



Published in final edited form as:

*Neuron*. 2017 September 27; 96(1): 228–241.e5. doi:10.1016/j.neuron.2017.09.005.

## Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex

Alexa Tompary<sup>1</sup> and Lila Davachi<sup>1,2</sup>

<sup>1</sup>Department of Psychology, New York University, New York, NY, 10003, USA

<sup>2</sup>Center for Neural Science, New York University, New York, NY, 10003, USA

### Summary

Structured knowledge is thought to form, in part, through the extraction and representation of regularities across overlapping experiences. However, little is known about how consolidation processes may transform novel episodic memories to reflect such regularities. In a multi-day fMRI study, participants encoded trial-unique associations that shared features with other trials. Multi-variate pattern analyses were used to measure neural similarity across overlapping and non-overlapping memories during immediate and one-week retrieval of these associations. We found that neural patterns in the hippocampus and medial prefrontal cortex represented the featural overlap across memories, but only after a week. Furthermore, after a week, the strength of a memory's unique episodic reinstatement during retrieval was inversely related to its representation of overlap, suggesting a trade-off between the integration of related memories and recovery of episodic details. These findings suggest that consolidation-related changes in neural representations support the gradual organization of discrete episodes into structured knowledge.

### Keywords

memory consolidation; human fMRI; hippocampus; medial prefrontal cortex; pattern similarity

### Introduction

Our semantic knowledge is a highly structured network of associations that are, at least in some part, learned through the extraction and consolidation of common features across many episodic experiences. However, our understanding of how memories of discrete episodic events are transformed into structured information over time is crude at best. From a neuroscientific perspective, successful episodic memory retrieval is thought to be initially

---

Corresponding Author/Lead Contact: Lila Davachi, Department of Psychology, New York University, 6 Washington Place, Room 871A, New York, NY 10003, lila.davachi@nyu.edu.

**Author Contributions:** Conceptualization, A.T. and L.D.; Methodology, A.T. and L.D.; Software, A.T.; Formal Analysis, A.T.; Investigation, A.T.; Resources, L.D.; Writing – Original Draft, A.T.; Writing – Review & Editing, A.T. and L.D.; Visualization, A.T. and L.D.; Supervision, L.D.; Funding Acquisition, L.D.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

supported by the hippocampus, but then may gradually be supported by distributed cortical regions through incremental, coordinated reactivation of memories across the hippocampus and cortex (Alvarez and Squire, 1994; Nadel et al., 2000). Evidence for such a mechanism has been explored in rodent replay studies (Pavlidis and Winson, 1989; Wilson and McNaughton, 1994) and in human neuroimaging research, by measuring how changes in resting state connectivity after new learning relate to later memory (Tambini et al., 2010; Tambini and Davachi, 2013; Schlichting and Preston, 2014; Tompary et al., 2015). Such hippocampal-cortical dialogue has been hypothesized to enable the slow extraction of statistical regularities common across overlapping episodic events (McClelland et al., 1995). However, how this process transforms the neural traces of episodic memories over the course of consolidation remains unknown.

Behavioral research in rodents and in humans provides compelling evidence that the structure of episodic memories changes with time. In a recent experiment, rodents learned a set of platform locations that were sampled from a predetermined distribution of locations. After one day, the animals tended to navigate to specific platform locations, but after thirty days, their swim patterns more closely matched the underlying probability distribution of all platform locations (Richards et al., 2014). Prior work has also shown that rodents begin to generalize context-specific behaviors to novel environments with time (Wiltgen and Silva, 2007). These findings suggest that recent memories are composed of distinct episodes, but remote memories become transformed and integrated into more generalized representations of related information. In humans, behavioral work has shown that rule acquisition and use is more evident with a temporal delay (Sweegers and Talamini, 2014). Similarly, other work suggests that sleep enhances transitive inference behavior (Ellenbogen et al., 2007; Lau et al., 2010) and benefits the extraction and generalization of statistical regularities across motor and acoustic patterns (Wagner et al., 2004; Durrant et al., 2011, 2013; Batterink and Paller, 2017). However, few studies to date have shed light on how the underlying neural representations of memories with shared features are transformed over time. In the present study, we examined whether neural representations of memories with overlapping features become more similar after a period of consolidation.

There is evidence that the medial prefrontal cortex (mPFC) likely plays an important role in the transformation of episodic memories over time, given its established involvement in two distinct mnemonic processes: retrieval of consolidated memories and encoding- or retrieval-mediated integration. First, increased mPFC activation has been associated with retrieval of remote memories (Sterpenich et al., 2009; Takashima et al., 2006) and retrieval of memories stabilized through sleep or spaced learning (Sterpenich et al., 2007; Takashima et al., 2007). Furthermore, multivariate patterns of activity in mPFC have been shown to be more discriminable for remote autobiographical memories than for recent memories (Bonnici et al., 2012).

However, an entirely distinct line of work has implicated the mPFC in tasks that require or explicitly instruct the online integration of information with shared content. Activation of the mPFC and its connectivity with the hippocampus increases when encoding episodes containing stimuli that overlap with recently learned trials (Kuhl et al., 2010; Zeithamova et al., 2012; Schlichting and Preston, 2016). Activation of mPFC is also related to retrieval-

mediated integration and updating of existing memories both in humans and rodents (Tse et al., 2007, 2011; Zeithamova and Preston, 2010). Furthermore, increased hippocampal-mPFC connectivity has been observed during the retrieval of memories with regularities across episodes (Sweegers et al., 2014). Taken together, these results provide converging evidence that mPFC involvement in retrieval is modulated both by the age of the memory and the necessity of integration computations during learning (Preston and Eichenbaum, 2013). However, little work has examined how consolidation may promote such restructuring of overlapping memories in mPFC over time.

How hippocampal memory representations change over the course of consolidation is more contentious. Relational memory theory proposes that the hippocampus is uniquely equipped to encode and represent information with relational links (Cohen and Eichenbaum, 1993; Eichenbaum, 1999). Consistent with this theory, a large body of neuropsychological and neuroimaging work has implicated hippocampal processes in the encoding of associations across unique features of an experience (Ryan et al., 2000; Davachi et al., 2003; Ranganath et al., 2003; Staresina and Davachi, 2006, 2008; see Davachi, 2006 for a review). Furthermore, hippocampal activation during new learning appears to relate to the later integration of overlapping events (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010; Kuhl et al., 2010; Schlichting et al., 2014; Schlichting and Preston, 2016). But these findings may simply be an extension of the role of hippocampal processes in supporting the initial formation of episodic memories. In other words, the reactivation of older memories during new learning may create a new, integrated memory trace through the same associative binding processes that are thought to support new episodic associative encoding. But it is unclear whether and how those overlapping associations are stored and represented in the hippocampus as part of an enduring memory trace, and how they might change with consolidation.

Furthermore, theoretically, the development of hippocampal relational nodes is not easily reconciled with complementary learning systems (CLS) theory, in which the proposed role of hippocampus is to store detailed, orthogonalized episodic memories, from which relational information is slowly extracted and represented in cortical regions (McClelland et al., 1995). According to CLS, memories rich in episodic or contextual details continue to be reinstated in the hippocampus, while schematic and generalized information come to be represented in cortical regions (Frankland and Bontempi, 2005; Nadel and Moscovitch, 1997; Winocur et al., 2010). Consistent with the notion that hippocampal representations maintain episodic specificity, recent imaging studies have shown that multivariate representations of specific memories are reinstated in the hippocampus during successful remembering (Tomparly et al., 2016; Mack and Preston, 2016). With respect to univariate activation, some work highlights that hippocampal activation during retrieval decreases with the age of the memory (Takashima et al., 2006, 2009; Watanabe et al., 2012). At the same time, other work demonstrates that hippocampal activation is related to remote retrieval for memories that retain distinct episodic elements (Viard et al., 2007; Harand et al., 2012; Sterpenich et al., 2009). Furthermore, neural patterns in the hippocampus have been shown to carry information about distinct episodes for both recent and remote autobiographical memories (Bonnici et al., 2012, 2013), suggesting that the hippocampus may continue to represent episodic features of memories as they are transformed over time, consistent with

multiple trace theory (Nadel et al., 2000). Although there have been attempts to reconcile the dual roles of the hippocampus in integrating memories of overlapping episodes and separating memories of distinct episodes (Kumaran and McClelland, 2012; Schapiro et al., 2017), further research is needed to reconcile these seemingly disparate hippocampal computations that occur during learning, and how the resulting memory traces may change with consolidation.

In the present study, we employed multi-variate pattern analyses to ask whether memories come to be represented more similarly to other memories with overlapping content over time. In this experiment, participants encoded trial-unique objects paired with one of four repeating scenes, such that multiple objects were studied with the same scene ('overlapping memories') and others were paired with a different scene ('non-overlapping memories'). Participants were later scanned during cued associative retrieval of individual scenes associated with each object. Critically, memory for half of the object-scene associations was tested immediately after learning (recent memories), and memory for the other half was tested after a week delay (remote memories, Figure 1A). Manipulating the time between encoding and retrieval allowed us to attribute any differences between the two retrieval periods to the additional influence of consolidation processes on the retrieval of remote memories. Our analysis approach is different from past work using multivariate patterns to characterize episodic events: rather than computing the neural similarity of items within and across categories (LaRocque et al., 2013), we computed the neural similarity between distinct memories with and without overlapping features. This allowed us to quantify the degree to which multi-voxel representations of recent and remote episodic memories reflect their commonalities.

We asked two distinct but complementary questions: (1) how does the neural representational similarity between overlapping and non-overlapping memory representations change over time, and (2) how do these changes relate to the reinstatement of specific episodic information captured during encoding? We predicted that ongoing consolidation processes would promote greater representational similarity in mPFC between overlapping memories, compared to non-overlapping memories. We predicted that this effect would be evident only for remote memories, as they will have undergone more consolidation. We also explored neural similarity in the hippocampus over time, as it is less clear how hippocampal memory representations evolve as a result of ongoing consolidation processes. We then asked how neural similarity between overlapping and non-overlapping memories relates to measures of episodic reinstatement to probe whether the restructuring of overlapping memories promotes or interferes with the maintenance of features that are unique to each memory. Finally, to more closely target consolidation processes, we investigated the relationship between the restructuring of overlapping memories and changes in functional connectivity during rest periods after learning.

## Results

### Memory performance

On average, participants correctly chose the scene studied with the object cue for 94.2% (SD: 6.0%) of the recently learned pairs, and for 54.1% (SD: 14.6%) of the remotely learned

pairs. As expected, participants' performance was significantly better for recent compared to remote memories ( $t_{(21)}=15.05$ ,  $p<10 \times 10^{-13}$ ) but, importantly, the percentage of correctly recalled scenes was significantly greater than chance (25%) for both retrieval tests (recent:  $t_{(21)}=38.16$ ,  $p<0.001$ ; remote:  $t_{(21)}=8.79$ ,  $p<0.001$ ; Figure 1C). Object recognition was computed with  $d'$ . Recognition was above chance (50%) both for recent memories (mean  $d'$ : 6.28, SD: 1.28;  $t_{(21)}=22.93$ ,  $p<0.001$ ) and for remote memories (mean  $d'$ : 2.31, SD: 1.28;  $t_{(21)}=13.81$ ,  $p<0.001$ ). As with cued recall, object recognition was also significantly greater for recent memories relative to remote memories ( $t_{(21)}=17.92$ ,  $p<0.001$ ).

### Retrieval similarity over time

To examine the consolidation-dependent reorganization of overlapping memory representations, two measures of retrieval similarity were calculated during both recent (immediately after encoding) and remote (1 week after encoding) cued associative retrieval (Figure 2). The multivariate pattern of activation evoked by each trial-unique object during retrieval was correlated with (1) the patterns of all other objects studied with the same scene (overlapping similarity) and (2) the patterns of all other objects studied with a different scene (non-overlapping similarity). These two correlations were computed separately for each object whose scene was confidently remembered (HC correct) in each retrieval session.

