HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2021 August 19.

Published in final edited form as:

Neuroimage. 2021 August 01; 236: 118033. doi:10.1016/j.neuroimage.2021.118033.

Reinstatement of item-specific contextual details during retrieval supports recombination-related false memories

Alexis C. Carpenter[a,*], **Preston P. Thakral**[a,b], **Alison R. Preston**[c], **Daniel L. Schacter**[a]

[a]Department of Psychology and Center for Brain Science, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, United States

[b]Department of Psychology and Neuroscience, Boston College, United States

[c]Center for Learning and Memory and Department of Psychology, University of Texas at Austin, United States

Abstract

Flexible retrieval mechanisms that allow us to infer relationships across events may also lead to memory errors or distortion when details of one event are misattributed to the related event. Here, we tested how making successful inferences alters representation of overlapping events, leading to false memories. Participants encoded overlapping associations ('AB' and 'BC'), each of which was superimposed on different indoor and outdoor scenes that were pre-exposed prior to associative learning. Participants were subsequently tested on both the directly learned pairs ('AB' and 'BC') and inferred relationships across pairs ('AC'). We predicted that when people make a correct inference, features associated with overlapping events may become integrated in memory. To test this hypothesis, participants completed a final detailed retrieval test, in which they had to recall the scene associated with initially learned 'AB' pairs (or 'BC' pairs). We found that the outcome of inference decisions impacted the degree to which neural patterns elicited during detailed 'AB' retrieval reflected reinstatement of the scene associated with the overlapping 'BC' event. After *successful* inference, neural patterns in the anterior hippocampus, posterior medial prefrontal cortex, and our content-reinstatement region (left inferior temporal gyrus) were more similar to the overlapping, yet incorrect 'BC' context relative to after *unsuccessful* inference. Further, greater hippocampal activity during inference was associated with greater reinstatement of the incorrect, overlapping context in our content-reinstatement region, which in turn tracked contextual misattributions during detailed retrieval. These results suggest recombining memories

[*]Corresponding author. alexiscarpenter@g.harvard.edu (A.C. Carpenter).

Credit author statement

Alexis Carpenter: Writing- Original draft preparation, Conceptualization, Methodology, Investigation, Data Analysis. Preston Thakral: Writing- Reviewing and Editing, Conceptualization, Methodology, Investigation, Data Analysis. Alison Preston and Daniel Schacter: Writing-Reviewing and Editing, Conceptualization, Methodology.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118033.

Data/Code availability

Data and materials are available upon direct request and are subject to anonymization to protect the privacy of participants. IRB permissions were not obtained to allow data to be uploaded to an online repository. As such, requestors must have approval from their IRB and acknowledge the source of the data in any reports using the data. Additionally, a Data Usage Agreement (DUA) is required before any data and/or materials will be made available.

during successful inference can lead to misattribution of contextual details across related events, resulting in false memories.

Keywords

Associative inference; Episodic memory; False memory; Hippocampus; Retrieval

Episodic memory supports our ability to retrieve distinct elements of past experiences (Tulving, 1983). In addition, a growing body of evidence suggests that episodic memory also allows us to recombine such elements to create novel episodes that have not been directly experienced (e.g., Moscovitch et al., 2016; Thakral et al., 2019). For example, according to the *constructive episodic simulation hypothesis* (Schacter and Addis 2007a, 2007b, 2020) flexibly retrieving and recombining elements of past experiences is critical for our ability to imagine or simulate events that may occur in the future. In addition to simulating possible future events, such constructive episodic processes (Schacter, 2012) have been shown to support inferential retrieval (Preston et al., 2004; Zeithamova et al., 2012a; Zeithamova and Preston, 2010), means-end problem solving (Jing et al., 2016; Madore and Schacter, 2014; Sheldon et al., 2011), and divergent creative thinking (Madore et al., 2015).

However, the constructive episodic simulation hypothesis also holds that the functional benefits of flexible retrieval and recombination may be accompanied by a cost: susceptibility to memory errors such as source misattribution and false recognition that can result from mistakenly combining elements of distinct past experiences (Schacter and Addis, 2007a, 2007b, 2020; for related views, see Dudai and Carruthers, 2005; Suddendorf and Corballis, 2007). That is, while such constructive processes support a range of adaptive mnemonic functions, they may also leave memory prone to error or distortion (cf., Bartlett, 1932; Brainerd and Reyna, 2005; Loftus et al., 1978; Howe, 2011; McClelland, 1995; Roediger, 1996; Schacter, 2001; Schacter et al., 2011, 2021).

Using a modified associative inference paradigm, Carpenter and Schacter (2017) directly tested the key claim of the constructive episodic simulation hypothesis that the same flexible retrieval processes that are used to combine elements of distinct episodes into functionally useful, novel representations, may also produce memory errors. Associative inference is an adaptive process that supports our ability to reactivate and recombine past episodes in order to infer a novel relationship that has not been directly experienced (e.g., Zeithamova and Preston, 2010). In previous versions of the modified associative inference paradigm, during the first session participants were asked to learn partially overlapping ‘AB’ and ‘BC’ events comprised of a unique ‘A’ or ‘C’ person and a shared ‘B’ object superimposed on an indoor or outdoor scene. Importantly, these scene contexts contained at least ten contextual details that were contradictory across the overlapping ‘AB’ and ‘BC’ events (e.g., event ‘AB’ contained a white couch and event ‘BC’ contained a brown couch – see Fig. 1 for example images). Participants were instructed to learn both the direct person-object relationships (i.e., ‘AB’ and ‘BC’), the indirect relationships between two people based on the shared object (i.e., ‘AC’) and the contextual details of the scenes associated with each event (e.g., the color of the couch). Results revealed significantly higher rates of false memories (i.e., trials

where participants chose the contextual detail from the overlapping event and misattributed its source to the currently cued event) after successful associative inference compared to both after unsuccessful inference and before successful associative inference (Carpenter and Schacter, 2017).

Carpenter and Schacter (2017) argued that flexible retrieval and recombination mechanisms active during the test of directly learned/associative inference trials may result in false memories because inferring the relationship across the 'AB' and 'BC' events requires participants to both reactivate distinct 'AB' and 'BC' episodes and further flexibly recombine the nonoverlapping 'A' and 'C' items. During such flexible retrieval, contextual details from one event may be more fully bound to the overlapping, yet incorrect source, resulting in memory errors associated with heightened cross-episode binding (cf., Bridge and Voss, 2014a, 2014b) as a result of flexible retrieval and recombination processes. Supporting the role of retrieval-based processes in false memories, past work has shown that reinstatement or reminders of past events during new learning can result in memory errors, where details from one event are misremembered as having come from an alternate event (Hupbach et al., 2008, 2007; Hupbach et al., 2009; Gershman et al., 2013).

While past behavioral results support the specific link between flexible retrieval mechanisms and both successful inference and false memories (Carpenter and Schacter, 2017; Carpenter and Schacter 2018a), nothing is known about the neural basis of this effect. Specifically, it is unknown whether the neural representation of the currently cued event becomes more similar to the overlapping, yet incorrect event context following successful associative inference as compared to unsuccessful inference. Such changes in representational similarity would be expected if during successful associative inference 'AB' and 'BC' representations are indeed reactivated and recombined to create a more integrated representation wherein contextual details are more fully, yet mistakenly bound to the overlapping, yet incorrect event. The main purpose of the present functional resonance magnetic imaging (fMRI) study is to test the novel prediction that the same flexible retrieval mechanisms that support successful inference decisions directly affect the neural representations of the original events during subsequent retrieval attempts.

Past research has shown that regions of the medial temporal lobe (MTL), namely the hippocampus (Preston et al., 2004; Zeithamova and Preston, 2010), and inferior frontal gyrus (IFG) are particularly important when individuals make successful inference judgments (i.e., 'AC' decisions; Zeithamova and Preston, 2010) relative to retrieving directly learned associations (i.e., 'AB' and 'BC'). These results suggest that the MTL and IFG play a unique role in flexibly recombining overlapping memories during inference. In particular, IFG may control the retrieval of and resolve interference between competing memory representations (Badre and Wagner, 2007; Oztekin et al., 2009). Taken together, the IFG may work in concert with the hippocampus to support successful inference when participants have not integrated overlapping 'AB' and 'BC' events during encoding (for a similar integrative encoding account, see Schlichting et al., 2015; Shohamy and Wagner, 2008; Zeithamova et al., 2012a).

Related work suggests that successful retrieval of event details is thought to involve the reinstatement of encoding-related activity in the hippocampus and other content-specific cortical regions (Slotnick and Schacter, 2006; Johnson and Rugg, 2007; Thakral et al., 2015; for reviews, see Danker and Anderson, 2010; Rugg et al., 2015; Slotnick, 2004). Further, successful memory decisions are associated with stronger item-specific reinstatement in both the hippocampus (and surrounding MTL regions) and content-specific cortical regions (Bird et al., 2015; Gordon et al., 2014; Kuhl and Chun, 2014; Lee et al., 2019; Mack and Preston, 2016; Oedekoven et al., 2017; Pacheco Estefan et al., 2019; Ritchey et al., 2013; Staresina et al., 2012; Tompary et al., 2016; Wing et al., 2015; for a review, see Xue, 2018). Finally, during retrieval, item-specific reinstatement in the hippocampus via pattern completion processes may drive subsequent reinstatement in content-specific cortical regions, supporting the successful retrieval of event details (Bosch et al., 2014; Pacheco Estefan et al., 2019; Gordon et al., 2014; Ritchey et al., 2013; Staresina et al., 2012; Tompary et al., 2016; see also Wing et al., 2015 for evidence regarding hippocampal activity during encoding driving subsequent cortical reinstatement during retrieval).

Reactivating related memories (i.e., ‘AB’) during new learning (i.e., ‘BC’) can promote memory integration via hippocampal-mPFC interactions that support successful associative inference (Zeithamova et al., 2012a). Further, after participants learn such related ‘AB’ and ‘BC’ associations, neural patterns in the anterior hippocampus and posterior mPFC show evidence of memory integration, such that patterns of neural activity for the non-overlapping ‘A’ and ‘C’ items become more similar to one another, relative to items from unrelated events (Schlichting et al., 2015). While the aforementioned studies focus on integration of related memories during and as a result of encoding partially-overlapping ‘AB’ and ‘BC’ associations, such integration/recombination mechanisms can also operate during successful inference judgments (i.e., during the inference test itself), connecting not only the non-overlapping elements that support successful inference (e.g., the man and the boy) but also surrounding contextual features (e.g., the color of the couch).

In sum, results of past work utilizing various fMRI analytic methods highlight the role of the MTL (specifically the anterior hippocampus), IFG and medial prefrontal regions in the retrieval and flexible recombination of overlapping, yet distinct memories in order to support successful associative inference (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a). Further, following the flexible retrieval and recombination of distinct episodes, the neural patterns for non-overlapping items (i.e., ‘A’ and ‘C’) become more similar to one another in the anterior hippocampus and posterior medial prefrontal cortex (Schlichting et al., 2015). Finally, greater item-specific reinstatement during retrieval tracks various aspects of participants’ memories from free-recall of event details (e.g., Oedekoven et al., 2017) to ratings of recognition memory confidence (Ritchey et al., 2013).

In the present fMRI study, we assessed how flexible retrieval/cross episode binding mechanisms that support successful associative inference affect the specific neural representations of the original event contexts. Further, we evaluated how reinstatement of the overlapping, yet incorrect event context following successful inference impacts the likelihood that participants misattribute such contextual details in memory. We did so by utilizing both univariate analyses and a representational similarity analysis (RSA) approach

with our modified associative inference paradigm. Thus, the goal of the current study is to target item-specific reinstatement of the contextual information that we hypothesize is retrieved and bound to the overlapping event during successful inferential retrieval. In line with this goal, we targeted the increased reinstatement of the overlapping, yet incorrect contextual scene information following successful inferential retrieval and further related the fidelity of such overlapping, yet incorrect contextual reinstatement to participants' subsequent false memory scores. To do so, we modified our previous associative inference paradigm to allow for the decoding of memory for specific event contexts by introducing a pre-exposure phase during the first study session.

During the scanned pre-exposure phase, prior to learning the overlapping 'AB' and 'BC' associations, participants viewed each of the 'AB' and 'BC' scene context images in isolation without the superimposed people (i.e., 'A' and 'C' items) and objects (i.e., 'B' item). Following the pre-exposure phase, participants learned the partially overlapping 'AB' and 'BC' event pairs outside of the scanner. Following a 24-hour delay, participants were scanned again while completing two sets of detail retrieval trials, testing their memory for the specific contextual event details, separated by the directly learned and associative inference tests.

Given past literature highlighting the role of the anterior hippocampus and posterior mPFC regions in both successful associative inference and memory recombination/integration, we limited our RSA results to three regions of interest (ROIs):[1] anterior hippocampus, posterior mPFC, and a 'content-reinstatement' region in the inferior temporal cortex, specifically the L. inferior temporal gyrus (L. ITG; see Fig. 5). We chose to focus on the L. ITG as our content-reinstatement region due to past work suggesting that this region may be sensitive to the reinstatement of the specific contextual details relevant to our paradigm (e.g., object information/objects in context - the color of the couch; Han et al., 2013; Kreigeskorte et al., 2008; Ranganath et al., 2004; Vaidya et al., 2002; Woloszyn and Sheinberg, 2009; for review, see Bar, 2004). Based on past work highlighting the role of the anterior hippocampus, posterior mPFC and L. IFG in flexibly retrieving and recombining previously learned relationships in order to support successful associative inference (e.g., Zeithamova and Preston, 2010), we chose to limit our univariate analyses to these three ROIs (see "Identifying ROIs for RSA.").

For each RSA ROI, during detail retrieval trials, we aimed to measure the event-specific reinstatement of contextual details that were mistakenly bound to the overlapping event context as a consequence of flexible retrieval mechanisms that support successful inference. Consistent with this goal, we correlated patterns of neural activity during detail retrieval trials with patterns of neural activity during the pre-exposure phase when participants

[1]Given the problem of multiple comparisons associated with a whole-brain searchlight approach and other related issues (see Etzel, Zacks & Braver, 2013) and that we had strong apriori hypotheses regarding the role of the anterior hippocampus, posterior mPFC and our content-reinstatement region (L. ITG) in our current task, we focused the RSA analyses to only three predetermined ROIs. Further, in order to highlight the specificity of the current results to our three predetermined ROIs, we conducted additional control analyses using posterior hippocampus and anterior mPFC ROIs. Thus, strong ROI-specific hypotheses based on past literature, in concert with anatomical control analyses showing region specificity, allow us to focus the results and discussion on specific and logical regions known to be involved in flexibly retrieving and recombining past events. Future, more exploratory, work should attempt to determine the role of other core network regions typically involved with episodic memory related tasks in successful associative inference and subsequent false memories.

passively viewed the overlapping, yet incorrect context (see Fig. 2). For example, during detail retrieval trials, participants were presented with person A_1 (e.g., man cue) in order to cue the retrieval of details related to context AB_1 (e.g., context with the white couch). In line with past research, reactivation or reinstatement of context AB_1, in response to the cue person A_1 may track participants' true memory performance (Mack and Preston 2016). Alternatively, in line with the goal of the current study, reinstatement of the incorrect context from the overlapping event (e.g., context BC_1 with the brown couch – overlapping, yet incorrect contextual reinstatement), in response to cue person A_1, may track participants' false memory performance as a result of reactivating and recombining the partially overlapping events during the directly learned/associative inference test (see Fig. 1).[2] We tested whether making inferences would promote integration of overlapping memories, leading to memory misattributions in which the spatial context from one event is retrieved when remembering a related memory.

