 = 0.27$ (Fig. 3b). There was also a significant type-by-phase interaction effect, such that scene evidence was significantly greater during CS+ exemplars compared with CS− exemplars in the first versus second half of conditioning, $F(1,16) = 5.46$, $p = .033$, $\eta_p^2 = 0.26$. There were no other main or interaction effects on scene evidence values. These results suggest that the selective effects of aversive learning on prior context reinstatement is strongest during earlier versus later phases of threat conditioning.

In the previous analysis, we found that scene-related neural context reinstatement was qualitatively the strongest when participants viewed negative items during the first half of conditioning (see Fig. 3b). We next asked if this Phase 1 context reinstatement is related to the selective consolidation of conceptually-related CS+ items. A Pearson's partial linear correlation revealed that the amount of scene evidence in left PPA on early-phase CS+ trials was significantly positively correlated with Phase 1 RME scores, partial $r(15) = 0.52$, $p = .031$. This brain-behavior relationship was only observed in relation to early-phase CS+ trials in left PPA, as no relationship was observed with scene evidence on any other trial type (CS+ or CS−), half of conditioning (Early or Late) or right PPA (all p's > 0.05). Critically, we also wanted to demonstrate that scene evidence was indeed a reflection of reinstatement processes rather than of simply viewing the current CS+ and CS− images during Phase 2. To this end, we also performed the same correlation analyses with CS+ memory bias scores for items from Phase 2 rather than Phase 1. None of

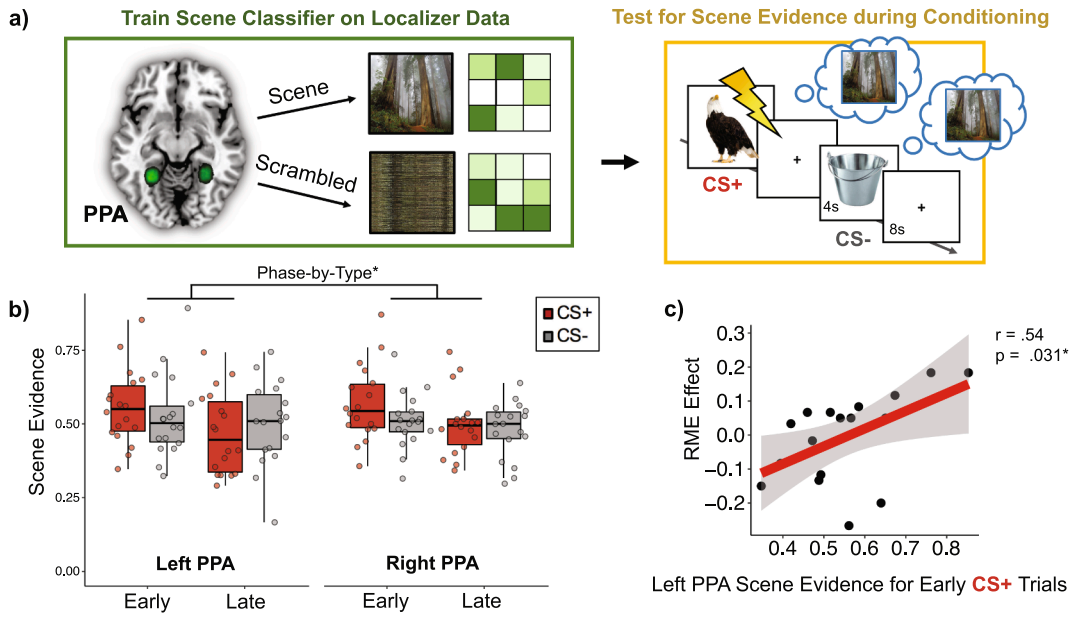

Fig. 3. Neural reinstatement of the prior mental context increased during aversive moments in the first half of fear conditioning, and this pattern was related to greater retroactive memory benefits for items from the CS+ category. (a) A multivoxel pattern classifier was trained on the localizer data to discriminate scene versus scrambled images in the left and right parahippocampal place area (PPA; top left green box). Green circles represent study-specific probabilistic left and right PPA masks across all participants. The scene classifier was then tested on fMRI data from the conditioning phase of the experiment (Phase 2; top right gold box). Because scenes were only presented during Phase 1 of the incidental encoding task, any scene evidence output by the classifier during the CS+ and CS− images during conditioning was interpreted as reinstatement of the prior mental context (blue thought bubbles). (b) Scene evidence output by the pattern classifier was greater for aversive items (CS+) compared to neutral (CS−) items during the first versus second half of conditioning. Colored boxplots represent 25th–75th percentiles of the data, the center line the median, and the error bars the s.e.m. Overlaid dots represent individual participants. (c) A partial linear correlation analysis revealed that aversive-learning induced retroactive memory effect (RME) was correlated with greater scene evidence on CS+ trials from the first half of conditioning. Lightning bolt indicates shock. *p < .05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

these correlations were significant (all p's > 0.05).