To ask whether mPFC comes to represent commonalities across memories, we first defined a region in mPFC modulated by retrieval of remote memories. To do this, we performed a univariate analysis that indexed activation during the remote retrieval session (Figure S1A). Voxels in this mPFC region exhibited greater activation during successful retrieval of remote memories relative to unsuccessful attempts at retrieval. We then submitted participants' retrieval similarity in this region to a 2 (Time: recent, remote)  $\times$  2 (Overlap: overlapping, non-overlapping) repeated-measures ANOVA, limiting the analysis to HC correct trials. This revealed a main effect of overlap ( $F_{(1,18)}=5.62$ ,  $p=0.03$ ), qualified by an interaction between time and overlap ( $F_{(1,18)}=7.33$ ,  $p=0.01$ ; Figure 3A). This interaction was driven by greater similarity amongst overlapping memories relative to non-overlapping memories during remote retrieval ( $t_{(18)}=2.55$ ,  $p=0.02$ ), and no difference in retrieval similarity during recent retrieval ( $t_{(18)}=0.79$ ,  $p=0.45$ ). The finding that mPFC shows greater similarity in retrieval patterns across overlapping memories compared to non-overlapping memories only after a week suggests that over time, neural patterns of memories in mPFC become organized according to their commonalities with other memories.

The same 2  $\times$  2 ANOVA also revealed an interaction in bilateral hippocampus ( $F_{(1,18)}=4.89$ ,  $p=0.04$ , Figure 3C). Interestingly, this interaction was also driven in part by a trend for *decreased* retrieval similarity for overlapping relative to non-overlapping trials ( $t_{(18)}=-1.78$ ,  $p=0.09$ ) during recent retrieval, perhaps suggestive of pattern separation of overlapping memories immediately. There was no reliable difference between overlapping and non-overlapping similarity for remote memories ( $t_{(18)}=1.66$ ,  $p=0.11$ ). A 2 (Overlap: overlapping, non-overlapping)  $\times$  2 (Time: recent, remote)  $\times$  2 (Hemisphere: right, left) ANOVA revealed no interaction with hemisphere (Table 1), suggesting that the relationship between feature overlap and time does not differ significantly in the right or left hippocampus.

There is a growing interest in functional specialization along the long axis of the hippocampus (Poppenk et al., 2013). In particular, past work suggests that anterior hippocampus is more involved in integrative computations while the posterior hippocampus is more likely to represent specificity in the environment. Based on this research, we investigated whether there were corresponding differences in retrieval similarity across the long axis of the hippocampus. We computed a 2 (Overlap: overlapping, non-overlapping)  $\times$  2 (Time: recent, remote)  $\times$  2 (Region: anterior, posterior) ANOVA and found an interaction between region and time ( $F_{(1,18)}=6.11$ ,  $p=0.02$ ) in addition to a strong interaction between overlap and time ( $F_{(1,18)}=20.93$ ,  $p<0.001$ ; Table 1). To unpack these effects, we computed a 2 (Overlap: overlapping, non-overlapping)  $\times$  2 (Time: recent, remote) ANOVA separately for patterns in anterior and posterior hippocampus. In posterior hippocampus, there was a no main effect of time ( $F_{(1,18)}=2.50$ ,  $p=0.13$ ) and a main effect of overlap ( $F_{(1,18)}=11.93$ ,  $p=0.002$ ), qualified by an interaction ( $F_{(1,18)}=19.72$ ,  $p<0.001$ ). This interaction was driven by greater overlapping similarity relative to non-overlapping similarity for remote ( $t_{(18)}=4.13$ ,  $p<0.001$ ) but not recent ( $t_{(18)}=0.02$ ,  $p=0.98$ ) memories, and greater retrieval similarity for remote trials relative to recent trials that were overlapping ( $t_{(18)}=2.96$ ,  $p<0.001$ ) but not non-overlapping ( $t_{(18)}=0.22$ ,  $p=0.83$ ). Thus, posterior hippocampus looked qualitatively similar, but with stronger effects, than the results from the whole hippocampus.

In anterior hippocampus, there was a marginal main effect of time ( $F_{(1,18)}=3.42$ ,  $p=0.08$ ) and a main effect of overlap ( $F_{(1,18)}=7.71$ ,  $p=0.01$ ), qualified by an interaction ( $F_{(1,18)}=8.10$ ,  $p=0.01$ ). Consistent with posterior hippocampus and mPFC, retrieval similarity was significantly greater for overlapping relative to non-overlapping items for remote ( $t_{(18)}=3.06$ ,  $p=0.006$ ) but not recent ( $t_{(18)}=-0.34$ ,  $p=0.74$ ) memories. However, in contrast to these other regions, retrieval similarity in anterior hippocampus *decreased* over time for non-overlapping trials ( $t_{(18)}=-2.82$ ,  $p=0.01$ ) but not overlapping trials ( $t_{(18)}=-0.72$ ,  $p=0.48$ ).

We next asked whether the time-dependent reorganization across memories was selective to mPFC and hippocampus, or if other brain regions, specifically those in the classic retrieval network (Rugg and Vilberg, 2013), showed similar results. We chose posterior medial cortex (PMC), a critical region in the recollection network that has been shown to reinstate encoding patterns during retrieval (Bird et al., 2015; Kuhl et al., 2011; Long et al., 2016; Chen et al., 2017) to serve as a control region. Critically, we defined this region from the same contrast as the mPFC ROI. We applied the same 2 (Overlap: overlapping, non-overlapping)  $\times$  2 (Time: recent, remote) ANOVA to HC correct trials and found no effects or interactions (all  $F$ 's  $< 0.05$ , all  $p$ 's  $> 0.82$ , Figure 3B). Importantly, because we defined PMC from the same contrast as mPFC, this finding serves as an effective control and suggests that using this particular univariate contrast did not bias our multivariate results. When we expanded this region to include all voxels in PMC regardless of retrieval success, there were still no effects of time or overlap (all  $F$ 's  $< 0.237$ , all  $p$ 's  $> 0.14$ ). Thus, while PMC was more engaged during successful remote retrieval relative to unsuccessful attempts, patterns of activation in this region did not reflect overlap, as was observed in mPFC and hippocampus.

Prior work has demonstrated that univariate activation in mPFC and hippocampus changes with consolidation (Takashima et al., 2006; Sterpenich et al., 2007). Based on this, we examined whether overall univariate activation in these regions was related to, or could

explain, the changes in similarity that emerged over time. We found no significant differences in activation during successful retrieval over time (Figure S1B, S1C). Furthermore, activation did not reliably influence the relationship between retrieval similarity, overlap, and time (Table S1). Finally, all pair-wise tests of retrieval similarity were confirmed with non-parametric permutation tests (Table S2).

The cued retrieval task required successful retrieval of the scene that was associated with each target object. Thus, it is unclear whether the time-dependent changes in retrieval similarity track how strategic retrieval processes change over time, or whether they might still be evident without demands on associative retrieval. To examine this, we computed the same similarity analyses on object recognition trials, where reactivation of the overlapping scene is not required or necessary to perform the task. We found a similar but weaker pattern of effects in the hippocampus and mPFC (Figure S2, Table S3). This suggests that retrieval similarity in this experiment is sensitive to a mixture of information: signal corresponding perhaps more to the memory trace itself, as well as signal corresponding to the explicit retrieval of the corresponding scene image.

### Encoding similarity

In prior work using the associative inference (AB-BC) paradigm, memory representations in the hippocampus have been shown to be reinstated when encoding a new event that contains features from a prior event (Schlichting et al., 2014). While we did not observe difference in retrieval similarity across recent memories, there is still the possibility that this effect emerged during learning and contributed to the remote retrieval effects. To assess this possibility, we computed overlapping and non-overlapping similarity across encoding. We found evidence of pattern separation signals in hippocampus – specifically, greater non-overlapping similarity relative to overlapping similarity – and no effects in mPFC. Furthermore, encoding similarity did not relate to retrieval similarity on a trial-by-trial basis (Figure S3).

However, past work using similar approaches has found that other regions are able to represent overlapping memories during encoding (Xue et al., 2010; Ward et al., 2013; Xiao et al., 2017) or during retrieval immediately after (Kuhl and Chun, 2014). To find evidence for this in our own experiment, we examined pattern similarity in PPA, a ventral temporal region that codes for scene information (Epstein and Kanwisher, 1998), and in exploratory searchlights. We found that visually-sensitive regions exhibited greater similarity for overlapping trials relative to non-overlapping trials, with no reliable changes over time (Figure S4, Table S4). This hints at how different features of memories may be represented in distinct regions: regions that may be more sensitive to specific visual content show evidence of representational overlap at both recent and remote time points, but memory representations in the hippocampus and mPFC appear to undergo transformations and come to represent overlap in memory more strongly after a delay.

### Episodic reinstatement during retrieval

Past research has found that encoding-retrieval similarity (ERS), measured as the similarity between the neural pattern of a paired associate during encoding and the pattern evoked by

its successful retrieval, has been related to successful memory retrieval (Staresina et al., 2012; Ritchey et al., 2013; Tompariy et al., 2016). Our results so far show that memories become reorganized with time, raising the question of whether encoding patterns associated with each memory are still reinstated during remote retrieval despite such reorganization. To address this question, we computed ERS for each object-scene pair by correlating each trial's encoding pattern, averaged across the three encoding presentations, with the pattern evoked by its later retrieval ('same-memory ERS'). To identify the extent to which ERS reflects reinstatement signals that may be shared across different memories, we correlated the average correlation between the pattern of each retrieval trial and all patterns of encoding trials that shared the same scene ('same-scene ERS'). By comparing these two measures, we were able to isolate the extent to which memory-specific information is reinstated during retrieval (Figure 4A, left). Based on past findings that episodic memory representations are reinstated specifically in the right hippocampus (Mack and Preston, 2016; Tompariy et al., 2016), we examined whether neural patterns in right hippocampus remained sensitive to the reinstatement of individual memories after a week.

We first focused on remote retrieval. A 2 (Accuracy: HC correct, incorrect)  $\times$  2 (ERS: same-memory, same-scene) ANOVA revealed an interaction between accuracy and ERS ( $F_{(1,18)}=5.94$ ,  $p=0.03$ ), and a trend for a main effect of accuracy ( $F_{(1,18)}=3.59$ ,  $p=0.07$ ) but not ERS ( $F_{(1,18)}=0.42$ ,  $p=0.52$ ; Figure 4A, right). The interaction was driven by greater same-memory ERS for remembered pairs relative to forgotten pairs during remote retrieval ( $t_{(18)}=2.25$ ,  $p=0.04$ ), consistent with the past work described above. Furthermore, the interaction was driven by greater same-memory ERS relative to same-scene ERS in remembered pairs ( $t_{(18)}=2.47$ ,  $p=0.02$ ), suggesting that the right hippocampus exhibits memory-specific reinstatement at this time point.

To investigate whether there were changes in successful memory reinstatement over time, we computed all ERS values for HC correct trials into a 2 (Time: recent, remote)  $\times$  2 (ERS: same-memory, same-scene) ANOVA. We found a trending interaction between time and ERS ( $F_{(1,18)}=3.36$ ,  $p=0.08$ ), a trending main effect of ERS ( $F_{(1,18)}=4.14$ ,  $p=0.06$ ), and no effect of time ( $F_{(1,18)}=0.09$ ,  $p=0.77$ ). Like in the above paragraph, this interaction was driven by greater same-memory ERS than same-scene ERS for remote memories ( $t_{(18)}=2.47$ ,  $p=0.02$ ) but not recent memories ( $t_{(18)}=0.11$ ,  $p=0.91$ ). Interestingly, there was also a reliable decrease in same-scene ERS over time ( $t_{(18)}=2.68$ ,  $p=0.02$ ), but no reliable difference in same-memory ERS over time ( $t_{(18)}=-0.75$ ,  $p=0.46$ ).

The modulation of memory-specific ERS by remote memory in right hippocampus remained significant when accounting for variability in univariate activation across trials (Table S5) and was not driven by variability in the number of trials in each condition across participants (Table S6). By contrast, in left hippocampus, mPFC, and PMC, ERS was not modulated by remote memory success and did not differ between recent and remote retrieval (all  $p$ 's  $> 0.34$ ). When considered with the retrieval similarity, these findings suggest that the hippocampus continues to reinstate details of individual memories, while also representing structure across memories.