1. Materials and methods

1.1. Participants

31 participants completed both sessions of the study in full (M_{age} = 21.10, SD = 2.21; $M_{education}$ = 14.90, SD = 2.07; 19 female). Participants were recruited via advertisements at Boston University and Harvard University. All participants were native English speakers, right-handed, had normal or corrected-to-normal vision, and no history of psychiatric or neurological disorders. Participants gave informed consent and were treated in accordance with guidelines approved by the committee on the use of human subjects at Harvard University and received pay for completing the study. Two participants were excluded from all behavioral and fMRI analyses because they did not have at least one triad in each bin (i.e., before vs. after, unsuccessful vs. successful inference), rendering the critical comparison of before vs. after successful vs. unsuccessful inference impossible. Specifically, one participant was excluded for low performance on the associative inference task resulting in insufficient successful inference triads and one participant was excluded for high performance on the associative inference task resulting in insufficient unsuccessful inference triads. Thus, 29 participants (M_{age} = 21.07, SD = 2.25; $M_{education}$ = 14.86, SD = 2.10; 19 female) with sufficient successful and unsuccessful inference triad numbers were included in all behavioral analyses. For one participant, one run of the detail retrieval trials before the directly learned/associative inference task was excluded from both behavioral and fMRI analyses due to experimenter error (run was repeated twice – thus, the repeated run was excluded from all analyses). For subsequent univariate and RSA analyses, one participant was excluded from each for having too few successful/unsuccessful inference trials (i.e., fewer than 15 trials in each bin; see Supplemental Table 1 for avg. trials included in RSA analyses) and/or directly learned/associative inference trials (i.e., fewer than 8 trials

[2]Please note that in the current study we define source misattributions and mistaken recombination/cross-episode binding as instances where participants remember or reinstate contextual details from the overlapping 'BC' event and attribute such details to their memory for the currently cued ' AB' event, for example. Thus, in context of the current detail retrieval task, we define reinstatement of the overlapping, yet incorrect context as reinstatement of 'BC' scene details in response to the 'AB' event cue. While inferring that the man lives in a house with both a brown couch and a white couch (i.e., the second order inference) may indeed be useful in other contexts, the current detail retrieval task defines such responses as false memories where contextual details of the 'BC' event were mistakenly bound to the overlapping 'AB' event.

per bin). Further, five participants were excluded from subsequent univariate and RSA analyses due to excessive movement (≥ 3 mm translation or ≥ 3° rotation within runs). Thus, the remaining 23 participants (M_{age} = 21.09, SD = 2.43; $M_{education}$ = 14.91, SD = 2.31; 15 female) were included in all fMRI analyses. All critical RSA analyses are based on within-subject correlations and a sample size of 23 participants is consistent with sample sizes of past work both using similar paradigms (Schlichting et al., 2015; Zeithamova et al., 2012a; van Kesteren et al., 2020) and analyses (see Liang and Preston, 2017; Mack and Preston, 2016; Tompary et al., 2016; Tompary and Davachi, 2017).

1.2. Summary of procedure

Participants came into the lab for two sessions separated by a 24-hour delay (the procedures here follow our prior studies, Carpenter and Schacter, 2017, 2018a, 2018b). All experimental sessions were run using PsychoPy2 (v1.80.03). During the first session, participants completed six runs of the pre-exposure phase while in the scanner followed by the 'AB' and 'BC' encoding phases outside of the scanner. During the second session, participants completed all tasks while in the scanner. Participants completed one half (i.e., six runs) of the detail retrieval trials prior to completing three runs of the directly learned and associative inference trials. After the test of directly learned and associative inference trials, participants were given a short, approximately five-minute break inside the scanner, while the experimenter conditionalized the second half of the detail retrieval trials based on each participant's performance on the directly learned trials for each triad. That is, due to time constraints, after the directly learned/associative inference test, participants were only tested on detail retrieval questions that corresponded to triads for which they got the directly learned trials correct given that only these triads could be used in subsequent analyses. After participants completed the second half (i.e., six runs) of the detail retrieval trials, they were debriefed and compensated for their participation in the study (see Fig. 1).

1.3. Pre-exposure phase

Participants completed six runs of the pre-exposure task. Stimuli consisted of 96 still color images depicting indoor and outdoor scenes (e.g., an office or a park; subtending 9.19° by 6.84° in visual angle) that would later be used as the event contexts for partially overlapping 'AB' and 'BC' pairs. Each run consisted of 64 'AB' or 'BC' event contexts without the superimposed people or objects. Each 'AB' and 'BC' event context was presented for two seconds and repeated four times across runs. Participants were instructed to view each image and attend to the details of the image during the two second viewing period. After each image, participants were given two seconds to make a task-irrelevant pleasantness rating on a scale from 1 to 4 (1 = very unpleasant, 4 = very pleasant). Pleasantness ratings were included as an attentional check during the pre-exposure phase. Pre-exposure trials for which a participant did not respond to the pleasantness rating were excluded from all analyses. The pleasantness rating period was followed by a four second fixation period.

1.4. AB and BC encoding

Following the pre-exposure phase, participants completed the 'AB' and 'BC' encoding phases outside of the scanner. Stimuli consisted of 96 still color images depicting everyday life events (e.g., man walking the dog). Color images of common objects (e.g., toy truck)

and people were superimposed on outdoor and indoor scenes. Scenes were counterbalanced across participants such that each scene was used equally often for both 'AB' and 'BC' pairs. Using Adobe Photoshop CC 2015, 96 partially overlapping pairs (48 'AB' pairs, 48 'BC' pairs – 48 total ABC triads[3]) were constructed. Overlapping 'AB' and 'BC' pairs were constructed such that two people ('A' and 'C') shared an association with an overlapping object ('B'; i.e., one ABC triad; see Fig. 1).

Participants first completed the 'AB' encoding task which consisted of the 48 'AB' images, followed by the 'BC' encoding task which consisted of the 48 'BC' images. Each image was randomly presented for 10 s within their respective encoding block (i.e., 'AB' encoding and 'BC' encoding). Participants were instructed to learn both the direct associations (i.e., 'AB' and 'BC') and the indirect associations (i.e., 'AC') along with the contextual scene information presented. Following each image, participants were asked to provide a judgment of learning (JOL) on a scale from 1 to 4 (1 = definitely forget, 4 = definitely remember). JOLs were collected in order to ensure participants' attention during the encoding phase.

1.5. Directly learned and associative inference trials

While in the scanner, participants completed the first half of the detail retrieval trials, and following that, were tested on directly learned ('AB' and 'BC') and associative inference ('AC') trials. During each directly learned trial, a single person (e.g., an 'A' or 'C' person) was presented at the top of the screen and two choice objects were presented at the bottom of the screen (e.g., two 'B' objects from different ABC triads). On the associative inference trials, a cue person ('A') was presented along with two choice people at the bottom of the screen (i.e., the correct 'C' person from the ABC triad and a lure 'C' person from another triad). Participants were instructed on associative inference trials that the association between the cue ('A') and the correct choice ('C') was indirect, mediated through an object ('B') that shared an association with both the cue and the correct choice during encoding. Participants were given four seconds to choose the item that they remembered was in some way related to the cue person (i.e., either directly or indirectly) or respond 'neither' if they remembered that the cue person had not been directly or indirectly related to either of the answer choices. Trials where participants did not respond within the four second response period were excluded from all analyses (2% of trials). Participants completed three runs each consisting of 32 directly learned trials and 16 associative inference trials. The presentation order of the trials was pseudorandomized within runs with the constraint that 'AC' associative inference trials were shown before their corresponding 'AB' and 'BC' directly learned trials in order to ensure that participants were not able to form an association between 'A' and 'C' people during test based on the co-occurrence of answer choices. Each directly learned and associative inference trial was followed by a variable fixation period with an average of four seconds (see Fig. 1).

[3]As a part of a previous unpublished study, eight participants were asked to rate the distinctiveness of each 'AB' and 'BC' image from a larger set of 60 ABC triads previously created for another study on a scale from 1 to 9 (1 = not at all distinctive, 9 = extremely distinctive). The current set of 48 ABC triads were chosen from the larger set of 60 ABC triads based on the distinctiveness ratings of this previous group of participants. That is, we chose the 48 most distinctive ABC triads ($M_{distinctive}$ = 4.18, SE = 0.56) for the current study from a set of 60 ABC triads that had been constructed for a previous study ($M_{indistinctive}$ = 3.01, SE = 0.30; $t(7) = 3.15$, $p = .016$, mean difference = 1.17, 95% CI [0.29, 2.05], $d = 1.12$).

1.6. Detail retrieval

Ten detail retrieval questions were constructed for each of the 48 ABC triads (five questions related to image ‘AB’ and five questions related to image ‘BC’). Detail questions were directly related to contextual details that were present but contradictory in the ‘AB’ and ‘BC’ scenes and did not reference the overlapping ‘B’ object. A cutout of the cue person (i.e., either ‘A’ or ‘C’) was presented to the right of each detail question in order to indicate which scene context the question was referring to. Each detail retrieval trial consisted of a six second ‘remember’ period followed by a four second ‘response’ period. During the ‘remember’ period, participants viewed the question prompt and the cue person and were asked to recall the currently cued event scene context in as much detail as possible. Following the six second ‘remember’ period, participants were given four response options: the correct item, a misinformation item, an unrelated foil item and a 100% unsure option. The misinformation item consisted of information from the overlapping image in the triad (e.g., if the detail question were related to the ‘AB’ image, the misinformation item would be a contradicting detail from the ‘BC’ image, such as a brown couch when a white couch had appeared in the ‘AB’ image). Foil items were details that were not presented in either of the overlapping images (e.g., gray couch). Each detail retrieval trial was followed by a four second fixation period.

Participants completed the ten detail retrieval trials for one half of the 48 ABC triads split into six runs before being tested on the directly learned and associative inference trials. Each run consisted of 40 detail retrieval trials corresponding to either a previously learned ‘AB’ or ‘BC’ image. As noted in the Summary of Procedure section, trials for each run of the alternate half of detail retrieval trials tested after the directly learned/associative inference test were conditionalized based on participants’ performance on the directly learned and associative inference task ($M_{trials\ per\ run}$ = 21.30, SE = 0.44; see Fig. 1).

1.7. Coding of triad and memory type

Consistent with previous work using the modified associative inference paradigm and false memory tasks (Carpenter and Schacter, 2017, 2018a, 2018b), successful inference triads were defined as triads for which participants were correct in their responses on both the directly learned and associative inference trials. That is, they were able to successfully recognize the ‘AB’ and ‘BC’ pairs and were further able to retrieve and recombine these events in order to infer the indirect ‘AC’ relationship. Alternatively, unsuccessful inference triads were defined as triads where participants were correct in their response on the directly learned trials but were incorrect in their response on the associative inference trial (i.e., chose the incorrect option or ‘neither’). That is, they were able to successfully recognize the ‘AB’ and ‘BC’ pairs but were not able to retrieve and recombine these events in order to infer the indirect ‘AC’ relationship.

Within successful and unsuccessful inference triad bins both before and after the directly learned/associative inference test, false memories were defined as detail questions for which participants chose the misinformation detail from the overlapping event and attributed this detail to their memory of the currently cued event (e.g., brown couch; see Fig. 2). True memories were defined as detail questions for which participants chose the correct detail

from the currently cued event and attributed this detail to their memory for the currently cued event (e.g., white couch). Foil memories were defined as detail questions for which participants chose the foil detail (i.e., a detail that was not present in either the currently cued or the overlapping event) and attributed this detail to their memory for the currently cued event (e.g., gray couch). Unsure memories were defined as detail questions for which participants chose the '100% Unsure' response option, indicating that they were 100% unsure in their memory for the context associated with the currently cued event. See Supplemental Figure 1 for overall rates of true, false, foil and unsure memory responses.

1.8. fMRI acquisition and preprocessing

Functional and anatomic images were acquired at the Harvard Center for Brain Science using a 3-Tesla Siemens Prisma scanner with a 32-channel head coil. Anatomic images were acquired with a magnetization-prepared rapid gradient echo sequence (matrix size of 256 × 256, 1 mm^3 resolution, 176 slices). Functional images were acquired with a multiband echo-planar imaging sequence (TR = 2 s, TE = 30 milliseconds, matrix size of 136 × 136, 84 slices - 3 slices acquired simultaneously, 1.5 mm^3 resolution, multiband factor of 3). Slices were auto-aligned to an angle 20° toward coronal from anterior- posterior commissure alignment.

fMRI data were analyzed using Statistical Parametric Mapping (SPM12, Wellcome Department of Cognitive Neurology, London, UK). Functional image preprocessing for each task (i.e., pre-exposure, detail-before, directly learned/associative inference, detail-after) consisted of slice-time correction (using the first slice as the reference), spatial realignment, and normalization into Montreal Neurological Institute space using the TPM template supplied by SPM12 (no resampling). Following normalization, functional images were smoothed with a 3 mm full-width-half-maximum (FWHM) Gaussian smoothing kernel. Anatomical images were also normalized.

1.9. Univariate analysis of fMRI data

Univariate analysis during the directly learned/associative inference test was conducted using a two-stage mixed effects general linear model (GLM). In the first stage, there were four trial types of interest: correct inference (M_{trials} = 29.78, SE = 1.01), incorrect inference (M_{trials} = 17, SE = 0.97), correct directly learned (M_{trials} = 70.87, SE = 1.90), incorrect directly learned (M_{trials} = 23.57, SE = 1.91). There was one additional trial type of no interest which comprised excluded trials and trials without a response (2% of all trials; M_{trials} = 2.78, SE = 0.53). A four second boxcar function that onset concurrently with the directly learned or associative inference trial was used to model neural activity. The associated BOLD response was modeled by convolving the boxcar functions with a canonical hemodynamic response function to yield regressors in a GLM. Six movement-related regressors (three for rotation and three for rigid-body translation) and regressors modeling each scan run were also entered into the design matrix.

In the second stage, the participant-specific parameter estimates for the four events of interest were entered into a one-way repeated measures ANOVA with participants modeled as a random-effect. An individual voxel threshold of $p < .005$ was employed and corrected

for multiple comparisons to $p < .05$ with a cluster extent threshold of 21 voxels (for full details on this method of correction, see Slotnick, 2017; Slotnick et al., 2003; for recent studies employing this method of correction, see Bowen and Kensinger, 2017; Ford and Kensinger, 2017; Kark and Kensinger, 2019; Thakral et al., 2020). The cluster extent threshold was computed using a Monte Carlo simulation with 10,000 iterations with an estimated spatial autocorrelation of 4.40 mm (i.e., the FWHM of the image corresponding to the standard error of the model). This method of correction provides an appropriate balance of Type I and Type II errors, while maintaining an acceptable false-positive rate (Slotnick, 2017). We conducted a whole-brain univariate analysis by contrasting trials associated with correct inference > incorrect inference to identify regions associated with successful associative inference (see also, Zeithamova and Preston, 2010). Given past work and our hypotheses highlighting the complementary roles of the hippocampus, posterior mPFC and IFG during successful associative inference (see Introduction), we also conducted ROI analyses within these three regions. Specifically, the contrast of correct inference > incorrect inference was used to identify activity in each of the aforementioned ROIs (i.e., those regions associated with successful associative inference). Activity within each ROI was then extracted and interrogated to identify which regions were more involved with successful associative inference compared to successful directly learned retrieval or whether these regions support both successful associative inference and directly learned retrieval to a similar extent. Parameter estimates from these ROIs were extracted using MarsBaR (v0.44 http://marsbar.sourceforge.net/index.html; Brett et al., 2002) and subjected to a 2 (trial type: directly learned vs. associative inference) x 2 (accuracy: correct vs. incorrect) repeated measures ANOVA (note that this ANOVA is independent of the procedure used to identify the neural activity). To ensure selectivity of the hippocampal ROI, the correct inference > incorrect inference contrast was inclusively with an anatomical bilateral anterior hippocampus mask generated using the Wake Forest University PickAtlas tool (WFU PickAtlas v3.0.5; http://fmri.wfubmc.edu/software/pickatlas; Maldjian et al., 2003).