In summary, these results show that neural reinstatement of prior mental contexts is greater during early compared to late phases of aversive learning when arousal responses may have been highest (indexed by greater amygdala activation; see Supplementary Materials). Our results also suggest that reactivation of the prior neutral context

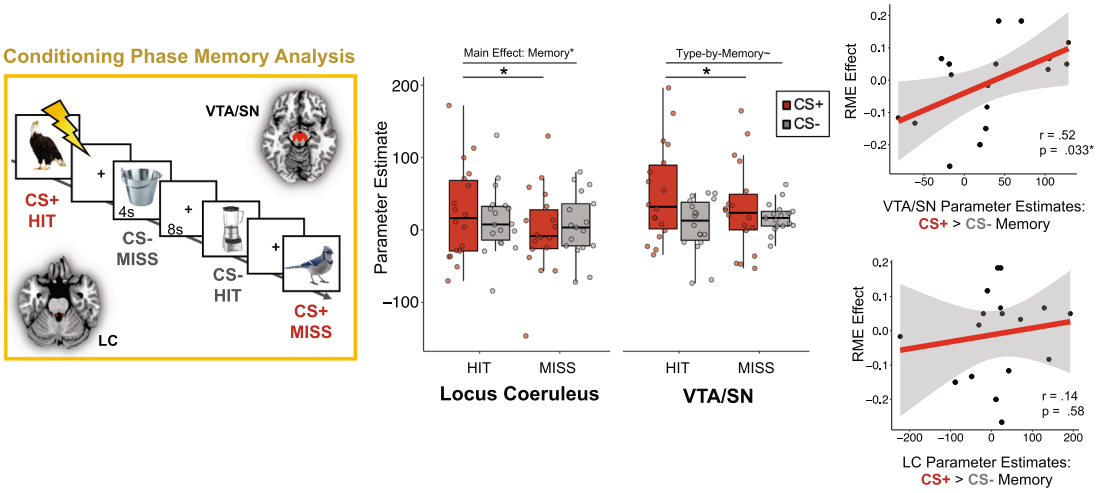

Fig. 4. Effects of catecholaminergic nuclei activation on aversive memories, and their relationship with the retroactive memory benefit for conceptually-related items. (Left Panel) A subsequent memory general linear modeling (GLM) analysis was performed for items that were incidentally encoded during the conditioning phase of the experiment. Each trial from conditioning was sorted by CS type (CS+ or CS−) and whether it was remembered 24 h later (hit or miss). BOLD signal was extracted from an anatomical atlas-defined VTA/SN and locus coeruleus (LC) mask for each of the conditioning-phase trials. (Middle Panel) VTA/SN activation was significantly greater when participants successfully encoded CS+ items (dark red bar), and was also more engaged during encoding of CS+ compared to CS− items during conditioning. LC activation was significantly greater during successful item encoding, which was primarily driven by memory enhancement effects for CS+ items. Colored boxplots represent 25th–75th percentiles of the data, the center line the median, and the error bars the s.e.m. Overlaid dots represent individual participants. (Right Panel) Activation-related aversive memory enhancement scores for the VTA/SN (top) and LC (bottom) were computed by subtracting encoding-related parameter estimates for CS− trials (i.e., hit minus miss) from encoding-related parameter estimates for CS+ trials. The VTA/SN aversive memory enhancement measure was positively correlated with the magnitude of the retroactive memory benefit (RME) across participants, but LC scores were not. *p < .05. ~ p < .01. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

during earlier aversive moments may serve to selectively strengthen recent, conceptually-related memories. Additionally, the relationship between scene classifier evidence and memory was not seen for items encoded in Phase 2, lending credence to our interpretation that these indeed reflect context reinstatement effects.

3.3. Aversive memory encoding during conditioning

In the next analysis targeting 'online' effects of aversive learning, we first examined if neuromodulatory activity during encoding of CS+ exemplars during threat conditioning was increased and related to memory for those items (Fig. 4, leftmost panel). Furthermore, we were especially interested in whether threat-related encoding patterns in the VTA/SN and LC during Phase 2 were also related to the selective and retroactive memory benefit.

Brainstem ROI analyses revealed a significant main effect of CS Type on VTA/SN BOLD signal, $F(1,16) = 5.82$, $p = .014$, $\eta_p^2 = 0.27$, with activation being significantly greater when participants viewed CS+ items compared to CS− items during conditioning (Fig. 4, middle panel, red bar). Furthermore, activation of VTA/SN differentiated CS+ items that would later be remembered from those that would be forgotten, but did not do so for CS− items. Planned paired t-tests revealed that VTA/SN BOLD signal was significantly higher when participants successfully encoded CS+ items, $t(17) = 2.26$, $p = .019$ (one-tailed), but not when they successfully encoded CS− items, $t(16) = -0.98$, $p = .17$ (one-tailed), during Phase 2. This pattern was qualified by a marginally significant type-by-memory interaction effect, $F(1,16) = 3.64$, $p = .066$, $\eta_p^2 = 0.19$, with CS+ trials leading to greater encoding-related VTA/SN activation than CS− trials.

For the LC, there was a significant main effect of Memory Outcome on BOLD signal, with participants exhibiting greater LC BOLD signal for items that were subsequently remembered, $F(1,16) = 3.52$, $p = .040$, $\eta_p^2 = 0.18$ (one-tailed). Separate planned follow-up t-tests on LC activation revealed a significant main effect of memory for Phase 2 CS+ items, with LC BOLD signal being greater for CS+ items that were subsequently remembered compared to forgotten, $t(17) = 1.84$, $p = .042$ (one-tailed). This main effect was not significant for Phase 2 CS− items, $t(16) = 0.32$, $p = .27$ (one-tailed). We did not observe any other main or interaction effects on LC activation (p's > 0.05).

Importantly, across participants, the magnitude of the Phase 2 aversive memory enhancement supported by VT/SN activation was significantly positively correlated with the extent of RME across participants, partial $r(17) = 0.52$, $p = .033$ (Fig. 4, rightmost panel). By contrast, there was no significant correlation between LC aversive memory activation and RME scores, partial $r(17) = 0.14$, $p = .58$. These findings suggest that VTA/SN activation not only selectively promotes encoding of aversive material but also enhances the selective consolidation of recently encoded related information. Thus, the local effects of aversive learning on VTA/SN encoding processes also appear to converge with the selective and ongoing consolidation of overlapping memories. While LC activation was associated with enhanced memory encoding in general, especially for aversive items, this modulation was not associated with the retroactive memory benefit for CS+ exemplars from Phase 1.

