OXFORD

Cerebral Cortex, 2023, 1–18
https://doi.org/10.1093/cercor/bhad179
Original Article

Memory reactivation and suppression modulate integration of the semantic features of related memories in hippocampus

Neal W Morton [1,*], Ellen L. Zippi [2], Alison R. Preston [1,3,4]

[1]Center for Learning and Memory, The University of Texas at Austin, Austin, TX 78712, United States,
[2]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 95064, United States,
[3]Department of Psychology, The University of Texas at Austin, Austin, TX 78712, United States,
[4]Department of Neuroscience, The University of Texas at Austin, Austin, TX 78712, United States

*Corresponding author: Center for Learning and Memory, The University of Texas at Austin, 1 University Station Stop C7000, Austin, TX 78712-0805, USA. Email: nmorton@uwm.edu

Encoding an event that overlaps with a previous experience may involve reactivating an existing memory and integrating it with new information or suppressing the existing memory to promote formation of a distinct, new representation. We used fMRI during overlapping event encoding to track reactivation and suppression of individual, related memories. We further used a model of semantic knowledge based on Wikipedia to quantify both reactivation of semantic knowledge related to a previous event and formation of integrated memories containing semantic features of both events. Representational similarity analysis revealed that reactivation of semantic knowledge related to a prior event in posterior medial prefrontal cortex (pmPFC) supported memory integration during new learning. Moreover, anterior hippocampus (aHPC) formed integrated representations combining the semantic features of overlapping events. We further found evidence that aHPC integration may be modulated on a trial-by-trial basis by interactions between ventrolateral PFC and anterior mPFC, with suppression of item-specific memory representations in anterior mPFC inhibiting hippocampal integration. These results suggest that PFC-mediated control processes determine the availability of specific relevant memories during new learning, thus impacting hippocampal memory integration.

Key words: RSA; prefrontal cortex; hippocampus; fMRI; episodic memory.

Introduction

The ability to form connections between related, but distinct, events is thought to be critical for a number of cognitive abilities such as reasoning (Schlichting and Preston 2015), spatial navigation (Tolman 1948), and semantic learning (Landauer and Dumais 1997; Anderson and McCulloch 1999; Rao and Howard 2008). Evidence from neuroimaging studies suggests that exposure to events that overlap with prior experience triggers reactivation of related memories (Zeithamova et al. 2012a; Molitor et al. 2021). Reactivated memories may then be updated with new information to create an integrated memory trace that contains information about both experiences (Zeithamova et al. 2012b; Schlichting and Preston 2015). By combining information from both experiences, integrated memory traces may form a cognitive map encoding relationships among the different experiences, including relationships that have not been directly observed (Morton et al. 2017, 2020).

Recent evidence also suggests, however, that memory reactivation does not always lead to integration and may trigger an active differentiation process that reduces interference between memories (Poppenk and Norman 2014; Ritvo et al. 2019; Molitor et al. 2021). During active differentiation, reactivated memories are thought to be pruned to remove features from the representation that are shared among competing events, thus reducing the potential for interference (Norman et al. 2007). During encoding of events that overlap with existing memories, active differentiation may therefore suppress reactivation of elements from competing memories (Wimber et al. 2015; Schlichting et al. 2022).

Both integration and differentiation have been observed within hippocampus (HPC) (Schlichting et al. 2015; Molitor et al. 2021). The degree to which component memories are represented more similarly in HPC predicts how likely participants will be able to infer connections between memories (Molitor et al. 2021). Prefrontal cortex may influence how overlapping memories are encoded, as it is thought to bias HPC to either retrieve or suppress memories related to current experience (Eichenbaum 2017). Recent empirical work suggests that medial prefrontal cortex (mPFC) interacts with HPC to select relevant memories for retrieval based on the current context (Varga et al. In press; Navawongse and Eichenbaum 2013; Place et al. 2016). Together with HPC, mPFC has been proposed to mediate integration of new content with reactivated memories, leading to the formation of interconnected representations that represent commonalities across events (Varga et al. In press; van Kesteren et al. 2010b; Tse et al. 2011; Zeithamova et al. 2012a; Schlichting et al. 2015; Schlichting and Preston 2016; Wikenheiser and Schoenbaum 2016). However, whereas mPFC has often been studied as a single functional region, a recent study suggests that there may be distinct functional subregions within mPFC. Although posterior mPFC (pmPFC) forms integrated representations of related

Received: May 8, 2019. **Revised:** April 27, 2023. **Accepted:** April 28, 2023

memories, anterior mPFC (amPFC) instead forms differentiated representations (Schlichting et al. 2015), suggesting that these subregions may have distinct roles in representing memory content. In contrast to pmPFC, ventrolateral prefrontal cortex (vlPFC) is thought to resolve interference between competing memories by suppressing prior experience during new learning (Kuhl et al. 2012a; Preston and Eichenbaum 2013; Wimber et al. 2015); however, the mechanism by which vlPFC inhibits specific memories remains unclear. Furthermore, although research has demonstrated that both mPFC and vlPFC influence memory encoding, little is known about how these control regions interact during encoding to determine whether related events ultimately become integrated or differentiated in HPC.

During new learning, PFC control processes are thought to impact HPC representations by inhibiting memory retrieval in HPC (Levy and Anderson 2012) or selecting specific memories to be reactivated in HPC (Navawongse and Eichenbaum 2013; Rajasethupathy et al. 2015). Prefrontal control of HPC retrieval processes therefore may modulate what information is encoded with new memories. During encoding of an event that overlaps with an existing related memory, the related memory may either be reactivated, providing an opportunity to combine information from both events into an integrated memory, or suppressed, so that a differentiated memory is formed and interference between memories is minimized (Fig. 1A–C) (Richter et al. 2016; Schlichting et al. 2022). These opposing representational strategies are thought to be supported by distinct subregions along the long axis of the HPC (Poppenk et al. 2013; Morton et al. 2017; Brunec et al. 2018), with posterior HPC (pHPC) forming differentiated representations of related events, whereas anterior HPC (aHPC) integrates information across episodes (Fig. 1D; Collin et al. 2015; Schlichting et al. 2015).

To determine how PFC control mechanisms mediate the reactivation or suppression of memories to impact memory integration in aHPC, we used a version of the associative inference paradigm (Figs. 1A and 2). We trained participants to learn an initial set of pairs (AB); each pair consisted of a famous person or landmark and a common object. After participants learned each pair, fMRI data were collected as they learned a set of object pairs. The object pairs included overlapping pairs (BC), each of which shared one object with one of the initial (AB) pairs, and non-overlapping pairs (XY), which included two new objects. Participants were not instructed about the overlap among pairs and were simply instructed to learn the new object pairs. Finally, participants were tested on their ability to infer indirect associations between items that were never seen together directly (AC), but that shared a common associate (B). Here, we focus on how brain activity during encoding predicted later performance on the inference test. We hypothesized that reactivation of the previous memory would facilitate memory integration (Fig. 1B), whereas suppression of the related memory would prevent integration, causing the initial and overlapping pairs to be stored as differentiated memories (Fig. 1C). We predicted that successful memory integration during overlapping pair (BC) encoding would facilitate performance on the inference (AC) test (Zeithamova et al. 2012a).

Although previous studies have examined how reactivation of related memories can influence new learning, many have used coarse measures to track the processes involved, mainly focusing on reactivation of category-level information of related experiences (Kuhl et al. 2012b; Zeithamova et al. 2012a; Molitor et al. 2021). Although memory reactivation may involve reinstatement of activity patterns that are specific to the category of remembered stimuli, reactivation of category-specific activity patterns

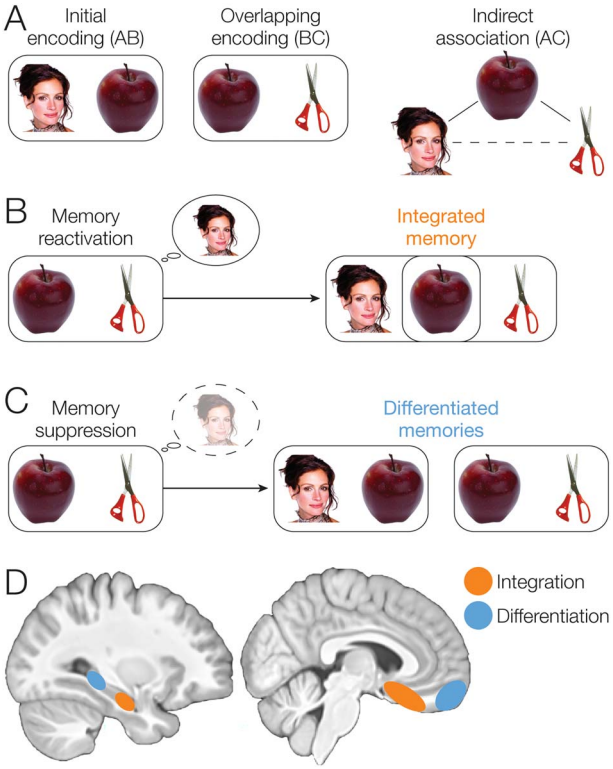

Fig. 1. Associative inference task overview and predictions. (A) After participants learn a set of initial (AB) pairs, followed by overlapping (BC) pairs, participants are given a surprise inference test to measure their knowledge of indirect (AC) associations. (B) During overlapping encoding, the initial memory may become reactivated, supporting formation of an integrated memory containing features of both pairs. Integrated memories will then facilitate later retrieval of the indirect association between A and C items. (C) In contrast, the initial memory may be suppressed during overlapping encoding, causing the memories of the two pairs to be differentiated. (D) We predicted that aHPC and pmPFC would promote integration of related memories, whereas pHPC and amPFC would promote formation of differentiated memories.

does not necessarily mean that a specific memory has been reactivated. To obtain a more specific index of memory reactivation, we used pattern similarity analysis to test for reactivation of perceptual and semantic information related to specific items in memory (Fig. 3). Before the memory task, we measured patterns of brain activity during presentation of each initial (A) item. We created a perceptual template for each item based on the pattern of activity elicited by perception of each item (Fig. 3A). We then used these templates to test for reactivation or suppression of item-specific activity during encoding of overlapping associations (Wimber et al. 2015; Mack and Preston 2016). This approach allowed us to examine how reactivation and suppression in prefrontal cortex and medial temporal lobe relate to subsequent ability to infer associations between initial (A) items and indirectly related (C) items.

The pattern similarity analysis allowed us to track reinstatement of the perceptual templates of previously experienced event elements (A items) during overlapping pairs (BC). However, successful retrieval of a memory may also involve activation of semantic information (i.e., general knowledge about an item that is not specific to a single episode) (Tulving 1972; Morton et al. 2021) related to its component elements. Reactivation of existing semantic knowledge associated with previous experiences may facilitate new learning (Van Kesteren et al. 2012; Gilboa and Marlatte 2017; Liu et al. 2017). To detect reactivation of semantic

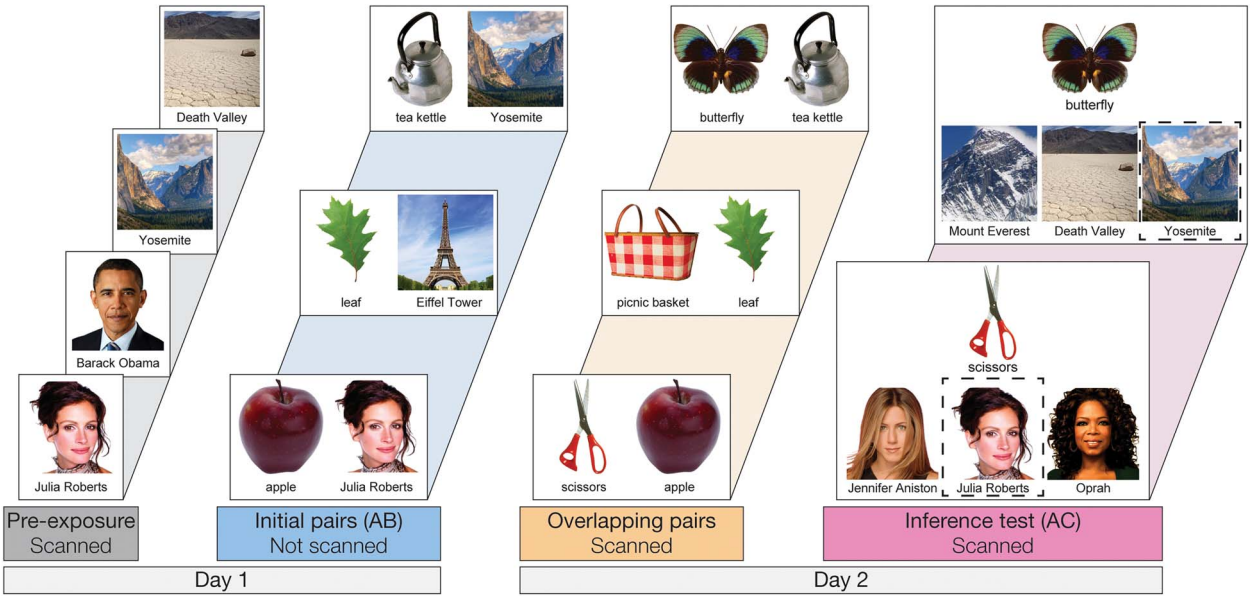

Fig. 2. On day 1, participants were first exposed to all famous person and place (A) items during scanning. They then learned a set of initial (AB) pairs associating each person and place to an object, over four study-test cycles. On day 2, participants learned a set of overlapping (BC) and non-overlapping (XY; not shown) object pairs. Finally, participants completed a surprise inference test of indirect (AC) associations. Outside the scanner, participants were tested on the object-object associations (BC and XY), and then tested on the initial (AB) pairs again.

features of individual A items during overlapping pair encoding (BC), we leveraged a recently developed method, wiki2vec, for quantifying the semantic similarity of well-known stimuli (Morton et al. 2021) (Fig. 3B). This method uses natural language processing of the Wikipedia article corresponding to a given item, representing the content of the article as a high-dimensional vector. We used wiki2vec to derive vectors for each of the items used in the study, including 60 famous people, 60 famous places, and 360 common objects. We then used the similarity between the vectors corresponding to a given pair of items as an index of their semantic similarity (Fig. 7A; Supplementary Figs. S2–S4). Prior work has shown that the wiki2vec model successfully predicts human participant ratings of conceptual similarity and correlates with multiple veridical features such as the occupation, age, and gender of famous people and the category and geographical location of famous locations (Morton et al. 2021). To track the reactivation of semantic knowledge related to initial pair memories during overlapping encoding, we created a model-based representational dissimilarity matrix (RDM; Kriegeskorte et al. 2008a) quantifying the pattern of trial dissimilarity that would be predicted for a representation of reactivated A item information. We then used the semantic reactivation model to search for brain regions that showed this same dissimilarity pattern, reflecting reactivation of A item semantic features during overlapping BC events.

