**Title:** Bridges, not walls: the hippocampus builds narrative memories across distant events

**Authors:** Brendan I. Cohn-Sheehy,[1-3]* Angelique I. Delarazan,[5] Zachariah M. Reagh,[5] Jordan E. Crivelli-Decker,[3-4] Kamin Kim,[3] Alexander J. Barnett,[3] Jeffrey M. Zacks,[5] Charan Ranganath[2-4]

**Affiliations:**

[1]M.D./Ph.D. Program, University of California, Davis, Sacramento, CA, USA.

[2]Neuroscience Graduate Group, University of California, Davis, Davis, CA, USA.

[3]Center for Neuroscience, University of California, Davis, Davis, CA, USA.

[4]Department of Psychology, University of California, Davis, Davis, CA, USA.

[5]Department of Psychological and Brain Sciences, Washington University, 1 Brookings Drive, St. Louis, MO, USA.

*Corresponding Author: Brendan I. Cohn-Sheehy

## Abstract

Life's events are scattered throughout time, yet they often coalesce to form a larger narrative in memory. Prior research suggests that the hippocampus, which supports memory for past events, can sometimes integrate distant events into unified memories. However, the conditions which lead to hippocampus-dependent memory integration are unclear. Here, we used functional brain imaging to test whether the ability to form a larger narrative (narrative coherence) drives hippocampus-dependent memory integration. During encoding of fictional stories, the hippocampus supported patterns of activity, including activity at boundaries between events, which were more similar between distant events that formed one coherent narrative, compared with events taken from unrelated narratives. Twenty-four hours later, the hippocampus preferentially supported detailed recall of coherent narrative events, through reinstatement of hippocampal activity patterns from encoding. These findings reveal a key function of the hippocampus: the dynamic integration of events into a narrative structure for memory.

## Introduction

It is well-established that although experience unfolds over a continuous timeline, memory for past experiences, or episodic memory (*1*), can be segmented into individual units of time called "events" (*2*). For example, say you are in the middle of reading a scientific article when you receive a phone call from your friend Sandra. Sandra proceeds to vent about an upcoming date, with a man who might be very charming or, alternatively, quite boorish. If you recall Sandra's call later on, you would find it difficult to also recall your time reading the scientific article—information about two adjacent events can become segregated in memory (*3–8*). However, event segmentation may not provide the only organization for memory. For example, you might subsequently encounter Sandra at a café, slapping an unnamed man as she walks out. Although Sandra appears in temporally separated events, they coalesce to form a *narrative*: a larger unit of information which encompasses multiple events, and in which one's understanding of each event is dependent on information from other events (*9–14*).

The ability to organize events into a narrative is beneficial for memory. In a recent series of experiments (*9*), we compared memory for events that could be integrated into a larger narrative, to memory for events that shared a character, but for which the character's actions could not be integrated into a narrative. We found that, when temporally separated events could be assimilated into a larger narrative, they were recalled in greater detail, compared with temporally separated events that merely shared a recurring character. This finding suggested that episodic memory might be organized at a level above and beyond individual events, on the basis of *narrative coherence*: the degree to which individual units of information can be interrelated within a single narrative representation (*10*, *11*). In the present study, we tested the possibility

that narrative coherence shapes the way the human brain encodes and retrieves events in memory.

Little is known about how the brain assembles narratives from individual events, but there is reason to think that the hippocampus plays a critical role. Many findings support the general idea that the hippocampus can build memories that integrate information across overlapping experiences (*15–21*), although this might not depend on narrative coherence. Several studies have suggested that when one encounters information that overlaps with a previous event (e.g. a recurring "B" item in separated "A-B" and "B-C" pairs), this overlap can trigger the hippocampus to form associations between temporally-distant experiences (*17, 18, 20, 21*). Some evidence supports a potential mechanism for memory integration during encoding: encountering a new event might trigger the hippocampus to reinstate activity patterns which correspond with a prior overlapping event, such that information from the prior event becomes embedded into memory for the new event (*15, 20, 22*). If the hippocampus can combine information across events during memory encoding, this might make it easier to recall both events later on.