3.4. Post-encoding hippocampal functional connectivity results

In the next analysis, we examined whether aversive learning biases post-encoding hippocampal functional connectivity. Four separate 2 (Hemisphere: left hippocampus, right hippocampus) × 2 (Rest Phase: preconditioning rest, postconditioning rest) ANCOVAs with Shock Category as a covariate revealed that, on average, aversive learning did not significantly alter hippocampal functional connectivity with the LC, VTA/SN, CS− cortex, or CS+ cortex (all p's > 0.05).

While we did not observe any main carryover effects of aversive learning on overall hippocampal connectivity, we were primarily interested in whether variability in experience-dependent hippocampal functional connectivity changes was related to the magnitude of RME effects. In these linear correlation analyses, hippocampal pre-to-post aversive learning functional connectivity values were collapsed across hemispheres, given we did not observe any interactions between these values and brain hemisphere in the prior analysis.

Consistent with our main hypotheses, increased post-encoding hippocampal coupling with CS+ category-selective cortex was positively correlated with Phase 1 RME scores, partial $r(15) = 0.62$, $p = .0082$, whereas changes in hippocampal-CS− category-selective cortex functional connectivity were not, partial $r(15) = -0.13$, $p = .61$ (Fig. 5). A William's test for dependent correlations indicated that these two brain-behavior correlations were also significantly different from each other, $t = -2.95$, $p < .01$. This finding suggests that aversive learning may enhance the selective consolidation of conceptually-related stimuli by biasing post-encoding hippocampal connectivity to target sensory processing regions associated with the aversive category. When examining connectivity patterns with brainstem nuclei, we did not observe any significant correlations between hippocampal changes in functional coupling with the LC, partial $r(15) = 0.12$, $p = .64$, or VTA/SN, partial $r(15) = 0.34$, $p = .18$.

To ensure that these brain-behavior relationships were specifically driven by aversive learning, we performed additional control analyses. For these regressions, we queried hippocampal-cortical functional connectivity changes from baseline rest period to the rest period following Phase 1 encoding, before any aversive learning had occurred (preconditioning; see Fig. 1). Because neither category of information was differentially salient during Phase 1 encoding, we did not expect there to be any relationships between hippocampal connectivity patterns and RME scores. Indeed, the control analyses revealed no significant relationships between RME scores and changes in hippocampal functional coupling with CS− category-selective cortex, partial $r(15) = -0.36$, $p = .15$, CS+ cortex, partial $r(15) = -0.40$, $p = .11$, or the difference between the two (Williams test: $t = 0.11$, $p < .91$). Furthermore, there were no significant correlations between RME scores and hippocampal functional connectivity changes with the LC, partial $r(15) = 0.019$, $p = .94$, or VTA/SN, partial $r(15) = -0.34$, $p = .18$. These analyses help verify that memory-related biases in hippocampal connectivity were indeed experience-dependent.

3.5. Mediation analysis

So far, the results have revealed that both online (conditioning-phase VTA/SN activation and classifier evidence during CS+) and offline (pre-to-post conditioning hippocampal-cortical functional connectivity) brain measures relate to the retroactive memory effect. In the next analysis, we examined whether these brain measures were also correlated each other. Pearson's correlation coefficient analyses revealed a significant correlation between VTA/SN aversive learning-related encoding activation and pre-to-post learning changes in hippocampal-CS+ category-selective cortex functional connectivity, partial $r(14) = 0.50$, $p = .048$. By contrast, the degree of early neural context reinstatement in left PPA on CS+ trials (when reinstatement was qualitatively strongest) did not correlate with either of these measures (p's > 0.05). These findings provided initial evidence that the online increases in VTA/SN activation may interact with subsequent consolidation processes to influence memory selectivity.

To test this relationship more directly, we examined if these aversive learning-related biases in post-encoding hippocampal-cortical functional connectivity could account for the association between online VTA/SN encoding-related activation and RME behavioral scores. Consistent with this possibility, we found that, across participants, increased pre-to-post aversive learning changes in hippocampal-CS+ category-selective cortex connectivity mediated the relationship between Phase 2 VTA/SN aversive encoding effects and Phase 1 RME scores (ACME = 0.00047; $p = .042$; Fig. 6).