In addition to measuring reactivation of semantic features of individual A items, the semantic vector model also allowed us to measure neural activity related to memory integration. We used the item semantic vectors to estimate the representational dissimilarity structure that would be expected if a given brain region integrated information from the indirectly related items (A and C) on each overlapping encoding trial (Fig. 3B). This integration model then allowed us to identify brain regions involved in integrating information from the two distinct episodes at the semantic level.

Using our measures of perceptual item reactivation, semantic feature reactivation, and semantic feature integration (Fig. 3), we tested how memory reactivation and suppression during encoding influence memory integration. We hypothesized that PFC memory control processes would modulate the availability of semantic and perceptual features of the initial (AB) memories during encoding of the overlapping (BC) pairs. We predicted that pmPFC would reactivate information related to the initial A items, facilitating formation of integrated memories in aHPC with information about both the initial and overlapping events. In contrast, we predicted that vlPFC and amPFC would inhibit reactivation of A item memories, preventing memory integration in aHPC (Fig. 1D). Finally, we hypothesized that memory integration during overlapping encoding would facilitate performance on the inference (AC) test. Therefore, we predicted that A item reactivation and semantic integration in HPC and mPFC during overlapping encoding would predict better performance on the later inference test, whereas A item suppression during overlapping encoding would predict worse performance.

Materials and methods

Experimental model and subject details

Participants

Thirty-five participants from the University of Texas at Austin area participated in the study (19 female, mean age: 22.3 years, range: 18–30). Participants provided informed consent, and the experimental procedure was approved by the University of Texas at Austin IRB. Three participants were excluded due to excessive motion, one due to handedness concerns, and one due to performance on the inference task that was not significantly above chance (binomial test; $p > 0.05$). This resulted in 30 included participants (16 female, mean age: 22.3 years, range: 18–30, right-handed).

Method details

Stimuli

Stimuli included color photographs of 60 famous faces (30 female, 30 male), 60 famous places (30 manmade, 30 natural), and 360

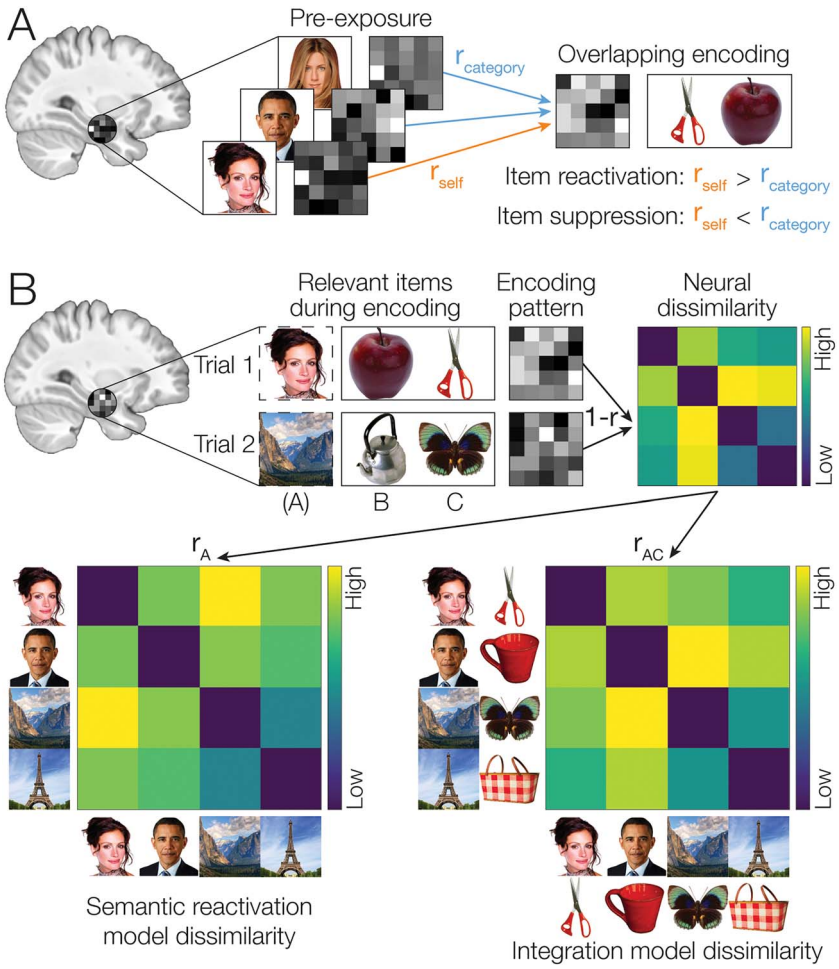

Fig. 3. (A) To detect reactivation of item-specific information during encoding of overlapping pairs, patterns of activity during encoding were correlated with perceptual item templates measured during the pre-exposure phase. We used a searchlight analysis to locate brain areas showing either reactivation or suppression of overlapping memories. When self-correlations (correlation with the specific item associated with the current pair) were greater than within-category correlation, we took that as evidence of reactivation of item-specific activity. Conversely, if self-correlations were less than within-category correlation, then this would be evidence that the overlapping memory was selectively suppressed. (B) We used RSA to detect activation of semantic information related to different items during overlapping encoding. Activity patterns during encoding may represent information related to either of the two presented items (B, C), to the related item (A), or to combinations of items (e.g. AC). To determine what information was represented by a given brain area during encoding, we calculated an RDM comparing patterns observed on different trials. We searched for brain areas that represented semantic features of reactivated (A) items by comparing neural dissimilarity to a model of semantic reactivation where only the A item was represented on each trial. We also searched for areas demonstrating memory integration by comparing neural dissimilarity to an integration model that combined features from the indirectly related (A and C) items.

common objects. All stimuli were sized to 300 × 300 pixels. In all experiment phases (except the localizer, which used different stimuli), pictures were presented with the stimulus name just below them. Famous stimuli were selected to maximize both familiarity and visual and semantic distinctiveness. Familiarity of the individual stimuli was based on mean familiarity ratings from a recent study with 20 participants age 18–30 years (Morton and Polyn 2017), who each rated 256 face and 256 place stimuli on a four-point scale.

Experimental design

Participants completed two fMRI sessions, separated by approximately 24 hours (Fig. 2). Stimuli were presented using Psychophysics Toolbox 3.0.11 (Brainard 1997) (Table 1). On the first day, participants were scanned during localizer and pre-exposure tasks designed to measure brain activity patterns related to stimulus category and individual stimuli. Outside of the scanner, they then learned 120 initial face-object and scene-object (AB) pairs over four study-test cycles. On the second day, participants learned 120 overlapping (BC) and 60 non-overlapping (XY) object–object pairs during scanning, followed by a surprise inference (AC) test. Participants were then tested on the directly learned BC and XY object pairs. Finally, participants completed a final test of the AB pairs. A previous study used the localizer and pre-exposure tasks to examine neural representations of semantic knowledge (Morton et al. 2021); here, we focus on the associative learning tasks.

Localizer

Participants were scanned during a task designed to measure brain activity related to faces, scenes, objects, and rest. Participants were presented with blocks of color photographs of unfamiliar faces (36 female, 36 male), unfamiliar scenes (36 man-made, 36 natural), and common objects (72 objects, distinct from those presented in the memory task). Each of four scanning runs included six 20-s blocks (one each of female faces, male faces,

Table 1. Key resources used in the study.

Resource	Source	Identifier
Deposited Data		
Wikipedia	Wikimedia Foundation	https://dumps.wikimedia.org
word2vec	Google (Mikolov et al. 2013)	https://code.google.com/archive/p/word2vec/
Software and Algorithms		
MATLAB 2012B	MathWorks	https://www.mathworks.com
PsychToolbox 3.0.11	Brainard 1997	http://psychtoolbox.org
FSL 5.0.9	FMRIB (Jenkinson et al. 2012)	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
ANTs 2.1.0	Avants et al. 2008	https://stnava.github.io/ANTs/
AFNI 16.1.20	NIMH (Cox 1996)	https://afni.nimh.nih.gov
PyMVPA 2.6.0	Hanke et al. 2009	http://www.pymvpa.org
Wikipedia Extractor 2.6	Giuseppe Attardi	https://github.com/attardi/wikiextractor
NLTK 3.2.3	Bird et al. 2009	http://www.nltk.org

manmade scenes, and natural scenes, and two object blocks), with 18 s of rest at the beginning and end. During each block, participants viewed 10 stimuli (1.6-s duration, 0.4-s inter-stimulus interval) while performing a one-back task. Participants indicated via button press whether each stimulus was new or a repeat (each block contained one repeat).

Pre-exposure

Participants were scanned during presentation of all 120 famous face and scene stimuli. There were six runs of the viewing task; each run included 40 stimuli (10 from each subcategory of female, male, natural, and manmade) that were each presented twice. Presentation order was randomized within each run. Each stimulus was presented in two randomly selected runs, for a total of four presentations per stimulus. To ensure sustained attention, participants performed a detection task unrelated to the presented pictures, indicating whether a circle that appeared in the middle of each stimulus was yellow or cyan. Each stimulus was presented for 2 s; the colored circle was presented at 0.25–0.75 s after stimulus onset. The inter-stimulus interval was random, ranging from 2 to 6 s, with intervals sampled according to an exponential distribution (4:2:1 ratio of 2, 4, and 6 s, respectively). Participants were asked to respond via button press while the stimulus was onscreen. Accuracy on the detection task was near ceiling (mean: 0.987).

Initial learning

Outside of the scanner, participants were presented with 60 face-object pairs and 60 scene-object initial (AB) pairs. They learned the initial pairs over four study-test repetitions, with a different random order of pairs each time. Participants were instructed to learn each pair by creating a story or mental image. On each repetition, participants were presented with each pair for 3.5 s, with a 0.5-s inter-stimulus interval. The face or scene was always presented on the right, and the object was always presented on the left. After each study phase, participants were tested on each pair with a three-alternative forced-choice test. Participants were shown a B item on the top of the screen, with three potential A items on the bottom, and given 3.5 s to respond via button press. The three choices were all drawn from the same subcategory (female, male, manmade, or natural). After 3.5 s, participants were again presented with the correct pair for 1 s. There was an inter-stimulus interval of 0.5 s between each test trial.

Overlapping and non-overlapping pair learning

Participants were scanned during learning of a set of 180 object-object pairs. There were 120 overlapping (BC) pairs, which shared an object with one of the AB pairs learned on the first day, and 60 non-overlapping pairs that included two new objects. For overlapping pairs, the C object was always presented on the left. Participants were instructed to learn each pair by creating a story or mental image and were not informed about the overlap between some of the pairs. Unlike the initial learning, participants were only shown each pair once. On each trial, an object pair was presented for 3.5 s, with an 8.5 s inter-stimulus interval. Pair learning trials were spread over six scanning runs, with 30 trials in each run (20 BC pairs, with 5 pairs associated with each of the four sub-categories of A items, and 10 XY pairs).

Inference test

After the overlapping learning phase, participants were told that some of the pairs they had just learned shared an object with the pairs they had learned the previous day. They were then instructed on the inference task, which required inferring an indirect relationship between a C item (an object) and an A (face or scene) item, based on their shared association with a B item (also an object). Participants completed a practice trial with an example set of pairs to ensure understanding. During scanning, participants were tested on inference judgments with a three-alternative forced-choice test. Here, we focused on control processes operating during encoding; analysis of scan data from the test phase will be reported elsewhere. Participants were shown a C item at the top of the screen, with three potential A items on the bottom, all drawn from the same subcategory. Each test display was shown on the screen for 5.5 s, and participants were instructed to respond via button press during this period. Trials were separated by an 8.5-s inter-stimulus interval. The 120 test trials were spread over five scanning runs, with 24 trials in each run (six from each of the four sub-categories of A items).

Direct pair tests

Outside the scanner, participants were tested on the overlapping (BC) and non-overlapping (XY) pairs. The test procedure was similar to the inference test, but with an inter-stimulus interval of 0.5 s. C/Y items were used as cues, and participants were instructed to indicate the correct B/X item. BC and XY tests were randomly intermixed. Finally, participants were again tested on the initial (AB) pairs, to assess whether any forgetting of these pairs had occurred between days. The test procedure was the same as the tests of the overlapping and non-overlapping pairs, with a 5.5-s test display and 0.5-s inter-stimulus interval.

Image acquisition

Imaging data were acquired on a 3.0-T Siemens Skyra MRI at the University of Texas at Austin Biomedical Imaging Center (RRID:SCR_021898). A T1-weighted 3-D MPRAGE volume (TR: 1.9 s, TE: 2.43 ms, flip angle: 9°, FOV: 256 mm, 192 slices, 1mm^3 voxels) was acquired on each day for coregistration and parcellation. Two oblique coronal TSE T2-weighted volumes were acquired perpendicular to the main axis of the HPC (TR: 13.15 s, TE: 82 ms, flip angle: 150°, 384 × 384 matrix, 60 slices, 0.4 × 0.4 mm^2 in-plane resolution, 1.5 mm through-plane resolution) to facilitate localization of activity in the medial temporal lobe. High-resolution whole-brain functional images were acquired using a T2*-weighted multiband-accelerated EPI pulse sequence developed by the Center for Magnetic Resonance Research at the University of Minnesota (TR: 2 s, TE: 31 ms, flip angle: 73°, FOV: 220 mm, 75 slices, matrix: 128 × 128, 1.7 mm^3 voxels, multiband factor: 3, GRAPPA factor: 2, phase partial Fourier: 7/8). We acquired a field map on each day (TR: 589 ms, TE: 5 and 7.46 ms, flip angle: 5°, matrix: 128 × 128, 60 slices, 1.5 × 1.5 × 2 mm^3 voxels) to allow for correction of magnetic field distortions.

Image processing

Data were preprocessed with an analysis pipeline (Morton 2022) using tools from FMRIB Software Library (FSL) 5.0.9 (Jenkinson et al. 2012) and Advanced Normalization Tools (ANTs) 2.1.0 (Avants et al. 2008) (Table 1). T1 scans were corrected for bias field using *N4BiasFieldCorrection* (Tustison et al. 2010), co-registered using *antsRegistration* and *antsApplyTransforms*, scaled using a single multiplicative value, and averaged. FreeSurfer 5.3.0 (Dale et al. 1999) was used to automatically segment cortical and subcortical areas based on the average T1 volume. A brain mask was created by dilating the reconstructed cortex estimated from FreeSurfer and intersecting with the automatically generated mask from FreeSurfer. The brain mask was used to remove non-brain tissue from the T1 images. The *buildtemplateparallel* program from ANTs was used to create a group-level T1 template from brain-extracted MPRAGE scans from the individual participants (rigid initial target, 3 affine iterations, 10 nonlinear iterations). The resulting template was registered to the FSL 1-mm MNI template brain using affine registration implemented in ANTs, to obtain a final template.