1.10. Identifying ROIs for RSA

RSA was conducted within three ROIs identified by the above univariate analysis: 1) bilateral anterior hippocampus, 2) left inferior temporal gyrus (L. ITG), and 3) posterior segment of the mPFC (i.e., the subcallosal gyrus). We chose to focus on the L. ITG as our content-reinstatement region for RSA given past work suggesting that this region is involved in the reinstatement of contextual information relevant to the current paradigm (i.e., objects/objects in context – the color of the couch; Han et al., 2013; Kreigeskorte et al., 2008; Ranganath et al., 2004; Vaidya et al., 2002; Woloszyn and Sheinberg, 2009; for review, see Bar, 2004).

We did not have a clear hypothesis as to the specific role of the L. IFG during the retrieval of contextual details *after* successful associative inference and therefore excluded this region as an ROI for the RSA analyses. Moreover, prior findings do not speak to a specific role of the L. IFG in either event separation *or* integration effects after successful inference (Schlichting et al., 2015). We note that we did hypothesize that the L. IFG may be involved with the controlled retrieval of, and resolving interference between, competing memory representations (see above; Badre and Wagner, 2007; Oztekin et al., 2009) *during*

the directly learned/associative inference test, and therefore included L. IFG in only the univariate analysis.

Due to the insufficient number of voxels in our univariate-defined functional ROIs for RSA analyses (i.e., < 103 voxels in each of three ROIs; Misaki et al., 2010), we defined the three ROIs anatomically.[4] Note that a similar pattern of results were observed using functional ROIs but were not significant. The bilateral hippocampus and L. ITG were defined as the L. and R. hippocampus and L. ITG labels, respectively, of the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) as implemented in the WFU PickAtlas Tool (Maldjian et al., 2003). The subcallosal gyrus was defined using the Talairach Daemon Labels (Lancaster et al., 1997, 2000) because the AAL atlas does not define this region. We chose the subcallosal gyrus as our posterior mPFC ROI given that this region overlapped with both our univariate results and past work demonstrating this region's role in event integration following successful associative inference (Schlichting et al., 2015). Given that previous studies have hypothesized that there are functional and representational differences along the hippocampal long axis (Collin et al., 2015; Frank et al., 2019) and our hypothesis implicating the anterior portion of the hippocampus in flexible recombination and cross-episode binding mechanisms (see Introduction), we segmented both the left and right hippocampus into three parts of approximately equal length (anterior: $y = -4$ to -18, middle: $y = -19$ to -29, posterior: $y = -30$ to -40; Collin et al., 2015). We did not have any hypotheses with respect to hemispheric differences in the hippocampus and therefore combined the most anterior third of the left and right hippocampus into a single bilateral anterior hippocampal ROI. The number of voxels within each ROI was 2028 voxels in the anterior hippocampus, 6732 voxels in the L. ITG, and 1135 voxels within the subcallosal gyrus. Fig. 5a illustrates each of the ROIs.

1.11. RSA of fMRI data

Analyses were conducted using the Princeton MVPA Toolbox (https://code.google.com/p/princeton-mvpa-toolbox/) and custom MATLAB scripts. Functional data from each ROI were preprocessed prior to RSA (for similar preprocessing steps, see Kuhl and Chun, 2014; Thakral et al., 2019). First, functional image preprocessing was conducted as described above with the exception of spatial smoothing. Second, the data from each ROI were de-trended to remove linear and quadratic trends, and z-scored across voxels within each scanning session. Third, estimates of the voxel-wise BOLD response for each pre-exposure and detail retrieval trial were obtained by averaging the z-transformed BOLD signal between TRs 2–3 (i.e., the expected peak of the hemodynamic response) following the onset of each pre-exposure image and detail retrieval cue, respectively. The single trial estimates for each of the two sets of the detail retrieval trials (i.e., detail-before and detail-after) were concatenated with the corresponding pre-exposure trials (i.e., pre-exposure trials from triads that were tested during detail-before vs. detail-after sessions respectively), such that all relevant trials from the two tasks (i.e., detail retrieval and pre-exposure) were included in the detail-before and detail-after sessions. Single trial estimates for voxels in each ROI for each

[4]Although we opted to utilize anatomical ROIs, an alternative approach would be to loosen the individual voxel threshold and inflate the original ROIs. However, we chose to take an anatomical ROI approach as the voxel size is predetermined resulting in less experimenter degrees of freedom (i.e., the choice of threshold and voxel size).

set of detail sessions (i.e., detail-before and detail-after sessions) were then z-scored across both trials and voxels. The resulting z-transformed values were used in the RSA.

We used RSA to assess the similarity between patterns of neural activity during detail retrieval trials after successful associative inference and those when participants viewed the overlapping, yet incorrect context image during the pre-exposure phase. For each participant and each detail retrieval trial, we correlated activity patterns associated with the detailed retrieval of the currently cued event (e.g., event AB_1) with the average pattern associated with viewing the overlapping, yet incorrect event (e.g., event BC_1) during the pre-exposure phase (i.e., r_{match}; Note: patterns for each unique context were averaged across all presentations of said context in the pre-exposure phase). For example, when cued with the man in the blue shirt, our goal was to quantify the degree to which participants reinstated the overlapping, yet incorrect event context depicting the living room with the brown couch (i.e., r_{match}, see Fig. 1). We contrasted these r_{match} correlations with $r_{mismatch}$ correlations between the activity patterns associated with the detailed retrieval of the currently cued event (e.g., event AB_1) and average patterns associated with viewing all other unrelated context images (e.g., event BC_4) that were from triads in the same bin (i.e., before vs. after directly learned/associative inference test, successful vs. unsuccessful inference triads). For example, when cued with the man in the blue shirt, $r_{mismatch}$ correlations reflect the degree to which participants reinstated all other unrelated event contexts from the same bin (e.g., the bowling alley context, see Fig. 1).

For each participant, each ROI and each bin, we calculated a pattern similarity score (r_{match} - $r_{mismatch}$), which represents the item-specific reinstatement of overlapping, yet incorrect contextual details during retrieval (see Fig. 2b; for similar logic see Schlichting et al., 2015). Correlations were Fisher z-transformed before statistical analyses were conducted.

To determine how flexible retrieval/cross-episode binding mechanisms supporting successful associative inference and subsequent false memories affects the neural representations of the retrieved events, pattern similarity scores for each participant were then subjected to three (one for each ROI: anterior hippocampus, L. ITG, posterior mPFC) 2 (time: before vs. after) x 2 (inference: successful vs. unsuccessful) repeated measures ANOVA. Increased pattern similarity during detail retrieval after successful inference compared to after unsuccessful inference would be expected if flexible recombination during the directly learned/associative inference task, which supports successful associative inference, also led participants to mistakenly transfer and bind contextual details across event boundaries (e.g., details from event ‘AB’ mistakenly bound to event ‘BC’).[5] That is, pattern similarity scores (i.e., r_{match} – $r_{mismatch}$) reflect the reinstatement of the overlapping, yet incorrect contextual details independent of any pattern similarity that may be attributable to: 1) general successful

[5]While past work has shown that reinstating details does indeed support successful memory decisions (e.g., Mack & Preston, 2016), we did not expect differences in reinstatement results quantifying retrieval of the correct context after successful relative to unsuccessful inference. Thus, we would not predict that reinstatement of the context directly related to the currently cued event would differ as a consequence of inference. Statistically, quantifying reinstatement of contextual details from the currently cued event would require comparing reinstatement scores to zero for each condition, which is not easily interpretable because baseline similarity can be driven by various factors (e.g., vascularity; Haynes, 2015; Bhandari, Gagne & Badre, 2018; see Footnote 9). Further, if we were to include correlations reflecting true memory reinstatement (e.g., correlations between BC retrieval and BC pre-exposure), such results would not impact the interpretation of our key results highlighting the reinstatement of the overlapping, yet incorrect event (e.g., brown couch) because the key false memory finding is comparing reinstatement across conditions.

inference and/or repeated-retrieval related pattern similarity because 'mismatch' correlations are only performed between triads from the same bin as the currently cued event and thus, act as a proxy for both general inference and repeated-retrieval related pattern similarity; 2) encoding of the overlapping 'AB' and 'BC' relationships and inferring the 'AC' relationship because pre-exposure trials occurred prior to 'AB' and 'BC' encoding and did not include any of the superimposed people or objects critical for learning such relationships; 3) perceptually-driven similarity between the pre-exposure phase and the detail retrieval trials because patterns were correlated during the 'remember' period of the detail retrieval trials where only the question and the cue person were presented on the screen, neither of which were present during the pre-exposure phase. Thus, pattern similarity results reported in the current study represent the *item-specific* reinstatement of overlapping, yet incorrect *contextual details* during the detail retrieval portions of the task.

2. Behavioral results

2.1. Directly learned and associative inference trials

First, we evaluated participants' accuracy on directly learned and associative inference trials. Participants responded correctly on 75% of directly learned trials (SE = 0.02) and 63% of associative inference trials (SE = 0.02). Consistent with prior work using similar associative inference paradigms (Carpenter and Schacter, 2017; Carpenter and Schacter, 2018a; Carpenter and Schacter, 2018b; Zeithamova and Preston, 2010), we found significantly longer reaction times (RTs) on associative inference ($M_{inference}$ = 2429 msec, SE = 51) compared to directly learned trials (M_{direct} = 2099 s, SE = 41; $t(28) = 9.76$, $p < .001$, mean difference = 329, 95% CI [260, 399], $d = 1.81$), suggesting that there is an additional recombination-related retrieval mechanism required for successful inference under single-trial learning conditions.[6]

2.2. False memory

In order to determine how flexible recombination during retrieval supports both successful inference and subsequent false memories, false memory scores were subjected to a 2 (time: before vs. after) x 2 (inference: successful vs. unsuccessful) repeated measures ANOVA.

[6]In order to get adequate trial numbers for fMRI analyses, we doubled the number of triads participants were asked to learn relative to previous studies (24 vs. 48 triads; Carpenter & Schacter, 2017), thus increasing the difficulty of the associative inference task. Consequently, reaction time results from the directly learned/associative inference task showed higher levels of non-compliance/guessing on some successful associative inference trials reflected by a higher proportion of successful inference triads showing a negative reaction time (RT) difference. That is, for a subset of successful inference triads, participants responded significantly faster on the associative inference trial than on the corresponding directly learned trials. Under the current experimental conditions (i.e., single-trial learning with limited encoding time), it is highly unlikely that the indirect inference relationships would be easier to retrieve/more readily available than those relationships that participants directly learned. If participants were performing the associative inference task as instructed, one would expect RT differences to be zero (i.e., if the overlapping 'AB' and 'BC' representations were integrated during encoding) or positive (i.e., if the overlapping 'AB' and 'BC' representations were recalled and recombined during test). Thus, such 'successful inference' triads with largely negative RT differences likely reflect guessing/non-compliance on the more effortful recall-based inference trial compared to the more recognition-based directly learned trials. As a result of such non-compliance/guessing in the current study, we excluded any 'successful inference' triads where the difference in RTs on correct inference and correct directly learned trials for the triad was more than two standard deviations below the mean. This exclusionary criterion was performed for triads that were tested both before and those triads tested after the directly learned and associative inference test and resulted in 10 outlier triads across all 29 participants being excluded from all analyses with no single participant losing more than 3 triads total (2% of total successful inference triads). Thus, all reported behavioral and fMRI results in the current study reflect only triads where participants indeed took the time necessary to either retrieve the previously integrated ABC representation or retrieve and flexibly recombine the previously learned 'AB' and 'BC' relationships in order to infer the indirect 'AC' relationship during test rather than including outlier triads with RT patterns that likely suggest guessing/non-compliance.

Results revealed a time by inference interaction, $F(1,28) = 5.61$, $p = .025$, $\eta_p^2 = 0.17$, no main effect of time, $F(1,28) < 1$, $p > .250$, $\eta_p^2 = 0.007$, and no main effect of inference, $F(1,28) \langle 1, p \rangle .250$, $\eta_p^2 = 0.02$ (see Fig. 3). Subsequent paired t-tests revealed that the interaction was largely driven by higher false memory scores after successful inference ($M_{successful} = 0.34$, SE = 0.01) compared to after unsuccessful inference ($M_{unsuccessful} = 0.31$, SE = 0.01; $t(28) = 2.12$, $p = .043$, mean difference = 0.03, 95% CI [0.001, 0.06], $d = 0.39$). There was a trend toward a significant difference between false memory scores after successful inference compared to before successful inference, such that participants showed marginally higher false memory scores after ($M_{after} = 0.34$, SE = 0.01) compared to before successful inference ($M_{before} = 0.32$, SE = 0.01; $t(28) = 1.85$, $p = .076$, mean difference = 0.02, 95% CI [−0.002, 0.04], $d = 0.34$). Critically, there were no significant differences in false memory scores either before successful inference ($M_{successful} = 0.32$, SE = 0.01) compared to before unsuccessful inference ($M_{unsuccessful} = 0.33$, SE = 0.02; $t(28) \langle 1 \rangle$, p) .250, mean difference = 0.01, 95% CI [−0.02, 0.05], $d = 0.11$) or before ($M_{before} = 0.33$, SE = 0.02) compared to after unsuccessful inference ($M_{after} = 0.31$, SE = 0.01; $t(28) = 1.48$, $p = .15$, mean difference = 0.03, 95% CI [−0.01, 0.06], $d = 0.27$; see Fig. 3 for behavioral results).[7]

2.3. True memory

True memory scores were subjected to an ANOVA identical to that reported for false memory scores. Results revealed a significant main effect of time, $F(1,28) = 8.62$, $p = .007$, $\eta_p^2 = 0.24$, no significant main effect of inference, $F(1,28) \langle 1, p \rangle .250$, $\eta_p^2 = 0.001$, and a significant time by target interaction, $F(1,28) = 4.80$, $p = .037$, $\eta_p^2 = 0.15$. Subsequent paired t-tests revealed that the interaction was driven by a significant difference in true memory scores after ($M_{after} = 0.40$, SE = 0.01) compared to before unsuccessful inference ($M_{before} = 0.33$, SE = 0.02; $t(28) = 3.29$, $p = .003$, mean difference = 0.07, 95% CI [0.03, 0.11], $d = 0.61$). There was a trend toward a significant difference between true memory scores before successful inference ($M_{successful} = 0.36$, SE = 0.01) compared to before unsuccessful inference ($M_{unsuccessful} = 0.33$, SE = 0.02; $t(28) = 1.83$, $p = .078$, mean difference = 0.03, 95% CI [−0.003, 0.06], $d = 0.34$). Critically, there was no significant difference in true memory scores after successful inference ($M_{successful} = 0.37$, SE = 0.01) compared to after unsuccessful inference ($M_{unsuccessful} = 0.40$, SE = 0.01; $t(28) = 1.27$, $p = .21$, mean difference = 0.02, 95% CI [−0.01, 0.06], $d = 0.24$) or after ($M_{after} = 0.37$, SE = 0.01) compared to before successful inference ($M_{before} = 0.36$, SE = 0.01; $t(28) = 1.05$, $p > .250$, mean difference = 0.02, 95% CI [−0.02, 0.05], $d = 0.19$).

2.4. Foil memory

Foil memory scores were subjected to an ANOVA identical to that reported for false and true memory scores. Results revealed a trend toward a significant main effect of time, $F(1,28) = 4.01$, $p = .055$, $\eta_p^2 = 0.13$, no significant main effect of inference, $F(1,28) \langle 1, p \rangle .250$, $\eta_p^2 < 0.001$, and no significant time by inference interaction for foil memory scores, $F(1,28) < 1$, $p > .250$, $\eta_p^2 = 0.004$. Importantly, foil memory scores were similar both before (M_{before}

[7]See Supplemental Figure 2 for a reaction time-based approach relating recombination-related RT differences on the directly learned/associative inference test to participants' false memory scores on the detail retrieval task.