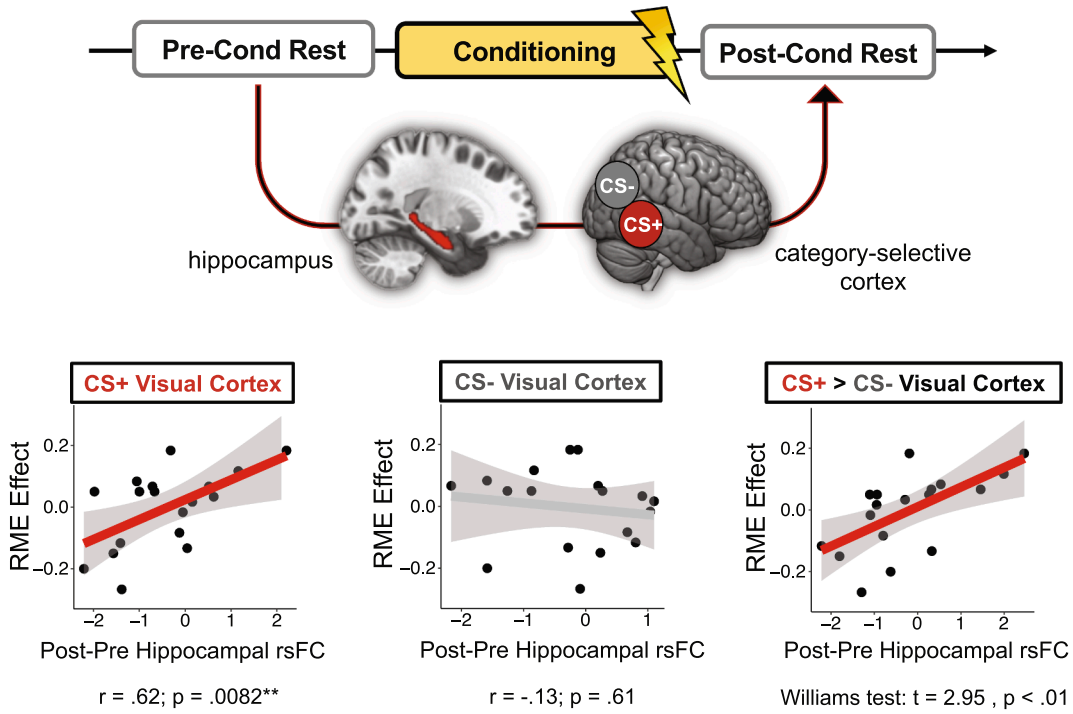

Fig. 5. Aversive learning-dependent changes in post-encoding hippocampal functional connectivity relate to across-participant variability in the retroactive memory effect (RME). Conditioning-dependent changes in hippocampal resting-state functional connectivity (rsFC) with CS+ (aversive) and CS− category-selective cortex were assessed by subtracting connectivity z-stats for pre-conditioning rest from connectivity z-stats from post-conditioning rest. Across participants, greater aversive learning-related RME effects were associated with greater hippocampal functional coupling with CS+ category-selective cortex (red line) but not CS− category-selective cortex (gray line). These hippocampal functional connectivity patterns were also significantly different from each other, indicating that, following aversive learning, greater selective retroactive memory benefits relate to a shift in hippocampal coupling towards CS+ cortex and away from CS− cortex. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

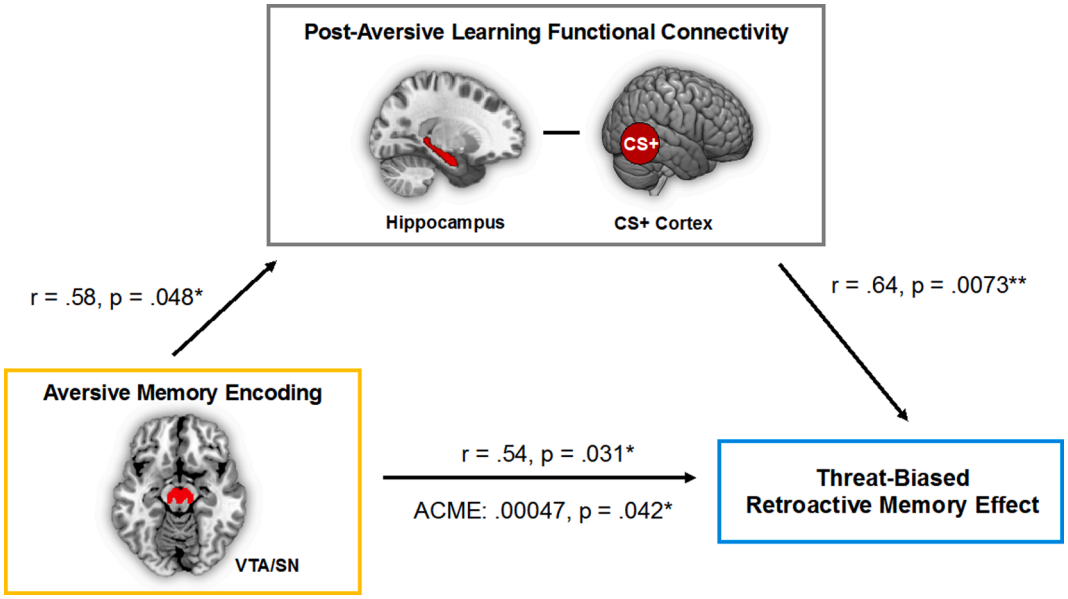

Fig. 6. Post-encoding hippocampal functional connectivity with CS+ category-selective cortex mediates the relationship between VTA/SN aversive encoding activation and the retroactive memory benefit. Path values represent partial Pearson's correlation coefficients after controlling for the effects of Shock Category across participants (shocked on animals or shocked on tools). *p < .05; **p < .01.

To determine the specificity of this mediation effect to Phase 1 memory biases, we also performed the same correlation analyses, but this time targeting aversive memory biases for Phase 2 items. This memory measure (like RME) was computed by subtracting corrected recognition scores between Phase 2 CS+ and CS− items. In contrast to the brain-behavior relationships identified for Phase 1 memory, we did not find any significant associations between Phase 2 aversive learning-related memory biases and either the VTA/SN effect, partial r(14) = 0.29, p = .28, or hippocampal-CS+ connectivity, partial r(15) = 0.28, p = .27, across participants. Moreover, the same type of mediation

analysis as before was not significant (ACME = 0.00011, p = .66). Together, these findings suggest that offline hippocampal processes may stabilize putative dopaminergic learning 'tags' for salient concepts. Consistent with models of behavioral tagging, this consolidation process may only be necessary for preserving weaker memories of conceptually-related items encountered close in time. By contrast, aversive representations may be sufficiently enhanced during initial encoding, as encoding-related VTA/SN activation was related to better memory for Phase 2 CS+ compared to Phase 2 CS− images.

4. Discussion

Decades of research has focused on understanding why and how aversive experiences are both vividly and enduringly remembered. Emerging empirical evidence suggests that aversive experiences can also selectively influence the consolidation of previously encountered neutral information that is conceptually-related to the aversive event (Dunsmoor et al., 2015; Hennings et al., 2021). Here, we find that both inter-related and independent mechanisms occurring at the time of a strong, aversive learning event and continuing into post-encoding time periods promote these retroactive memory benefits.