Functional scans were corrected for motion through alignment to the center volume using MCFLIRT, using spline interpolation. Functional scans were unwarped using a modified version of *epi_reg* from FSL, which was adapted to use boundary-based registration implemented in FSL, followed by ANTs to refine registration between functional scans and the T1. Registration refinement was done over 20 iterations, with an unwarped functional target image updated on each iteration (using FSL's *fugue* tool) according to the latest registration. The brain mask derived from FreeSurfer was projected into native functional space and used to remove non-brain tissue from the unwarped scans. Average brain-extracted unwarped functional scans were registered to a single reference scan (the first study-phase scan on the second day) using ANTs. After calculating all transformations, motion correction, unwarping, and registration to the reference functional scan were applied to the raw functional data using B-spline interpolation, in two steps to minimize interpolation. The bias field for the average functional image in each run was estimated for each scan using N4 bias correction implemented in ANTs and removed by dividing the timeseries by a single estimated bias field image. Functional time series were high-pass filtered (128-s FWHM) and smoothed at 4-mm FWHM using FSL's SUSAN tool (Smith and Brady 1997).

Regions of interest

Based on prior work in a task similar to ours (Schlichting et al. 2015), we focused our analysis on anatomical regions of interest (ROIs) in mPFC, inferior frontal gyrus/insula (IFG), HPC, perirhinal cortex (PRC), and parahippocampal cortex (PHC). To create group-level ROIs, FreeSurfer segmentations of individual subjects were projected to group template space. An IFG/insula mask was created for each subject in template space, including the pars opercularis, pars orbitalis, pars triangularis, and insula regions. Subject masks were averaged to create a probability map for IFG/insula. This map was thresholded at 0.5 (to include voxels labeled as IFG/insula for at least half of subjects) to create an initial mask. This mask was then dilated by smoothing with a sigma of 1 mm and thresholding at 0.01. We used a previously defined mPFC mask, which was drawn by hand on a 1-mm MNI template brain image based on a cytoarchitectonic atlas (Öngür et al. 2003; Price and Drevets 2010; Schlichting et al. 2015). We used ANTs to transform this mask to our template space using SyN registration (Avants et al. 2010). We also used previously defined medial temporal lobe regions that were drawn manually on a 1-mm MNI template brain (Liang and Preston 2017), which we transformed to our template space using SyN registration.

Quantification and statistical analysis

Estimation of item-level activation patterns

Patterns of activation associated with individual A items (during the pre-exposure phase) and individual pair encoding trials were estimated under the assumptions of the general linear model (GLM) using a least squares-separate (LS-S) approach (Mumford et al. 2012). Parameter estimate images were calculated for each of the 40 items presented in each pre-exposure scan. Each 2-s item presentation was convolved with a canonical (double gamma) hemodynamic response function from FSL. Each item was estimated using a separate model; the two presentations of an item within each scan were modeled as a single regressor, with presentations of other items modeled as a separate regressor. Additional regressors of no interest included six motion parameters and their temporal derivatives, framewise displacement, and DVARS (Smyser et al. 2010). Additional regressors were created to remove timepoints with excessive motion (defined as greater than 0.5 mm of framewise displacement and greater than 0.5% change in BOLD signal for DVARS), as well as one timepoint before and two timepoints after each high-motion frame. High-pass temporal filtering (128-s FWHM) was applied to the regressors. Individual stimulus activity parameter estimates were calculated using a GLM implemented in custom Python routines (Morton and Zippi 2022). Each voxel's activity was z-scored across stimuli within run. Finally, normalized activity patterns for each item were averaged across the two scans in which that item appeared, resulting in 120 estimated item activity patterns. A similar method was used to estimate activity related to each individual trial during the scanned overlapping encoding phase. For each of the six study runs, activation related to each of the 30 trials was estimated using a GLM using LS-S. Activation for each voxel was z-scored across stimuli within run.

Searchlight analyses

All searchlights were conducted within each subject's native functional space with a radius of 3 voxels, using PyMVPA 2.6.5 (Hanke et al. 2009) and custom Python routines (Morton and Zippi 2022). To assess significance, we used a permutation testing method (Stelzer et al. 2013). For each searchlight, a null statistic

based on 100 sets of permuted indices was estimated for each subject (details of the permutation are specified for each analysis below). Actual and permuted statistics were transformed to group space using ANTs. To estimate a group-level null distribution, we sampled from the subject-level permutations 100,000 times, calculating an average across all subjects on each permutation. We then compared the actual average statistic to this null distribution for each voxel, and thresholded to include only voxels with $P < 0.01$ (one-sided test). For cluster correction, we estimated spatial smoothness of the data based on residuals from the trial-level GLM of the study phase (in this case, we used a single LS-A model; Mumford et al. 2012). Residuals from the native-space GLM models for each run were transformed to template space using ANTs. For a given ROI, smoothness was estimated using the *3dFWHMx* tool from AFNI 16.1.0 (Cox 1996) with the autocorrelation function method, and averaged across all volumes, runs, and subjects. Finally, a cluster size threshold was determined for each ROI based on this estimated smoothness, using 3dClustSim with 2,000 iterations, one-sided thresholding at $P < 0.01$, and familywise alpha of 0.05.

Measurement of item reactivation and suppression

To measure A item reactivation within a given searchlight sphere during overlapping encoding, we compared encoding patterns to the famous face and scene item patterns measured during the pre-exposure phase (Fig. 3A). For each BC encoding trial, we calculated the correlation between the BC trial pattern and the pattern for the A item corresponding to that trial. We compared these self-similarity values to a baseline of correlations between the BC trial pattern and all A items within the same category (face or scene) as the A item associated with that BC pair. Within a given searchlight sphere, we tested whether there was greater self-similarity (r_{self}) than within-category similarity (r_{within}). The difference statistic $r_{self} - r_{within}$ was calculated separately for each category and then averaged across categories. A null distribution of the $r_{self} - r_{within}$ statistic was estimated by permuting study trial item labels within category. We also tested for areas with evidence of item suppression, defined as $r_{within} > r_{self}$. We ran a separate searchlight to test for areas that showed greater A item reactivation during overlapping encoding trials for which inference was subsequently correct than for subsequently incorrect trials. For each sphere, we calculated an interaction: $(r_{self,correct} - r_{within,correct}) - (r_{self,incorrect} - r_{within,incorrect})$. The interaction statistic was calculated separately within each category and then averaged across categories. A null distribution was estimated by permuting the correct vs. incorrect labels within category.

Follow-up analyses examined clusters that showed significant reactivation or suppression at the group level. We reverse-normalized the cluster masks to each participant's native functional space. Each individual mask was then dilated by one voxel and intersected with the participant's gray matter mask defined by FreeSurfer. We then calculated item reactivation based on pattern similarity within each mask, separately for encoding trials where the subsequent inference test was correct and for subsequently incorrect inference trials. We also measured category reactivation, defined as $r_{within} - r_{between}$, where r_{within} is the average correlation between the encoding-trial pattern and the patterns observed during pre-exposure for each of the items in the same category as the A item associated with that trial (excluding the A item itself), and $r_{between}$ is the correlation between the encoding trial pattern and all items in the different category from the A item (Fig. S1).

Model of semantic similarity

Using RSA, we tested for patterns of activity related to different event elements related to the overlapping encoding trials, including the initial (A) item, the linking (B) item, and the new (C) item. To develop predictions for the representational dissimilarity of different items, we used the wiki2vec embedding method (Morton et al. 2021) to quantify the semantic similarity between all of our celebrity, famous landmark, and common object stimuli. The wiki2vec method uses text from the Wikipedia page for each item with natural language processing to generate a high-dimensional vector representation that reflects general knowledge about each item (Table 1). There are a number of advantages of using Wikipedia for this purpose. Text on Wikipedia is updated on a regular basis and has a large number of articles, and therefore provides information about virtually any famous person or place stimulus, as well as many different types of common objects. We used a dump of Wikipedia text obtained soon before the start of data collection, which should be relatively reflective of knowledge that was current at the time of the study. Finally, Wikipedia contains information about people, places, and things in a similar format, making it possible to compute a single vector space with information about all of these types of entities.

Although Wikipedia articles are highly tailored to the subject of each article, and therefore contain a relatively precise sample of words related to different items, each article contains a smaller number of words than is typically used for similarity-estimation techniques such as latent semantic analysis (Landauer and Dumais 1997). Therefore, we used another publicly available resource, Google's word2vec representations, to estimate vector representations for each item (Mikolov et al. 2013). The word2vec model is based on a large text corpus derived from Google News. We used a publicly available set of 300-dimensional representational vectors, which were created using a continuous bag-of-words algorithm applied to words and phrases from the Google News corpus.

To construct our model of semantic similarity, we first obtained a Wikipedia XML dump from three months prior to the start of the experiment, on 2015 February 19. The download was then converted to plain text using Wikipedia Extractor. We then selected a Wikipedia page for each stimulus. The page was chosen to match the sense of each item based on the picture presented with that item; for example, the page used for "bat" was the one corresponding to the animal, not the type of sports equipment. We used natural language processing to generate a set of words corresponding to the Wikipedia article for each item. Named entities, nouns, verbs, and adjectives were extracted from each of the article texts using the Python-based Natural Language Toolkit (NLTK) (Bird et al. 2009). Recognized named entities, such as "Eiffel Tower," were treated as a single unit for subsequent processing. Words were replaced with their lemma to increase consistency across words in an article and across articles. This preprocessing of each article resulted in a "bag of words" corresponding to each item; that is, only term frequency was considered for subsequent processing.

For the set of words associated with each article, we searched for a match in the word2vec corpus. We constructed a vector for each article by summing all the word2vec vectors for each word, weighted by the absolute frequency of that word in the article. The item vectors were then used to construct semantic models of different combinations of event elements during the encoding phase. Prior work has shown that the wiki2vec model successfully predicts human judgments of conceptual similarity and is sensitive to veridical features such as occupation, age,

and gender of famous people and the category and geographic location of famous places (Morton et al. 2021).

Models of semantic reactivation and integration

Given our vector-space model of individual items, we were able to construct models of dissimilarity for different aspects of learned stimuli. Each model corresponds to a distinct RDM predicting the dissimilarity between different overlapping (BC) pair encoding trials. For example, we could form a model RDM corresponding to initial (A) items by looking up the model vector corresponding to the A item associated with a given BC trial and calculating correlation distance across trials for those vectors. This model would be distinct from a model based on, for example, the B item associated with each BC trial, allowing us to distinguish between regions representing different event elements. We could also construct models corresponding to composites of multiple event elements. For example, we formed predicted dissimilarity for a region that represents information related to both parts of the overlapping pairs (i.e. the B and C item for that trial) by adding the B and C vector for each trial and calculating dissimilarity across trials based on those composite vectors.

We used our semantic model to construct two models of overlapping trial dissimilarity, reflecting reactivation (A) and integration (A and C). To detect areas showing reactivation of semantic information related to the initial memory, rather than representation of the current episode, we contrasted neural correlation with the A model, relative to correlation with the composite BC model. To detect integration of the two separate (AB and BC) episodes, we contrasted a model with attributes of the distinct parts of both episodes (A and C) with a model of the common aspect of both episodes (B). For each model, we calculated Spearman correlation between the model and neural dissimilarity within a given searchlight sphere. We did this separately for subsequently correct and incorrect items and tested whether the contrast between models was greater for correct than incorrect items. A null distribution was estimated by permuting items within the correct and incorrect bins, relative to the models of dissimilarity.

Results

Learning of initial and overlapping pairs supports cross-event inference

On the first day of the experiment, participants learned the 120 initial (AB) pairs well, with performance approaching ceiling after four study-test repetitions (Fig. 4A). After a single presentation of the overlapping (BC) pairs on the second day (approximately 24 hours after the first session), all included participants performed above chance on the inference (AC) test (mean: 78.2%, SEM: 2.7%; all participants $P < 0.009$, binomial test). Response time on correct inference test trials was 2947 ms (SEM: 90 ms; Fig. 4B).

Test accuracy was greater for both the initial (AB) and the overlapping (BC) pair tests compared with the inference (AC) tests (AB at the end of initial learning: mean 98.9%, SEM 0.2%, $t(29) = 7.66$, $p = 2.0 \times 10^{-8}$, Cohen's $d = 1.95$; BC: mean 92.7%, SEM 1.7%, $t(29) = 9.56$, $p = 1.9 \times 10^{-10}$, $d = 1.166$). These results suggest that participants sometimes learn both the AB and BC pairs of a given triad but still fail to correctly infer an association between the A and C items. Test accuracy was also greater on the non-overlapping (XY) pair tests compared with AC tests (XY: mean 86.4%, SEM 2.1%, $t(29) = 7.30$, $p = 4.9 \times 10^{-8}$, $d = 0.61$). Test accuracy was greater for the overlapping (BC) pairs than for the non-overlapping (XY) pairs ($t(29) = 5.05$, $p = 2.2 \times 10^{-5}$, Cohen's $d = 0.603$). Response times on correct BC trials were also faster than correct XY trials (BC: mean 1859 ms, SEM 67 ms; XY: mean 2,249 ms, SEM 73 ms; $t(29) = 10.2$, $p = 4.4 \times 10^{-11}$, $d = 1.02$). Finally, the final test of the initial (AB) pairs showed near-ceiling performance (mean: 99.5%, SEM 0.2%), confirming that participants did not forget the initial pairs between days.

Suppression of related memories in posterior hippocampus modulates memory reactivation in medial temporal lobe cortex

We hypothesized that memory for each overlapping (BC) pair would be modulated by whether the initial (AB) pair was reactivated or suppressed during encoding (Fig. 1). Although accuracy in retrieving BC associations was high overall (Fig. 4A), suggesting that BC associations were consistently learned, we predicted that the manner in which BC associations were stored in memory would predict performance on the inference (AC) test. Specifically, we predicted that during overlapping pair (BC) learning, reactivation of a specific face or scene (A item) from the initial (AB) pair would promote integration of the initial (AB) and overlapping (BC) pairs, leading to better performance on the inference (AC) test. Conversely, we predicted that suppression of the A item during overlapping encoding would inhibit integration of the initial and overlapping pairs, leading to worse performance on the inference test. We measured reactivation and suppression by comparing activation patterns during encoding of overlapping pairs with the perceptual templates for each item, measured during the pre-exposure phase (Fig. 3A). Using a searchlight analysis, we first isolated brain regions that showed activation patterns with a greater correlation with the A item associated with each overlapping encoding trial (r_{self}), compared with correlation with all other items from the same category as the A item ($r_{category}$; Fig. 3A). We used a second searchlight analysis to isolate brain regions showing selective suppression of the A item, relative to other items from the same category (i.e., $r_{self} < r_{category}$). We searched within bilateral a priori ROIs that included IFG, mPFC, and HPC, as these regions have previously been shown to reflect learning of item relationships in a similar associative inference task (Schlichting et al. 2015); we also included PHC and PRC, which have been shown to reactivate specific memories during retrieval (Mack and Preston 2016). Results were cluster-corrected with a threshold of $\alpha = 0.05$ within each of our ROIs.