= 0.24, SE = 0.01) compared to after successful inference (M_{after} = 0.22, SE = 0.02; $t(28)$ = 1.37, p = .18, mean difference = 0.02, 95% CI [−0.01, 0.05], d = 0.25) and before (M_{before} = 0.24, SE = 0.01) compared to after unsuccessful inference (M_{after} = 0.22, SE = 0.01; $t(28)$ = 1.72, p = .097, mean difference = 0.03, 95% CI [−0.005, 0.06], d = 0.32). Further, there were no significant differences in foil memory scores after successful ($M_{successful}$ = 0.22, SE = 0.02) compared to unsuccessful inference ($M_{unsuccessful}$ = 0.22, SE = 0.01; $t(28) \langle 1, p \rangle$ .250, mean difference = 0.003, 95% CI [−0.03, 0.04], d = 0.03) or before successful ($M_{successful}$ = 0.24, SE = 0.01) compared to unsuccessful inference ($M_{unsuccessful}$ = 0.24, SE = 0.01; $t(28) \langle 1, p \rangle$ .250, mean difference = 0.004, 95% CI [−0.03, 0.04], d = 0.05).

2.5. Unsure memory

Unsure memory scores were subjected to an ANOVA identical to that reported for false, true, and foil memory scores. Results revealed no significant main effects of time, $F(1,28)$ = 1.30, $p > .250$, η_p^2 = 0.044, or inference, $F(1,28)$ = 3.06, p = .091, η_p^2 = 0.098, and no significant time by inference interaction for unsure memory scores, $F(1,28) \langle 1, p \rangle$ .250, $\eta_p^2 < 0.001$. Thus, unsure memory scores were similar both before (M_{before} = 0.08, SE = 0.02) and after (M_{after} = 0.07, SE = 0.01) successful inference and before (M_{before} = 0.09, SE = 0.02) and after (M_{after} = 0.08, SE = 0.02) unsuccessful inference (see Supplemental Figure 1 for overall rates of true, false, foil and unsure memory).

3. fMRI results

3.1. Univariate activity in anterior hippocampus and prefrontal regions supports successful associative inference

Successful associative inference related activity identified with the correct inference > incorrect inference contrast was observed in numerous brain regions including the anterior hippocampus, posterior mPFC and left IFG (see Fig. 4; see also Supplemental Table 2 for a full list of regions). Importantly, these same three regions have been repeatedly identified by past work using similar associative inference paradigms (Preston et al., 2004; Zeithamova and Preston, 2010; Schlichting et al., 2015) and are the focus for the ROI analyses reported here.

In order to determine whether activity in these regions supports successful associative inference beyond what is necessary for the successful retrieval of directly learned associations, we extracted activity from the three ROIs noted above (i.e., hippocampus, L. IFG, posterior mPFC) and subjected these parameter estimates to three 2 (trial type: directly learned vs. associative inference) x 2 (accuracy: correct vs. incorrect) repeated measures ANOVAs. Within our hippocampal ROI, results of the correct inference > incorrect inference contrast revealed two clusters in the left hippocampus (x = −16, y = −10, z = −18, spatial extent from y = −8 to −14, 30 voxels and x = −27, y = −7, z = −24, spatial extent from −5 to −10, 24 voxels) and one cluster within the right hippocampus (x = 36, y = −8, z = −16, spatial extent from y = −8 to −20, 29 voxels). The contrast of correct inference > incorrect inference also revealed two clusters within the L. IFG (x = −45, y = 30, z = −7, 26 voxels and x = −26, y = 34, z = −7, 22 voxels) and three clusters within the posterior mPFC (x = 4, y = 11, z = −16, 58 voxels and x = −3, y = 6, z = −14, 21 voxels

and $x = -6$, $y = 16$, $z = -22$, 23 voxels). We failed to find any evidence for differences in the results amongst the clusters within each ROI (e.g., amongst the three clusters within the hippocampus; trial type by accuracy Fs $\langle$ 1, ps $\rangle$ 0.250), thus clusters within each ROI were combined to form three single ROIs (i.e., bilateral anterior hippocampus, L. IFG, and posterior mPFC; see Fig. 4a).

A significant trial type by accuracy interaction was found in the hippocampus, $F(1,22) = 22.32$, $p < .001$, $\eta_p^2 = 0.50$, the left IFG, $F(1,22) = 12.49$, $p = .002$, $\eta_p^2 = 0.36$, and the posterior mPFC, $F(1,22) = 16.95$, $p < .001$, $\eta_p^2 = 0.44$. In order to determine if there were any differences across regions in the pattern of results for the hippocampus, left IFG and posterior mPFC, beta values from these regions were subjected to a 3 (region: hippocampus vs. left IFG vs. posterior mPFC) x 2 (trial type: directly learned vs. associative inference) x 2 (accuracy: correct vs. incorrect) repeated measures ANOVA. Results revealed a significant trial type by accuracy interaction, $F(1,22) = 38.88$, $p < .001$, $\eta_p^2 = 0.64$, but critically, no significant region by trial type by accuracy interaction, $F(2,44) < 1$, $p > .250$, $\eta_p^2 = 0.04$ (see Supplemental Tables 2 and 3 for full tables of correct > incorrect inference and correct > incorrect directly learned contrasts). Across regions, results revealed greater activation during correct inference compared to correct directly learned trials, $t(22) = 4.21$, $p < .001$, mean difference = 0.11, 95% CI [0.06, 0.17], $d = 0.88$ (see Fig. 4b).[8] Note that the interaction for our L. IFG ROI was largely driven by no significant difference for correct directly learned compared to incorrect directly learned, $t(22) = 1.34$, $p = .20$, mean difference = 0.08, 95% CI [−0.04, 0.21], $d = 0.28$, and a decrease in activity for incorrect inference compared to incorrect directly learned, $t(22) = -3.37$, $p = .003$, mean difference = −0.23, 95% CI [−0.36, −0.09], $d = 0.71$, rather than an increase in activity for correct inference compared to correct directly learned despite the significant difference reported across regions.

While the current study narrowly focuses only on three ROIs reliably identified in past work using similar associative inference paradigms, future more exploratory work should attempt to identify how other corenetwork regions may be involved in successful associative inference. That is, we do not argue that the anterior hippocampus, posterior mPFC and L. IFG are the *only* regions important for successful associative inference. Rather, we argue that the current study using different instructions, stimuli and study-test delays is able to identify the same regions highlighted in past work as being important for the flexible retrieval and recombination of past information in support of successful associative inference.

[8]We chose to define our univariate clusters using the correct > incorrect inference contrast because we wanted to identify regions that are important for successful associative inference. Importantly, identifying our clusters using this contrast does not introduce circularity into the trial type by accuracy ANOVAs because these regions may also be involved in retrieving the directly learned ‘AB’ and ‘BC’ associations resulting in a main effect of trial type with no significant trial type by accuracy interaction. Thus, the purpose of the ANOVAs was to determine whether these regions were similarly involved in both successful associative inference and the retrieval of directly learned associations.

3.2. Item-Level reinstatement of overlapping, yet incorrect contextual details after successful associative inference

We hypothesized that after successful associative inference, neural patterns in the anterior hippocampus, L. ITG, and posterior mPFC would be more similar to neural patterns when participants viewed the overlapping, yet incorrect event context compared to after unsuccessful inference or before successful inference, reflecting the successful inference-dependent reinstatement of contextual details from the overlapping, yet incorrect event. For each participant, ROI, and bin, we calculated a pattern similarity score (r_{match} - $r_{mismatch}$), which represented the item-specific reinstatement of overlapping, yet incorrect contextual details during retrieval (see Fig. 2b) and subjected participants' pattern similarity scores to three (one for each ROI: anterior hippocampus, L. ITG, posterior mPFC; see Fig. 5a for anatomical masks) x 2 (time: before vs. after) x 2 (inference: successful vs. unsuccessful) repeated measures ANOVAs (see Supplemental Figure 3 for RSA results split by r_{match} and $r_{mismatch}$).

In line with the role of the anterior hippocampus in the rapid binding of event details both within (Eichenbaum and Cohen, 2001; Hannula and Ranganath, 2008; Shimamura, 2010) and across event boundaries (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a), we found evidence for item-specific reactivation of the overlapping, yet incorrect event context in the anterior hippocampus. Specifically, the ANOVA conducted on the pattern similarity scores revealed a significant time by inference interaction, $F(1,22) = 6.12$, $p = .022$, $\eta_p^2 = 0.22$, with greater pattern similarity scores after successful associative inference compared to after unsuccessful associative inference, $t(22) = 3.18$, $p = .004$, mean difference = 0.003, 95% CI [0.001, 0.006], $d = 0.65$. Critically, there was no significant difference in pattern similarity scores before successful inference compared to before unsuccessful inference, $t(22) < 1$, $p > .250$, mean difference = 0.0002, 95% CI [−0.002, 0.002], $d = 0.04$.[9]

The foregoing results support the hypothesis that during successful associative inference flexible recombination/cross-episode binding mechanisms linked to the anterior hippocampus may result in the mistaken binding of contextual details from event to the overlapping, yet incorrect source. Additional evidence for the reinstatement of contextual details from the overlapping, yet incorrect event may manifest in content-reinstatement regions similarly to how reinstatement of correct event details in such regions supports successful retrieval. In line with this hypothesis, results revealed a significant time by inference interaction, $F(1,22) = 7.90$, $p = .010$, $\eta_p^2 = 0.26$, in our content-reinstatement region (i.e., L. ITG). Subsequent *t*-tests revealed greater pattern similarity after successful associative inference compared to after unsuccessful associative inference, $t(22) = 2.33$, $p = .029$, mean difference = 0.002, 95% CI [0.0003, 0.004], $d = 0.48$. Further, results

[9]Our analytic approach of examining predicted differences in correlations across bins relative to the magnitude of individual correlations vs. zero is consistent with past RSA studies of episodic memory (e.g., Ritchey et al., 2013; Kuhl & Chun 2014; Wing et al., 2015). We believe that tests within a given bin relative to 0 are not easily interpretable because baseline similarity can be driven by various factors (e.g., vascularity; Haynes, 2015; Bhandari, Gagne & Badre, 2018). To control for such non-specific differences and directly test our hypotheses of greater reinstatement of overlapping yet, incorrect contextual details after successful relative to unsuccessful inference, we chose to compare the magnitude of the correlation across bins (e.g., before vs. after and successful vs. unsuccessful inference) and to not include the results of t-tests vs. 0.

revealed greater pattern similarity after successful associative inference compared to before successful associative inference, $t(22) = 3.26$, $p = .004$, mean difference = 0.003, 95% CI [0.001, 0.005], $d = 0.68$. Finally, there was no significant difference in pattern similarity scores before successful inference compared to before unsuccessful inference, $t(22) = 1.69$, $p = .11$, mean difference = 0.002, 95% CI [−0.0005, 0.005], $d = 0.35$.

In our final ROI, the posterior mPFC, the ANOVA revealed a significant time by inference interaction, $F(1,22) = 4.94$, $p = .037$, $\eta_p^2 = 0.18$. Subsequent *t*-tests revealed a trend toward greater pattern similarity after successful associative inference compared to after unsuccessful associative inference $t(22) = 1.92$, $p = .068$, mean difference = 0.004, 95% CI [−0.0003, 0.008], $d = 0.40$. Identical to results in the anterior hippocampus and L. ITG, results revealed no significant difference in pattern similarity scores before successful inference compared to before unsuccessful inference $t(22) = 1.09$, $p > .250$, mean difference = 0.002, 95% CI [−0.002, 0.005], $d = 0.23$.[10] Taken together, results show that after successful associative inference, when participants attempt to retrieve contextual details associated with the currently cued event, neural patterns are more similar to when participants were viewing the overlapping yet, incorrect event context relative to all other event contexts that were also from successful inference triads after the directly learned/associative inference test (see Fig. 5b for results; see Supplemental Results and Supplemental Figure 4 for RSA Control Analyses).

3.3. Reinstatement in anterior hippocampus correlates with L. ITG

The anterior hippocampus has been hypothesized to support successful retrieval by driving the reinstatement of encoding-related cortical activity in response to a partial event cue (i.e., pattern completion; Bosch et al., 2014, Pacheco Estefan et al., 2019; Gordon et al., 2014; Ritchey et al., 2013; Staresina et al., 2012; Tompary et al., 2016). Hippocampally-driven cortical reinstatement of such event details during retrieval has further been shown to track participants' memories for various aspects of an event (c.f., Gordon et al., 2014). In line with this hypothesis, we asked whether pattern similarity to the overlapping, yet incorrect event in the anterior hippocampus during retrieval affected the reinstatement of contextual details that were mistakenly bound to the overlapping, yet incorrect event in the L. ITG (i.e., our hypothesized content-reinstatement ROI) as a result of successful associative inference. That is, while previous results evaluate RSA effects within each ROI, the current results aim to understand how the hippocampus and L. ITG (our content-reinstatement region) *interact* in support of the retrieval of contextual details from the overlapping, yet incorrect event.

To test this across-region relationship, we first calculated the pattern similarity score ($r_{match} - r_{mismatch}$) for each trial within each bin in both anterior hippocampus and L. ITG ROIs. As reported above, the trial-wise pattern similarity scores reflect the similarity in the pattern of neural activity when, for example, participants are cued to retrieve contextual details associated with event AB_1 and when participants viewed event BC_1 context during the

[10] In order to determine if pattern similarity scores differed as function of ROI, we subjected participants' pattern similarity scores to a 3 (region: anterior hippocampus vs. L. ITG vs. posterior mPFC) x 2 (time: before vs. after inference) x 2 (inference: successful vs. unsuccessful) repeated measures ANOVA. Importantly, results revealed a significant time by inference interaction, $F(1,22) = 15.55$, $p = .001$, $\eta_p^2 = 0.41$, but no significant region by time by inference interaction, $F(2,44) < 1$, $p > .250$, $\eta_p^2 = 0.02$, suggesting that the overall patterns of results in our three ROIs were not significantly different from one another.

pre-exposure phase, relative to all other 'BC' event contexts from the same bin. Thus, trial-wise pattern similarity scores here reflect representational overlap between the currently cued event and the overlapping, yet incorrect event context. Next, for each participant and each bin, we correlated pattern similarity scores in the anterior hippocampus with pattern similarity scores in the L. ITG during the detail retrieval task (see Fig. 6a).

Results revealed that for successful inference triads both before and after the directly learned/associative inference test, there was a significant positive relationship between pattern similarity scores in the anterior hippocampus and the L. ITG (*before successful inference:* $t(22) = 2.19$, $p = .039$, mean difference = 0.07, 95% CI [0.004, 0.14], $d = 0.46$; *after successful inference:* $t(22) = 2.99$, $p = .007$, mean difference = 0.12, 95% CI [0.04, 0.20], $d = 0.62$). There was a trend toward a significant hippocampus-ITG relationship for unsuccessful inference triads before the directly learned/associative inference test, $t(22) = 1.88$, $p = .074$, mean difference = 0.09, 95% CI [−0.009, 0.18], $d = 0.39$, and no significant relationship for unsuccessful inference triads after the directly learned/associative inference test, $t(22) \langle 1, p \rangle .250$, mean difference = 0.03, 95% CI [−0.08, 0.14], $d = 0.12$. That is, a significant hippocampus-ITG relationship during retrieval was found for successful inference triads where the overlapping 'AB' and 'BC' event representations were either successfully integrated during encoding or flexibly recombined during retrieval (see Fig. 6b).

3.4. Univariate hippocampal effects correlate with context reinstatement in L. ITG

We hypothesized that flexible recombination/cross-episode binding mechanisms active during the directly learned/associative inference test result in contextual details being mistakenly bound to the overlapping, yet incorrect event context. Further, we hypothesized that the degree to which these misbound contextual details are reinstated during subsequent retrieval attempts should track with participants' false memory scores. In order to test the first element of our hypothesis, we correlated the strength of our univariate activity effects in the anterior hippocampus during the directly learned/associative inference test (i.e., correct inference > incorrect inference relative to correct directly learned > incorrect directly learned) with the subsequent strength of the pattern similarity effects in our hypothesized content reinstatement region (i.e., L. ITG) during the detail retrieval task (i.e., successful inference after > unsuccessful inference after relative to successful inference before > unsuccessful inference before). Results revealed that the strength of univariate effects in the anterior hippocampus was positively correlated with the degree to which neural patterns in the L. ITG became more similar to the overlapping, yet incorrect event context after successful inference relative to unsuccessful inference, $r = 0.43$, $p = .041$ (see Fig. 7a).