### Relationship between retrieval similarity and ERS

Interestingly, the hippocampus showed evidence both for memory-specific reinstatement and for the consolidation-dependent reorganization of related memories. To directly assess the relationship between these two effects in the hippocampus, we entered trial-level estimates of memory-specific ERS (same-memory – same-scene), time (recent, remote), and retrieval similarity (overlapping – non-overlapping) into a mixed-effects linear regression. For both anterior and posterior hippocampus, this model revealed an interaction between ERS and time (anterior:  $\chi^2=7.42$ ,  $p=0.006$ , posterior:  $\chi^2=8.66$ ,  $p=0.003$ , Figure 4B). Focusing on remote memory, we found an inverse relationship between ERS in right hippocampus and retrieval similarity in both anterior and posterior hippocampus (anterior:  $\chi^2=5.32$ ,  $p=0.02$ , posterior:  $\chi^2=5.44$ ,  $p=0.02$ ), such that trials with greater ERS in right hippocampus exhibited a smaller difference in retrieval similarity for overlapping relative to non-overlapping memories. This relationship was not apparent when applying the same model to recent memory (anterior:  $\chi^2<0.001$ ,  $p=0.99$ , posterior:  $\chi^2=0.085$ ,  $p=0.77$ ), and ERS in right hippocampus did not relate to restructuring in mPFC ( $\chi^2=0.016$ ,  $p=0.90$ ). These findings highlight a potential trade-off between memory-specific reinstatement, perhaps reflecting the fidelity of detailed episodic recovery, and consolidation-dependent memory restructuring evident across related memories. Such a trade-off suggests that that over time, memories whose patterns more closely match their initial encoding state may be less likely to be integrated with related memories.

### Relationship between retrieval similarity and rest connectivity

Prior work has shown that early indicators of memory consolidation can be measured during immediate post-encoding rest periods (Tambini et al., 2010; 2013; Schlichting and Preston, 2014; Tompary et al., 2015). Thus, we aimed to test to what extent post-encoding connectivity may be related to long-term memory reorganization. Our first approach was to query whether there were global changes in connectivity with the hippocampus and mPFC as a result of encoding. To do this, we conducted exploratory seed analyses by entering the average timecourses of hippocampal and mPFC activation into separate voxel-wise GLMs for each rest scan. When comparing post-encoding connectivity (averaged over the three post-encoding rest scans) against baseline pre-encoding connectivity, we found evidence for significant increases in hippocampal connectivity with regions including but not limited to: left middle frontal gyrus, left inferior frontal gyrus, and left lateral occipital cortex (Figure 5B, left). There were also increases in mPFC connectivity with anterior cingulate gyrus, left supramarginal gyrus, and others (Figure 5B, right). For a full list of regions identified in these two analyses, see Table S7.

Having established that encoding induced lingering changes in connectivity with these regions during rest, we developed a more targeted ROI-based approach to identify whether such encoding-related changes in rest connectivity related to changes in retrieval similarity over time. To do this, we correlated the timecourse of activation between the hippocampus (whole, anterior, and posterior) and mPFC before and after encoding for each participant (Figure 5A). We then related the change in connectivity after encoding (post-encoding – pre-encoding) to the difference in retrieval similarity (overlapping – non-overlapping) across participants. This approach revealed a positive correlation between mPFC-anterior

hippocampus connectivity and retrieval similarity in anterior hippocampus (Figure 5C). This relationship was present for remote memories ( $r_{(17)}=2.56$ ,  $p=0.02$ ) and not recent memories ( $r_{(17)}=-0.04$ ,  $p=0.96$ ). Specifically, participants with a greater increase in mPFC-anterior hippocampus connectivity immediately after learning exhibited a greater difference in similarity for overlapping versus non-overlapping memories (i.e. more restructuring) after a week. This relationship between connectivity and retrieval similarity was numerically but not significantly greater for remote memories relative to recent (William's Test:  $t_{(18)}=1.66$ ,  $p=0.11$ ). No other ROI pairs exhibited a relationship between connectivity and remote retrieval similarity (all  $p$ 's>0.26).

Together these exploratory findings suggest that immediate post-learning changes in connectivity may reflect a consolidation mechanism that plays an active role in shaping memories over time, in a way that prioritizes their commonalities.

## Discussion

In the present study, we examined how representations of individual memories are transformed with consolidation. We found that memory representations in medial prefrontal cortex and in the hippocampus came to represent commonalities across memories only after a period of consolidation. Specifically, the neural patterns evoked during retrieval of overlapping memories were more similar to each other relative to patterns evoked by non-overlapping memories. Critically, this was evident only for memories retrieved one week after encoding. At the same time, reinstatement of encoding patterns was still evident in right hippocampus at one week, as indexed by greater encoding-retrieval similarity (ERS) for object-scene pairs remembered with high confidence relative to forgotten pairs.

These findings provide evidence in support of the theory that cortical regions come to store an extracted and transformed version of episodic memory traces over time (McClelland et al., 1995; Winocur et al., 2010). To date, most consolidation research supporting this theory has shown that mPFC becomes increasingly engaged during remote versus recent retrieval, such that there is greater activation in mPFC and greater mPFC-hippocampal connectivity when retrieving consolidated memories relative to newly formed ones (Sterpenich et al., 2007, 2009, Takashima et al., 2006, 2007; Sweegers et al., 2014). Additionally, one study has shown that personal autobiographical memories can be successfully decoded in ventral mPFC, with greater accuracy for older compared to newer memories (Bonnici et al., 2012). Taken together, this past research provides compelling evidence of the growing involvement of cortical regions in storing and retrieving memories over the course of consolidation. However, little was known about whether this transformation is sensitive to the content of memories and whether memories with featural overlap are represented differently than memories without overlap. Here, we show that mPFC represents the central tendencies across episodic experiences, but only after a period of consolidation.

We also found greater similarity for overlapping memories relative to non-overlapping memories in the hippocampus, and, again, this difference was only present for remote memories. This finding is striking given theoretical and empirical work concerning the function of the hippocampus in memory consolidation. Most theories posit that the

hippocampus orthogonalizes incoming information in order to avoid interference from related past experiences (O'Reilly and McClelland, 1994). Indeed, there is now a growing body of evidence for pattern separation signals in the hippocampus during encoding (Leutgeb et al., 2007; Bakker et al., 2008; LaRocque et al., 2013; Favila et al., 2016; Chanales et al., 2017) and work showing that variability in neural patterns across repeated testing is related to long-term retention (Wirebring et al., 2015). In the present experiment, hippocampal similarity across encoding was lower for overlapping trials relative to non-overlapping trials, consistent with this theoretical and empirical work. On the other hand, Cohen and Eichenbaum's relational memory theory (1993) hypothesizes that the hippocampus may support access to related memories through relational 'nodes' or neural ensembles that link overlapping memories. Research consistent with this theory finds that retrieval of overlapping memories is associated with hippocampal univariate activation during transitive inference (Heckers et al., 2004; Greene et al., 2006) and associative inference tasks (Preston et al., 2004) with greater activity for trials with features that overlapped with prior encoding trials. One potential explanation for why this is not evident in our results is that the difference in the necessity or expectation of integration at encoding may dictate whether stimuli are integrated or separated (Richter et al., 2016). Another factor that may bias encoding processes towards integration is the memory strength of past overlapping episodes. For instance, in an associative inference design, Schlichting and colleagues (2015) found that neural patterns in a hippocampal cluster was biased towards integration only if the AB pairs were strongly learned in a separate block before participants were presented with BC pairs. Interestingly, in our study, encoding patterns in the hippocampus represented overlapping memories more distinctively than non-overlapping memories, but then after a week of consolidation, retrieval patterns came to reflect integration. More work is needed to reconcile discrepancies between separation and integration signals in the hippocampus during encoding and how and when such signals may shift over the course of consolidation.

What time-dependent mechanisms may support the restructuring of overlapping memories? One possibility is that through active consolidation mechanisms, such as coordinated replay of memories in hippocampus and cortex, the associative links across memories with shared features may become strengthened. This strength could emerge through the distribution and representation of these links in cortical regions (Figure 6, bold lines). While few studies to date have investigated what dictates the reactivation or replay of specific features of memories, it may be that the overlapping components across different memories are prioritized over one-time, episode-specific features. The findings from the present experiment provide preliminary evidence for such a mechanism, as measured by changes in connectivity during periods of awake rest. Specifically, we found that connectivity with the hippocampus and mPFC was strengthened after encoding, relative to a pre-encoding baseline. Further, across participants, the encoding-related change in connectivity between mPFC and anterior hippocampus related to the representation of overlap in anterior hippocampal after one week. This suggests that restructuring of memories over time may be driven in part by consolidation mechanisms.

Forgetting may also play a role. Specifically, forgetting of unique details that differentiate related events may also result in the merging of overlapping memories over time. Such

episode-unique features may be more likely to be represented in the hippocampus (Figure 6, faded dots). Selective forgetting of these details may be adaptive because it enables the creation of models of past experiences that are not over-fitted; rather, by selectively retaining overlapping information and forgetting episode-unique information, our past experiences can be sculpted into a less detailed but more generalizable model that can inform decisions about new experiences in the current environment (Richards and Frankland, 2017). However, the fact that we see immediate increases in post-encoding connectivity that relate to memory restructuring in hippocampus at one week suggests that measurable, active consolidation processes also play a role. Most likely, both consolidation mechanisms and selective forgetting work together to shape memories over time, but future work is needed to measure the how each of these processes separately contributes to time-dependent changes in long-term memory representations.

We also found evidence that neural patterns in right hippocampus, but not the mPFC, reflected the successful reinstatement of episodic memories after one week of consolidation. This finding extends the growing body of work using encoding-retrieval similarity (ERS) approaches to assess episodic reinstatement during memory retrieval in the hippocampus (Mack and Preston, 2016; Tompariy et al., 2016) as well as in sensory regions across cortex (Johnson et al., 2009; Staresina et al., 2012; Ritchey et al., 2013; Kuhl and Chun, 2014; Danker et al., 2016). Furthermore, by computing ERS in addition to similarity across retrieval, we were able to quantify different aspects of representational information for each memory. The comparison of these two measurements revealed an interesting trade-off within remote memories: episodes whose neural patterns during retrieval better matched their initial encoding experience exhibited lower similarity with other overlapping memories. This consolidation-dependent trade-off between episodic reinstatement and the merging of related memories with consolidation in the hippocampus raises interesting questions concerning how different elements of memories are shaped with consolidation.

Moreover, these two processes may operate independently, such that the relative strength of memory traces in hippocampus and cortex determine the elements of memory that can be retrieved, as predicted by trace transformation theory (Winocur et al 2010). For example, neural patterns in the hippocampus are able to support successful decoding of vivid and perceptually rich autobiographical memories for at least 10 years (Bonnici et al., 2013), suggesting that the hippocampus retains some differentiated episodic information well after encoding. At the same time, schema research provides additional evidence for the idea that stronger cortical representations come to represent generalized memories. Anatomically, the ventral aspect of the mPFC region reported in our experiment partially overlaps with several other clusters found to be involved in schema-supported processing (van Kesteren et al., 2010a, 2014). In this body of work, mPFC activation during encoding is enhanced for content belonging to a schema (van Kesteren et al., 2010b, 2014; Bein et al., 2014) and is modulated by the retrieval of schematic information (van Kesteren et al., 2010a). While these studies do not consider how schematic memory interacts with episodic reinstatement, recent neuropsychological studies have demonstrated that damage to ventral mPFC reduces the schematic influence of associated information during episodic memory tests (Warren et al., 2014; Spalding et al., 2015), suggesting that ventral mPFC plays a necessary role in modulating episodic memory reinstatement for recent experiences. Future work, ideally

using behavioral tests that are sensitive to the recovery of different elements of episodic memories, is necessary to investigate whether and how reinstatement of unique details in the hippocampus would interfere with the retrieval of more generalized, flexible information represented in the hippocampus and in cortex.