We found evidence for item-specific reactivation during overlapping pair encoding in right PHC, left PRC, and right IFG/insula (Fig. 5A–B, Table 2), and evidence for A item suppression in right pHPC (Fig. 5C, Table 2). Because many previous studies have instead used reactivation of category-general patterns to index memory reactivation, we also examined whether activity in these regions reflected the category of the reactivated A items. PHC demonstrated significant reactivation of category activity, in addition to item-specific activity (Fig. S1). In contrast, PRC and IFG/insula showed only evidence for item-specific reactivation, with no reliable category reactivation. Although item-specific activity in pHPC was suppressed, pHPC demonstrated significant category-level reactivation. This dissociation between item-specific and category-specific activity in pHPC is consistent with prior work suggesting that activation patterns related to an item may be modulated independently from broader category-specific activation (Wimber et al. 2015) and provides further evidence that reactivation of category-specific activation patterns in a region does not necessarily imply reactivation of a specific memory.

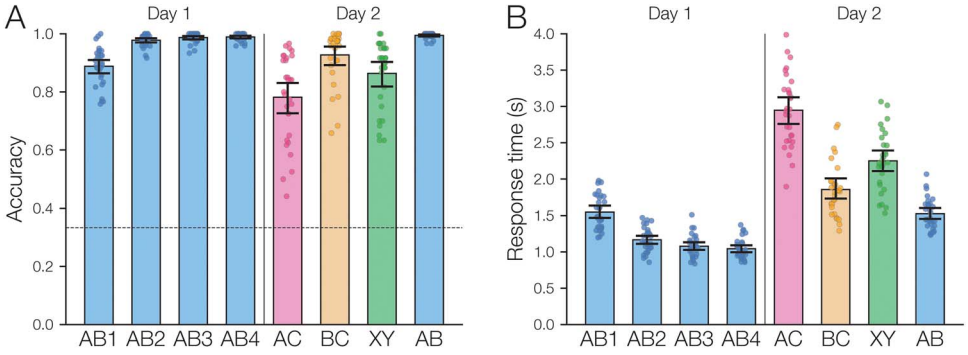

Fig. 4. (A) Accuracy on 3AFC (three-alternative forced-choice) association tests. Participants reached near-ceiling performance by the end of four study-test repetitions of the initial (AB) pairs. After learning the BC and XY pairs, participants performed above chance on tests of inference (AC), overlapping pairs (BC), and non-overlapping pairs (XY). Participants still remembered the initial (AB) pairs well the second day. The horizontal line shows chance performance ($0.3\overline{3}$). (B) Response time on association tests, for correct trials only. Error bars indicate 95% bootstrap confidence intervals. Points indicate individual participants.

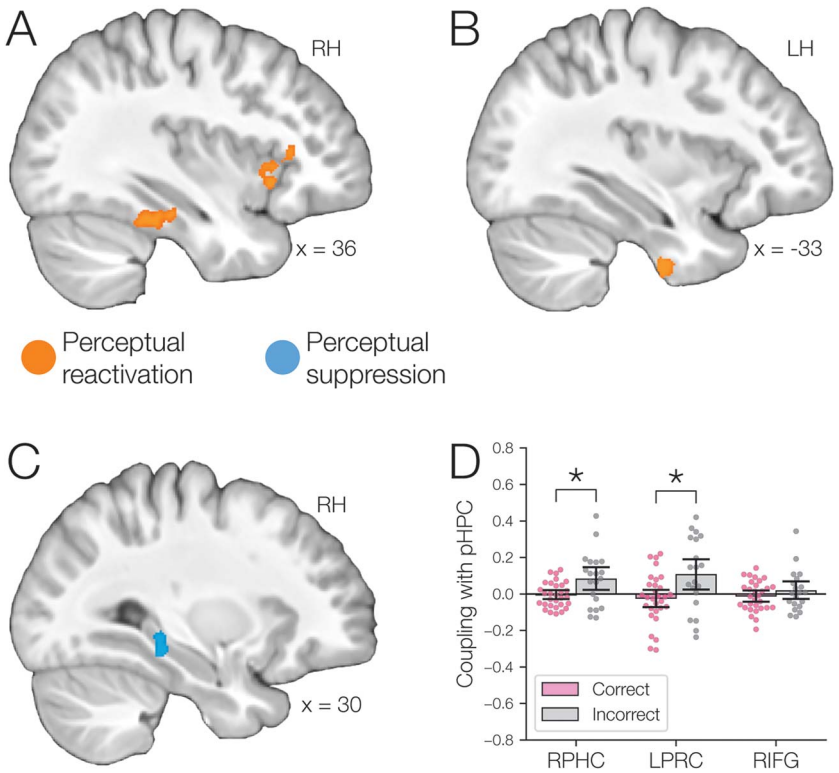

Fig. 5. (A) Right parahippocampal cortex (RPHC) and right inferior frontal gyrus/insula (RIFG) showed evidence of A item reactivation during overlapping pair encoding, based on correlation with perceptual templates derived from the pre-exposure phase. (B) Left perirhinal cortex (LPRC) also showed evidence of A item reactivation. (C) In contrast, right pHPC showed evidence of A item suppression. (D) Stronger coupling of pHPC with PHC and PRC predicted worse performance on the subsequent inference test: item activation in these regions was more strongly correlated for overlapping pair trials for which inference was subsequently incorrect relative to correct trials. Error bars indicate 95% bootstrap confidence intervals. Points indicate individual participants. *$P < 0.05$. See also Fig. S1.

Table 2. Peak MNI coordinates for reactivation/suppression searchlight clusters.

Effect	Area	Volume (mm^3)	X	Y	Z
Item reactivation	R IFG/insula	1,503	39	19	−5
	R PHC	978	38	−35	−21
	L PRC	355	−32	−12	−42
Item suppression	R pHPC	422	28	−33	−6
Item reactivation for correct inference	R aIFG	888	45	34	−14
	R amPFC	568	3	66	−17

Because HPC is thought to drive activity in cortical areas during episodic memory retrieval (Griffiths et al. 2018), we hypothesized that memory suppression in pHPC would prevent memory reactivation in cortex, making those memories less available. We used representational coupling analysis (Kriegeskorte et al. 2008a, 2008b) to examine whether the individual regions evincing A item reactivation or suppression interact during overlapping pair encoding. For each region that demonstrated A item reactivation or suppression, we calculated r_{self} for each trial. We then used robust regression implemented in statsmodels 0.13.2 (Seabold and Perktold 2010) with Tukey biweight normalization to estimate the relationship between r_{self} across trials, within each pair of regions. For each subject, the slope was calculated separately for each category (faces, scenes) and then averaged to obtain an estimate of how much A item reactivation was related across regions on a trial-by-trial basis. Here, a positive slope indicates that the amount of reactivation of the two regions correlates positively over trials, whereas negative slopes indicate a negative relationship between their reactivation strengths. We found significant positive relationships between each of the regions that demonstrated item-specific reactivation (right PHC and left PRC: mean slope 0.068, SEM 0.033, $t(29) = 2.08$, $p = 0.047$, Cohen's $d = 0.379$; right PHC and right IFG: mean slope 0.048, SEM 0.020, $t(29) = 2.36$, $p = 0.026$, $d = 0.432$; left PRC and right IFG: mean slope 0.035, SEM 0.014, $t(29) = 2.41$, $p = 0.023$, $d = 0.439$), but no overall relationship between item suppression in pHPC and the other regions (all $p > 0.05$). These results suggest that PRC, PHC, and IFG/insula are functionally coupled during encoding of overlapping associations such that they tend to reinstate item-specific representations on the same trials.

We next examined whether representational coupling between pHPC and the cortical memory reinstatement areas during overlapping encoding predicted performance on the inference test. We hypothesized that memory suppression in pHPC would inhibit memory reactivation in other areas during encoding, preventing memory integration and leading to worse inference performance. We separated overlapping encoding trials based on whether the corresponding subsequent inference decision was correct or incorrect. We then contrasted the slope (calculated separately for each category and then averaged) assessing the relationship between item-level reinstatement between each pair of regions during overlapping pair trials for correct and incorrect inferences. Ten participants were excluded from this analysis because they had fewer than five trials per category in one of the subsequent inference conditions. We found that greater coupling between pHPC and the medial temporal cortical regions, PHC and PRC, predicted lower accuracy on the inference test (Fig. 5D; pHPC and PHC: mean slope correct–incorrect −0.092, SEM 0.033, $t(19) = 2.78$, $p = 0.012$, Cohen's $d = 0.621$; pHPC and PRC: mean slope correct–incorrect −0.150, SEM 0.044, $t(19) = 3.39$, $p = 0.0031$, $d = 0.759$). We did not observe inference-related representational coupling differences between pHPC and IFG (mean slope correct–incorrect −0.029, SEM 0.032, $t(19) = 0.89$, $p = 0.382$, $d = 0.200$). These results suggest that suppression of related memories in pHPC during overlapping event encoding may inhibit memory reactivation in medial temporal lobe cortex, preventing the initial (AB) and overlapping (BC) memories from becoming integrated.

Memory reactivation and suppression in prefrontal cortex predict subsequent inference performance

We hypothesized that reactivation of specific A items during overlapping encoding would predict improved performance on the inference test, whereas suppression would predict lower inference performance. To examine whether A item reactivation on individual overlapping encoding trials predicted subsequent performance on the corresponding inference tests, we carried out a searchlight analysis within our ROIs. Within each searchlight sphere, we calculated the average similarity between the activity pattern observed during encoding of each trial and the associated A item during the pre-exposure phase, r_{self}, and the average similarity to other A items from the same category, r_{within}. We computed an item reactivation score by calculating $r_{self} - r_{within}$ for each category and averaging over categories. We tested whether item reactivation was greater during overlapping encoding trials for which the corresponding inference was correct than for trials leading to incorrect inferences.

We observed a significant interaction between A item reactivation during overlapping encoding and subsequent inference in anterior IFG (aIFG; Fig. 6A, Table 2) and amPFC (Fig. 6B, Table 2). During overlapping pair trials for which the inference was subsequently correct, we observed significant A item reactivation in aIFG (Fig. 6C; mean difference between self and category similarity: 0.017, SEM: 0.005, $t(29) = 3.67$, $p = 0.00097$, Cohen's $d = 0.670$), whereas significant A item suppression was observed for overlapping trials for which inference was incorrect (mean: −0.037, SEM: 0.016, $t(27) = 2.28$, $p = 0.031$, $d = 0.431$; two participants were excluded because there were not enough incorrect events to estimate item reactivation/suppression for both categories). Item-specific activation during overlapping encoding also varied in amPFC as a function of subsequent inference. In amPFC, we observed significant A item suppression on overlapping trials for which inference was subsequently incorrect (mean: −0.056, SEM: 0.021, $t(27) = 2.68$, $p = 0.013$, $d = 0.506$). Activation in amPFC uniquely reflected suppression, as item-level reactivation in amPFC was not observed for subsequently correct inferences (mean: −0.005, SEM: 0.005, $t(29) = 1.04$, $p = 0.31$, $d = 0.189$). These results suggest that aIFG is involved with both memory reactivation and memory suppression, on different trials, whereas amPFC is primarily involved in memory suppression.

Anterior IFG and amPFC have been proposed to be components of a control network that guides memory retrieval in HPC (Brown et al. 2016; Badre and Nee 2018), suggesting that these regions may work together to control memory retrieval during overlapping encoding. Because of the proposed functional connection between aIFG and amPFC, we hypothesized that the amount of item reactivation/suppression on individual trials would be correlated across these regions. We used representational coupling analysis to determine how reactivation and suppression of prior memories in these PFC regions relate on individual overlapping encoding trials. We used robust regression to estimate the slope relating r_{self} in aIFG with r_{self} in amPFC on individual trials. The slope was estimated separately for each category and then averaged across categories. The amount of A item reactivation/suppression on each overlapping pair trial (measured using r_{self}) was correlated between aIFG and amPFC (mean slope: 0.142, SEM: 0.026, $t(29) = 5.50$, $p = 6 \times 10^{-6}$, $d = 1.005$). The representational coupling among these regions did not vary between subsequently correct and incorrect trials (correct trial slope mean: 0.129, SEM 0.025; incorrect trial slope mean: 0.160, SEM: 0.065; $t(19) = 0.21$, $p = 0.84$, $d = 0.047$). These results suggest that these prefrontal regions are engaged both on encoding trials during which memory integration occurs and on trials where a differentiated memory is formed. By either reactivating or suppressing the memory of the related A item, aIFG

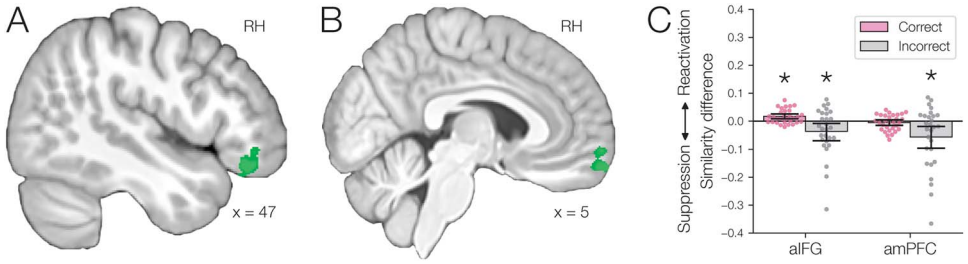

Fig. 6. (A) Anterior IFG showed greater A item reactivation ($r_{self} > r_{within}$) during overlapping pair trials with subsequently correct inference compared with trials with subsequently incorrect inference. (B) Anterior mPFC also showed a significant interaction between A item reactivation and subsequent inference. (C) In aIFG, there was both significant A item reactivation for subsequently correct inference trials and significant item suppression ($r_{within} > r_{self}$) for incorrect trials. In contrast, amPFC only showed evidence of A item suppression during overlapping trials for which inferences were later incorrect, with no item reactivation. Error bars indicate 95% bootstrap confidence intervals. Points indicate individual participants. *$P < 0.05$.

and amPFC may modulate whether encoding is biased toward memory integration or formation of differentiated memories.

mPFC represents semantic features of related memories during encoding

Although the analysis of item-specific memory reactivation described in the above sections suggests that reactivation of perceptual templates of overlapping events affects encoding of new memories, memory encoding is also thought to be affected by existing semantic knowledge (van Kesteren et al. 2010a, 2010b; Liu et al. 2017). Existing knowledge may provide a framework to speed learning of new associations (Tse et al. 2007, 2011; McKenzie et al. 2014). We hypothesized that semantic features associated with the initial pairs would, if reinstated during encoding of overlapping pairs, facilitate integration of the initial and overlapping memories. Previous studies suggest that mPFC facilitates encoding of events associated with prior knowledge (Zeithamova et al. 2012a; Schlichting and Preston 2015; Liu et al. 2017; Masís-Obando et al. 2022) and may help support subsequent inference of indirect associations between events (Spalding et al. 2018); however, the relationship between prior knowledge activation in mPFC during encoding and subsequent inference has not been observed directly.