3.5. Context reinstatement in L. ITG correlates with behavioral false memory effects

Next, we correlated the strength of the pattern similarity effects in our hypothesized content-reinstatement region (i.e., L. ITG) with the strength of our behavioral false memory effects (i.e., successful inference after > unsuccessful inference after relative to successful inference before > unsuccessful inference before) in order to determine whether successful inference related changes in overlapping, yet incorrect context reinstatement in the L. ITG were indeed related to participants' false memory scores. Results revealed that the degree to which

neural patterns in the L. ITG became more similar to the overlapping, yet incorrect event context after successful inference compared to unsuccessful inference relative to before was positively correlated with participants' false memory effects $r = 0.51$, $p = .012$ (see Fig. 7b), suggesting that reinstatement of contextual details from the overlapping, yet incorrect event may be responsible for successful inference-related changes in participants' false memory scores.

3.6. Content-Reinstatement mediates the relationship between flexible retrieval mechanisms and false memories

As a final analysis, we examined whether univariate effects in the anterior hippocampus, representing the degree of recombination/cross-episode binding during the directly learned/associative inference test, indirectly affects participants' detail memory responses via the cortical reinstatement of contextual details from overlapping events. The goal of the current analysis was to link the univariate results of the directly learned/associative inference test and the RSA results from the separate detail retrieval test. That is, the following mediation analysis aimed to reveal the relationship *across* the two tasks that participants were asked to complete, rather than understanding the mechanisms at play during each individual task (see above results and Fig. 6 for a discussion of how the anterior hippocampus may drive the reinstatement of the overlapping, yet incorrect event in the L. ITG *during* the detail retrieval test).

In order to assess the relationship between univariate activity effects in the anterior hippocampus, overlapping, yet incorrect context reinstatement effects and behavioral false memory effects, we subjected univariate activity effects from our bilateral anterior hippocampus ROI from the directly learned/associative inference test and participants' false memory effects to a mediation analysis with pattern similarity effects from the detail retrieval task within our content-reinstatement region (i.e., L. ITG), posterior mPFC, and bilateral anterior hippocampus as our three potential mediators (see Fig. 8).

The mediation analysis was conducted via the Multilevel Mediation and Moderation toolbox with 10,000 bootstrap samples (Wager et al., 2009; Atlas et al., 2010). The independent variable was correct inference > incorrect inference (relative to correct directly learned > incorrect directly learned) univariate activity from our bilateral anterior hippocampus ROI. Pattern similarity effects in the L. ITG, posterior mPFC and bilateral anterior hippocampus were included as our mediating variables and behavioral false memory effects were included as our dependent variable. Significant mediation was identified by the interaction of path a (univariate effects to pattern similarity effects) and path b (pattern similarity effects to false memory effects). Results revealed a significant indirect/mediation effect relating univariate anterior hippocampal activity effects during the directly learned/associative inference test with subsequent behavioral false memory effects when this relationship was mediated by pattern similarity effects in the L. ITG, mediation effect $ab = 0.11$ (0.06), $p = .02$. No other potential indirect pathways relating univariate activity during the directly learned/associative inference task to false memory effects from the detail retrieval task (e.g., univariate to posterior mPFC or univariate to anterior hippocampus) were significant, all $ps > .250$ (see Fig. 8).

4. Discussion

The current results provide direct neural evidence that 1) *specific* contextual details from an overlapping, yet incorrect event are reinstated during retrieval, resulting in false memories and 2) the same hippocampally-dependent flexible recombination mechanisms that support an adaptive function (i.e., successful inference) *increase* the likelihood that such misbound contextual details are reinstated during subsequent retrieval attempts.

We highlight *five* key findings of the current study. First, univariate results corroborate past studies and provide evidence for the involvement of the anterior hippocampus, posterior mPFC, and L. IFG regions in successful associative inference. Second, a neurally derived measure of trial-wise pattern similarity to the overlapping, yet incorrect event in the anterior hippocampus, posterior mPFC and L. ITG was greater after successful compared to unsuccessful inference. Third, the degree of reinstatement of overlapping, yet incorrect contextual details in the anterior hippocampus was positively correlated with the degree of reinstatement in our content-reinstatement region (i.e., the L. ITG). Fourth, the degree to which the incorrect, but related scene was reinstated in the L. ITG tracked participants' false memory effects, with greater reinstatement effects associated with stronger memory misattribution effects. Fifth, the univariate effects in the anterior hippocampus during the directly learned/associative inference task were positively correlated with the degree of successful inference-related changes in the reinstatement of contextual details from the overlapping, yet incorrect event in our content-reinstatement region during the detail retrieval task. Thus, in line with past work highlighting hippocampal-cortical interactions supporting correct memory responses (e.g., Gordon et al., 2014), in the current study, patterns of hippocampal activity during retrieval may drive the reinstatement of *misbound* contextual details in content-sensitive cortical regions. Futher, the degree of overlapping, yet incorrect context reinstatement in such content-reinstatement regions may result in the misattribution of such misbound details to participants' memory for the currently cued event. While the across-subject correlations and mediation analysis should be considered exploratory given the current sample size (e.g., Fritz and MacKinnon, 2007), all critical univariate and RSA analyses were performed *within-subject* and are well powered to test our hypothesis (see Liang and Preston, 2017; Mack and Preston, 2016; Tompary et al., 2016; Tompary and Davachi, 2017 which employed similar analyses and sample sizes).

4.1. Hippocampal and prefrontal retrieval processes support successful associative inference

In line with past work by Zeithamova and Preston (2010), we found univariate evidence for the involvement of anterior hippocampus, posterior mPFC and L. IFG regions in successful associative inference. The anterior hippocampus has been implicated in the flexible retrieval and rapid binding of associative information both within (Eichenbaum and Cohen, 2001; Hannula and Ranganath, 2008; Shimamura, 2010) and across event boundaries (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a). Specifically, in line with the current results, past studies using a similar associative inference task have linked the anterior hippocampus to the flexible reactivation and recombination of discrete 'AB' and

'BC' event representations in order to infer the relationship between the non-overlapping 'A' and 'C' items (Preston et al., 2004; Zeithamova and Preston, 2010).

Prefrontal regions including the posterior mPFC and L. IFG have been implicated in the integration of incoming information with existing knowledge structures and interference resolution for similar or competing items in memory, respectively. Specifically, past work has suggested a role for posterior mPFC regions in the integration of information into existing knowledge schemas during new learning (e.g., Bonasia et al., 2018). Schemas are organized knowledge frameworks related to a particular subject or event that support our ability generalize across event boundaries to extract the general or most common features of multiple related events (Bartlett, 1932). In a similar vein as integrated/recombined representations supporting successful associative inference, schemas allow for relationships between common event elements that have not been directly experienced together. That is, in a novel context, schemas may provide a framework by which expectations can be drawn based on past experiences with similar or conceptually-related contexts. As suggested by past work (Zeithamova et al., 2012b), such schema-based generalization and abstraction across event boundaries may rely on similar processes and/or representations that support successful associative inference (Bowman and Zeithamova, 2018; Schlichting et al., 2015; Spalding et al., 2018; Tse et al., 2011; van Kesteren et al., 2010b, 2010a; Zeithamova et al., 2012a; for a similar view, see also Nieuwenhuis and Takashima, 2011). In line with a role of the posterior mPFC in schema-based generalization and abstraction supporting memory integration, our posterior mPFC ROI was indeed similar to those reported in past work evaluating the effects of schema congruency/incongruency on associative memory (van Buuren et al., 2014) and memory integration (van Kesteren et al., 2020).

Finally, IFG regions have been implicated in the controlled retrieval of and interference resolution among competing memory representations (Badre and Wagner, 2007; Oztekin et al., 2009). Consistent with controlled retrieval/interference resolution interpretation of L. IFG function, we found greater L. IFG activity for correct inference compared to incorrect inference trials potentially because successful associative inference requires the reactivation and manipulation of similar, partially overlapping 'AB' and 'BC' representations and presumably requires greater interference resolution than the retrieval of a single directly learned representation. Taken together, the current univariate results implicate a key role for the anterior hippocampus, posterior mPFC, and L. IFG in the flexible use of previously learned representations stored in memory to learn novel associations among items that were never directly experienced together. Further, the current results replicate past work using a similar associative inference paradigm despite using different encoding instructions, more complex stimuli, and differing study- test delays (see Zeithamova and Preston, 2010).

4.2. Recombination-related contextual reinstatement in the hippocampus, posterior mPFC and content-reinstatement region

The current results extend past work relating the reinstatement of encoding-related patterns during retrieval to participants' memory decisions (e.g., Mack and Preston, 2016). Specifically, past work has highlighted both the relationship between neural reinstatement and hippocampal-cortical interactions in support of successful memory retrieval (for a

review, see Xue, 2018). The current results extend such findings to false memories for specific contextual details that were mistakenly bound to the currently cued event as a direct consequence of flexible retrieval processes that support successful inference.

During the detail retrieval task, we found greater neural pattern similarity between the currently cued event and the overlapping, yet incorrect context after successful inference compared to after unsuccessful inference in the anterior hippocampus, posterior mPFC and our content-reinstatement region (i.e., the L. ITG). Critically, we correlated memory-based patterns of activity during the detail retrieval task with neural patterns when participants viewed the overlapping, yet incorrect event context during the pre-exposure phase, which occurred prior to participants learning the overlapping 'AB' and 'BC' associations, and quantified our pattern similarity effects by taking correlations from the same relative to different event triads. That is, pattern similarity effects reported in the current study reflect the item-specific reinstatement of the overlapping, yet incorrect event context, independent of any general successful inference related processes or any perceptual similarities between the 'encoding' (i.e., pre-exposure) and retrieval phases.

We hypothesized that hippocampally-dependent flexible recombination and cross-episode binding mechanisms that support successful associative inference would result in a more integrated hippocampal representation on subsequent retrieval attempts, which would further result in the mistaken reinstatement of event elements from the overlapping, yet incorrect event context via hippocampally-driven cortical reinstatement mechanisms. In line with this hypothesis, during the detail retrieval task for successful inference triads, we found a significant positive relationship between pattern similarity scores in the anterior hippocampus and the L. ITG. This finding suggests that pattern similarity effects in the anterior hippocampus may result in the reinstatement of overlapping, yet incorrect contextual details in content-selective cortical regions potentially via erroneous pattern completion processes (whereby elements of overlapping, yet incorrect context are mistakenly reinstated in response to the cue person).

Such false memory results are consistent with past research demonstrating that, under certain circumstances, false memories can be accompanied by the false reactivation of content-sensitive cortical regions (e.g., Aminoff et al., 2008; Kahn et al., 2004; Karanian and Slotnick, 2017, 2018; Kurkela and Dennis, 2016; Slotnick and Schacter, 2004). They also fit with work showing that the reinstatement or reminders of past contextual information during new learning, can result in source misattributions where new information is mistakenly remembered as having come from the original context (Hupbach et al., 2008, 2007, 2009; Gershman et al., 2013). Such studies show that the same regions active during encoding may come online both for the retrieval of true and false memories and also during new learning, resulting in source misattributions. By contrast, the current results demonstrate that false memories can be supported by the *item-specific* reinstatement of contextual details (for related work see also Liang and Preston, 2017; Kim et al., 2019) and further, that flexible retrieval-related changes in false contextual reinstatement track such changes in participants' false memory scores.[11]

Importantly, the current study highlights strong ROI-specific hypotheses based on past literature, which allows us to narrowly focus the results and discussion on specific and logical regions known to be involved in flexibly retrieving and recombining past events. Future, more exploratory, work should attempt to determine the role of other core network regions typically involved with episodic memory related tasks in successful associative inference and subsequent false memories.

4.3. Relating flexible recombination mechanisms to the neural reinstatement of contextual details and behavioral false memories

Given past work suggesting that flexible retrieval mechanisms may come at a cost - namely, the misattribution of contextual details from one event to the overlapping yet incorrect event - we hypothesized that hippocampally-dependent flexible retrieval processes active during the directly learned/associative inference test may drive subsequent reinstatement of contextual details from the overlapping, yet incorrect event in our content-reinstatement region. Further, we hypothesized that inference-dependent changes in contextual reinstatement in our content-reinstatement region may drive participants' behavioral false memory effects. In line with this hypothesis, we found that individual differences in inference-related univariate hippocampal activity, representing flexible recombination/cross-episode binding mechanisms during the directly learned/associative inference test, were positively correlated with the change in inference-related reinstatement of contextual details from the overlapping, yet incorrect event in our content-reinstatement region. Further, the change in successful inference-related reinstatement of contextual details from the overlapping, yet incorrect event in our content-reinstatement region was positively correlated with participants' behavioral false memory effects. In sum, these results suggest that the greater the degree to which participants recombined the partially overlapping 'AB' and 'BC' events in order to infer the relationship between the non-overlapping 'A' and 'C' elements, the greater the reinstatement of specific contextual details that were mistakenly bound to the overlapping, yet incorrect event as a result of successful associative inference. Further, the reinstatement of contextual details during subsequent retrieval attempts may drive the reported pattern of behavioral false memory effects.

In a mediation analysis aimed at linking anterior hippocampal univariate activity effects during the directly learned/associative inference test and behavioral false memory effects via pattern similarity effects in our three ROIs, we found a significant indirect effect of univariate activity in the anterior hippocampus during the directly learned/associative inference task on subsequent false memory scores via the reinstatement of contextual details from the overlapping, yet incorrect event in the L. ITG (i.e., our content-reinstatement region). Such results build on past work highlighting the relationship between memory errors and recombining elements of distinct episodic or autobiographical memories (e.g., Burt et al., 2004; Devitt et al., 2015; Odegard and Lampinen, 2004).

[11]The ideal comparison to determine whether reinstatement of the overlapping, yet incorrect tracks with participants' false memory scores would be at the trial-level rather than across conditions. However, if we were to split each condition into true, false and foil memory responses we would not have sufficient trials to compare reinstatement results among memory response types. Namely, our cutoff for inclusion in the RSA analyses is 15 trials per condition and limiting our analyses to only false memory responses would result in the majority of participants being excluded due to low trial counts. Future work, potentially using a paradigm resulting in higher rates of false memory responses should attempt to elucidate trial-level reinstatement and behavioral false memory relationships.

While results of the current mediation analysis should be considered exploratory given the small sample size for an across-subjects mediation effect (Fritz and MacKinnon, 2007), they are in line with the results of the previously reported correlations and suggest that recombination/cross-episode binding-related activity may be related to subsequent changes in pattern similarity in regions that are important for reinstating encoding-related perceptual information during retrieval. Future research should attempt to clarify the role of flexible retrieval processes in the reinstatement of subsequent event details using larger sample sizes and a task more suited for classic mediation analyses.

Conclusion

Together, our findings suggest that hippocampally-dependent flexible recombination/cross-episode binding mechanisms support successful associative inference and these same flexible retrieval processes result in the neural representations of the original event becoming more similar to the overlapping, yet incorrect context during subsequent retrieval attempts. Further, the degree to which these overlapping, yet incorrect contextual details were later reinstated after successful inference compared to after unsuccessful inference (relative to reinstatement effects before successful compared to unsuccessful inference) in content-reinstatement regions tracked participants false memory effects. These findings suggest that the false memory effects reported here may be the result of the mistaken binding of contextual details from the overlapping yet incorrect event context to the currently cued event as a result of successful associative inference.