One limiting factor in our experiment stems from the comparison of high-confident correct memories retrieved immediately after encoding versus after a week delay. Memories retrieved immediately after encoding may be a mix of strong and weak memories, some of which would have been forgotten if they had been tested a week after encoding. In contrast, memories that were successfully retrieved after a week were likely limited to those that were the strongest and most enduring. With the present design, it is unclear whether the stronger memories retrieved immediately after encoding might also show evidence for overlap (e.g. exhibit greater overlapping versus overlapping similarity). If this were the case, an argument could be made that only traces of strong memories show representational overlap, without undergoing any time- or consolidation-dependent transformations. Alternatively, it may be that retrieval strength alone is not related to the representation of overlap, and instead some consolidation-dependent transformation is necessary for this representation to emerge. Indeed, one could argue that the strongest memories may be those with the most distinctive episodic content and hence be *less likely* to show overlap with other related memories. More research is needed to adjudicate between these two interpretations.

We found that over time, memory representations in anterior and posterior hippocampus diverged, such that after a week of consolidation, neural patterns of non-overlapping memories were less similar in anterior hippocampus, while neural patterns of overlapping memories were more similar in posterior hippocampus. This finding is difficult to reconcile with some recent findings implicating anterior hippocampus in memory integration, not separation. In humans, anterior hippocampus exhibits greater activation during the encoding (Shohamy and Wagner, 2008) and retrieval of overlapping associations (Heckers et al., 2004; Preston et al., 2004; Greene et al., 2006). Furthermore, neural patterns in anterior hippocampus have been shown to reflect learning of overlapping and sequential associations (Schapiro et al., 2012; Schlichting et al., 2015). In rodents, receptive fields in dorsal hippocampus are smaller than ones found in ventral hippocampus (Kjelstrup et al., 2008). Taken together, this research suggests that anterior hippocampus may be more involved in integrative computations while the posterior hippocampus represents specificity in the environment (Poppenk et al., 2013). However, our results are more consistent with theories that situate hippocampal computations within larger functional networks. Anatomically, perirhinal cortex (PRC) and parahippocampal cortex (PHC) show preferential connectivity with anterior and posterior hippocampus, respectively, through connections with lateral and medial entorhinal cortex (Burwell and Amaral, 1998; Suzuki and Amaral, 1994; Witter et al., 2000). A model of medial temporal lobe function based on this anatomy suggests that anterior hippocampus and PRC are part of a larger anterior-temporal network that is sensitive to specific items, concepts and their salience, while posterior hippocampus and PHC are part of a larger posterior-medial network that may be more sensitive to spatial and temporal contexts (Davachi, 2006; Ranganath and Ritchey, 2012). Within this framework, one may predict a greater sensitivity to distinct features of memories in anterior hippocampus, and a bias towards representing overlapping information in posterior

hippocampus, particularly when the overlapping information content are scenes - consistent with our findings. While these exploratory anterior-posterior differences should be interpreted with caution, they are suggestive of two opposing functions within the hippocampus that may drive the emergence of structure across memories: integration of related memories and separation of distinct memories.

In summary, the present results demonstrate that neural representations of related memories merge with consolidation. These findings raise new questions about the features of episodic memories that are prioritized by consolidation mechanisms. While more work on this topic needed, we suggest that the reorganization of overlapping memory representations may play an important role in the creation of structured general knowledge.

## Star Methods

### Contact for Reagent and Resource Sharing

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Lila Davachi (ld44@nyu.edu).

### Experimental Model and Subject Details

**Participants**—Twenty-two students from New York University (10 female, mean age: 26.8, range: 21 – 34) participated in the experiment. All participants were right-handed native English speakers with normal or corrected to normal vision. The New York University Institutional Review Board approved all consent protocols.

### Method Details

**Experiment design**—Participants completed two fMRI sessions separated by one week. In the first session, participants were scanned while encoding pairs of objects and scenes, and then performing object recognition and scene recall tests for half of the pairs. In the second session one week later, they completed the same recognition and recall tests on the other half of the pairs (Figure 1A).

**Encoding**—Participants were presented with images of 128 objects, each paired with one of four scenes: beach, city, bedroom or jungle (Figure 1B). They were explicitly instructed to associate each object-scene pair by vividly imagining the object interacting in the scene. The object-scene pairings were randomized across participants, and the order of the pairs was pseudo-randomized such that no back-to-back pair shared the same scene. Participants studied each pair three times across encoding, split across six encoding scans. Each presentation of all 128 pairs was pseudo-randomly divided over two 10.5 m scans, such that the number of trials containing each scene associate was equated across the two scans. Participants viewed each pair for 3 s, and then rated the vividness of their mental image on a scale of 1 to 4 (1: 'not vivid', 4: 'very vivid') using an MRI-compatible button box. The response mappings were counterbalanced across participants. Participants were given the option of pressing the thumb button to indicate that they failed to create a vivid scenario of the object in the scene. The response window lasted for 2 s and was followed by a jittered fixation period lasting 3, 4.5 or 6 s.

**Rest scans**—The experiment began with an 8 m baseline rest scan, followed by the first presentation of the pairs across two encoding scans. Each of the three presentations of the stimulus set was followed by another 8 m resting state scan, for a total of 6 encoding scans interleaved with 4 rest scans. During the rest scans, participants fixated a small black dot in the center of a gray screen and were instructed to remain awake and think about whatever they like.

**Object recognition**—After the final rest scan, participants completed an object recognition test in two 10.5 m scans. They viewed 64 of the encoded objects intermixed with 64 novel foils and were asked to endorse each object as old or new. The five response options included a measure of confidence ('confident old', 'unsure old', 'unsure new', 'confident new', 'don't know') and the response mappings were counter-balanced across participants. On each trial, the object cue was presented for 3 s, and then the response options were displayed during a 2 s window. Each trial was followed by a jittered fixation period ranging from 3 to 6 s. All trials included in the retrieval and recognition similarity findings were correctly recognized during this test (with trials included in retrieval similarity analyses requiring high-confident scene retrieval in addition to successful recognition).

**Scene recall**—The object recognition task was followed by the scene recall task, which comprised two 6 m scans. Participants viewed 64 of the encoded objects and were asked to choose which of the four scenes had been associated with each object. As in the object recognition, the stimulus presentation lasted for 3 s. During the following 2 s response window, participants viewed a prompt with the response options ('beach', 'city', 'bedroom', 'jungle', 'don't know'), which were counter-balanced across participants. After this response, participants were prompted to judge the confidence of their choice on a scale of 1 to 4 (1: 'not confident', 4: 'very confident'). These responses were collapsed into two bins (1-2: low-confident, 3-4: high-confident). This second response window lasted for 2 s and was followed by a jittered fixation period ranging from 3 to 6 s.

**Remote retrieval session**—Participants returned to the scanner one week later and completed the object recognition and scene recall tasks on the other 64 object-scene pairs that had not been tested in the first session. All timing, test order, and stimulus presentation parameters were identical to the memory tests from the first session.

Participants then completed two 10 m localizer scans. During the localizer scans, participants viewed 5 blocks each of faces, scenes, objects and scrambled images. Each block contained 12 images that appeared on screen for 1.5 s. Two out of every 12 images repeated back-to-back, and participants were instructed to press a button when they noticed an immediate repetition of any image. The order of the blocks was randomly generated for each participant and each block was separated by a 12 s fixation period.

**fMRI parameters**—All scanning was performed using a 3T Siemens Allegra MRI system with a whole-head coil. Visual stimuli were projected onto a screen that was viewed through a mirror attached to the participant's head coil. Functional echo-planar imaging (EPI) scans were oriented to intersect the anterior and posterior commissures (2000-ms TR, 15-ms TE, flip angle=82°, FOV=192×240, 34 slices, 3-mm isotropic voxels). For both sessions, a

customized calibration scan was collected using the same slice prescription as the EPI scans for use as an in-plane spin-density image as well as an estimate of any inhomogeneities in the magnetic field. At the end of the second scan, a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence (1 mm isotropic voxels, 176 sagittal slices) was collected.

**Preprocessing**—All scans underwent the same preprocessing steps using FSL (FEAT: <http://www.fmrib.ox.ac.uk/fsl>). The first 6 volumes of each EPI were discarded to allow for scanner stabilization. Then, each scan was slice-time corrected, realigned to correct for motion within each run, and smoothed. Data to be used for similarity analyses were smoothed with a 3mm FWHM Gaussian kernel. For data to be used in univariate or connectivity analyses, a 6mm kernel was applied. The data were high-pass filtered at 0.01 Hz to remove low-frequency drifts in signal.

**Regions of interest**—We identified mPFC and PMC using a univariate contrast at remote retrieval, which identified voxels with greater activation for high-confident (HC) correct relative to incorrect trials (see Supplemental Data). PMC was further constrained by masking the voxels in this contrast by the PMC region defined by a probabilistic atlas (Shirer et al., 2012; Chen et al., 2016). Bilateral hippocampus was anatomically defined using FSL's automatic subcortical segmentation protocol (FIRST). The hippocampus was segmented along its long axis by dividing the number of coronal slices in each hemisphere into three sections. The most anterior third of the coronal slices was designated as anterior hippocampus, and the most posterior third of the coronal slices was designated as posterior hippocampus. The localizer scans were used to functionally define PPA (see Localizer section of Methods). All ROIs were resampled, masked to exclude voxels outside of the brain, and aligned with the functional volumes.

### Quantification and Statistical Analysis

**Retrieval activation**—To investigate how univariate activation related to successful retrieval over time, the retrieval scans were entered into voxel-wise GLMs (FEAT). Trials were modeled with 3 s boxcars locked to the onset of each trial and convolved with FEAT's hemodynamic response function (HRF). Three regressors were included in each GLM to account for source recall accuracy: (1) high-confident correct responses, (2) low-confident correct responses, and (3) incorrect, missing, or 'don't know' responses. To account for potential artifacts from head motion, the 6 motion regressors derived from the motion correction procedure were included in each GLM along with their temporal derivatives, as well as stick function regressors to account for sudden head movements. These stick functions were generated by FSL's Motion Outliers algorithm, which identifies large displacements in head position by measuring the difference in intensity between each volume and the preceding volume.

The resulting statistical maps were aligned to MNI space by concatenating a rigid-body transformation from each functional run to the participant's MPRAGE with a non-linear transformation from the MPRAGE to MNI space with a 10 mm warp resolution. These aligned images were entered into group-level analyses and corrected for multiple



comparisons using cluster-mass thresholding ( $p < 0.05$ , cluster-forming threshold  $z = 2.3$ ). Average  $t$  values were extracted across all ROIs separately for the recent and remote retrieval sessions.

**Localizer**—Two localizer scans were used to functionally define PPA. Each scan was entered into a voxel-wise GLM with regressors corresponding to the four categories of stimuli, their temporal derivatives, and the same motion regressors that were derived for retrieval scans. Each of the four regressors of interest was modeled as 18 s boxcars locked to the onset of each block and convolved with Feat's HRF. Parameter estimates from the two scans were averaged together. Bilateral PPA was defined from a contrast of scenes > faces within each participant. Each ROI was created by growing an 8-mm sphere around the most scene-selective voxel in each hemisphere of the posterior parahippocampal gyrus.

**Pattern similarity estimates**—All preprocessed encoding, recognition, and retrieval scans were modeled in separate GLMs in each participant's native space. For each trial, a separate regressor was generated, using a 3 s boxcar at the onset of the trial and convolved with Feat's HRF (Mumford et al., 2012). This resulted in 2 GLMs each for remote retrieval, recent retrieval, recent recognition, and remote recognition, each with 32 boxcar regressors, and 6 GLMs for encoding, each with 64 boxcar regressors. For recognition scans, an extra regressor was included, with 3 s boxcars that corresponded to the onsets of all novel foils in each run. The equivalent regressors that were included to account for head motion in the univariate activation GLMs were included in these models as well.