To track the reactivation of semantic features associated with A items during overlapping pair encoding, we used representational similarity analysis (RSA; Fig. 3B). Using data from Wikipedia and Google News, we used the wiki2vec method (Zippi et al. 2020; Morton et al. 2021) to construct a vector-space model of all celebrities, landmarks, and common objects in our stimulus set (see Materials and Methods for details). This model maps each item in the experiment to a point in a high-dimensional space reflecting the general knowledge associated with each item and allows us to estimate the semantic similarity of any pair of items (Fig. 7A). Composite vectors of multiple items can also be constructed by summing the different component vectors. For example, to construct a model of the BC pair scissors–apple, we added the vectors corresponding to scissors and apple from the model to form a composite. Vectors corresponding to different individual trials can then be compared using correlation distance to estimate a full RDM (Kriegeskorte et al. 2008a). This RDM can then be compared with dissimilarity across trials observed in a given brain region (Fig. 3B).

If a brain region, such as mPFC, represents semantic features related to the reactivated A item from the initial episode, the RDM for that region comparing different overlapping pair trials should correlate with the RDM for the A item model (estimated based on the items presented to each individual subject). We contrasted this semantic reactivation model with a baseline model, B + C, representing just semantic features related to the overlapping pair items being presented on the screen during each trial. We used a searchlight analysis to identify areas within our ROIs where semantic reactivation of A items during overlapping encoding predicted inference performance, testing for areas that showed a higher correlation with the A model than the B + C model, weighted by inference performance (i.e., having a greater difference for correct than incorrect trials).

We found evidence for a selective representation of semantic features related to the reactivated A item during overlapping encoding, which predicted subsequent inference performance, in pmPFC (Fig. 7B, Table 3). Although pmPFC showed evidence of reactivation of semantic details, there was no significant perceptual template reactivation in that region (mean $r_{self} - r_{within} = -0.005$, SEM: 0.005, $t(29) = 1.13$, $p = 0.27$, Cohen's $d = 0.206$). These findings suggest that pmPFC is more involved in representing conceptual features of related memories, rather than reactivation of item-specific perceptual activity, during overlapping encoding. These findings are consistent with previous studies implicating pmPFC in memory integration (Zeithamova et al. 2012a; Schlichting et al. 2015) and provide evidence that pmPFC facilitates formation of integrated memories by activating relevant prior knowledge during new learning.

Integration of semantic features in aHPC and PRC predicts subsequent inference

Our analysis of perceptual and semantic reactivation of related items during overlapping encoding demonstrates that a network of regions, including medial temporal lobe cortex and pmPFC, make information about related events available during overlapping encoding. We hypothesized that semantic information about the initial event could then be combined with information about the overlapping event to form an integrated memory. To test this hypothesis, we constructed an integration model by creating a composite of semantic features related to the A and C items on each overlapping pair trial and calculating the predicted trial RDM for an integrated representation (Fig. 3B). We contrasted the semantic AC integration model RDM with a baseline RDM based on the semantic features related to the B item, which we predicted would be included in any memory of the overlapping episode, regardless of whether it was stored as an integrated or differentiated memory. To identify regions where AC integration supported subsequent inference, we used a searchlight analysis to test for a higher correlation with the AC integration model

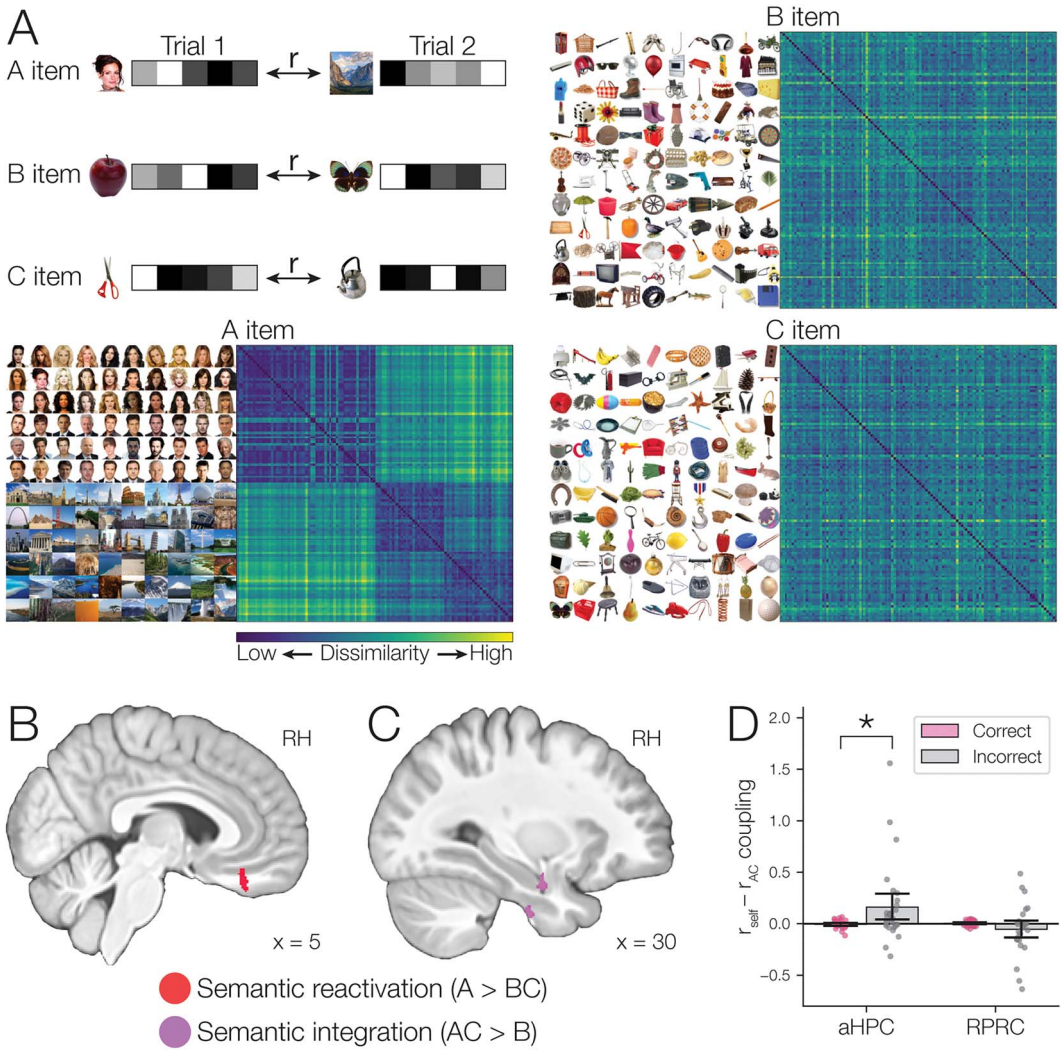

Fig. 7. (A) A model of semantic similarity was used to generate a vector representation of each item that reflects its semantic features. Each encoding trial had three semantic vectors, corresponding to different items (A, B, and C). The model allowed us to predict, for each participant, the similarity structure across trials that should be associated with semantic representations of different items. Matrices illustrate predicted dissimilarity of different encoding trials for one participant, with photographs indicating the relevant item on each trial. (B) During encoding of overlapping (BC) pairs, we tested for reactivation of semantic features related to the A item (r_A) associated with each trial, contrasted with features related to the presented items (r_{BC}). We found that pmPFC showed evidence of reactivation of semantic features of the A item associated with each trial, which selectively predicted inference performance ($r_A - r_{BC}$ greater for correct subsequent inference than incorrect subsequent inference). (C) To test for areas involved in memory integration during overlapping encoding, we contrasted an integration model representing both the indirectly related A and C items (r_{AC}) with a baseline model representing just the B item (r_B). Pattern similarity in aHPC and PRC showed evidence of greater memory integration on subsequently correct inference trials, relative to incorrect trials, suggesting that these regions are involved in integrating indirectly associated items. (D) Item activation in amPFC was correlated with integration in aHPC, only on encoding trials for which the subsequent inference judgment was incorrect. Trials with greater item suppression in amPFC were associated with a decrease in integration in aHPC. Error bars represent 95% bootstrap confidence intervals. Points indicate individual participants. *$P < 0.05$. See also Figs. S2–S5.

Table 3. Peak MNI coordinates for semantic searchlight clusters.

Effect	Area	Volume (mm^3)	X	Y	Z
Semantic reactivation (A > BC)	R pmPFC	352	4	35	−24
Semantic integration (AC > B)	R aHPC	294	31	−5	−22
	R PRC	241	30	−13	−37

compared with the baseline model, weighted by subsequent inference performance. We found evidence for integrated AC representations being formed during overlapping encoding in right aHPC (Fig. 7C, Table 3). This finding is consistent with evidence from prior work that aHPC represents the relationships between distinct events (Collin et al. 2015; Schlichting et al. 2015) and further provides evidence that the formation of integrated memories may involve composing a representation that combines semantic features of both events. We also found evidence of integrated AC representations during overlapping encoding in right PRC

(Fig. 7C, Table 3); this may reflect integration of features related to the indirectly related (A and C) items. To further examine the contribution of A and C item information, we carried out two additional searchlights. The first searchlight tested for greater A item semantic correlation compared with B item semantic correlation, weighted by inference performance, and the second searchlight tested for greater C item semantic correlation compared with B item semantic correlation, weighted by inference performance. The searchlights identified voxels that overlapped each other and the original clusters, suggesting that semantic information related to both A and C (relative to B) is predictive of subsequent inference performance (Fig. S5).

Item suppression in anterior mPFC correlates with suppressed integration in HPC

Medial PFC has been theorized to shape encoding by controlling the availability of prior memories during new learning (Schlichting and Preston 2015). We hypothesized that if a prior memory is suppressed by amPFC, that memory will be inhibited during overlapping encoding, preventing formation of an integrated memory. Consistent with this account, we found that suppression of item patterns in amPFC during overlapping encoding predicted decreased performance on the inference test (Fig. 6C). We hypothesized that item suppression in amPFC might impact encoding of overlapping pairs by inhibiting memory integration in right aHPC and right PRC (Fig. 7C). To test this hypothesis, we first calculated A item reactivation (i.e., self-similarity between encoding activity in amPFC and the perceptual template) for each trial as before (Fig. 6B). We also calculated a measure of semantic AC integration as before (Fig. 7C) but estimating it for individual trials. For each trial, we calculated the pattern dissimilarity of that trial (separately for the aHPC and PRC ROIs) to every other trial, resulting in an observed dissimilarity vector for each trial. We then compared this observed dissimilarity vector to the predicted dissimilarity vector for that trial based on the AC integration model (Fig. 3B) using Spearman's correlation. This correlation provided an estimate of the degree of semantic AC integration occurring on each trial. Finally, for each subject, we used robust regression to estimate the trial-by-trial relationship between A item reactivation in amPFC and AC integration in aHPC and PRC. We then tested whether the relationship varied depending on subsequent inference accuracy. One participant was excluded from this analysis because they had fewer than five incorrect inference trials.

Suppression of A item activity during overlapping encoding in amPFC correlated with decreased AC integration in aHPC on subsequently incorrect trials (Fig. 7D; mean slope: 0.161, SEM: 0.070, $t(28) = 2.28$, $p = 0.031$, Cohen's $d = 0.424$), but not subsequently correct trials (mean slope: −0.006, SEM 0.007, $t(29) = 0.81$, $p = 0.43$, $d = 0.148$). Given the presence of outlier slope values in the incorrect condition, we also assessed slope using a nonparametric Wilcoxon test; we observed similar results ($p = 0.022$). The relationship between amPFC and aHPC was significantly greater for incorrect trials than for correct trials (mean difference: 0.166, SEM: 0.070, $t(28) = 2.36$, $p = 0.026$, $d = 0.438$; Wilcoxon $p = 0.013$). In contrast, there was no significant relationship between A item reactivation in amPFC and AC integration in right PRC on either incorrect (mean slope: -0.054, SEM: 0.044, $t(28) = 1.22$, $p = 0.24$, $d = 0.219$) or correct (mean slope: 0.004, SEM: 0.005, $t(29) = 0.88$, $p = 0.39$, $d = 0.160$) trials, and no significant difference between correct and incorrect trials (mean difference: 0.058, SEM: 0.045, $t(28) = 1.28$, $p = 0.21$, $d = 0.238$). These results provide evidence that amPFC may influence encoding in HPC by inhibiting the co-activation of information about the indirectly related A and C items in aHPC. Suppression of related memories during new encoding may inhibit memory integration, leading to decreased ability to subsequently infer relationships among overlapping events.

Discussion

We measured brain activity during encoding of material that overlapped with prior experience, allowing us to track reactivation and suppression of specific elements of related events in a trial-by-trial manner during encoding. We found evidence for separate brain networks that dynamically promote or inhibit integration of overlapping memories (Fig. 1D). Memory integration was supported by aHPC and pmPFC, whereas memory differentiation was supported by pHPC and amPFC. During new event encoding, pmPFC reactivated semantic knowledge related to prior, overlapping memory elements, and aHPC integrated semantic features of both the previous memory and the new experience into a combined representation. Participants learned both the initial and overlapping pairs with high accuracy, but accuracy in inferring indirect associations was lower (Fig. 4A), suggesting that participants sometimes learn the individual pairs but do not form integrated memories. Consistent with our prediction that memory integration would promote inference, reactivation and integration of prior semantic knowledge predicted individuals' subsequent ability to reason about the relationship among events. On overlapping encoding trials with greater semantic reactivation in pmPFC and semantic integration in aHPC, participants were more likely to later correctly infer indirect associations between items from overlapping memories (Fig. 7B–C).

We also found evidence for a dissociable set of regions that inhibit memory integration by suppressing related memories during new learning. During overlapping event encoding, pHPC representations of specific, related memory elements were suppressed. Suppression in pHPC at the representational level correlated with decreased reactivation of prior memory elements in medial temporal lobe cortex. This interaction was stronger on trials for which participants later failed to retrieve the indirect association between memories, suggesting that memory suppression in pHPC inhibits memory integration. Furthermore, we found evidence that vlPFC influences whether overlapping memories are stored as integrated or distinct memories. At the level of individual item representations, memory reactivation/suppression in vlPFC correlated with the amount of memory suppression in amPFC. Memory suppression in amPFC, in turn, correlated with suppressed semantic integration in aHPC. Based on these results, we propose that, after HPC reactivates a memory of a related event, control processes in vlPFC and amPFC are recruited to select among memory contents (Fig. 8A–B). Prefrontal cortex then controls whether memories become integrated in aHPC (Fig. 8C–D).