More generally, and in line with the tenets of the constructive episodic simulation hypothesis discussed at the outset (Schacter and Addis, 2007a, 2007b, 2020), our results provide novel neuroimaging evidence that directly links flexible retrieval and recombination processes with memory errors that result from adaptive uses of those processes, which in our paradigm involve supporting successful associative inference. Accordingly, these results also lend novel neural support to the broader idea that memory errors and distortions are produced by adaptive constructive processes (Schacter, 2012) that support diverse functions, including future event simulation, semantic processing, and memory updating (e.g., Chadwick et al., 2016; Dewhurst et al., 2016; Howe, 2011; Howe and Garner, 2018; Schacter et al., 2011; for a recent review, see Schacter et al., 2021). We think that future studies that elucidate neural basis of such effects will contribute importantly to our understanding of the constructive nature of memory and cognition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Institute of Mental Health grant MH060941 and National Institute on Aging grant AG08441 awarded to Daniel L. Schacter. This research was carried out in whole at the Harvard Center for Brain Science and involved the use of instrumentation supported by the Shared Instrumentation Grant Program, specifically, grant number S10OD020039. We thank the University of Minnesota Center for Magnetic Resonance Research for their work on the SMS-BOLD sequence used in the current study and Aleea Devitt, Nadia Brashier, Ethan Harris, Sarah Kalinowski and Jyotika Bindra for assistance with data collection. A brief version of this study was presented as a poster at the Cognitive Neuroscience Society's 27th annual meeting.

References

Aminoff E, Schacter DL, Bar M, 2008. The cortical underpinnings of context-based memory distortion. J. Cogn. Neurosci20 (12), 2226–2237. doi:10.1162/jocn.2008.20156. [PubMed: 18457503]

Atlas LY, Bolger N, Lindquist MA, Wager TD, 2010. Brain mediators of predictive cue effects on perceived pain. J. Neurosci30 (39), 12964–12977. doi:10.1523/JNEUROSCI.0057-10.2010. [PubMed: 20881115]

Badre D, Wagner AD, 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia45, 2883–2901. doi:10.1016/j.neuropsychologia.2007.06.015. [PubMed: 17675110]

Bar M, 2004. Visual objects in context. Nat. Rev. Neurosci5, 617–629. doi:10.1038/nrn1476. [PubMed: 15263892]

Bartlett FC, 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge.

Bird CM, Keidel JL, Ing LP, Horner AJ, Burgess N, 2015. Consolidation of complex events via reinstatement in posterior cingulate cortex. J. Neurosci35 (43), 14426–14434. doi:10.1523/JNEUROSCI.1774-15.2015. [PubMed: 26511235]

Bhandari A, Gagne C, Badre D, 2018. Just above chance: is it harder to decode information from prefrontal cortex hemodynamic activity patterns?J. Cogn. Neurosci30 (10), 1473–1498. doi:10.1162/jocn_a_01291. [PubMed: 29877764]

Bonasia K, Sekeres MJ, Gilboa A, Grady CL, Winocur G, Moscovitch M, 2018. Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. Neurobiol. Learn. Mem153, 26–39. doi:10.1016/j.nlm.2018.02.017. [PubMed: 29474955]

Bosch AE, Jehee JFM, Fernandez G, Doeller CF, 2014. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. J. Neurosci34 (22), 7493–7500. doi:10.1523/JNEUROSCI.0805-14.2014. [PubMed: 24872554]

Bowen HJ, Kensinger EA, 2017. Recapitulation of emotional source context during memory retrieval. Cortex91, 142–156. doi:10.1016/j.cortex.2016.11.004. [PubMed: 27923474]

Bowman CR, Zeithamova D, 2018. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. J. Neurosci38 (10), 2605–2614. doi:10.1523/JNEUROSCI.2811-17.2018. [PubMed: 29437891]

Brainerd CJ, Reyna VF, 2005. The Science of False Memory. Oxford University Press doi:10.1093/acprof:oso/9780195154054.001.0001.

Brett MA, Valabregue R, Poline JB, 2002. Region of interest analysis using an SPM toolbox (abstract). Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan, 16 Available on CD-ROM in *NeuroImage*.

Bridge DJ, Voss JL, 2014a. Hippocampal binding of novel information with dominant memory traces can support both memory stability and change. J. Neurosci34, 2203–2213. doi:10.1523/JNEUROSCI.3819-13.2014. [PubMed: 24501360]

Bridge DJ, Voss JL, 2014b. Active retrieval facilitates across-episode binding by modulating the content of memory. Neuropsychologia63, 154–164. doi:10.1016/j.neuropsychologia.2014.08.024. [PubMed: 25173711]

Burt CDB, Kemp S, Conway M, 2004. Memory for true and false autobiographical event descriptions. Memory12, 545–552. doi:10.1080/09658210344000071. [PubMed: 15615313]

Carpenter AC, Schacter DL, 2017. Flexible retrieval: when true inferences produce false memories. J. Exp. Psychol.: Learn., Mem. Cogn43, 335–349. doi:10.1037/xlm0000340. [PubMed: 27918169]

Carpenter AC, Schacter DL, 2018a. False memories, false preferences: flexible retrieval mechanisms supporting successful inference bias novel decisions. J. Exp. Psychol.: Gen147 (7), 988–1004. doi:10.1037/xge0000391. [PubMed: 29419307]

Carpenter AC, Schacter DL, 2018b. Flexible retrieval mechanisms supporting successful inference produce false memories in younger but not older adults. Psychol. Aging33 (1), 134–143. doi:10.1037/pag0000210. [PubMed: 29494184]

Chadwick MJ, Anjum RS, Kumaran D, Schacter DL, Spiers HJ, Hassabis D, 2016. Semantic representations in the temporal pole predict false memories. Proc. Natl. Acad. Sci113 (36), 10180–10185. doi:10.1073/pnas.1610686113. [PubMed: 27551087]

Collin SH, Milivojevic B, Doeller CF, 2015. Memory hierarchies map onto the hippocampal long axis in humans. Nat. Neurosci18 (11), 1562–1564. doi:10.1038/nn.4138. [PubMed: 26479587]

Danker JF, Anderson JR, 2010. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol. Bull136, 87–102. doi:10.1037/a0017937. [PubMed: 20063927]

Devitt AL, Monk-Fromont E, Schacter DL, Addis DR, 2015. Factors that influence the generation of autobiographical memory conjunction errors. Memory24, 204–222. doi:10.1080/09658211.2014.998680. [PubMed: 25611492]

Dewhurst SA, Anderson RJ, Grace L, van Esch L, 2016. Adaptive false memory: imagining future scenarios increases false memories in the DRM paradigm. Mem. Cognit44 (7), 1076–1084. doi:10.3758/s13421-016-0620-0.

Dudai Y, Carruthers M, 2005. The Janus face of Mnemosyne. Nature434, 567. doi:10.1038/434567a. [PubMed: 15800602]

Eichenbaum H, Cohen NJ, 2001. From Conditioning to Conscious Recollection: Memory systems of the Brain. Oxford University Press, Oxford.

Etzel JA, Zacks JM, Braver TS, 2013. Searchlight analysis: promise, pitfalls, and potential. Neuroimage78, 261–269. doi:10.1016/j.neuroimage.2013.03.041. [PubMed: 23558106]

Ford JH, Kensinger EA, 2017. Age-related reversals in neural recruitment across memory retrieval phases. J. Neurosci37 (20), 5172–5182. doi:10.1523/JNEUROSCI.0521-17.2017. [PubMed: 28442537]

Frank LE, Bowman CR, Zeithamova D, 2019. Differential functional connectivity along the long axis of the hippocampus aligns with differential role in memory specificity and generalization. J. Cogn. Neurosci31 (12), 1958–1975. doi:10.1162/jocn_a_01457. [PubMed: 31397613]

Fritz MS, Mackinnon DP, 2007. Required sample size to detect the mediated effect. Psychol. Sci18 (3), 233–239. doi:10.1111/j.1467-9280.2007.01882.x. [PubMed: 17444920]

Gershman SJ, Schapiro AC, Hupbach A, Norman KA, 2013. Neural context reinstatement predicts memory misattribution. J. Neurosci15, 8590–8595. doi:10.1523/JNEUROSCI.0096-13.2013.

Gordan AM, Rissman J, Kiani R, Wagner AD, 2014. Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cereb. Cortex24 (12), 3350–3364. doi:10.1093/cercor/bht194. [PubMed: 23921785]

Han X, Berg AC, Samaras D, Leung HC, 2013. Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions. Neuroimage73 (8), 8–15. doi:10.1016/j.neuroimage.2013.01.055. [PubMed: 23380167]

Hannula DE, Ranganath C, 2008. Medial temporal lobe activity predicts successful relational memory binding. J. Neurosci28 (1), 116–124. doi:10.1523/JNEUROSCI.3086-07.2008. [PubMed: 18171929]

Haynes JD, 2015. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron87, 257–270. doi:10.1016/j.neuron.2015.05.025. [PubMed: 26182413]

Howe ML, Garner SR, 2018. Can false memories prime alternative solutions to ambiguous problems? Memory26 (1), 96–105. doi:10.1080/09658211.2017.1332226. [PubMed: 28553746]

Howe ML, 2011. The adaptive nature of memory and its illusions. Curr. Dir. Psychol. Sci20, 312–315. doi:10.1177/0963721411416571.

Hupbach A, Gomez R, Hardt O, Nadel L, 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn. Mem14, 47–53. doi:10.1101/lm.365707. [PubMed: 17202429]

Hupbach A, Gomez R, Nadel L, 2009. Episodic memory reconsolidation: updating or source confusion?Memory17, 502–510. doi:10.1080/09658210902882399. [PubMed: 19468955]

Hupbach A, Hardt O, Gomez R, Nadel L, 2008. The dynamics of memory: context- dependent updating. Learn. Mem15, 574–579. doi:10.1101/lm.1022308. [PubMed: 18685148]

Jing HG, Madore KP, Schacter DL, 2016. Worrying about the future: an episodic specificity induction impacts problem solving, reappraisal, and well-being. J. Exp. Psychol.: Gen145, 402–418. doi:10.1037/xge0000142. [PubMed: 26820166]

Johnson JD, Rugg MD, 2007. Recollection and the reinstatement of encoding-related cortical activity. Cereb. Cortex17 (11), 2507–2515. doi:10.1093/cercor/bhl156. [PubMed: 17204822]

Kahn I, Davachi L, Wagner AD, 2004. Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. J. Neurosci24 (17), 4172–4180. doi:10.1523/JNEUROSCI.0624-04.2004. [PubMed: 15115812]

Karanian JM, Slotnick SD, 2017. False memories for shape activate the lateral occipital complex. Learn. Mem24, 552–556. doi:10.1101/lm.045765.117. [PubMed: 28916630]

Karanian JM, Slotnick SD, 2018. Confident false memories for spatial location are mediated by V1. Cogn. Neurosci9, 139–150. doi:10.1080/17588928.2018.1488244. [PubMed: 29898628]

Kark SM, Kensinger EA, 2019. Post-encoding amygdala-visuosensory coupling is associated with negative memory bias in healthy young adults. J. Neurosci39 (16), 3130–3143. doi:10.1523/JNEUROSCI.2834-18.2019. [PubMed: 30760626]

Kim G, Norman KA, Turk-Browne NB, 2019. Neural overlap in item representations across episodes impairs context memory. Cereb. Cortex29 (6), 2682–2693. doi:10.1093/cercor/bhy137. [PubMed: 29897407]

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA, 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron60 (6), 1126–1141. doi:10.1016/j.neuron.2008.10.043. [PubMed: 19109916]

Kuhl BA, Chun MM, 2014. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. J. Neurosci34 (23), 8051–8060. doi:10.1523/JNEUROSCI.4328-13.2014. [PubMed: 24899726]

Kurkela KA, Dennis NA, 2016. Event-related fMRI studies of false memory: an activation likelihood estimation meta-analysis. Neuropsychologia81, 149–167. doi:10.1016/j.neuropsychologia.2015.12.006. [PubMed: 26683385]

Lancaster JL, Summerln JL, Rainey L, Freitas CS, Fox PT, 1997. The Talairach Daemon, a database server for Talairach Atlas Labels. Neuroimage5, 238–242.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT, 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp10, 120–131. doi:10.1002/1097-0193(200007)10:3<120::aid-hbm30>3.0.co;2-8. [PubMed: 10912591]

Lee SH, Kravitz DJ, Baker CI, 2019. Differential representations of perceived and retrieved visual information in hippocampus and cortex. Cereb. Cortex29 (10), 4452–4461. doi:10.1093/cercor/bhy325. [PubMed: 30590463]

Liang JC, Preston AR, 2017. Medial temporal lobe reinstatement of content-specific details predicts source memory. Cortex91, 67–78. doi:10.1016/j.cortex.2016.09.011. [PubMed: 28029355]

Loftus EF, Miller DG, Burns HJ, 1978. Semantic integration of verbal information into a visual memory. J. Exp. Psychol. [Hum. Learn.]4 (1), 19–31. doi:10.1037/0278-7393.4.1.19.

Mack ML, Preston AR, 2016. Decisions about the past are guided by reinstatement of specific memories in the hippocampi. Neuroimage127, 144–157. doi:10.1016/j.neuroimage.2015.12.015. [PubMed: 26702775]

Madore KP, Schacter DL, 2014. An episodic specificity induction enhances means-end problem solving in young and older adults. Psychol. Aging29, 913–924. doi:10.1037/a0038209. [PubMed: 25365688]

Madore KP, Addis DR, Schacter DL, 2015. Creativity and memory: effects of an episodic-specificity induction on divergent thinking. Psychol. Sci26, 1461–1468. doi:10.1177/0956797615591863. [PubMed: 26205963]

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH, 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage19 (3), 1233–1239. doi:10.1016/s1053-8119(03)00169-1. [PubMed: 12880848]

McClelland JL, 1995. Constructive memory and memory distortions: a parallel-distributed processing approach. In: Schacter DL (Ed.), Memory Distortion. Harvard University Press, Cambridge, MA, pp. 69–90.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N, 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage53 (1), 103–118. doi:10.1016/j.neuroimage.2010.05.051. [PubMed: 20580933]

Moscovitch M, Cabeza R, Winocur G, Nadel L, 2016. Episodic memory and beyond: the hippocampus and neocortex in transformation. Annu. Rev. Psychol67, 105–134. doi:10.1146/annurev-psych-113011-143733. [PubMed: 26726963]

Nieuwenhuis ILC, Takashima A, 2011. The role of the ventromedial prefrontal cortex in memory consolidation. Behav. Brain Res218, 325–334. doi:10.1016/j.bbr.2010.12.009. [PubMed: 21147169]

Odegard TN, Lampinen JM, 2004. Memory conjunction errors for autobiographical events: more than just familiarity. Memory12, 288–300. doi:10.1080/09658210244000621. [PubMed: 15279433]

Oedekoven CSH, Keidel JL, Berens SC, Bird CM, 2017. Reinstatement of memory representations for lifelike events over the course of a week. Sci. Rep7, 14305. doi:10.1038/s41598-017-13938-4. [PubMed: 29084981]

Oztekin I, Curtis CE, McElree B, 2009. The medial temporal lobe and the left inferior prefrontal cortex jointly support interference resolution in verbal working memory. J. Cogn. Neurosci10, 1967–1979. doi:10.1162/jocn.2008.21146.

Pacheco Estefan D, Sanchez-Fibla M, Duff A, Principe A., Rocamora R, Zhang H, Axmacher N, Verschure PFMJ, 2019. Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. Nat. Commun10 (1), 2255. doi:10.1038/s41467-019-09569-0. [PubMed: 31113952]

Preston AR, Shrager Y, Dudokovic NM, Gabrieli JD, 2004. Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus14 (2), 148–152. doi:10.1002/hipo.20009. [PubMed: 15098720]

Ranganath C, Cohen MX, Dam C, D'Esposito M, 2004. Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. J. Neurosci24 (16), 3917–3925. doi:10.1523/JNEUROSCI.5053-03.2004. [PubMed: 15102907]

Ritchey M, Wing EA, LaBar KS, Cabeza R, 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb. Cortex23 (3), 2818–2828. doi:10.1093/cercor/bhs258. [PubMed: 22967731]

Roediger HL, 1996. Memory illusions. J. Mem. Lang35, 76–100. doi:10.1006/jmla.1996.0005.

Rugg MD, Johnson JD, Uncapher MR, 2015. Encoding and retrieval in episodic memory: insights from fMRI. In: Addis DR, Barense M, Duarte A (Eds.), Wiley Handbooks in Cognitive Neuroscience. The Wiley Handbook on the Cognitive Neuroscience of Memory, pp. 84–107.