However, overlapping information does not always result in memory integration—in fact, longstanding evidence suggests that when events share overlapping features (e.g. a recurring character), these events are more difficult to remember (*23, 24*). That is, overlapping events can compete with each other during memory retrieval (*23, 24*). Furthermore, some findings suggest that the hippocampus might hyper-differentiate between overlapping events (*25, 26*), and results from another study suggest that the hippocampus might simultaneously support the ability to segregate and integrate events that share overlapping features (*17*). Recent models have attempted to describe the conditions under which the hippocampus supports memory integration versus differentiation (*15, 25, 27–29*). These models do not address any role for narrative

coherence—though, as our recent behavioral findings suggest (*9*), the ability to form a narrative might determine whether or not events become integrated in memory.

Recent evidence suggests that narrative coherence modulates activity in the hippocampus (*30*, *31*). Milivojevic et al. (*31*) used functional magnetic resonance imaging (fMRI) to analyze patterns of hippocampal activity evoked by pairs of temporally adjacent events. Although events were initially unrelated to each other, participants were instructed that some event pairs would form a narrative with a subsequently-presented, linker event—when events could become linked, hippocampal activity patterns became more similar between events (*31*). This finding suggested that, after one initially encodes events, patterns of activity in the hippocampus might become modified to support information about a larger narrative. However, overlapping activity patterns are not equivalent to memory integration per se. If similar activity patterns reflect that events are integrated in memory, we would also expect hippocampal activity patterns would play a beneficial role when memory is assessed (i.e. during recall of events). No study to date has demonstrated whether narrative coherence modulates hippocampus-dependent integration of temporally distant events.

An intriguing possibility is that the hippocampus might integrate events into a coherent narrative, but only at specific moments during experience. Several recent studies (*32*–*34*) have found that the hippocampus becomes particularly active when people perceive transitions between events, or *event boundaries* (*2*, *35*). Although the significance of boundary-evoked activation is unclear, it might indicate that the hippocampus supports retrieval of information from past events in order to inform encoding of a new event (*22*, *36*). For instance, running into your friend Sandra would not only mark the start of a new event, but might also lead you to retrieve Sandra's previous event from memory and integrate it with the new event (*2*, *22*, *37*, *38*).

In support of this idea, recent electroencephalography (EEG) findings suggested that, at an event boundary, neurophysiological activity recapitulates information about a preceding, temporally-adjacent event (*39*, *40*). On the other hand, event boundaries provide a basis for segmenting events in memory (*2*, *35*). In other words, event boundaries might act as "walls" between temporally adjacent events—or, they might also serve as optimal points for the hippocampus to build "bridges" between temporally distant events.

Building on prior findings showing that narratives modulate activity in the hippocampus (*30*, *31*, *41*), we sought to determine whether hippocampal activity, including activity at event boundaries, supports the integration and subsequent recall of temporally distant events that form a larger, coherent narrative. Functional magnetic resonance imaging (fMRI) was used to image brain activity while participants listened to fictional stories (**Fig. 1A**), in which recurring characters appeared in pairs of temporally separated events. Critically, these stories were written to include pairs of events that could form one Coherent Narrative, and pairs involving an overlapping character in Unrelated Narratives (see **Fig. 1B** for examples). These temporally separated events were embedded in longer stories which did not involve these recurring characters. For example, one four-minute story (Story 2 in **Fig. 1A**) followed a protagonist who, by chance, encountered the characters Sandra and Johnny, and in a subsequent story (Story 4 in **Fig. 1A**), a different protagonist also encountered Sandra and Johnny. In one version of these stories, the two events involving Sandra could be integrated into a Coherent Narrative (e.g., two events that describe a single dating experience) whereas the events involving Johnny were parts of Unrelated Narratives (e.g., searching for a lost recipe in Event 1, evading financial troubles in Event 2). However, both Coherent and Unrelated Narrative events were meaningfully distinct from the surrounding plot, such that events involving recurring characters could only form a

larger narrative in the Coherent Narrative condition. Importantly, different versions of the stimuli were constructed and pseudo-randomly assigned to different subjects (see **Methods**), in order to minimize content-driven differences between Coherent and Unrelated Narrative events. Participants returned one day later, and they were scanned during recall of the events involving each recurring character (**Fig. 1C**).