Ventrolateral PFC has been proposed to represent task goals (Badre and Nee 2018) and resolve interference between conflicting memories by selecting task-relevant memories (Thompson-Schill et al. 2005; Kuhl et al. 2007, 2012a). Previous work has demonstrated that right vlPFC shows increased activation during tasks that require suppressing memory retrieval (Benoit and Anderson 2012) or selectively retrieving one item over another (Kuhl et al. 2007; Wimber et al. 2015). Here, we found evidence that right vlPFC influences memory reactivation and suppression—in the absence of explicit task instructions—modulating trial-by-trial what memory contents are available during new learning. Furthermore, whereas previous studies have

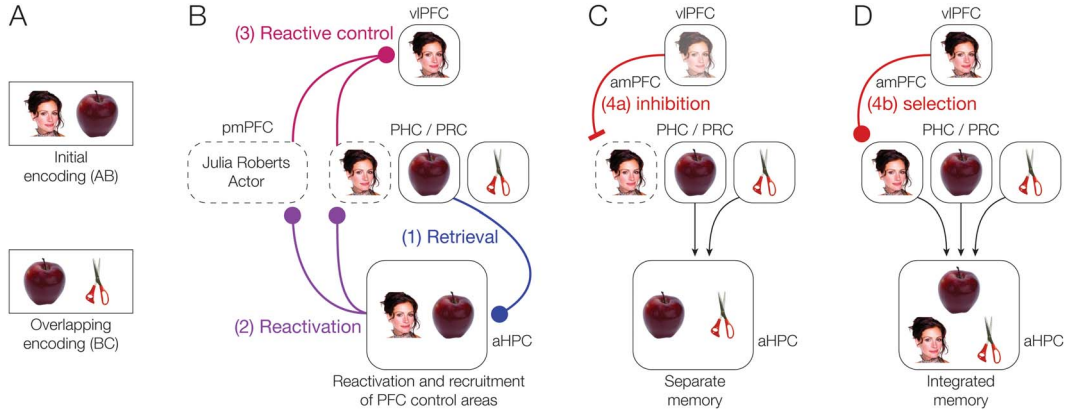

Fig. 8. Conceptual summary of proposed medial temporal lobe and prefrontal contributions to memory reactivation, selection, and integration. (A) Example initial (AB) and overlapping (BC) pairs. (B) During presentation of an overlapping pair, the individual items are represented in PHC and PRC. The related initial memory is retrieved in aHPC (step 1), causing reactivation of A item perceptual (PHC/PRC) and semantic (pmPFC) information (step 2). The conflict between the previous memory and the new event triggers reactive control mechanisms in PFC (step 3). Control mechanisms then determine whether an integrated or differentiated memory is formed in aHPC. (C) If vlPFC inhibits the initial memory by suppressing activity specific to the A item in amPFC (step 4a), connections via PHC/PRC will cause the A item to also be suppressed in aHPC. As a result, a differentiated memory will be formed for the overlapping pair, distinct from the memory of the initial pair. (D) If control mechanisms instead select for the A item (step 4b), an integrated memory will be formed in aHPC.

shown that the magnitude of vlPFC activation correlates with memory suppression in other areas (Kuhl et al. 2007; Wimber et al. 2015), here, we used pattern analysis to demonstrate that vlPFC itself represents item-specific memory contents. Critically, we measured item-specific activity in vlPFC during viewing of individual items before the memory task; this activity was then reactivated during encoding of overlapping pairs the next day. This finding demonstrates that vlPFC exhibits stable item-specific activity in humans, consistent with recent work in non-human primates (Ghazizadeh et al. 2018), and further suggests that item-specific activity reflects the selection of specific memories to be retrieved or suppressed. Selective retrieval of task-relevant memories is thought to facilitate integration of information across multiple events (Schlichting and Preston 2015), whereas memory suppression is thought to be involved in an active differentiation process that limits interference between competing memories (Hulbert and Norman 2014; Ritvo et al. 2019). Our results suggest that right vlPFC may be involved in selecting between these two competing goals, with memory reactivation in vlPFC facilitating memory integration, and memory suppression inhibiting integration. Critically, in our task, there is no explicit instruction of whether participants should integrate or differentiate the overlapping memories. Our results suggest that right vlPFC activity patterns are sensitive to trial-level fluctuation in integration and differentiation, but the cause of these trial-level differences remains unclear. An important goal for future research will be to clarify the role of right vlPFC by examining under what circumstances it promotes integration or differentiation in the absence of explicit encoding instructions.

Medial PFC, like vlPFC, has been implicated in memory control processes (Preston and Eichenbaum 2013; Rajasethupathy et al. 2015; Benoit et al. 2016; Eichenbaum 2017; Schmidt et al. 2019). In particular, mPFC is thought to represent prior knowledge that can be reactivated to guide behavior in new situations (Varga et al. In press; Ghosh and Gilboa 2014; Moscovitch et al. 2016; van Kesteren et al. 2016; Wikenheiser et al. 2017; Baldassano et al. 2018; Zhou et al. 2019). The mPFC has further been proposed to facilitate encoding of events associated with prior knowledge (Zeithamova et al. 2012a; Schlichting and Preston 2015; Liu et al. 2017; Spalding et al. 2018; Masís-Obando et al. 2022). For instance, rodent research has shown that prior knowledge of a spatial layout (i.e., a spatial schema) facilitates acquisition of new related memories and speeds their consolidation (Tse et al. 2007); such facilitated learning critically depends on mPFC activity and its interactions with HPC (Tse et al. 2011). In the current study, the initial pairs that participants learned were composed of common objects paired with famous people and places. We used a recently developed and validated method (Morton et al. 2021) to model the semantic knowledge associated with these real-world people and places, allowing us to quantify the reactivation of this knowledge during encoding of the overlapping object-object pairs and determine how prior knowledge activation affects new learning. We found evidence that pmPFC reactivates knowledge associated with individual memory elements from the initial pairs, and that this reactivation of prior knowledge facilitates integration of the initial and overlapping pairs. Our findings thus suggest that pmPFC represents detailed semantic knowledge associated with relevant memories, which may help to provide a framework to facilitate encoding of new, related events.

Anterior and posterior mPFC have different cytoarchitectonic structure and connectivity patterns (Cavada 2000; Öngür et al. 2003), which indicate that they may play distinct roles in cognition (Roy et al. 2012). Recent memory studies suggest that amPFC serves a distinct functional role in representing overlapping events compared with pmPFC (Schlichting et al. 2015; Schlichting and Preston 2016). A study using a similar associative task to the present paradigm found evidence that amPFC forms differentiated representations of overlapping events after learning, whereas pmPFC forms integrated traces (Schlichting et al. 2015). Differentiated memories may be formed through an active differentiation process that decreases overlap between competing memories (Hulbert and Norman 2014; Ritvo et al. 2019). In the present study, we found evidence that amPFC may actively differentiate overlapping memories by suppressing the competing memories during encoding. On overlapping encoding trials with increased memory suppression in amPFC, participants performed worse on a later inference test that required combining information across the initial and overlapping events. Furthermore, we found that vlPFC and amPFC are representationally coupled

during overlapping encoding, suggesting that these regions may communicate to resolve interference between overlapping, but distinct, events. Previous work suggests that vlPFC and mPFC are components of a network recruited during memory suppression that renders specific memories less likely to be retrieved in the future (Anderson et al. 2016). Here, we find that vlPFC may operate during encoding to select or suppress related memories through its interactions with amPFC. We found that selection of related memories during encoding promotes formation of integrated memories, whereas suppression leads to formation of differentiated memories, without necessarily causing forgetting of the suppressed memories.

Along with the representational dissociations we observed within mPFC, our data further indicate distinct representational processes within the HPC during overlapping event encoding. Studies in both rodents and humans have demonstrated that there is a representational gradient within HPC, with pHPC representing local information, and aHPC representing global information about contexts as a whole (Poppenk et al. 2013; Strange et al. 2014; Eichenbaum 2017; Morton et al. 2017; Brunec et al. 2018, 2019). aHPC, which has bidirectional connections with pmPFC (Barbas and Blatt 1995; Cavada 2000), has been implicated in representing memories that integrate information from multiple events (Collin et al. 2015; Schlichting et al. 2015; Mack et al. 2018; Baraduc et al. 2019). Recent work has demonstrated that, after learning of overlapping events, indirectly related items elicit similar representations in aHPC (Schlichting et al. 2015; Molitor et al. 2021). However, it is unclear from previous studies how these integrated representations are composed during learning. Prior work has demonstrated that, when participants view familiar stimuli, HPC activates representations that are sensitive to individual items (Quiroga et al. 2005) and their associated semantic features (Morton et al. 2021). Prior work has also found that PRC, which is functionally coupled with aHPC (Libby et al. 2012), represents item-specific conceptual features (Martin et al. 2018) and relationships between items (Sakai and Miyashita 1991; Schapiro et al. 2012; Pudhiyidath et al. 2022). Here, we found evidence that integrated memories incorporate semantic features related to both of the indirectly related items, both in HPC and PRC.

The incorporation of semantic features into the integrated memory may facilitate flexible retrieval, allowing integrated memories to be retrieved on the basis of a partial cue. For example, if one knows that a famous musician was indirectly associated with a common tool, these known semantic features could be used to help target retrieval of the relevant integrated memory. Over time, the process of integrating features of related items may also facilitate learning of concept representations that capture general knowledge about item categories (Mack et al. 2016, 2018; Bowman and Zeithamova 2018; Morton and Preston 2021) and schema representations that reflect general knowledge about contexts (Masís-Obando et al. 2022). The HPC is anatomically and functionally connected with large-scale cortical networks (Witter et al. 2000; Libby et al. 2012) that represent distinct features of stimuli (Huth et al. 2012; Deniz et al. 2019; Morton et al. 2021). HPC is thought to drive memory reactivation in these cortical networks (Ritchey et al. 2015; Cooper and Ritchey 2019), which may then support decision making based on retrieved semantic knowledge (Ranganath and Ritchey 2012).

In contrast to aHPC, pHPC is thought to form differentiated memories of individual events (Schlichting et al. 2015). Similar to amPFC, we found evidence that the formation of differentiated memories in pHPC may be supported by suppression of related memories during encoding. Theoretical work suggests that overlapping memories may become differentiated through a process of active differentiation, wherein overlapping memories are retrieved but then inhibited in a process that reduces competition between memories (Ritvo et al. 2019). Interestingly, pHPC evinced reactivation of category-specific activation patterns while also demonstrating suppression of item-specific activation. We propose that, in our study, pHPC may have retrieved the overlapping memory but then minimized competition between memories by suppressing features that are specific to that memory (i.e., item-specific features), while maintaining other features (i.e., category-specific features). This possibility should be investigated in future research to further clarify how memory competition affects different features associated with overlapping memories.

Suppression of item-specific activation patterns in pHPC may modulate the availability of related memories during overlapping encoding. On trials during which medial temporal lobe cortical areas were representationally coupled with pHPC, participants were more likely to later fail to make correct cross-event inferences, suggesting that memory suppression in pHPC is related to inhibition of memory integration. However, it is currently unclear why pHPC influences memory availability in medial temporal lobe cortex on some trials more than others. One possibility is that PFC control processes modulate the relative influence of pHPC and aHPC in mediating memory reactivation. When aHPC is inhibited, activation of items in medial temporal lobe cortex might be driven mainly by pHPC, causing a decrease in reactivation of related event elements in medial temporal lobe cortex. A goal for future research will be to examine whether PFC control processes modulate the effect of suppression in pHPC, either directly or through modulation of activity in aHPC.

Together, the present findings reveal that distinct control mechanisms are involved in determining whether overlapping events are stored as differentiated or integrated memories (Fig. 8). We found evidence that right vlPFC, which has been implicated in selecting between different potential task goals (Badre and Nee 2018), may activate or suppress item templates to select whether a given encoding trial is focused on integrating the current event with the prior memory or forming a new, differentiated memory. We propose that right vlPFC may affect encoding by modulating memory activation in amPFC. Memory suppression in amPFC may then affect encoding in aHPC by inhibiting the formation of integrated memories combining semantic features associated with the two events. In contrast, we found evidence that pmPFC supports the memory integration process by representing relevant semantic knowledge. Our results thus demonstrate that complementary mechanisms are involved in determining how related events are represented in memory, and that these mechanisms are sensitive not only to the perceptual features of the individual events, but also their conceptual meaning. In doing so, our data provide evidence for how activation of semantic knowledge influences new encoding, thus testing the fundamental predictions of leading neurobiological theories of memory and knowledge representation (Wang and Morris 2010; Ghosh and Gilboa 2014; Moscovitch et al. 2016; Eichenbaum 2017).

Acknowledgments

The authors thank Bernie Gelman and Robert Molitor for help with data collection, Dasa Zeithamova for help with experimental

design and analysis, Jackson Liang and Katie Guarino for help with defining anatomical masks, and Michael Mack, Katherine Sherrill, and Christine Coughlin for valuable discussions.

CRediT authors statement

Neal W Morton (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Ellen L. Zippi (Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Validation, Writing—review and editing), Alison R. Preston (Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review and editing).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This research was supported by the National Science Foundation (CAREER Award BCS 1056019 to A.R.P.) and the National Institute of Mental Health (NIH-NIMH R01 MH100121 to A.R.P. and NIH-NIMH National Research Service Award F32 MH114869 to N.W.M.) of the National Institutes of Health.

Conflict of interest statement: None declared.

Data and code availability

Code for generating wiki2vec vectors for an arbitrary set of well-known stimuli is publicly available (Zippi et al. 2020). Code implementing the reported analyses is also publicly available (Morton and Zippi 2022). De-identified data are available upon request to the corresponding author.

References

Anderson MC, McCulloch KC. Integration as a general boundary condition on retrieval-induced forgetting. *J Exp Psychol Learn Mem Cogn*. 1999:25(3):608–629.

Anderson MC, Bunce JG, Barbas H. Prefrontal–hippocampal pathways underlying inhibitory control over memory. *Neurobiol Learn Mem*. 2016:134(Pt A):145–161.

Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008:12(1):26–41.

Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC. The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*. 2010:49(3):2457–2466.

Badre D, Nee DE. Frontal cortex and the hierarchical control of behavior. *Trends Cogn Sci*. 2018:22(2):170–188.

Baldassano C, Hasson U, Norman KA. Representation of real-world event schemas during narrative perception. *J Neurosci*. 2018:38(45):9689–9699.

Baraduc P, Duhamel J-R, Wirth S. Schema cells in the macaque hippocampus. *Science*. 2019:363(6427):635–639.

Barbas H, Blatt GJ. Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus*. 1995:5(6):511–533.

Benoit RG, Anderson MC. Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*. 2012:76(2):450–460.

Benoit RG, Davies DJ, Anderson MC. Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proc Natl Acad Sci*. 2016:113:E8492–E8501.

Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. Sebastopol (CA): O'Reilly Media, Inc.; 2009.

Bowman CR, Zeithamova D. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *J Neurosci*. 2018:38:2605–2614.

Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997:10(4):433–436.

Brown TI, Gordon AM, Bailenson JN, Carr VA, Bowles B, Favila SE, LaRocque KF, Wagner AD. Prospective representation of navigational goals in the human hippocampus. *Science*. 2016:352(80):1323–1326.

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Liu ZX, Grady C, Rosenbaum RS, Winocur G, Barense MD, et al. Multiple scales of representation along the hippocampal anteroposterior Axis in humans. *Curr Biol*. 2018:28:2129–2135.e6.

Brunec IK, Robin J, Patai EZ, Ozubko JD, Javadi AH, Barense MD, Spiers HJ, Moscovitch M. Cognitive mapping style relates to posterior–anterior hippocampal volume ratio. *Hippocampus*. 2019:29(8):1–7.

Cavada C. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb Cortex*. 2000:10:220–242.

Collin SHP, Milivojevic B, Doeller CF. Memory hierarchies map onto the hippocampal long axis in humans. *Nat Neurosci*. 2015:18:1562–1564.