This procedure resulted in a separate map of  $t$  values for each trial during encoding, recognition, and retrieval. Then, for each trial, the resulting spatial pattern of activity across each ROI was extracted into a vector and z-scored. Similarity between different vectors was computed using Pearson correlations. All correlations were Fisher-transformed prior to statistical testing. Four measures of pattern similarity were computed: retrieval similarity (correlations amongst retrieval patterns), recognition similarity (correlations amongst recognition patterns), encoding similarity (correlations amongst encoding patterns), and encoding-retrieval similarity (correlations between encoding patterns and retrieval patterns).

**Retrieval similarity**—Retrieval similarity was computed for every HC correct retrieval trial whose corresponding object was also correctly recognized. Each trial's retrieval vector was correlated with (1) the retrieval vectors of all other objects with a shared scene (overlapping similarity), and (2) the retrieval vectors of all objects with a different scene (non-overlapping similarity). The resulting  $r$  values were then averaged to create one measure of overlapping similarity and one measure of non-overlapping similarity for each trial. To avoid inflated correlations as a function of temporal proximity with each scan (Mumford et al., 2014), correlations were limited to trials occurring in different scans.

These correlations were calculated separately for objects presented in each retrieval period (recent and remote). Importantly, only objects that were successfully recognized and whose scenes were recalled with high confidence were included in this analysis. Low-confident correct trials were excluded because at the group level, low-confident responses were more likely to be incorrect than correct during remote retrieval ( $t_{(18)} = -3.39$ ,  $p < 0.01$ ), and were

equally likely to be correct relative to incorrect during recent retrieval ( $t_{(18)}=0.36$ ,  $p=0.92$ ), suggesting that some unknown proportion of the low-confident correct responses were not based on intact memory and instead were guesses. Recognition similarity was computed in the same way as retrieval similarity, but instead included all recognition trials where participants endorsed the target object as 'old' with high confidence.

**Encoding similarity**—Encoding similarity was computed in a similar fashion as retrieval similarity. First, for each trial, the three vectors corresponding to each encoding presentation were averaged into one pattern. Then, overlapping and non-overlapping similarity scores were computed separately for each trial without considering subsequent memory or whether it would be tested immediately or after one week. In other words, the average pattern evoked by an object-scene pair during encoding was correlated with (1) the average encoding vectors of all objects with a shared scene and (2) the average encoding vectors of all objects with a different scene.

**Encoding-retrieval similarity**—Encoding-retrieval similarity (ERS; Xue et al., 2010; Staresina et al., 2012; Ritchey et al., 2013) was computed for every object-scene pair. To do this, each trial's average encoding vector was correlated with its corresponding retrieval vector, resulting in one ERS value for each object-scene pair ('same-memory ERS'). The correlations were then sorted by retrieval period (recent and remote) and memory performance (HC correct retrieval and correct recognition, versus incorrect recognition and/or retrieval) and averaged across trials within each condition for each participant.

As a control comparison, we computed ERS across trials with different objects but shared scenes ('same-scene ERS'). To do this, each trial's retrieval vector was correlated with the average encoding vector of each trial that shared the same scene and memory status (HC correct retrieval and correct recognition, versus incorrect retrieval). These correlations were averaged together to create a measure of same-scene ERS for each trial, then averaged across trials within each condition for each participant. To index the extent to which ERS captures the reinstatement of trial-specific information, same-scene ERS was subtracted from same-memory ERS for each trial, and for each participant ('memory-specific ERS').

**Searchlight analyses**—To explore the influence of overlap and time on pattern similarity outside of our *a priori* ROIs, we conducted whole-brain searchlight analyses on the four retrieval scans. These analyses were conceptually similar to the main retrieval similarity analysis, only conducted for spheres corresponding to each voxel in the brain instead of targeted ROIs. First, correlations between overlapping and non-overlapping trials were computed in each participant's native space. For each retrieval trial, a sphere with a 3-voxel radius was moved through every voxel throughout the brain (64 voxels per sphere). In spheres where at least 32 voxels were situated within the brain, voxels in the sphere were extracted and reshaped into a vector. Then, for each sphere, each the pattern corresponding to each retrieval trial was correlated with all retrieval patterns of trials studied with the same scene, and all retrieval patterns of trials studied with a different scene. These correlations were averaged to create a measure of overlapping and non-overlapping similarity for each trial. These measures were assigned to the middle voxel within each sphere, and then averaged across trials, resulting in whole-brain maps of overlapping and non-overlapping

similarity for each participant at each retrieval period. As with the ROI analysis, only trials whose objects were correctly recognized and whose associated scenes were remembered with high confidence were included, and only correlations of trials across runs were considered.

The resulting Fisher-transformed maps were aligned to MNI space with the same set of transformations used for the univariate retrieval analyses. Because the increased spatial blurring of the maps, caused by computing similarity across highly overlapping spheres, group-level statistics were conducted using FSL's `randomise` function. Reported clusters were identified using Threshold-Free Cluster Enhancement (TFCE) and were controlled for family-wise error rate ( $p < 0.05$ ).

**Rest connectivity analysis**—The rest scans were used to measure encoding-related changes in functional connectivity, as indexed by low-frequency correlations between ROI pairs (Albert et al., 2009; Tambini et al., 2010; Tompary et al., 2015). Preprocessed data from the four rest scans were entered into separate GLMs to model nuisance signals. As with all other scans, 6 motion regressors, their temporal derivatives, and stick functions accounting for sudden head movements were included. Additional regressors were included to account for nuisance signals from white matter tissue cerebral spinal fluid (CSF). To create these regressors, each participant's MPRAGE was segmented into separate masks comprising gray matter, white matter and CSF, using FSL's `FAST` function. The gray matter and CSF masks were aligned to each participant's functional volumes and then eroded using FSL's `fslmaths` function, ensuring that these masks did not contain voxels that partially overlapped with gray matter. Then, the average timecourse across all voxels in each mask was extracted from the preprocessed rest scans. These timecourses were entered into each run's GLM along with their temporal derivatives.

The residuals of these GLMs were bandpass-filtered, leaving signal ranging from 0.01 and 0.1 Hz, which is the frequency range known to correspond to correlations between gray matter regions in functional neuroimaging data (Cordes et al., 2001). Then, average mean time-course for every volume in each rest run was extracted for each ROI. These timecourses were then correlated (Pearson correlation), Fisher transformed, and entered into statistical tests.

**Statistical tests**—All correlations were Fisher transformed before being submitted to statistical tests. For the majority of group-level comparisons, where there was sufficient data in each bin for each participant, repeated-measures ANOVAs and paired t-tests were used to characterize the data. These statistics were replicated using non-parametric permutation tests to account for the different numbers of correlations that comprise the overlapping and non-overlapping similarity scores for each trial, and for the different numbers of trials that were remembered and forgotten in the case of the ERS analyses (see Supplemental Data). Three participants were excluded from all similarity analyses due to an insufficient number ( $< 10$ ) of remote memories that were both correctly recognized and whose scenes were remembered with high confidence.

Trial-level relationships between similarity measures or between similarity and univariate activation were tested with mixed-effects linear regressions using the lme4 package in R (<http://cran.r-project.org/web/packages/lme4/>). Significance was determined using model comparisons, resulting in  $\chi^2$  values and corresponding  $p$  values, or with likelihood ratio tests. Intercepts and slopes for each participant were specified as random effects, and model comparisons were conducted to determine which experimental variables, if any, were included as by-participant random factors.

**Non-parametric tests of retrieval similarity**—Each trial's overlapping similarity and non-overlapping similarity consisted of an average measure computed across pair-wise correlations with patterns from different trials. The number of pair-wise correlations for each trial varied over time and by memory performance across participants. Because correlations were only computed between trials across different retrieval runs, each retrieval trial was correlated with up to 8 patterns from overlapping trials and up to 24 patterns from non-overlapping trials, depending on how many of those trials were remembered with high confidence. To ensure that the variability in the number of correlations used for each trial cannot explain the interaction between overlap and time relating to retrieval similarity, we developed two control analyses.

First, we computed a permutation test by creating a null distribution of retrieval similarity for each participant, while retaining the number of comparisons entered into each trial's overlapping and non-overlapping similarity measure. To do this, we shuffled each run's vector of patterns across all trials within the run, such that each trial's pattern was assigned to a different trial's scene and memory status. We repeated this procedure 10,000 times per participant and re-computed overlapping and non-overlapping similarity for each HC correct trial at each permutation. These values were averaged across trials for each participant, resulting in null distributions of overlapping and non-overlapping similarity for each retrieval period. The values were then subtracted to create a null distribution of the difference between overlapping and non-overlapping similarity. We then calculated the true difference in overlapping and non-overlapping similarity for each participant and computed the z-score of the true difference relative to that participant's distribution of shuffled differences. We then submitted the z-scores to t-tests against zero for each retrieval session, where a reliable difference above zero indicates that across participants, the true difference between overlapping and non-overlapping similarity is greater than the shuffled distributions across participants.

Second, we developed a sub-sampling procedure to reduce the number of comparisons used to calculate non-overlapping similarity to match the number used to calculate overlapping similarity for each participant. Specifically, for each participant and retrieval session, the number of comparisons originally used to compute overlapping similarity for a given trial was identified, and then the same number of non-overlapping trials were randomly drawn to compute non-overlapping similarity. This procedure was repeated 10,000 times to compute a distribution of non-overlapping similarity on that trial. These distributions were averaged across trials separately for recent and remote retrieval, resulting in a distribution of non-overlapping similarity for each participant at each time-point. Then, each participant's true overlapping similarity score was computed as a z-score relative to that participant's

distribution of non-overlapping similarity. These z-scores were submitted to pair-wise t-tests against zero for each retrieval session, where a significant difference from zero indicates that true overlapping similarity is robustly different than the distribution of non-overlapping similarity.

**Non-parametric test of ERS**—When computing ERS as a function of remote memory in the right hippocampus, the number of HC correct and incorrect trials entered into the analysis varied across participants. We implemented a non-parametric test to confirm that the variability in the number of trials used to compute ERS across participants could not explain the modulation of ERS by memory. We generated null distributions of ERS for each condition within participants by shuffling each participant's memory performance across runs 10,000 times and re computing ERS separately for HC correct and incorrect trials. We then computed the z-score of each participant's true difference in ERS between HC correct and incorrect trials, relative to the shuffled distribution of differences. This z-score was submitted to a t-test against zero, where a significant difference above zero would indicate that the true difference in ERS between HC correct and incorrect trials is reliably greater than the difference computed using shuffled values.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to dedicate this paper to the memory of Howard Eichenbaum, a fearless pioneer and leader in the study of hippocampal function, always a mentor and, above all, a friend. We will miss your warmth, intellectual generosity and excitement for scientific discovery and debate. We thank Avi Chanale for help with analyses, David Clewett for comments on the manuscript, and Brice Kuhl for helpful discussions. This work was supported by the National Institute of Mental Health Grant MH074692 (L.D.), Dart Neuroscience (L.D.), and NSF Graduate Research Fellowship Program (A.T.).