Building on a series of experiments in rodents, the behavioral tagging model posits that a weak memory trace can be strengthened by a stronger, arousing event that engages the same neural pathways (Ballarini et al., 2009; Moncada & Viola, 2007). Existing behavioral tagging models rely on a strong arousing experience, such as novel context exploration, to induce a general, non-specific engagement of neural pathways. Because we know this effect can be targeted to semantically-related representations and not to others, we aimed to test if the similarity between the arousing category and recently encountered items might induce the reactivation of the preceding learning context. Using a 'context tagging' and multivoxel classification procedure, we found that Phase 1 context reinstatement, as indexed by classifier evidence of scene-related processing, was significantly more likely to occur during aversive images than mundane images during the first half of fear conditioning. This is consistent with the hypothesis that reactivation of prior mental states can be induced when aversive events are most novel, salient, and arousing.

By limitations of the design, our context tag was general to the entire Phase 1 block. This limited our ability to measure which parts of Phase 1 were selectively reactivated during conditioning. However, if this reactivation is a mechanism that promotes the selective strengthening of representations related to the aversive category, it should correlate with behavioral strengthening of those specific Phase 1 items. Indeed, we find that the amount of prior context reactivation at these early aversive moments was associated with the extent of the retroactive memory benefit.

A core feature of the behavioral tagging model is that DA and NE release are necessary for triggering the production of proteins that can transform weak learning tags into more enduring memory traces (Moncada, 2017; Moncada et al., 2011; Ritchey et al., 2016; Wang et al., 2010). Earlier work demonstrating that these neuromodulators can convert early-phase long-term potentiation (LTP) processes to a more persistent form of late-phase LTP laid important groundwork for experiments targeting memory expression (Frey & Morris, 1997; Straube et al., 2003). In addition to modulating consolidation processes, much work also shows that dopaminergic and noradrenergic activity promote the encoding of motivationally-relevant information in long-term memory (Cahill et al., 1994; O'Carroll et al., 2006; Rossato et al., 2009; Shohamy & Adcock, 2010; Strange et al., 2003). These converging lines of work suggest that aversive events activate catecholaminergic systems, and, by extension, facilitate the strongest modulation of synaptic plasticity for activated synapses. Aligning with this idea, we found that greater VTA/SN activation during encoding of aversive versus neutral stimuli also relates to selective consolidation of conceptually-related stimuli encountered several minutes earlier. Thus, strong dopaminergic activity during the encoding of CS+ exemplars may have selectively triggered consolidation processes in synaptic pathways related to and associated with the aversive learning-relevant information.

Retroactive memory benefits were also associated with hippocampal processes continuing into post-encoding time periods. Specifically, we found that individuals who showed greater hippocampal functional coupling with CS+ category-selective cortex after aversive learning also showed a larger memory benefit for CS+ category items from Phase 1 of encoding. This finding adds to a growing literature implicating post-encoding hippocampal-cortical functional coupling in the preferential retention of motivationally-significant memories (de Voogd et al., 2016; Murty et al., 2016). Our data expand upon this work by showing that post-encoding hippocampal connectivity may also be strengthening more remote representations encoded in the same neural pathways. Interestingly, we also find that these increases in hippocampal coupling with CS+ category-selective cortex mediated the relationship between online dopaminergic aversive encoding processes and the retroactive memory benefit. The long-term benefits of dopaminergic activity on memory have been previously linked to cellular consolidation mechanisms, including the stabilization of hippocampal plasticity (Lisman et al., 2011; Lisman & Grace, 2005) as well as persistent memory reactivation in hippocampal neurons (McNamara et al., 2014). In the same vein, hippocampal-cortical interactions are thought to selectively facilitate the storage of recent information that received a 'salience' or 'behavioral' tag at encoding (Moncada et al., 2015; Wang et al., 2010). The current results offer evidence that dopaminergic processes may provide such a 'relevance tag' for motivationally-relevant information, which is then stabilized by hippocampal processes to promote the preferential retention of weakly encoded but related memories.

Although our findings shed new light on how aversive learning influences the selectivity of memory consolidation, there are several limitations that warrant consideration. First, we had a modest number of participants, so additional work will be necessary to replicate and validate these effects. We underscore that the data used in the brain-behavior correlations is embedded within-subject across-trial subtraction. Although the number of participants is modest, these subtractions should help account for noise-related variability that might emerge across subjects, as well in item category-related effects. Second, in contrast to earlier work (Dunsmoor et al., 2015; Hennings et al., 2021; Patil et al., 2017), we did not find a significant aversive learning-related retroactive memory benefit in behavior. One reason for the null effect in our study may have been the introduction of neutral scene stimuli during the initial encoding phase, which may have altered the encoding of Phase 1 items. Third, due to equipment malfunction, we were unable to collect skin conductance measures as an endogenous index of threat acquisition. We do not believe, however, that this detracts from our interpretations concerning aversive learning- or salience-related effects on selective consolidation. In the current design, we weren't specifically interested in evaluating the efficacy of threat conditioning but rather how imbuing existing neutral memories with motivational significance affects their long-term consolidation. Thus, we were able to verify that participants learned the motivational significance of the stimuli using their shock expectancy ratings, which are considered a valid measure of human threat conditioning with strong face- and construct-validity (Boddez et al., 2013).