Cooper RA, Ritchey M. Cortico-hippocampal network connections support the multidimensional quality of episodic memory. *eLife*. 2019:8:e45591.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996:29(3):162–173.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*. 1999:9:179–194.

Deniz F, Nunez-Elizalde AO, Huth AG, Gallant JL. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J Neurosci*. 2019:39:7722–7736.

Eichenbaum H. Memory: organization and control. *Annu Rev Psychol*. 2017:68:19–45.

Ghazizadeh A, Hong S, Hikosaka O. Prefrontal cortex represents long-term memory of object values for months. *Curr Biol*. 2018:28:2206–2217.e5.

Ghosh VE, Gilboa A. What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*. 2014:53:104–114.

Gilboa A, Marlatte H. Neurobiology of schemas and schema-mediated memory. *Trends Cogn Sci*. 2017:21:618–631.

Griffiths BJ, Parish G, Roux F, Michelmann S, van der Plas M, Kolibius LD, Chelvarajah R, Rollings DT, Sawlani V, Hamer H, et al. Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. Proc Natl Acad Sci. 2019:116:21834–21842.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*. 2009:7:37–53.

Hulbert JC, Norman KA. Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cereb Cortex*. 2014:25:3994–4008.

Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*. 2012:76:1210–1224.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012:62:782–790.

Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2008a:2:1–28.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. Matching categorical object representations in inferior temporal cortex of Man and monkey. *Neuron*. 2008b:60:1126–1141.

Kuhl BA, Dudukovic NM, Kahn I, Wagner AD. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat Neurosci*. 2007:10(7):908–914.

Kuhl BA, Bainbridge WA, Chun MM. Neural reactivation reveals mechanisms for updating memory. *J Neurosci*. 2012a:32:3453–3461.

Kuhl BA, Rissman J, Wagner AD. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia*. 2012b:50(4):458–469.

Landauer TK, Dumais ST. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev*. 1997:104:211–240.

Levy BJ, Anderson MC. Purging of memories from conscious awareness tracked in the human brain. *J Neurosci*. 2012:32(47):16785–16794.

Liang JC, Preston AR. Medial temporal lobe reinstatement of content-specific details predicts source memory. *Cortex*. 2017:91:67–78.

Libby LA, Ekstrom AD, Ragland JD, Ranganath C. Differential connectivity of perirhinal and parahippocampal cortices within human hippocampal subregions revealed by high-resolution functional imaging. *J Neurosci*. 2012:32:6550–6560.

Liu ZX, Grady C, Moscovitch M. Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cereb Cortex*. 2017:27:1991–2009.

Mack ML, Preston AR. Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *NeuroImage*. 2016:127:144–157.

Mack ML, Love BC, Preston AR. Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proc Natl Acad Sci*. 2016:113:13203–13208.

Mack ML, Love BC, Preston AR. Building concepts one episode at a time: the hippocampus and concept formation. *Neurosci Lett*. 2018:680:31–38.

Martin CB, Douglas D, Newsome RN, Man LLY, Barense MD. Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *elife*. 2018:7:e3187.

Masís-Obando R, Norman KA, Baldassano C. Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *eLife*. 2022:11:e70445.

McKenzie S, Frank AJ, Kinsky NR, Porter B, Rivière PD, Eichenbaum H. Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*. 2014:83:202–215.

Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. In: *Proceedings of Workshop at ICLR*. 2013.

Molitor RJ, Sherrill KR, Morton NW, Miller AA, Preston AR. Memory reactivation during learning simultaneously promotes dentate gyrus/CA2,3 pattern differentiation and CA1 memory integration. *J Neurosci*. 2021:41:726–738.

Morton NW. 2022. *fPrep: Preprocessing of fMRI data*. [WWW Document]. Zenodo. https://doi.org/10.5281/zenodo.5904032.

Morton NW, Polyn SM. Beta-band activity represents the recent past during episodic encoding. *NeuroImage*. 2017:147:692–702.

Morton NW, Preston AR. Concept formation as a computational cognitive process. *Curr Opin Behav Sci*. 2021:38:83–89.

Morton NW, Zippi EL. 2022. *Bender study: pattern analysis of memory reactivation, suppression, and integration*. [WWW Document]. Zenodo. https://doi.org/10.5281/zenodo.6967582.

Morton NW, Sherrill KR, Preston AR. Memory integration constructs maps of space, time, and concepts. *Curr Opin Behav Sci*. 2017:17:161–168.

Morton NW, Schlichting ML, Preston AR. Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proc Natl Acad Sci*. 2020:117:201912338.

Morton NW, Zippi EL, Noh SM, Preston AR. Semantic knowledge of famous people and places is represented in hippocampus and distinct cortical networks. *J Neurosci*. 2021:41:2762–2779.

Moscovitch M, Cabeza R, Winocur G, Nadel L. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu Rev Psychol*. 2016:67:105–134.

Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*. 2012:59:2636–2643.

Navawongse R, Eichenbaum H. Distinct pathways for rule-based retrieval and spatial mapping of memory representations in hippocampal neurons. *J Neurosci*. 2013:33:1002–1013.

Norman KA, Newman EL, Detre G. A neural network model of retrieval-induced forgetting. *Psychol Rev*. 2007:114:887–953.

Öngür D, Ferry AT, Price JL. Architectonic subdivision of the human orbital and medial prefrontal cortex. *J Comp Neurol*. 2003:460:425–449.

Place R, Farovik A, Brockmann M, Eichenbaum H. Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nat Neurosci*. 2016:19:992–994.

Poppenk J, Norman KA. Briefly cuing memories leads to suppression of their neural representations. *J Neurosci*. 2014:34:8010–8020.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. Long-axis specialization of the human hippocampus. *Trends Cogn Sci*. 2013:17:230–240.

Preston AR, Eichenbaum H. Interplay of hippocampus and prefrontal cortex in memory. *Curr Biol*. 2013:23:R764–R773.

Price JL, Drevets WC. Neurocircuitry of mood disorders. *Neuropsychopharmacology*. 2010:35:192–216.

Pudhiyidath A, Morton NW, Duran RV, Schapiro AC, Momennejad I, Hinojosa-Rowland DM, Molitor RJ, Preston AR. Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *J Cogn Neurosci*. 2022:34:1736–1760.

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. *Nature*. 2005:435:1102–1107.

Rajasethupathy P, Sankaran S, Marshel JH, Kim CK, Ferenczi E, Lee SY, Berndt A, Ramakrishnan C, Jaffe A, Lo M, et al. Projections

from neocortex mediate top-down control of memory retrieval. *Nature*. 2015:526:653–659.

Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci*. 2012:13:713–726.

Rao VA, Howard MW. Retrieved context and the discovery of semantic structure. *Adv Neural Inf Process Syst*. 2008:20:1193–1200.

Richter FR, Chanales AJH, Kuhl BA. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage*. 2016:124:323–335.

Ritchey M, Libby LA, Ranganath C. Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. In: O'Mara S, Tsanov M, editors. *The connected hippocampus*. Amsterdam, Netherlands: Elsevier; 2015. pp. 45–64.

Ritvo VJH, Turk-Browne NB, Norman KA. Nonmonotonic plasticity: how memory retrieval drives learning. *Trends Cogn Sci*. 2019:23:726–742.

Roy M, Shohamy D, Wager TD. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci*. 2012:16:147–156.

Sakai K, Miyashita Y. Neural organization for the long-term memory of paired associates. *Nature*. 1991:354:152–155.

Schapiro AC, Kustner LV, Turk-Browne NB. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol*. 2012:22:1622–1627.

Schlichting ML, Preston AR. Memory integration: neural mechanisms and implications for behavior. *Curr Opin Behav Sci*. 2015:1:1–8.

Schlichting ML, Preston AR. Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiol Learn Mem*. 2016:134:91–106.

Schlichting ML, Mumford JA, Preston AR. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat Commun*. 2015:6(1):8151.

Schlichting ML, Guarino KF, Roome HE, Preston AR. Developmental differences in memory reactivation relate to encoding and inference in the human brain. *Nat Hum Behav*. 2022:6:415–428.

Schmidt B, Duin AA, Redish AD. Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision-making. *J Neurophysiol*. 2019:121(6):1981–2000.

Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. In: *Proceedings of the 9th Python in Science Conference*. 2010:57:92–96.

Smith SM, Brady JM. SUSAN–A new approach to low level image processing. *Int J Comput Vis*. 1997:23:45–78.

Smyser CD, Inder TE, Shimony JS, Hill JE, Degnan AJ, Snyder AZ, Neil JJ. Longitudinal analysis of neural network development in preterm infants. *Cereb Cortex*. 2010:20:2852–2862.

Spalding KN, Schlichting ML, Zeithamova D, Preston AR, Tranel D, Duff MC, Warren DE. Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. *J Neurosci*. 2018:38:3767–3775.

Stelzer J, Chen Y, Turner R. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage*. 2013:65:69–82.

Strange BA, Witter MP, Lein ES, Moser EI. Functional organization of the hippocampal longitudinal axis. *Nat Rev Neurosci*. 2014:15:655–669.

Thompson-Schill SL, Bedny M, Goldberg RF. The frontal lobes and the regulation of mental activity. *Curr Opin Neurobiol*. 2005:15:219–224.

Tolman EC. Cognitive maps in rats and men. *Psychol Rev*. 1948:55:189–208.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM. Schemas and memory consolidation. *Science*. 2007:316(80):76–82.

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RGM. Schema-dependent gene activation and memory encoding in neocortex. *Science*. 2011:333(80):891–895.

Tulving E. Episodic and semantic memory. In: Tulving E, Donaldson W, editors. *Organization of Memory*. Cambridge (MA): Academic Press; 1972 p. 381–402.

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010:29:1310–1320.

van Kesteren MTR, Fernandez G, Norris DG, Hermans EJ. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci*. 2010a:107:7550–7555.

van Kesteren MTR, Rijpkema M, Ruiter DJ, Fernandez G. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *J Neurosci*. 2010b:30:15888–15894.

van Kesteren MTR, Brown TI, Wagner AD. Interactions between memory and new learning: insights from fMRI multivoxel pattern analysis. *Front Syst Neurosci*. 2016:10:1–5.

Van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN. How schema and novelty augment memory formation. *Trends Neurosci*. 2012:35:211–219.

Varga NL, Morton NW, Preston AR. Schema, inference, and memory. In: Kahana MJ, Wagner AD, editors. *Oxford handbook of human memory*. Oxford (UK): Oxford University Press; In press.

Wang S-H, Morris RGM. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annu Rev Psychol*. 2010:61:49–79.

Wikenheiser AM, Schoenbaum G. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nat Rev Neurosci*. 2016:17:513–523.

Wikenheiser AM, Marrero-Garcia Y, Schoenbaum G. Suppression of ventral hippocampal output impairs integrated orbitofrontal encoding of task structure. *Neuron*. 2017:95:1197–1207.e3.

Wimber M, Alink A, Charest I, Kriegeskorte N, Anderson MC. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat Neurosci*. 2015:18:582–589.

Witter MP, Naber PA, Van Haeften T, Machielsen WCM, Rombouts SARB, Barkhof F, Scheltens P, Lopes da Silva FH. Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*. 2000:10:398–410.

Zeithamova D, Dominick AL, Preston AR. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*. 2012a:75:168–179.

Zeithamova D, Schlichting ML, Preston AR. The hippocampus and inferential reasoning: building memories to navigate future decisions. *Front Hum Neurosci*. 2012b:6:1–14.

Zhou J, Gardner MPH, Stalnaker TA, Ramus SJ, Wikenheiser AM, Niv Y, Schoenbaum G. Rat orbitofrontal ensemble activity contains multiplexed but dissociable representations of value and task structure in an odor sequence task. *Curr Biol*. 2019:0:897–907.e3.

Zippi EL, Morton NW, Preston AR. 2020. Modeling semantic similarity of well-known stimuli [WWW document]. OSF. https://doi.org/10.17605/OSF.IO/72APM.

Supplementary Material

Perirhinal cortex and inferior frontal gyrus/insula reactivate item-specific patterns, but not category-general patterns

We found that item-specific activity is reactivated during overlapping encoding in right PHC, left PRC, and right IFG/insula, and suppressed in right posterior HPC (Fig. 5A–C). Because previous studies have often used reactivation of category-specific activity as an index of memory reactivation, we also examined whether category-level information was reactivated in these regions. In addition to r_{self} (average correlation between the activity pattern observed on each overlapping encoding trial and the perceptual template for the indirectly related item) and r_{within} (average correlation to all other items in the same category), we also calculated $r_{between}$, the average correlation between a given encoding trial and all items in the other category. For example, if a given overlapping encoding trial BC was associated with an initial pair AB for which the A item was a face, we would calculate the average similarity between the encoding pattern and all scene perceptual templates. To determine whether a given region showed category reactivation, we tested whether r_{within} was greater than $r_{between}$.

We examined each region identified as demonstrating item-specific reactivation or suppression (Fig. 5A–C). Each significant group-level cluster was reverse-normalized to each individual's native space, dilated by one voxel, and intersected with a gray-matter mask for that subject based on their FreeSurfer cortical reconstruction. We calculated r_{self}, r_{within}, and $r_{between}$ for each subject within each ROI (Fig. S1). We found significant category-level reactivation in right PHC and pHPC (PHC: mean $r_{within}-r_{between}$: 0.0082, SEM: 0.0030, $t_{(29)}$=2.82, p=0.0086, Cohen's d=0.51; pHPC: mean: 0.0034, SEM: 0.0016, $t_{(29)}$=2.18, p=0.038, d=0.40). We found no evidence of category reactivation in left PRC (mean: 0.00084, SEM: 0.0011, $t_{(29)}$=0.75, p=0.46, d=0.14) or right IFG/insula (mean: −0.0011, SEM: 0.00060, $t_{(29)}$=1.88, p=0.070, d=0.34).

These findings suggest that item-specific reactivation may, in some regions, occur in the absence of category-level reactivation. In the case of pHPC, we observe both significant category reactivation and significant item suppression. These results underscore the importance of using item-specific patterns to measure memory reactivation and suppression. PRC and IFG/insula may emphasize features that are specific to the reactivated episodes, at the expense of features that are common to the category of the reactivated item.