## References

- Albert NB, Robertson EM, Miall RC. The Resting Human Brain and Motor Learning. *Curr Biol.* 2009; 19:1023–1027. [PubMed: 19427210]
- Alvarez P, Squire LR. Memory consolidation and the medial temporal lobe: A simple network model. *Proc Natl Acad Sci.* 1994; 91:7041–7045. [PubMed: 8041742]
- Bakker A, Kirwan CB, Miller M, Stark CEL. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science.* 2008; 319:1640–1642. [PubMed: 18356518]
- Batterink LJ, Paller KA. Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation. *Brain Lang.* 2017; 167:83–93. [PubMed: 26443322]
- Bein O, Reggev N, Maril A. Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia.* 2014; 64:320–330. [PubMed: 25301385]
- Bird CM, Keidel JL, Ing LP, Horner AJ, Burgess N. Consolidation of complex events via reinstatement in posterior cingulate cortex. *J Neurosci.* 2015; 35:14426–14434. [PubMed: 26511235]
- Bonnici HM, Chadwick MJ, Lutti A, Hassabis D, Weiskopf N, Maguire EA. Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. *J Neurosci.* 2012; 32:16982–16991. [PubMed: 23175849]
- Bonnici HM, Chadwick MJ, Maguire EA. Representations of recent and remote autobiographical memories in hippocampal subfields. *Hippocampus.* 2013; 23:849–854. [PubMed: 23749406]

- Burwell RD, Amaral DG. Perirhinal and postrhinal cortices of the rat: Interconnectivity and connections with the entorhinal cortex. *J Comp Neurol.* 1998; 391:293–321. [PubMed: 9492202]
- Chanales AJ, Oza A, Favila SE, Kuhl BA. Overlap among spatial memories triggers repulsion of hippocampal representations. *bioRxiv.* 2017:99226.
- Chen J, Honey CJ, Simony E, Arcaro MJ, Norman KA, Hasson U. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cereb Cortex.* 2016; 26:3428–3441. [PubMed: 26240179]
- Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci.* 2017; 20:115–125. [PubMed: 27918531]
- Cohen, N., Eichenbaum, H. *Memory, Amnesia, and the Hippocampal System.* Cambridge (MA): The MIT Press; 1993.
- Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME. Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. *Am J Neuroradiol.* 2001; 22:1326–1333. [PubMed: 11498421]
- Danker JF, Tompary A, Davachi L. Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cereb Cortex* bhw146. 2016
- Davachi L. Item, context and relational episodic encoding in humans. *Curr Opin Neurobiol.* 2006; 16:693–700. [PubMed: 17097284]
- Davachi L, Mitchell JP, Wagner AD. Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci.* 2003; 100:2157–2162. [PubMed: 12578977]
- Durrant SJ, Taylor C, Cairney S, Lewis PA. Sleep-dependent consolidation of statistical learning. *Neuropsychologia.* 2011; 49:1322–1331. [PubMed: 21335017]
- Durrant SJ, Cairney SA, Lewis PA. Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb Cortex.* 2013; 23:2467–2478. [PubMed: 22879350]
- Eichenbaum H. The hippocampus and mechanisms of declarative memory. *Behav Brain Res.* 1999; 103:123–133. [PubMed: 10513581]
- Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP. Human relational memory requires time and sleep. *Proc Natl Acad Sci.* 2007; 104:7723–7728. [PubMed: 17449637]
- Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature.* 1998; 392:598–601. [PubMed: 9560155]
- Favila SE, Chanales AJH, Kuhl BA. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat Commun.* 2016; 7:11066. [PubMed: 27925613]
- Frankland PW, Bontempi B. The organization of recent and remote memories. *Nat Rev Neurosci.* 2005; 6:119–130. [PubMed: 15685217]
- Greene AJ, Gross WL, Elsinger CL, Rao SM. An fMRI analysis of the human hippocampus: Inference, context, and task awareness. *J Cogn Neurosci.* 2006; 18:1156–1173. [PubMed: 16839289]
- Harand C, Bertran F, La Joie R, Landeau B, Mézenge F, Desgranges B, Peigneux P, Eustache F, Rauchs G. The hippocampus remains activated over the long term for the retrieval of truly episodic memories. *PLoS ONE.* 2012; 7:e43495. [PubMed: 22937055]
- Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D. Hippocampal activation during transitive inference in humans. *Hippocampus.* 2004; 14:153–162. [PubMed: 15098721]
- Johnson JD, McDuff SGR, Rugg MD, Norman KA. Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron.* 2009; 63:697–708. [PubMed: 19755111]
- van Kesteren MTR, Rijpkema M, Ruiters DJ, Morris RGM, Fernández G. Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. *J Cogn Neurosci.* 2014; 26:2250–2261. [PubMed: 24702454]
- van Kesteren, MTR, van Rijpkema, M., Ruiters, DJ., Fernández, G. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *J Neurosci.* 2010a; 30:15888–15894. [PubMed: 21106827]

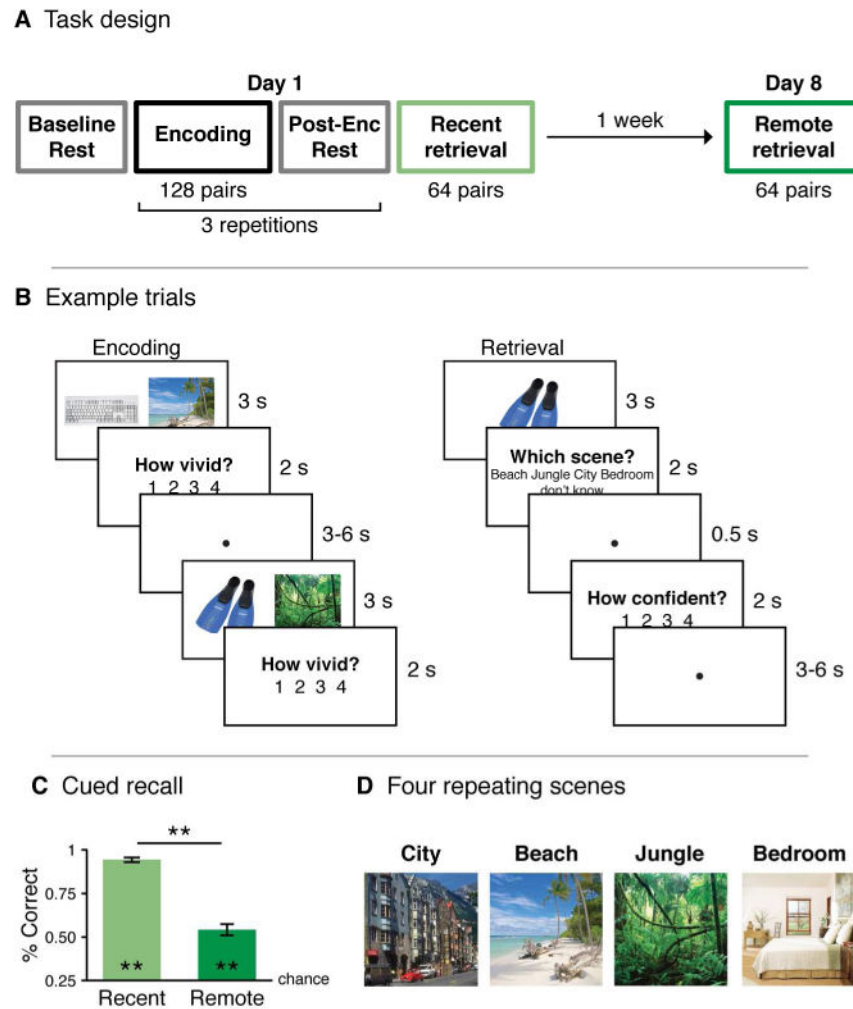
- van Kesteren, MTRvan, Fernández, G., Norris, DG., Hermans, EJ. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci.* 2010b; 107:7550–7555. [PubMed: 20363957]
- Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser MB. Finite scale of spatial representation in the hippocampus. *Science.* 2008; 321:140–143. [PubMed: 18599792]
- Kuhl BA, Chun MM. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J Neurosci.* 2014; 34:8051–8060. [PubMed: 24899726]
- Kuhl BA, Shah AT, DuBrow S, Wagner AD. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat Neurosci.* 2010; 13:501–506. [PubMed: 20190745]
- Kuhl BA, Rissman J, Chun MM, Wagner AD. Fidelity of neural reactivation reveals competition between memories. *Proc Natl Acad Sci.* 2011; 108:5903–5908. [PubMed: 21436044]
- Kumaran D, McClelland JL. Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychol Rev.* 2012; 119:573–616. [PubMed: 22775499]
- LaRocque KF, Smith ME, Carr VA, Withoft N, Grill-Spector K, Wagner AD. Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J Neurosci.* 2013; 33:5466–5474. [PubMed: 23536062]
- Lau H, Tucker MA, Fishbein W. Daytime napping: Effects on human direct associative and relational memory. *Neurobiol Learn Mem.* 2010; 93:554–560. [PubMed: 20176120]
- Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science.* 2007; 315:961–966. [PubMed: 17303747]
- Long NM, Lee H, Kuhl BA. Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *J Neurosci.* 2016; 36:12677–12687. [PubMed: 27821577]
- 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. [PubMed: 26702775]
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev.* 1995; 102:419–457. [PubMed: 7624455]
- 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. [PubMed: 21924359]
- Mumford JA, Davis T, Poldrack RA. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage.* 2014; 103:130–138. [PubMed: 25241907]
- Nadel L, Moscovitch M. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol.* 1997; 7:217–227. [PubMed: 9142752]
- Nadel L, Samsonovich A, Ryan L, Moscovitch M. Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus.* 2000; 10:352–368. [PubMed: 10985275]
- O'Reilly RC, McClelland JL. Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus.* 1994; 4:661–682. [PubMed: 7704110]
- Pavlidis C, Winson J. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J Neurosci.* 1989; 9:2907–2918. [PubMed: 2769370]
- Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. Long-axis specialization of the human hippocampus. *Trends Cogn Sci.* 2013; 17:230–240. [PubMed: 23597720]
- Preston A, Eichenbaum H. Interplay of hippocampus and prefrontal cortex in memory. *Curr Biol.* 2013; 23:R764–R773. [PubMed: 24028960]
- Preston AR, Shrager Y, Dudukovic NM, Gabrieli JDE. Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus.* 2004; 14:148–152. [PubMed: 15098720]
- Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci.* 2012; 13:713–726. [PubMed: 22992647]

- Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M. Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*. 2003; 42:2–13.
- Richards BA, Frankland PW. The Persistence and Transience of Memory. *Neuron*. 2017; 94:1071–1084. [PubMed: 28641107]
- Richards BA, Xia F, Santoro A, Husse J, Woodin MA, Josselyn SA, Frankland PW. Patterns across multiple memories are identified over time. *Nat Neurosci*. 2014; 17:981–986. [PubMed: 24880213]
- Richter FR, Chanales AJH, Kuhl BA. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage*. 2016; 124, Part A:323–335. [PubMed: 26327243]
- Ritchey M, Wing EA, LaBar KS, Cabeza R. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb Cortex*. 2013; 23:2818–2828. [PubMed: 22967731]
- Rugg MD, Vilberg KL. Brain networks underlying episodic memory retrieval. *Curr Opin Neurobiol*. 2013; 23:255–260. [PubMed: 23206590]
- Ryan JD, Althoff RR, Whitlow S, Cohen NJ. Amnesia is a deficit in relational memory. *Psychol Sci*. 2000; 11:454–461. [PubMed: 11202489]
- Schapiro AC, Kustner LV, Turk-Browne NB. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol CB*. 2012:1622–1627. [PubMed: 22885059]
- Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil Trans R Soc B*. 2017; 372:20160049. [PubMed: 27872368]
- Schlichting ML, Preston AR. Memory reactivation during rest supports upcoming learning of related content. *Proc Natl Acad Sci*. 2014; 111:15845–15850. [PubMed: 25331890]
- Schlichting ML, Preston AR. Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiol Learn Mem*. 2016; 134:91–106. [PubMed: 26608407]
- Schlichting ML, Zeithamova D, Preston AR. CA1 subfield contributions to memory integration and inference. *Hippocampus*. 2014; 24:1248–1260. [PubMed: 24888442]
- 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; 25:8151.
- Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. Decoding subject driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex*. 2012; 22:158–165. [PubMed: 21616982]
- Shohamy D, Wagner AD. Integrating memories in the human brain: Hippocampal midbrain encoding of overlapping events. *Neuron*. 2008; 60:378–389. [PubMed: 18957228]
- Spalding KN, Jones SH, Duff MC, Tranel D, Warren DE. Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *J Neurosci*. 2015; 35:15746–15751. [PubMed: 26609165]
- Staresina BP, Davachi L. Differential encoding mechanisms for subsequent associative recognition and free recall. *J Neurosci Off J Soc Neurosci*. 2006; 26:9162–9172.
- Staresina BP, Davachi L. Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. *J Cogn Neurosci*. 2008; 20:1478–1489. [PubMed: 18303974]
- Staresina BP, Henson RNA, Kriegeskorte N, Alink A. Episodic reinstatement in the medial temporal lobe. *J Neurosci*. 2012; 32:18150–18156. [PubMed: 23238729]
- Sterpenich V, Albouy G, Boly M, Vandewalle G, Darsaud A, Balteau E, Dang-Vu TT, Desseilles M, D'Argembeau A, Gais S, et al. Sleep-related hippocampo-cortical interplay during emotional memory recollection. *PLoS Biol*. 2007; 5:e282. [PubMed: 17958471]
- Sterpenich V, Albouy G, Darsaud A, Schmidt C, Vandewalle G, Vu TTD, Desseilles M, Phillips C, Degueldre C, Balteau E, et al. Sleep promotes the neural reorganization of remote emotional memory. *J Neurosci*. 2009; 29:5143–5152. [PubMed: 19386910]