Schacter DL, 2001. The Seven Sins of Memory: How the Mind Forgets and Remembers. Houghton Mifflin, Boston and New York.

Schacter DL, 2012. Adaptive constructive processes and the future of memory. Am. Psychol67, 603–613. doi:10.1037/a0029869. [PubMed: 23163437]

Schacter DL, Addis DR, 2007a. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos. Trans. R. Soc. B: Biol. Sci362, 773–786. doi:10.1098/rstb.2007.2087.

Schacter DL, Addis DR, 2007b. The ghosts of past and future. Nature445, 27. doi:10.1038/445027a. [PubMed: 17203045]

Schacter DL, Addis DR, 2020. Memory and imagination: perspectives on constructive episodic simulation. In: Abraham A (Ed.), The Cambridge Handbook of the Imagination. Cambridge University Press, Cambridge.

Schacter DL, Carpenter AC, Devitt AL, Thakral PP, 2021. Memory errors and distortion. In: Kahana MJ, Wagner AD (Eds.), The Oxford Handbook of Human Memory. Oxford University Press, New York.

Schacter DL, Guerin SA, St. Jacques PL, 2011. Memory distortion: an adaptive perspective. Trends Cogn. Sci15, 467–474. doi:10.1016/j.tics.2011.08.004. [PubMed: 21908231]

Schlichting ML, Mumford JA, Preston AR, 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun6, 8151. doi:10.1038/ncomms9151. [PubMed: 26303198]

Sheldon S, McAndrews MP, Moscovitch M, 2011. Episodic memory processes mediated by the medial temporal lobes contribute to open-ended problem solving. Neuropsychologia49, 2439–2447. doi:10.1016/j.neuropsychologia.2011.04.021. [PubMed: 21550352]

Shimamura AP, 2010. Hierarchical relational binding in the medial temporal lobe: the strong get stronger. Hippocampus20 (11), 1206–1216. doi:10.1002/hipo.20856. [PubMed: 20824723]

Shohamy D, Wagner AD, 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron60 (2), 378–389. doi:10.1016/j.neuron.2008.09.023. [PubMed: 18957228]

Slotnick SD, 2004. Visual memory and visual perception recruit common neural substrates. Behav. Cogn. Neurosci. Rev3, 207–221. doi:10.1177/1534582304274070. [PubMed: 15812107]

Slotnick SD, 2017. Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. Cogn. Neurosci8 (3), 150–155. doi:10.1080/17588928.2017.1319350. [PubMed: 28403749]

Slotnick SD, Schacter DL, 2004. A sensory signature that distinguishes true from false memories. Nat. Neurosci7, 664–672. doi:10.1038/nn1252. [PubMed: 15156146]

Slotnick SD, Schacter DL, 2006. The nature of memory related activity in early visual areas. Neuropsychologia44, 2874–2886. doi:10.1016/j.neuropsychologia.2006.06.021. [PubMed: 16901520]

Slotnick SD, Moo LR, Segal JB, Hart JR, 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn. Brain Res17 (1), 75–82. doi:10.1016/s0926-6410(03)00082-x.

Spalding KN, Schlichting ML, Zeithamova D, Preston AR, Tranel D, Duff MC, Warren DE, 2018. Ventromedial prefrontal cortex is necessary for normal as-sociative inference and memory integration. J. Neurosci38 (15), 3767–3775. doi:10.1523/JNEUROSCI.2501-17.2018. [PubMed: 29555854]

Staresina BP, Henson RN, Kriegeskorte N, Alink A, 2012. Episodic reinstatement in the medial temporal lobe. J. Neurosci32 (50), 18150–18156. doi:10.1523/JNEUROSCI.4156-12.2012. [PubMed: 23238729]

Suddendorf T, Corballis MC, 2007. The evolution of foresight: what is mental time travel, and is it unique to humans?Behav. Brain Sci30, 299–313. doi:10.1017/S0140525X07001975. [PubMed: 17963565]

Thakral PP, Madore KP, Schacter DL, 2020. The core episodic simulation network dissociates as a function of subjective and objective content. Neuropsychologia136, 107263. doi:10.1016/j.neuropsychologia.2019.107263. [PubMed: 31743681]

Thakral PP, Madore KP, Addis DR, Schacter DL, 2019. Reinstatement of event details during episodic simulation in the hippocampus. Cereb. Cortex doi:10.1093/cercor/bhz244.

Thakral PP, Wang TH, Rugg MD, 2015. Cortical reinstatement and the confidence and accuracy of source memory. Neuroimage109, 118–129. doi:10.1016/j.neuroimage.2015.01.003. [PubMed: 25583615]

Tompary A, Duncan K, Davachi L, 2016. High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. Hippocampus8, 995–1007. doi:10.1002/hipo.22582.

Tompary A, Davachi L, 2017. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. Neuron96 (1), 228–241. doi:10.1016/j.neuron.2019.12.020. [PubMed: 28957671]

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RG, 2011. Schema-dependent gene activation and memory encoding in the neocortex. Science333 (6044), 891–895. doi:10.1126/science.1205274. [PubMed: 21737703]

Tulving E, 1983. Ecphoric processes in episodic memory. Philos. Trans. R. Soc. London, Ser. B302, 361–371. doi:10.1098/rstb.1983.0060.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M, 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical

parcellation of the MNI MRI single-subject brain. Neuroimage15 (1), 273–278. doi:10.1006/nimg.2001.0978. [PubMed: 11771995]

Vaidya CJ, Zhao M, Desmond JE, Gabrieli JDE, 2002. Evidence for cortical encoding specificity in episodic memory: memory-induced reactivation of picture processing areas. Neuropsychologia40 (12), 2136–2143. doi:10.1016/S0028-3932(02)00053-2. [PubMed: 12208009]

van Buuren M, Kroes MC, Wagner IC, Genzel L, Morris RGM, Fernandez G, 2014. Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. J. Neurosci10 (34), 16662–16670. doi:10.1523/JNEUROSCI.2365-14.2014.

van Kesteren MT, Rijpkema M, Ruiter DJ, Fernandez G, 2010a. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. J. Neurosci30 (47), 15888–15894. doi:10.1523/JNEUROSCI.2674-10.2010. [PubMed: 21106827]

van Kesteren MT, Fernandez G, Norris DG, Hermans EJ, 2010b. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. Proc. Natl. Acad. Sci107, 7550–7555. [PubMed: 20363957]

van Kesteren MT, Rignanese P, Gianferrara PG, Krabbendam L, Meeter M, 2020. Congruency and reactivation aid memory integration through reinstatement of prior knowledge. Sci. Rep10, 4776. doi:10.1038/s41598-020-61737-1. [PubMed: 32179822]

Wager TD, Waugh CE, Lindquist M, Noll DC, Fredrickson BL, Taylor SF, 2009. Brain mediators of cardiovascular responses to social threat, Part I: reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. Neuroimage47, 821–835. doi:10.1016/j.neuroimage.2009.05.043. [PubMed: 19465137]

Wing EA, Ritchey M, Cabeza R, 2015. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. J. Cogn. Neurosci27 (4), 679–691. doi:10.1162/jocn_a_00740. [PubMed: 25313659]

Woloszyn L, Sheinberg DL, 2009. Neural dynamics in interior temporal cortex during a visual working memory task. J. Neurosci29 (17), 5494–5507. doi:10.1523/JNEUROSCI.5785-08.2009. [PubMed: 19403817]

Xue G, 2018. The neural representations underlying human episodic memory. Trends Cogn. Sci22 (6), 544–561. doi:10.1016/j.tics.2018.03.004. [PubMed: 29625850]

Zeithamova D, Dominick AL, Preston AR, 2012a. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron75, 168–179. doi:10.1016/j.neuron.2012.05.010. [PubMed: 22794270]

Zeithamova D, Preston AR, 2010. Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. J. Neurosci30, 14676–14684. doi:10.1523/JNEUROSCI.3250-10.2010. [PubMed: 21048124]

Zeithamova D, Schlichting ML, Preston AR, 2012b. The hippocampus and inferential reasoning: building memories to navigate future decisions. Front. Hum. Neurosci6, 70. doi:10.3389/fnhum.2012.00070. [PubMed: 22470333]

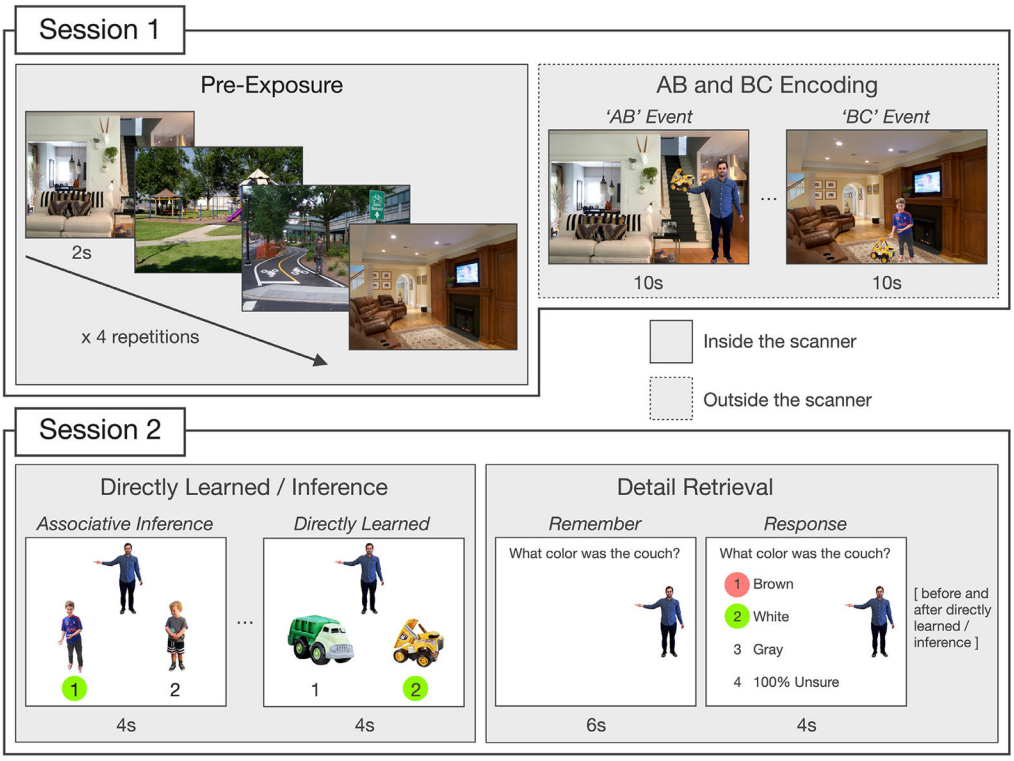

Fig. 1.

Schematic of experimental methods. Participants completed two sessions that were separated by a 24-hour delay. Session 1 consisted of two phases. During the pre-exposure phase, participants viewed each of the 'AB' and 'BC' contexts without the superimposed people and objects while in the scanner. During the AB and BC encoding phase participants learned overlapping 'AB' and 'BC' pairs outside of the scanner. For each event, participants were instructed to learn the direct relationships ('AB' and 'BC'), the indirect relationship ('AC'), and the event context details (e.g., the color of the couch). Following a 24-hour delay, participants completed Session 2, which consisted of three phases (two detail retrieval phases and one test of directly learned/associative inference trials). Participants completed one half of the detail retrieval trials before and completed the alternate half of the detail retrieval trials after the directly learned and associative inference test. During each detail retrieval trial participants first viewed the cue individual and the detail question (e.g., what color was the couch?) for six seconds. During this 'remember' period participants were instructed to think back to the currently cued context image and visualize the relevant contextual detail to the best of their ability. Following each six second 'remember' period, participants were presented with four possible answer choices: misinformation, true, foil and 100% unsure. The misinformation choice was the contradictory detail from the overlapping event (e.g., brown couch) and is circled in red. The correct choice was the true detail from the currently cued event (e.g., white couch) and is circled in green. The foil choice was a detail that was not present in either the currently cued or overlapping event (e.g., gray couch). Once the four possible answer choices appeared on the screen, participants were given four seconds to make their response. .

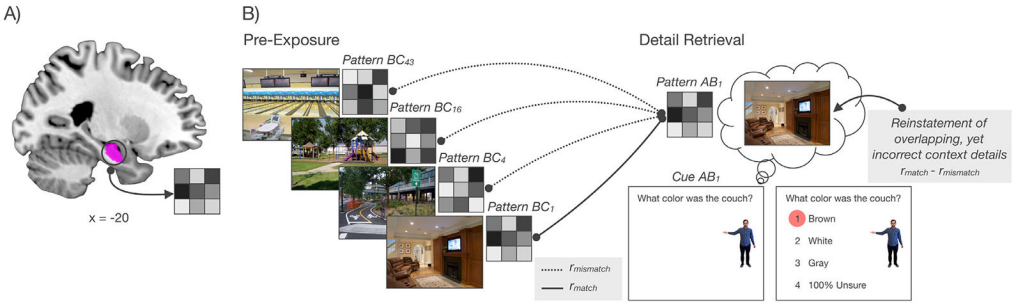

Fig. 2.

(A) Schematic of item-level reinstatement of overlapping, yet incorrect contextual details. A) For each anatomically defined ROI (i.e., bilateral anterior hippocampus – depicted above, L. ITG, bilateral subcallosal gyrus), the pattern of neural activity was extracted for every pre-exposure and detail retrieval trial. Patterns from the pre-exposure phase were averaged across all repetitions of the unique image. (B) Item-level reinstatement of overlapping, yet incorrect contextual details was measured by calculating the similarity between neural patterns during detail retrieval trials (e.g., AB_1 detail retrieval) and neural patterns when participants viewed the overlapping, yet incorrect event context during the pre-exposure phase (e.g., BC_1 pre-exposure; r_{match}) relative to neural patterns from the pre-exposure phase from other unrelated contexts coded in the same triad bin (e.g., successful inference vs. unsuccessful inference, before vs. after; $r_{mismatch}$). That is, if event ABC_1 were a successful inference triad from after the test of directly learned/associative inference trials, the neural patterns associated with AB_1 detail retrieval trials would be correlated with all other 'BC' pre-exposure patterns associated with successful inference triads whose detail retrieval questions also occurred after the test of directly learned/associative inference trials. r_{match} relative to $r_{mismatch}$ represents the item-specific reinstatement of overlapping, yet incorrect contextual details.

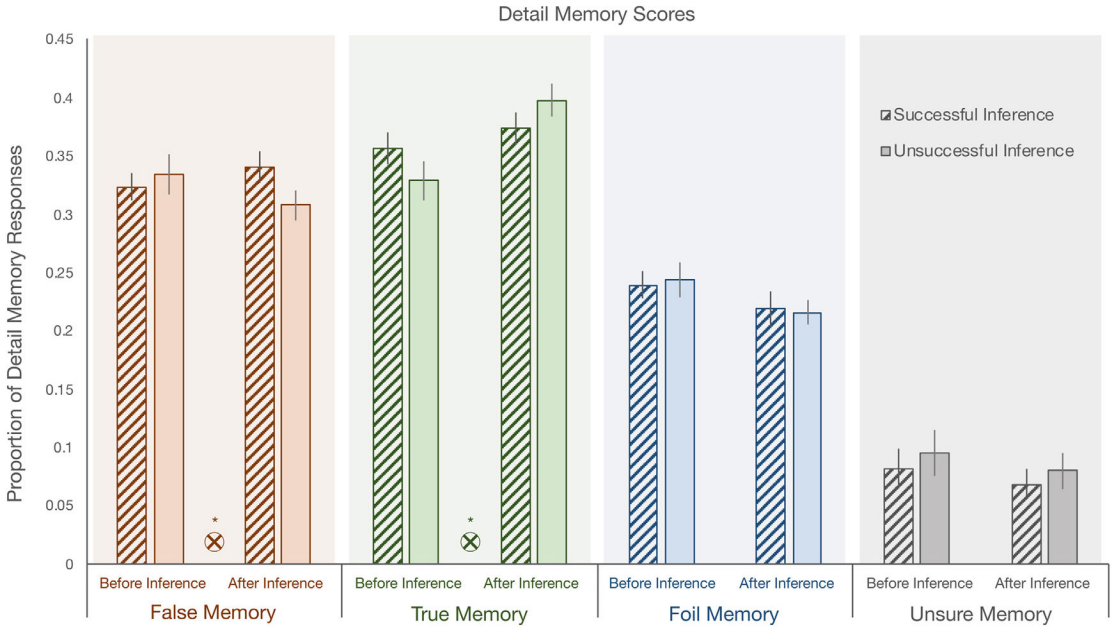

Fig. 3.