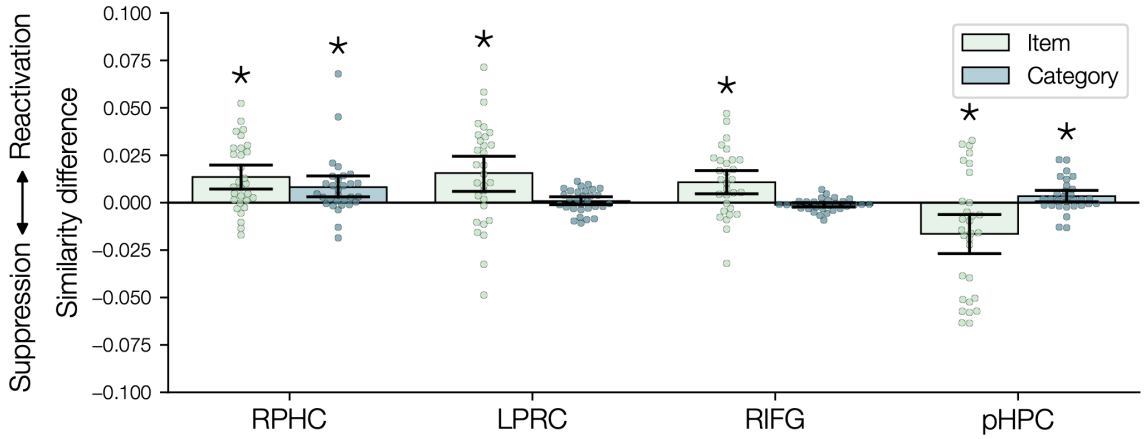

Fig. S1. Related to Fig. 5. Reactivation of item- and category-specific activity during overlapping encoding, in regions that showed evidence of item reactivation or suppression. Item reactivation is the difference between self-similarity and within-category similarity, while category reactivation is the difference between within-category similarity and between-category similarity. Note that regions were selected based on a significant difference between self-similarity and within-similarity, so estimates of this difference may be inflated. While each region showed evidence of item reactivation or suppression, category reactivation was only observed for RPHC and pHPC. RPHC: right parahippocampal cortex; LPRC: left perirhinal cortex; RIFG: right inferior frontal gyrus/insula; pHPC: posterior hippocampus. Error bars indicate 95% bootstrap confidence intervals. Points indicate individual participants. *: p<0.05.

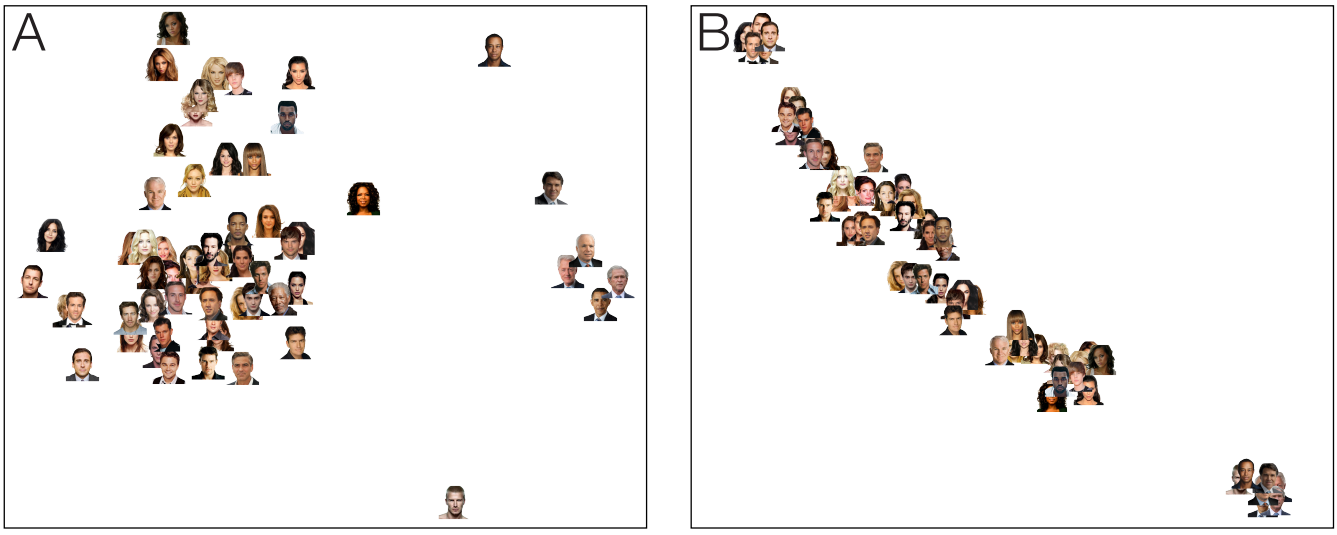

Fig. S2. Related to Fig. 7. (A) Two-dimensional MDS plot showing the relative semantic similarity of different person stimuli, according to the wiki2vec model. Items that are closer together have greater estimated similarity. (B) Relative positions of person stimuli based on *t*-SNE.

The semantic similarity model captures relationships between items within each category

Our wiki2vec model (Zippi et al. 2020; Morton et al. 2021) was used to estimate the semantic similarity between all items used in the experiment. To better understand what relationships are captured by the model, we first calculated the correlation distances between each pair of items based on their wiki2vec representations. We then used two-dimensional metric multidimensional scaling (MDS) and *t*-distributed stochastic neighbor embedding (*t*-SNE; perplexity=6, principal component analysis initialization) implemented in scikit-learn 1.0.2 (Pedregosa et al. 2011) to visualize the relative semantic similarity of items based on the model. MDS results in a low-dimensional representation that attempts to preserve the relative distances of all items (Torgerson 1952), while *t*-SNE emphasizes local neighborhoods of related items (Maaten and Hinton 2008). We examined these relationships separately for the different stimulus categories of famous people (Fig. S2), famous places (Fig. S3), and common objects (Fig. S4). We found that the model captures a number of divisions between different stimuli. People are divided into distinct clusters that reflect their occupation, such as politician, athlete, musician, and actor. There also appears to be a gradient between people who are only actors (e.g., Tom Cruise), people who both act and make music (e.g., Steve Martin), and musicians who have not done acting (e.g., Taylor Swift). Famous landmarks are

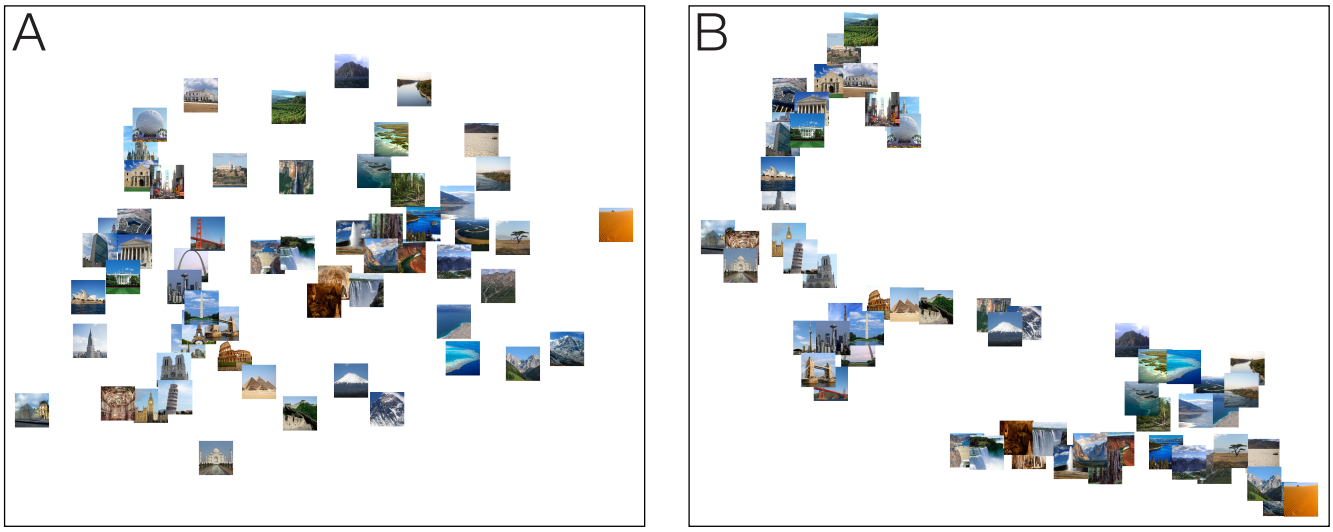

Fig. S3. Related to Fig. 7. (A) Two-dimensional MDS plot showing the relative semantic similarity of different place stimuli, according to the wiki2vec model. (B) Relative positions of place stimuli based on *t*-SNE.

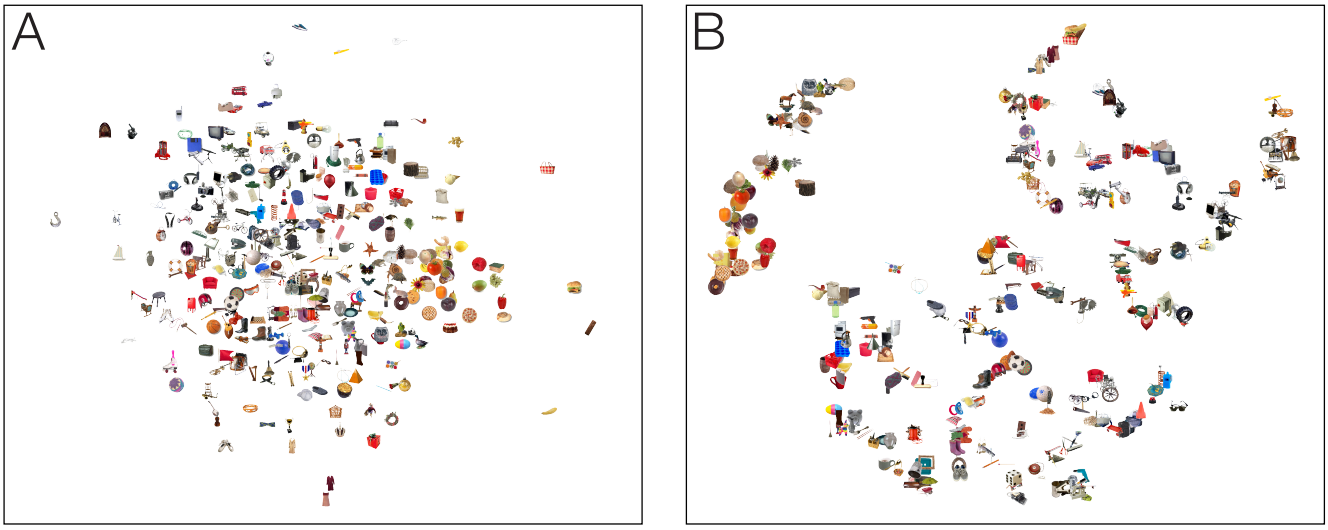

Fig. S4. Related to Fig. 7. (A) Two-dimensional MDS plot showing the relative semantic similarity of the object stimuli, according to the wiki2vec model. (B) Relative positions of scene stimuli based on *t*-SNE.

divided into manmade and natural landmarks, with finer relationships also apparent (for example, landmarks related to US politics, such as the White House, Supreme Court, Pentagon, Independence Hall, and the Headquarters of the United Nations are grouped together). Previous work has examined the wiki2vec model for these famous people and places in more detail and shown that wiki2vec accounts for both participant ratings of semantic similarity and objective semantic features such as categories of famous places and the age, gender, and occupation of famous people (Morton et al. 2021).

Here, we further found evidence that the wiki2vec model is sensitive to semantic features of common objects. The *t*-SNE projection, which emphasizes local relationships between similar items, identified many distinct clusters of objects (Fig. S4b); these clusters are less apparent in the MDS embedding, which emphasizes global relationships instead (Fig. S4a). The *t*-SNE plot exhibits clusters of foods, animals, instruments, and various categories of tools and household objects, demonstrating that wiki2vec provides a reasonable model of the semantic relatedness of common objects. Overall, the wiki2vec semantic model captures a number of relationships between items within the different categories, all within a common 300-dimensional space that allows vectors for different items to be combined to form composite representations.

Integrated semantic pattern activation reflects both reactivated initial items and new items

We found evidence that integrated semantic information was activated in hippocampus and perirhinal cortex during

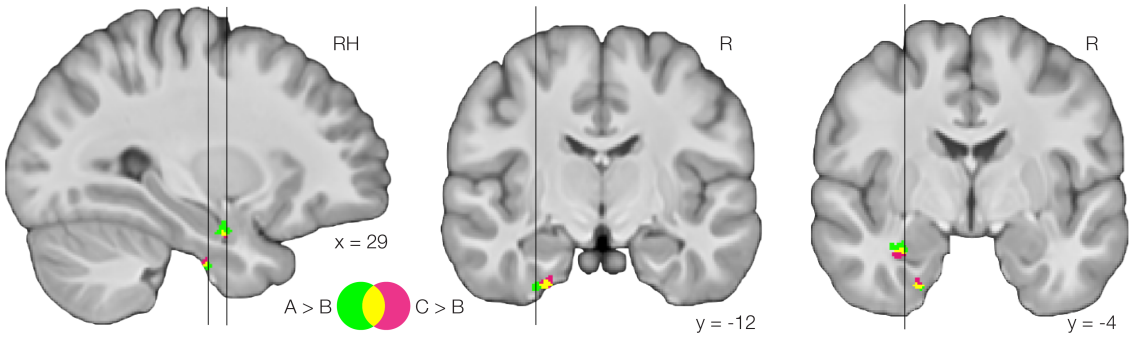

Fig. S5. Related to Fig. 7. Results of searchlight analyses testing for components of integrated representations during overlapping BC pair encoding. The A > B searchlight tested for a greater correlation with the A item model than the B item model, weighted by subsequent AC test accuracy. The C > B searchlight tested for a greater correlation with the C item model than the B item model, weighted by subsequent AC test accuracy. Significant voxels ($p<0.01$, permutation test, uncorrected) are shown for the two searchlights and their intersection. Both the hippocampal and perirhinal clusters exhibit an overlap between the A > B and C > B searchlights, suggesting that both components contribute to these clusters. Vertical lines indicate the corresponding cutting planes relating the sagittal and coronal views.

overlapping (BC) pair learning (Fig. 7C). The semantic integration searchlight identified areas where the difference between the correlation with the AC semantic model and the B item semantic model was greater for trials where the response on the subsequent AC test was correct. However, the AC semantic model is correlated with both the A item semantic model and the C item semantic model. This raises the possibility that clusters identified by the semantic integration searchlight might reflect representation of A or C information in isolation, rather than representing both simultaneously. To test whether both A item and C item information is activated during successful overlapping pair learning, we carried out two additional searchlight analyses. The first searchlight tested whether the difference between A item correlation and B item correlation was greater for subsequently correct trials. The second searchlight instead tested whether the difference between C item correlation and B item correlation was greater for subsequently correct trials. Voxelwise significance was assessed using the same procedure as the main semantic integration searchlight analysis. We found that, in both the hippocampus and perirhinal cortex clusters, there was overlap between significant voxels identified from the A > B and C > B searchlights. This suggests that the clusters identified by the main integration searchlight analysis (Fig. 7C) demonstrated activation of both A and C item information, rather than only reflecting one or the other.

References

Maaten L van der, Hinton G. 2008. Visualizing Data using t-SNE. J Mach Learn Res. 9:2579–2605.

Morton NW, Zippi EL, Noh SM, Preston AR. 2021. Semantic Knowledge of Famous People and Places Is Represented in Hippocampus and Distinct Cortical Networks. J Neurosci. 41:2762–2779.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011. Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res. 12:2825–2830.

Torgerson WS. 1952. Multidimensional scaling: I. Theory and method. Psychometrika. 17:401–419.

Zippi EL, Morton NW, Preston AR. 2020. Modeling semantic similarity of well-known stimuli [WWW Document]. URL https://osf.io/72apm/