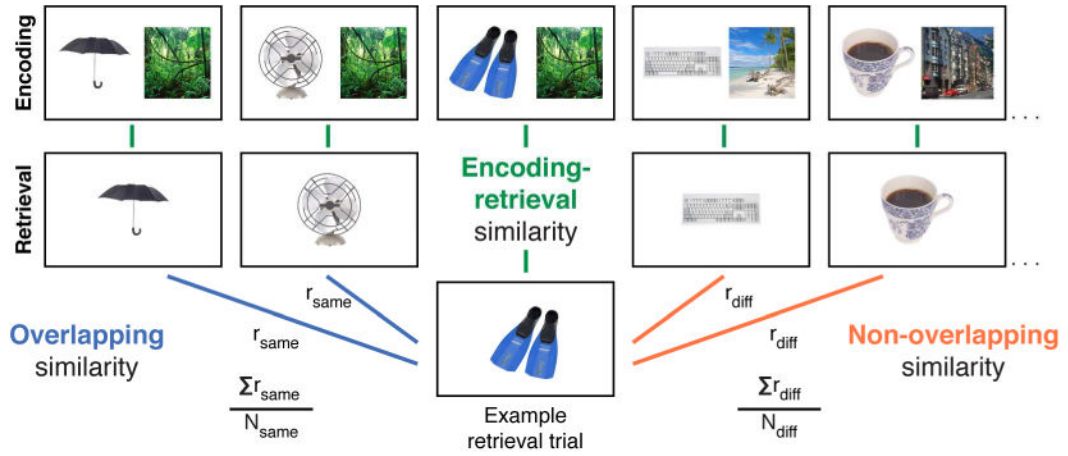


- Suzuki WA, Amaral DG. Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J Neurosci.* 1994; 14:1856–1877. [PubMed: 8126576]
- Sweegers CCG, Talamini LM. Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex.* 2014; 59:49–61. [PubMed: 25129237]
- Sweegers CCG, Takashima A, Fernández G, Talamini LM. Neural mechanisms supporting the extraction of general knowledge across episodic memories. *NeuroImage.* 2014; 87:138–146. [PubMed: 24215973]
- Takashima A, Petersson KM, Rutters F, Tendolkar I, Jensen O, Zwartz MJ, McNaughton BL, Fernández G. Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proc Natl Acad Sci U S A.* 2006; 103:756–761. [PubMed: 16407110]
- Takashima A, Nieuwenhuis ILC, Rijpkema M, Petersson KM, Jensen O, Fernández G. Memory trace stabilization leads to large-scale changes in the retrieval network: A functional MRI study on associative memory. *Learn Mem.* 2007; 14:472–479. [PubMed: 17622649]
- Takashima A, Nieuwenhuis ILC, Jensen O, Talamini LM, Rijpkema M, Fernández G. Shift from hippocampal to neocortical centered retrieval network with consolidation. *J Neurosci.* 2009; 29:10087–10093. [PubMed: 19675242]
- Tambini A, Davachi L. Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proc Natl Acad Sci.* 2013; 110:19591–19596. [PubMed: 24218550]
- Tambini A, Ketz N, Davachi L. Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron.* 2010; 65:280–290. [PubMed: 20152133]
- Tompary A, Duncan K, Davachi L. Consolidation of associative and item memory is related to post-encoding functional connectivity between the ventral tegmental area and different medial temporal lobe subregions during an unrelated task. *J Neurosci.* 2015; 35:7326–7331. [PubMed: 25972163]
- Tompary A, Duncan K, Davachi L. High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. *Hippocampus.* 2016; 26:995–1007. [PubMed: 26972485]
- Tse D, Langston RF, Takeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM. Schemas and memory consolidation. *Science.* 2007; 316:76–82. [PubMed: 17412951]
- Tse D, Takeuchi T, Takeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RGM. Schema-dependent gene activation and memory encoding in neocortex. *Science.* 2011; 333:891–895. [PubMed: 21737703]
- Viard A, Piolino P, Desgranges B, Chételat G, Lebreton K, Landeau B, Young A, Sayette VDL, Eustache F. Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: An fMRI study. *Cereb Cortex.* 2007; 17:2453–2467. [PubMed: 17204823]
- Wagner U, Gais S, Haider H, Verleger R, Born J. Sleep inspires insight. *Nature.* 2004; 427:352–355. [PubMed: 14737168]
- Ward EJ, Chun MM, Kuhl BA. Repetition Suppression and Multi-Voxel Pattern Similarity Differentially Track Implicit and Explicit Visual Memory. *J Neurosci.* 2013; 33:14749–14757. [PubMed: 24027275]
- Warren DE, Jones SH, Duff MC, Tranel D. False recall is reduced by damage to the ventromedial prefrontal cortex: Implications for understanding the neural correlates of schematic memory. *J Neurosci.* 2014; 34:7677–7682. [PubMed: 24872571]
- Watanabe T, Kimura HM, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, Miyashita Y, Konishi S. Functional dissociation between anterior and posterior temporal cortical regions during retrieval of remote memory. *J Neurosci.* 2012; 32:9659–9670. [PubMed: 22787051]
- Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. *Science.* 1994; 265:676–679. [PubMed: 8036517]
- Wiltgen BJ, Silva AJ. Memory for context becomes less specific with time. *Learn Mem.* 2007; 14:313–317. [PubMed: 17522020]
- Winocur G, Moscovitch M, Bontempi B. Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia.* 2010; 48:2339–2356. [PubMed: 20430044]

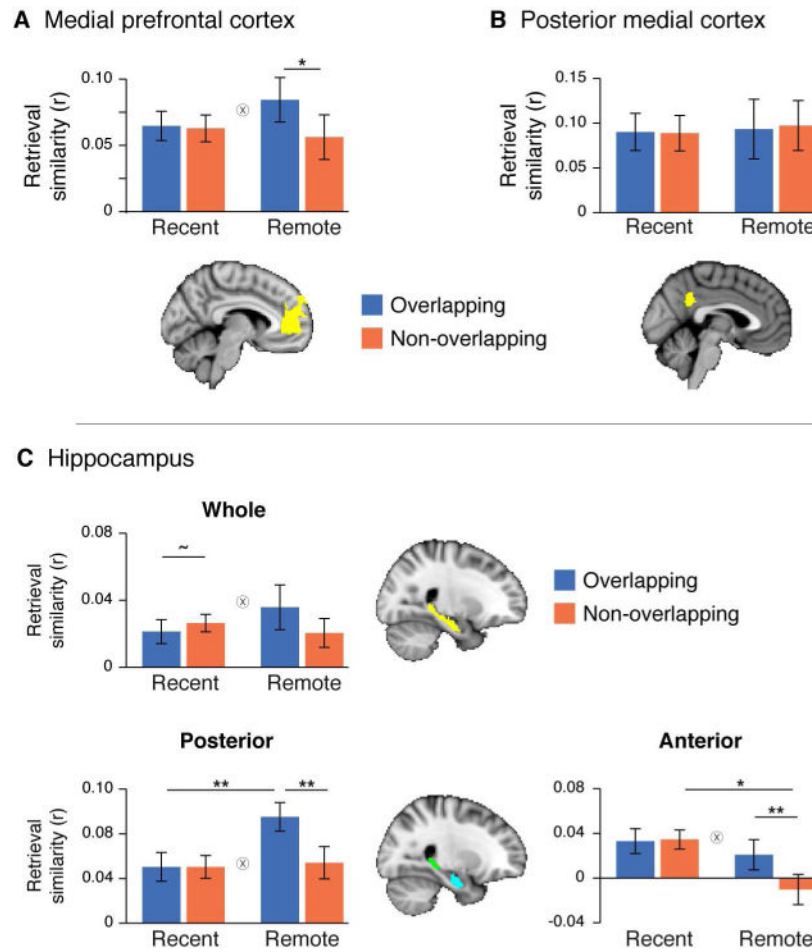
- Wirebring LK, Wiklund-Hörnqvist C, Eriksson J, Andersson M, Jonsson B, Nyberg L. Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *J Neurosci*. 2015; 35:9595–9602. [PubMed: 26134642]
- Witter MP, Wouterlood FG, Naber PA, Van Haefen T. Anatomical organization of the parahippocampal-hippocampal network. *Ann N Y Acad Sci*. 2000; 911:1–24. [PubMed: 10911864]
- Xiao X, Dong Q, Gao J, Men W, Poldrack RA, Xue G. Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *J Neurosci*. 2017; 37:2986–2998. [PubMed: 28202612]
- Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. Greater neural pattern similarity across repetitions is associated with better memory. *Science*. 2010; 330:97–101. [PubMed: 20829453]
- Zeithamova D, Preston AR. Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J Neurosci*. 2010; 30:14676–14684. [PubMed: 21048124]
- Zeithamova D, Dominick AL, Preston AR. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*. 2012; 75:168–179. [PubMed: 22794270]



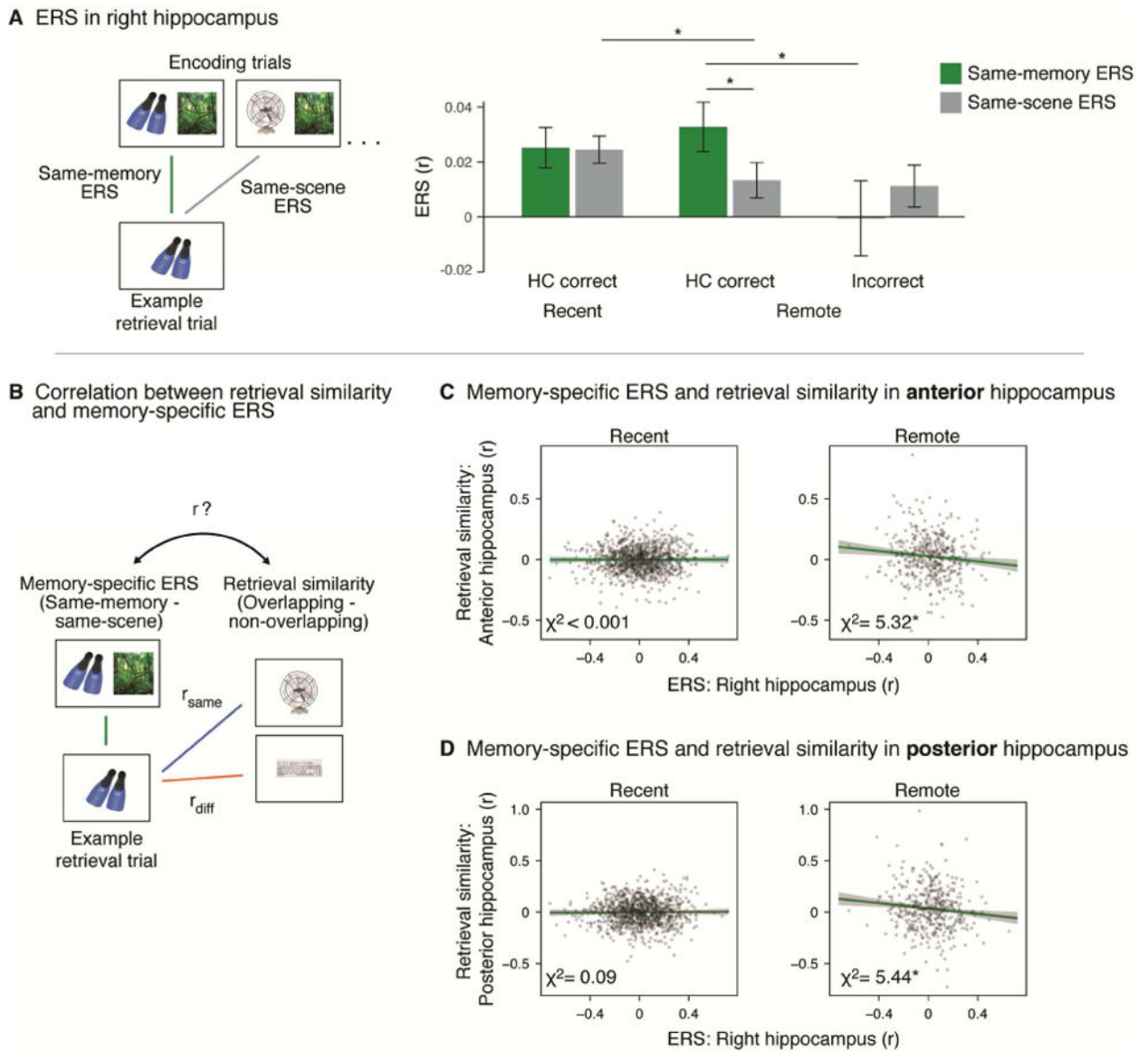
**Figure 1.** Experiment design. **A.** During the first session, participants encoded 128 object-scene pairs and then completed object recognition and scene recall tests for 64 of the pairs (recent retrieval). A week later, they returned to complete the same retrieval tasks for the other 64 pairs (remote retrieval). **B.** At encoding, participants viewed each object-scene pair and rated how vividly they imagined the object in the scene. During scene recall, participants were asked to choose which of the four scenes was studied with each object. **C.** Memory performance during the scene recall task. \*\* indicates  $p < 0.01$ . Error bars signify SEM. **D.** Images of the four scene associates.