Proportions of false, true, foil and unsure memory responses. Performance on detail retrieval trials was examined both before and after successful or unsuccessful inference. Importantly, only trials for which participants responded correctly on the directly learned trials and made either a correct or incorrect response on the associative inference trial were included in this analysis. Overall, participants true memory scores were significantly higher than false, foil and unsure memory scores. Importantly, overall false memory scores were also significantly higher than foil and unsure memory scores. Further, the false memory analysis of primary interest for the current study revealed a time by inference interaction where participants' false memory scores were significantly higher after successful inference compared to after unsuccessful inference. Such results suggest that flexible recombination during retrieval, which supports successful associative inference, may also lead to memory error or distortion where details of the overlapping, yet incorrect event context are reactivated and mistakenly bound to the currently cued event. Circled cross denotes time by inference interaction. * $p \leq 0.05$. Error bars represent ± 1 SEM.

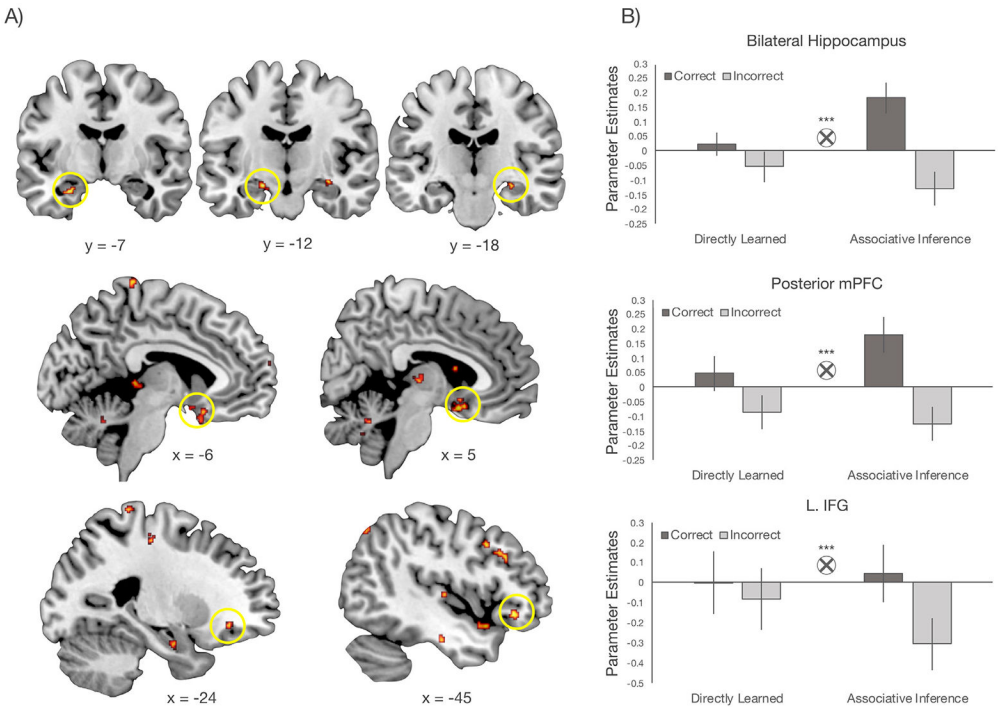

Fig. 4.

Univariate successful inference effects identified with the correct inference > incorrect inference contrast. (A) Parameter estimates were extracted from three ROIs identified by the correct inference > incorrect inference contrast: hippocampus ($x = -16$, $y = -10$, $z = -18$ and $x = -27$, $y = -7$, $z = -24$ and $x = 36$, $y = -8$, $z = -16$), posterior mPFC (i.e., subcallosal gyrus: $x = 4$, $y = 11$, $z = -16$ and $x = -3$, $y = 6$, $z = -14$ and gyrus rectus: $x = -6$, $y = 16$, $z = -22$) and L. IFG ($x = -45$, $y = 30$, $z = -7$and $x = -26$, $y = 34$, $z = -7$). Anterior hippocampal regions shown here are masked inclusively with the anatomically defined anterior hippocampus. (B) Parameter estimates for each ROI were subjected to a 2 (trial type: directly learned vs. associative inference) x 2 (accuracy: correct vs. incorrect) repeated measures ANOVA. Hippocampus, posterior mPFC and L. IFG regions showed significant trial type by accuracy interactions. Circled cross denotes time by inference interaction. * * * $p \leq 0.005$. Error bars represent ± 1 SEM.

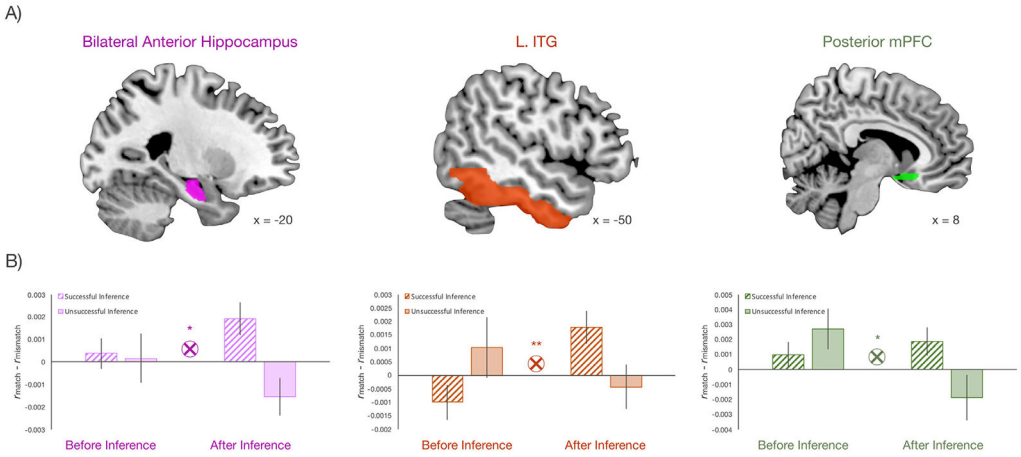

Fig. 5.

Results of representational similarity analysis using anatomically defined ROIs. (A) Analyses were conducted in three anatomically defined ROIs: bilateral anterior hippocampus, L. ITG, and posterior mPFC (i.e., subcallosal gyrus). (B) Pattern similarity scores were subjected to a 2 (time: before vs. after) x 2 (inference: successful vs. unsuccessful) repeated measures ANOVA. Results revealed a significant time by inference interaction in bilateral anterior hippocampus, L. ITG and the posterior portion of the mPFC suggesting that neural patterns during retrieval of contextual details following successful associative inference, become more similar to the overlapping, yet incorrect context compared to after unsuccessful inference. Thus, flexible recombination mechanisms that support successful associative inference also change the neural representations of the original events that allow for such successful inference. Circled cross denotes time by inference interaction. * * $p \leq 0.01$, * $p \leq 0.05$. Error bars represent ± 1 SEM. .

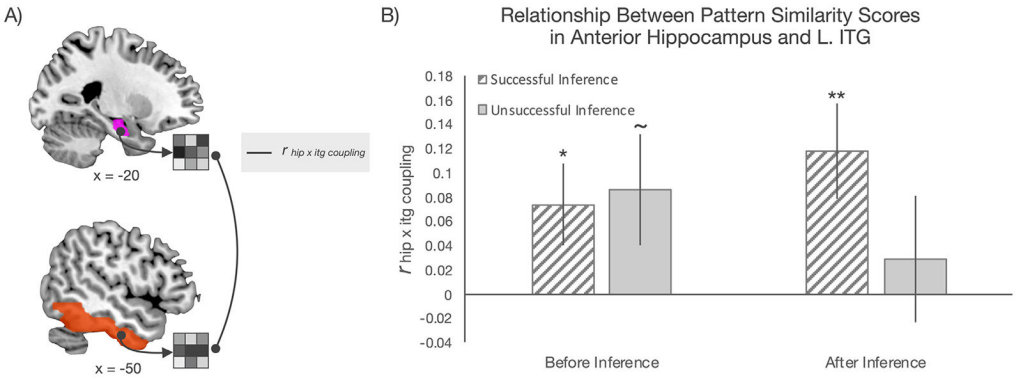

Fig. 6.

(A) Schematic of hypothesized relationship between pattern similarity scores in the anterior hippocampus and L. ITG during the detail retrieval task. (B) Results revealed that pattern similarity scores in the anterior hippocampus were positively correlated with pattern similarity scores in the L. ITG for successful inference triads tested both before and after the directly learned/associative inference test, suggesting that representational overlap in the anterior hippocampus as a result of successful associative inference may drive the subsequent reinstatement of contextual details that were mistakenly bound to the incorrect context in 'content-reinstatement' regions (i.e., L. ITG). No significant relationships were found for unsuccessful inference triads tested either before or after the directly learned/associative inference test. * * $p \leq 0.01$, * $p \leq 0.05$, ~ $p \leq 0.10$. Error bars represent ± 1 SEM. .

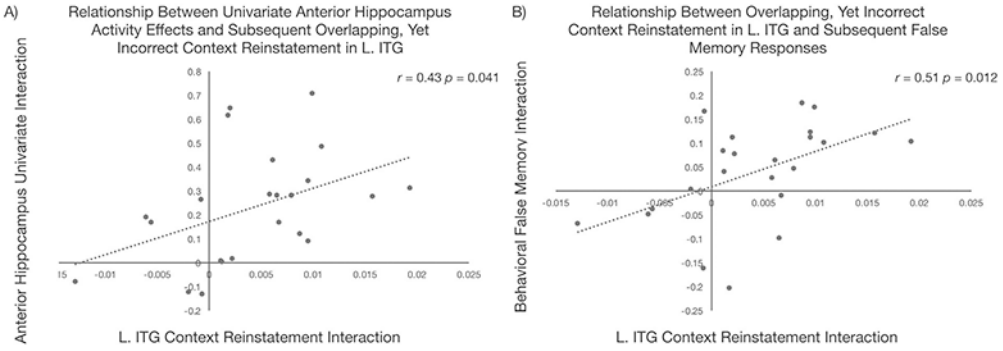

Fig. 7.

(A) Results of across-subject anterior hippocampus univariate and overlapping, yet incorrect context reinstatement in L. ITG correlation. Results revealed a significant positive relationship between the strength of univariate activity effects and subsequent overlapping, yet incorrect context reinstatement effects in the L. ITG, suggesting that the greater the flexible recombination/cross-episode binding mechanisms during correct compared to incorrect associative inference trials the greater the degree to which overlapping, yet incorrect contextual details are reinstated after compared to before successful associative inference relative to unsuccessful inference. (B) Results of across-subject overlapping, yet incorrect context reinstatement and behavioral false memory effects correlation. Results revealed a significant positive relationship between the strength of the overlapping, yet incorrect context reinstatement effects in the L. ITG and the strength of the behavioral false memory effects, suggesting that the degree to which overlapping, yet incorrect contextual details are reinstated after compared to before successful associative inference relative to unsuccessful inference supports the change in participants' attribution of such overlapping, yet incorrect misinformation details to the currently cued event after successful inference compared to before successful associative inference, relative to unsuccessful inference.

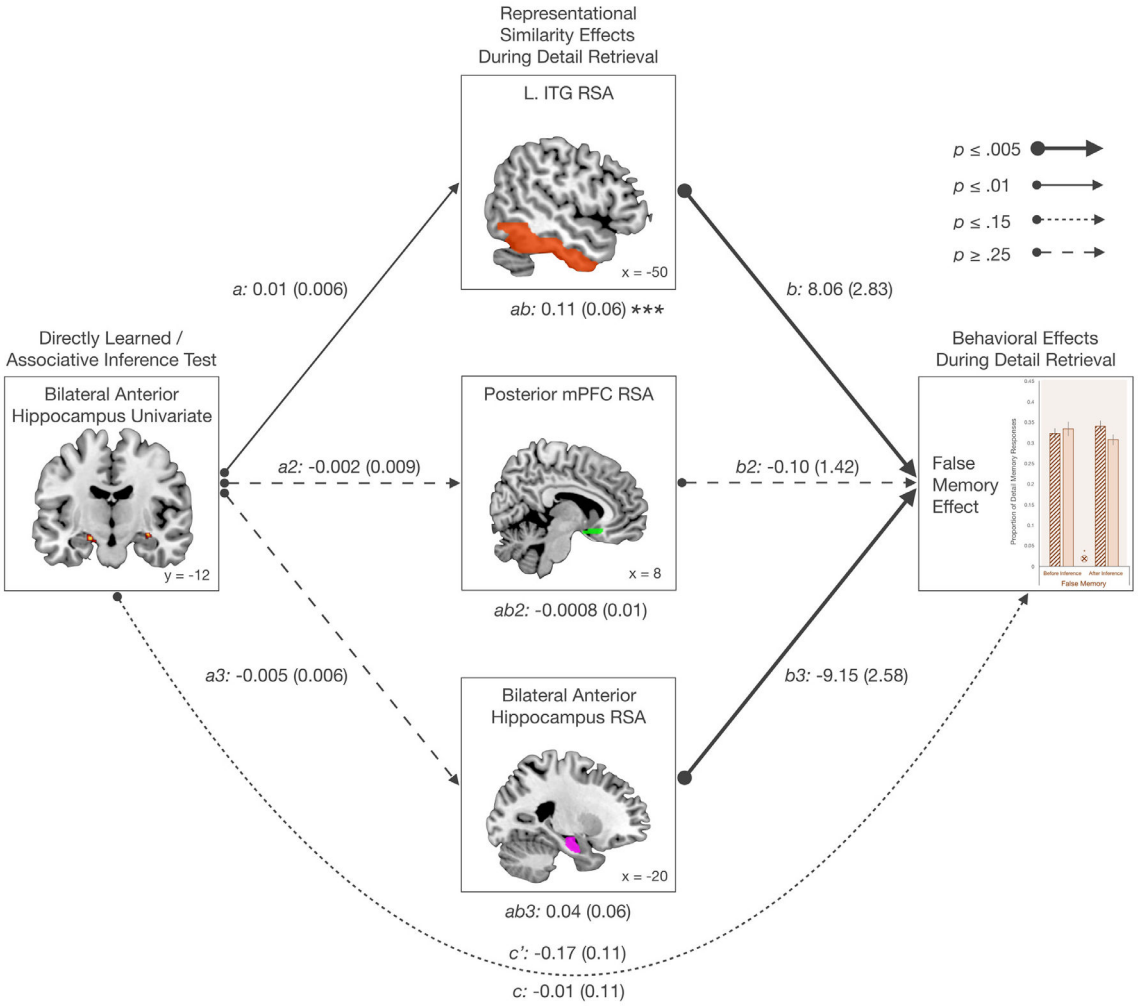

Fig. 8.

Depiction of exploratory mediation analysis linking univariate activity effects in bilateral anterior hippocampus during directly learned/associative inference test to subsequent changes in representational similarity during the detail retrieval task to the strength of the behavioral false memory effects. Numeric labels reflect standardized path coefficients (STE). Path thickness indicates the statistical significance of each direct effect. * * * indicates indirect effect of significance, $p = .02$. Results revealed a significant indirect effect of univariate activity during the directly learned/associative inference test on subsequent false memory effects via changes in representational similarity in our content reinstatement region – L. ITG (i.e., solid lines). Indirect effects via changes in representational similarity in bilateral anterior hippocampus and posterior mPFC were not significant (i.e., dashed lines). .