Another possibility for the null memory effect is that being in the MRI increased participants' baseline arousal to different degrees, which may have overshadowed the retroactive effects of the threat conditioning manipulation on Phase 1 encoding (Muehlhan et al., 2011). Despite not having direct measures of autonomic activation, however, we found that activation of arousal-related neuromodulatory systems (i.e., DA) was significantly higher for CS+ compared with CS− items during aversive learning, suggesting that CS+ images were indeed salient and processed as motivationally relevant (e.g., Shohamy & Adcock, 2010). Disentangling whether these ostensible DA effects – and

retroactive memory effects more broadly – relate to stimulus salience, emotional reactivity, attention, or physiological arousal is an important direction for future research. We also replicate previous fear conditioning fMRI work (LaBar et al., 1998) by showing that right amygdala activation is greater during early CS+ trials compared to late CS+ trials, the time-window when neural reinstatement of the prior context was most evident. This supports the idea that arousal induced by aversive events enhances the reinstatement of recent, overlapping memory representations in ways that may facilitate their storage in long-term memory.

In many ways, we believe individual differences are particularly interesting here, as it highlights differences in individuals' thresholds for aversive events to retroactively enhance memory. There are potentially many boundary conditions that determine whether a retroactive effect may occur, and, importantly, it seems to be the case that these conditions are met in some individuals and not others. One possibility supported by our data is that activation of the VTA/SN modulates the strength of the retroactive memory effect. The VTA/SN plays an important role in motivated attention and learning, including boosting memory encoding and consolidation of salient information. In rodents, it has also been shown to be essential for producing plasticity-related proteins that strengthen memory consolidation and stabilize recent memory traces (see Moncada, 2017). Weak activation in this region due to low motivation or attention to salient and/or aversive information should thereby yield less selective retroactive memory effects. It is also important to consider that behavioral tagging is a mechanism for boosting weak learning of recent information. Indeed, in humans the aversive learning-related retroactive memory effect only emerges for weak, single-shot encoding and disappears when encoding is strengthened through repetition (see Dunsmoor et al., 2015). Ensuring the initial stage of learning is "weak", however, is challenging to manipulate in human behavioral paradigms and is likely to vary across individuals as well as across learning trials.

Importantly, the main brain-behavior correlations observed in this study suggest that these 'online' and 'offline' mechanisms specifically function as *biasing* mechanisms for motivationally-relevant information. At first blush, the correlation results seem to suggest a symmetrical relationship wherein these neural processes can also benefit memory for information that fails to acquire new relevance. However, we believe a close examination of these results points to a specific effect of emotional learning on selective memory consolidation. First, the retroactive bias towards CS+ memory was only related to context reactivation patterns during CS+ stimuli but not CS− stimuli encountered during aversive learning. Second, in a similar finding, post-encoding hippocampal-cortical connectivity and retroactive memory biases was specifically driven by functional connectivity with CS+ cortex but not CS− cortex. This suggests that retroactive memory biases favor the CS+ stimuli *when* individuals show experience-dependent changes in hippocampal connectivity with cortical regions that process the motivationally-relevant category. Third, while the VTA/SN encoding mechanism was bidirectional and might therefore reflect a more general memory consolidation process, it was also more often engaged under aversive contexts. As such, this memory-enhancing mechanism is most likely to preference the storage of aversive-related information. In summary, it appears that *when* aversive moments enhance neuromodulation, enhance post-encoding communication between hippocampal-CS+ cortex pathways, or enhance context reinstatement, a specific memory advantage manifests for memories that acquire new relevance in the future.

The current findings may also inspire future studies geared towards understanding the other conditions under which the modulatory effects of aversive learning will spread to recent memories. For instance, existing theories posit that strong, arousing events will selectively enhance learning for information that engages a common neural substrate and is encountered close in time (Joels et al., 2006; Moncada & Viola, 2007). Thus, while we have focused on the spatial (neural overlap) convergence between two learning events, the temporal proximity of these events is also important for linking them together in memory. Indeed, recent work in rodents demonstrates that an aversive event will only become associated in memory with a recent, weaker learning event if they occur in close temporal proximity to each other (i.e., less than 6 h apart; Cai et al., 2016; Rashid et al., 2016).

It will also be important to identify any boundary conditions for the retroactive influence of aversive learning on recent memories; that is, determining the extent to which similarity between the negative and neutral stimuli drives this retroactive effect. The generalization of aversive responses can be driven by the amount of conceptual or perceptual resemblance between an affective stimulus and other neutral stimuli (Dunsmoor & Murphy, 2014; Verosky & Todorov, 2010). It has also been shown that post-encoding stress only modulates memory consolidation when it is administered in the same room as encoding, suggesting that the retroactive memory effects of physiological arousal and/or aversive events may also be constrained by the learning context and not just the overlap between the target memoranda (Sazma, McCullough, et al., 2019; Sazma, Shields, et al., 2019).

In summary, our findings reveal key evidence that multiple online and offline neural processes help to adaptively prioritize the consolidation of both salient and seemingly mundane information. Acquiring a better characterization of the factors that engage these memory mechanisms is central to understanding how aversive associations may spread and persist in post-traumatic stress disorder (PTSD) and phobias. At the same time, future studies may also inform how manipulating the conceptual or contextual overlap between to-be-encoded material and a stimulating event can be leveraged to benefit new learning.

CRediT authorship contribution statement

David Clewett: Conceptualization, Methodology, Writing – original draft. **Joseph Dunsmoor:** Conceptualization, Methodology, Writing – review & editing. **Shelby L. Bachman:** Writing – review & editing. **Elizabeth A. Phelps:** Conceptualization, Methodology, Writing – review & editing. **Lila Davachi:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Darren Yi for his assistance with data collection and programming. This project was funded by federal NIH grant R01 MH074692 to L.D. and by a fellowship on federal NIH grants T32 MH019524 and F32 MH114536 to D.C. Project contributions from S.B. were funded by NSF grant DGE-1842487 and NIH T32AG000037.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nlm.2021.107572.