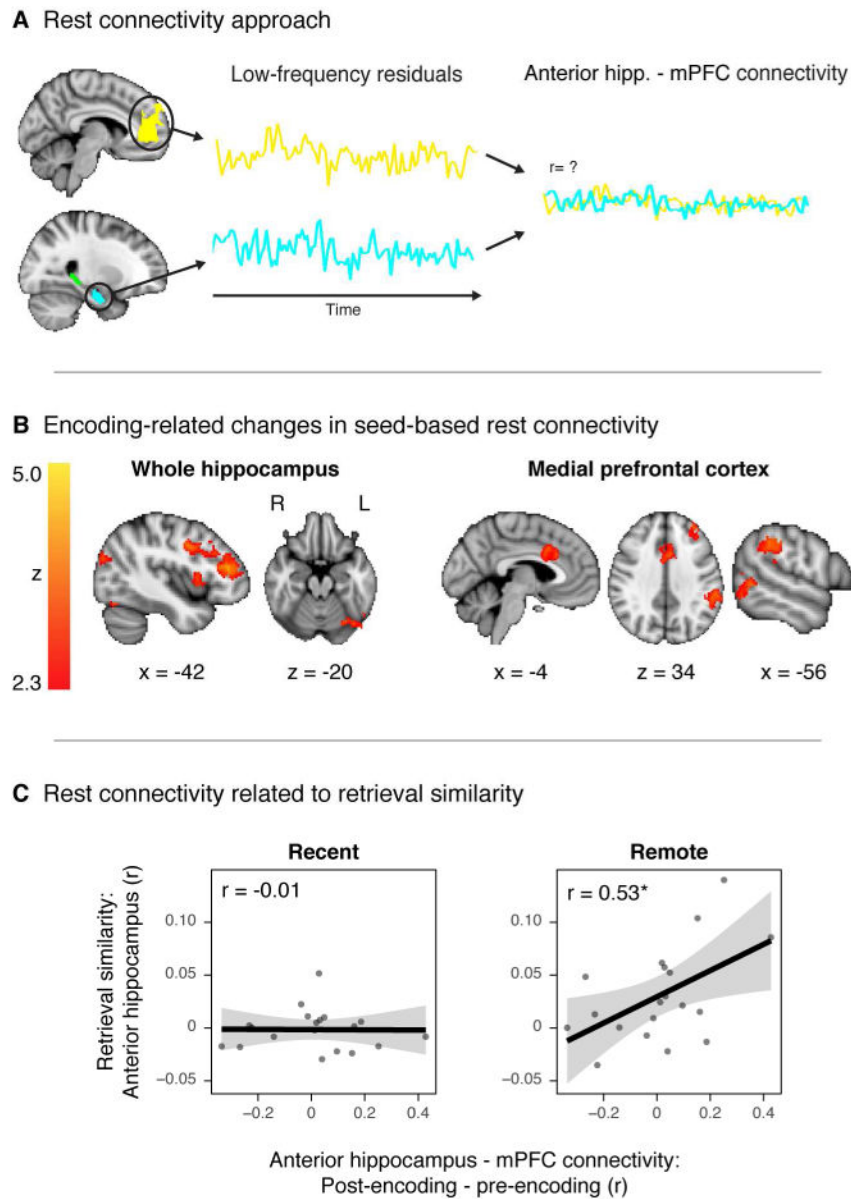
**Figure 2.** Analysis approach. Several similarity measures were computed for each trial. During retrieval, overlapping similarity (blue) was computed by correlating the pattern of activation evoked by each object-scene pair with all other patterns from pairs studied with the same scene, and then computing the average across those correlations. Non-overlapping similarity (orange) was computed by averaging all correlations between that pair's pattern of activation and the patterns of all pairs studied with a different scene. ERS (green) was computed for each pair by correlating its pattern at retrieval with its average pattern across its three encoding presentations.



**Figure 3.** Retrieval similarity. **A.** Retrieval similarity in mPFC. Similarity for overlapping trials was greater than similarity for non-overlapping trials during remote but not recent retrieval. **B.** Retrieval similarity in PMC. **C.** Retrieval similarity in hippocampus. In posterior hippocampus, similarity for overlapping trials increased over time. In anterior hippocampus, similarity for non-overlapping trials decreased over time. \*\* indicates  $p < 0.01$ . \* indicates  $p < 0.05$ . ~ indicates  $p < 0.10$ . Error bars signify SEM. ⊗ indicates significant interaction ( $p < 0.05$ ).



**Figure 4.** Encoding-retrieval similarity. **A.** Same-memory and same-scene ERS during recent and remote retrieval. In the right hippocampus, same-memory ERS was greater for HC correct trials relative to same-scene ERS and relative to incorrect trials during remote retrieval. \* indicates  $p < 0.05$ . Error bars signify SEM. **B.** Relationship between memory-specific ERS and retrieval similarity. Retrieval similarity was inversely correlated with ERS during remote retrieval, but not recent retrieval. Gray points represent all trials included in the analysis. Green lines represent the best fit line representing the relationship between retrieval similarity and memory-specific ERS. Gray ribbons signify 95% confidence intervals. \* indicates  $p < 0.05$ .



**Figure 5.** Rest connectivity. **A.** Rest connectivity approach. Rest scans were preprocessed, stripped of nuisance signals, and band-pass filtered. The mean residual signal was extracted from each ROI for each volume of each scan. Functional connectivity were measured either by entering the mean time-course of a seed region in a whole-brain voxelwise GLM (5B), or by correlating the mean time-courses of two regions (5C). **B.** Encoding-related changes in connectivity. Clusters indicate regions whose connectivity with the whole hippocampus (left) and mPFC (right) is greater after encoding relative to a pre-encoding baseline. Clusters survived correction for multiple comparisons using cluster-mass thresholding ( $p < 0.05$ , cluster-forming threshold  $z = 2.3$ ). **C.** Across participants, the change in connectivity between anterior hippocampus and mPFC (post-encoding – baseline rest) positively correlated with the average difference in retrieval similarity (overlapping – non-overlapping) in anterior

hippocampus, only for remote memories. Gray dots represent participants. Gray ribbons signify 95% confidence intervals. \* indicates  $p < 0.05$ .

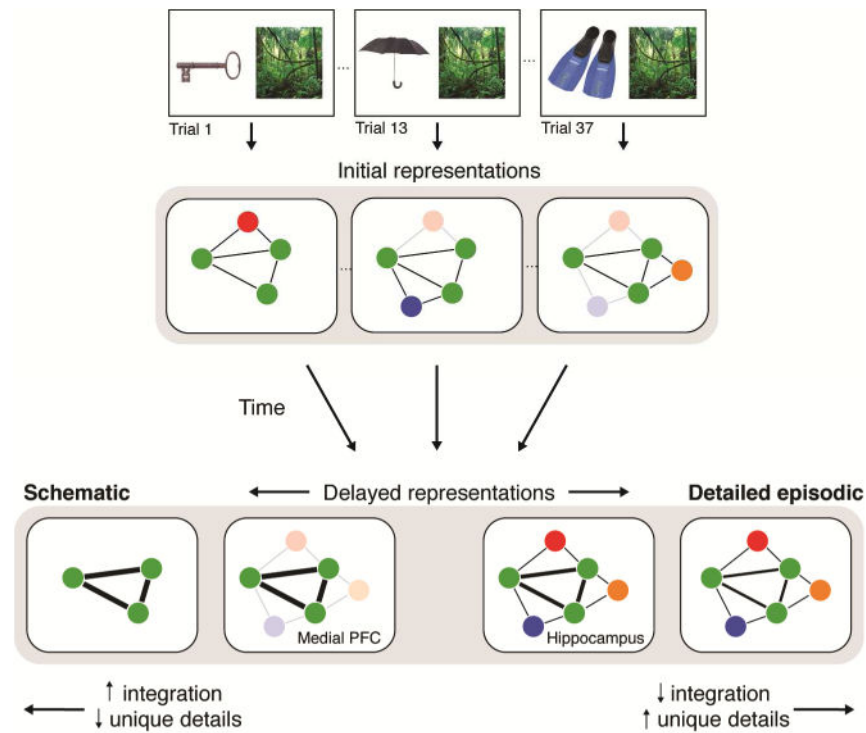
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 6.** Schematic of the theorized neural transformation of overlapping episodic memories. Three memories share an overlapping element (image of a rainforest). The representation of each memory consists of nodes representing features shared with other memories (green) or features unique to that memory (multi-colored). The thickness of line between two nodes represents the likelihood of coordinated activation of those nodes. Initially, the encoding and retrieval of each memory may recruit an overlapping subset of nodes as well as a distinct subset. Through consolidation mechanisms and other time-dependent processes, such as forgetting, memory representations may change along several dimensions: through loss of episodic details (fading or disappearance of nodes), or through strengthening of connections between overlapping features (thickening of lines between nodes). Schematic memories may lose the majority of unique nodes and retain strongly connected overlapping nodes. Memories that remain vividly episodic may retain unique nodes, but the connections between overlapping nodes may not be strengthened. Variations in how memories are transformed along these two dimensions may support the extraction and representation of gist-level or semantic memory over time.

**Table 1**

Results of repeated-measures ANOVAs predicting pattern similarity in the hippocampus. **Top:** Time (recent, remote), Overlap (overlapping, non-overlapping), Hemisphere (right, left), and all interactions were included as factors. **Bottom:** Time (recent, remote), Overlap (overlapping, non-overlapping), Region (anterior, posterior), and all interactions were included as factors.

<b>Effects: Hemisphere</b>	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>
Time	1	18	0.059	0.811
Overlap	1	18	2.041	0.170
Hemisphere	1	18	4.494	0.048*
Time × Overlap	1	18	6.895	0.017*
Time × Hemisphere	1	18	0.032	0.859
Overlap × Hemisphere	1	18	0.396	0.537
Time × Overlap × Hemisphere	1	18	0.118	0.735
<b>Effects: Long axis organization</b>	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>
Time	1	18	0.037	0.850
Overlap	1	18	13.030	0.002*
Region	1	18	11.770	0.003*
Time × Overlap	1	18	20.930	<0.001*
Time × Region	1	18	6.107	0.024*
Overlap × Region	1	18	1.063	0.316
Time × Overlap × Region	1	18	0.410	0.530

\* indicates  $p < 0.05$ .