References

Andreano, J. M., & Cahill, L. (2006). Glucocorticoid release and memory consolidation in men and women. *Psychological Science, 17*(6), 466–470. https://doi.org/10.1111/j.1467-9280.2006.01729.x

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences, 106*(34), 14599–14604.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage, 20*(2), 1052–1063. http://www.sciencedirect.com/science/article/pii/S105381190300435X.

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 201–206. https://doi.org/10.1016/j.jbtep.2012.08.003

Braun, E. K., Wimmer, G. E., & Shohamy, D. (2018). Retroactive and graded prioritization of memory by reward. *Nature Communications, 9*(1), 4886. https://doi.org/10.1038/s41467-018-07280-0

Buchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron, 20*(5), 947–957. https://doi.org/10.1016/s0896-6273(00)80476-6

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature, 371*(6499), 702–704. https://doi.org/10.1038/371702a0

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., … Lou, J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature, 534*(7605), 115–118.

de Voogd, L. D., Fernández, G., & Hermans, E. J. (2016). Awake reactivation of emotional memory traces through hippocampal–neocortical interactions. *Neuroimage, 134*, 563–572.

Dunsmoor, J. E., Kragel, P. A., Martin, A., & LaBar, K. S. (2014). Aversive learning modulates cortical representations of object categories. *Cerebral Cortex, 24*(11), 2859–2872. https://doi.org/10.1093/cercor/bht138

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25*(9), 1816–1821. https://doi.org/10.1177/0956797614535401

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature, 520*(7547), 345–348.

Dunsmoor, J. E., Prince, S. E., Murty, V. P., Kragel, P. A., & LaBar, K. S. (2011). Neurobehavioral mechanisms of human fear generalization. *Neuroimage, 55*(4), 1878–1888. https://doi.org/10.1016/j.neuroimage.2011.01.041

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment [Article]. *Nature, 392*(6676), 598–601. https://doi.org/10.1038/33402

Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature, 385*(6616), 533–536. https://doi.org/10.1038/385533a0

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience, 33*(20), 8590–8595.

Hennings, A. C., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2021). Emotional learning retroactively enhances item memory but distorts source attribution. *Learning & Memory, 28*(6), 178–186. https://doi.org/10.1101/lm.053371.120

Hermans, E. J., Kanen, J. W., Tambini, A., Fernandez, 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. *Cerebral Cortex, 27*(5), 3028–3041. https://doi.org/10.1093/cercor/bhw145

Joels, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: How does it work? *Trends in Cognitive Sciences, 10*(4), 152–158. https://doi.org/10.1016/j.tics.2006.02.002

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). How negative emotion enhances the visual specificity of a memory. *Journal of Cognitive Neuroscience, 19*(11), 1872–1887. <Go to ISI>://000250669900012.

Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., & Eckert, M. A. (2009). In vivo mapping of the human locus coeruleus [Article]. *Neuroimage, 47*(4), 1261–1267. https://doi.org/10.1016/j.neuroimage.2009.06.012

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*(1), 54–64. <Go to ISI>://000234139600016.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron, 20*(5), 937–945. <Go to ISI>://000073864700013.

Lisman, J., Grace, A. A., & Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in neurosciences, 34*(10), 536–547.

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory [Review]. *Neuron, 46*(5), 703–713. https://doi.org/10.1016/j.neuron.2005.05.002

Liu, D. L. J., Graham, S., & Zorawski, M. (2008). Enhanced selective memory consolidation following post-learning pleasant and aversive arousal [Article]. *Neurobiology of Learning and Memory, 89*(1), 36–46. https://doi.org/10.1016/j.nlm.2007.09.001

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*, 114–133. http://pps.sagepub.com/content/6/2/114.short.

McCullough, A. M., & Yonelinas, A. P. (2013). Cold-pressor stress after learning enhances familiarity-based recognition memory in men. *Neurobiology of Learning and Memory, 106*, 11–17. https://doi.org/10.1016/j.nlm.2013.06.011

McGaugh, J. L. (2013). Making lasting memories: Remembering the significant. *Proceedings of the National Academy of Sciences, 110*(Supplement 2), 10402–10407. http://www.pnas.org.idpproxy.reading.ac.uk/content/110/Supplement_2/10402.abstract.

McNamara, C. G., Tejero-Cantero, A., Trouche, S., Campo-Urriza, N., & Dupret, D. (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neuroscience, 17*(12), 1658–1660. https://doi.org/10.1038/nn.3843

Moncada, D. (2017). Evidence of VTA and LC control of protein synthesis required for the behavioral tagging process. *Neurobiology of Learning and Memory, 138*, 226–237. https://doi.org/10.1016/j.nlm.2016.06.003

Moncada, D., Ballarini, F., Martinez, M. C., Frey, J. U., & Viola, H. (2011). Identification of transmitter systems and learning tag molecules involved in behavioral tagging during memory formation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(31), 12931–12936. https://doi.org/10.1073/pnas.1104495108

Moncada, D., Ballarini, F., & Viola, H. (2015). Behavioral tagging: A translation of the synaptic tagging and capture hypothesis. *Neural Plasticity*, 1–21. https://doi.org/10.1155/2015/650780

Moncada, D., & Viola, H. E. (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: Evidence for a behavioral tagging. *The Journal of Neuroscience, 27*(28), 7476–7481. https://doi.org/10.1523/jneurosci.1083-07.2007

Muehlhan, M., Lueken, U., Wittchen, H. U., & Kirschbaum, C. (2011). The scanner as a stressor: Evidence from subjective and neuroendocrine stress parameters in the time course of a functional magnetic resonance imaging session. *International Journal of Psychophysiology, 79*(2), 118–126. https://doi.org/10.1016/j.ijpsycho.2010.09.009

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General, 143*(1), 15–20. https://doi.org/10.1037/a0031992

Murty, V. P., Shermohammed, M., Smith, D. V., Carter, R. M., Huettel, S. A., & Adcock, R. A. (2014). Resting state networks distinguish human ventral tegmental area from substantia nigra. *Neuroimage, 100*, 580–589.

Murty, V. P., Tompary, A., Adcock, R. A., & Davachi, L. (2016). Selectivity in post-encoding connectivity with high-level visual cortex is associated with reward-motivated memory. *Journal of Neuroscience*, 4015–4032.

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory, 88*(1), 40–47. https://doi.org/10.1016/j.nlm.2007.03.005

Nielson, K. A., Radtke, R. C., & Jensen, R. A. (1996). Arousal-induced modulation of memory storage processes in humans. *Neurobiology of learning and memory, 66*(2), 133–142.

Nielson, K. A., Yee, D., & Erickson, K. I. (2005). Memory enhancement by a semantically unrelated emotional arousal source induced after learning. *Neurobiology of Learning and Memory, 84*(1), 49–56. https://doi.org/10.1016/j.nlm.2005.04.001

O'Carroll, C. M., Martin, S. J., Sandin, J., Frenguelli, B., & Morris, R. G. M. (2006). Dopaminergic modulation of the persistence of one-trial hippocampus-dependent memory. *Learning & Memory, 13*(6), 760–769. https://doi.org/10.1101/lm.321006

Patil, A., Murty, V. P., Dunsmoor, J. E., Phelps, E. A., & Davachi, L. (2017). Reward retroactively enhances memory consolidation for related items. *Learn Mem, 24*(1), 65–69. https://doi.org/10.1101/lm.042978.116

Preuss, D., & Wolf, O. T. (2009). Post-learning psychosocial stress enhances consolidation of neutral stimuli. *Neurobiology of Learning and Memory, 92*(3), 318–326. https://doi.org/10.1016/j.nlm.2009.03.009

Rashid, A. J., Yan, C., Mercaldo, V., Hsiang, H.-L.-L., Park, S., Cole, C. J., De Cristofaro, A., Yu, J., Ramakrishnan, C., & Lee, S. Y. (2016). Competition between engrams influences fear memory formation and recall. *Science, 353*(6297), 383–387.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage, 23*(2), 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035

Ritchey, M., Murty, V. P., & Dunsmoor, J. E. (2016). Adaptive memory systems for remembering the salient and the seemingly mundane. *Behav Brain Sci, 39*, Article e221. https://doi.org/10.1017/S0140525X15001922

Rossato, J. I., Bevilaqua, L. R. M., Izquierdo, I., Medina, J. H., & Cammarota, M. (2009). Dopamine controls persistence of long-term memory storage. *Science, 325*(5943), 1017–1020. https://doi.org/10.1126/science.1172545

Sazma, M. A., McCullough, A. M., Shields, G. S., & Yonelinas, A. P. (2019). Using acute stress to improve episodic memory: The critical role of contextual binding. *Neurobiology of Learning and Memory, 158*, 1–8. https://doi.org/10.1016/j.nlm.2019.01.001

Sazma, M. A., Shields, G. S., & Yonelinas, A. P. (2019). The effects of post-encoding stress and glucocorticoids on episodic memory in humans and rodents. *Brain and Cognition, 133*, 12–23. https://doi.org/10.1016/j.bandc.2018.10.005

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences, 14*(10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002

Smeets, T., Giesbrecht, T., Jelicic, M., & Merckelbach, H. (2007). Context-dependent enhancement of declarative memory performance following acute psychosocial stress. *Biological Psychology, 76*(1–2), 116–123. https://doi.org/10.1016/j.biopsycho.2007.07.001

Southwick, S. M., Davis, M., Horner, B., Cahill, L., Morgan, C. A., Gold, P. E., … Charney, D. C. (2002). Relationship of enhanced norepinephrine activity during memory consolidation to enhanced long-term memory in humans. *American Journal of Psychiatry, 159*(8), 1420–1422.

Strange, B. A., Hurlemann, R., & Dolan, R. J. (2003). An emotion-induced retrograde amnesia in humans is amygdala- and beta-adrenergic-dependent. *Proceedings of the National Academy of Sciences of the United States of America, 100*(23), 13626–13631. https://doi.org/10.1073/pnas.1635116100

Straube, T., Korz, V., Balschun, D., & Frey, J. U. (2003). Requirement of beta-adrenergic receptor activation and protein synthesis for LTP-reinforcement by novelty in rat dentate gyrus. *The Journal of Physiology, 552*(3), 953–960. https://doi.org/10.1113/jphysiol.2003.049452

Tambini, A., & Davachi, L. (2019). Awake reactivation of prior experiences consolidates memories and biases cognition. *Trends in Cognitive Sciences, 23*(10), 876–890. https://doi.org/10.1016/j.tics.2019.07.008

Tambini, A., Ketz, N., & Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron, 65*(2), 280–290.

Tompary, A., Duncan, K., & Davachi, L. (2015). Consolidation of associative and item memory is related to post-encoding functional connectivity between the ventral tegmental area and different medial temporal lobe subregions during an unrelated task. *Journal of Neuroscience, 35*(19), 7326–7331.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science, 21*(6), 779–785. https://doi.org/10.1177/0956797610371965

Wang, S. H., Redondo, R. L., & Morris, R. G. (2010). Relevance of synaptic tagging and capture to the persistence of long-term potentiation and everyday spatial memory. *Proceedings of the National Academy of Sciences of the United States of America, 107*(45), 19537–19542. https://doi.org/10.1073/pnas.1008638107

Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: An introduction to methods*. Oxford University Press.