To determine whether hippocampal encoding is sensitive to narrative coherence, we

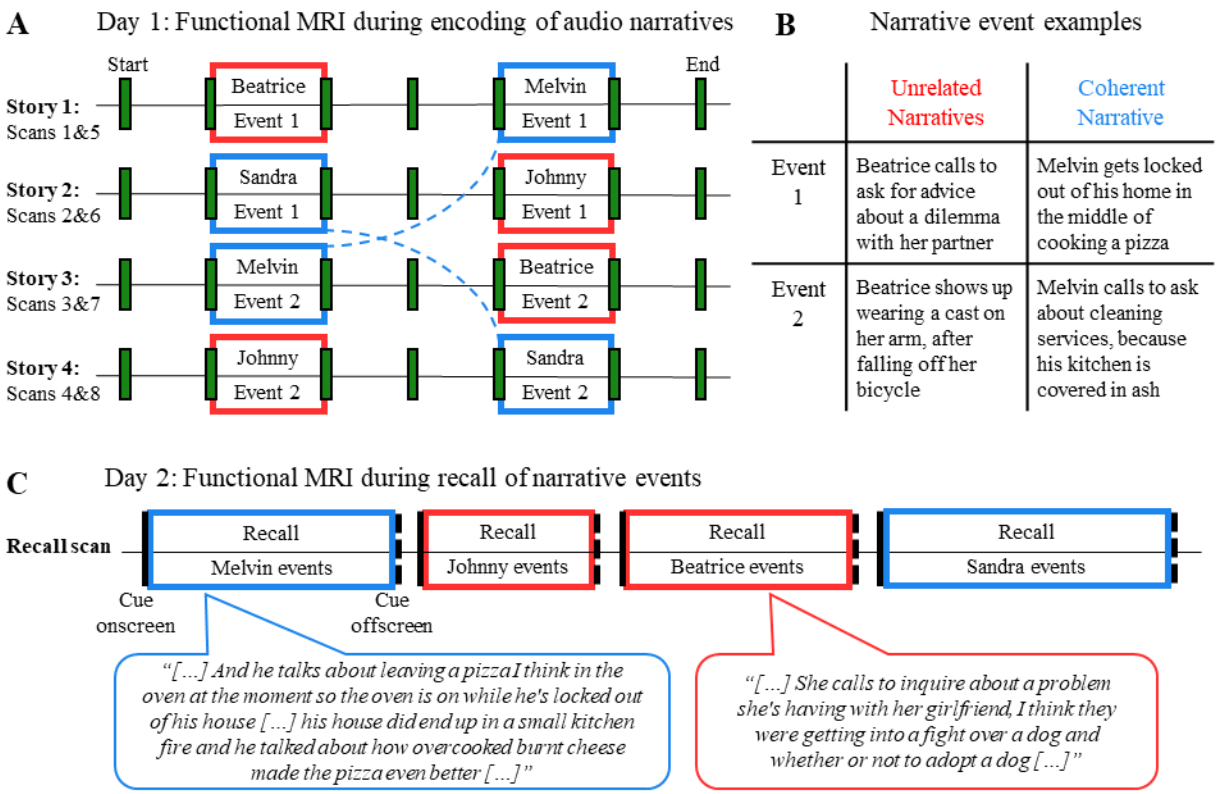


**Fig. 1: Experimental paradigm. (A)** Narrative stimuli presented during functional MRI: On Day 1, four fictional stories were presented serially during scanning, as audio clips (240s each), which were punctuated by discrete event boundaries (green bars). Events involving recurring characters (Beatrice, Melvin, Sandra, Johnny; 40s each) did not relate to the more continuous plot events surrounding them. For two of these recurring characters (Blue boxes and dashed lines), two temporally-distant events could form one Coherent Narrative (e.g. Sandra Events 1 and 2). In contrast, for two other recurring characters (Red boxes), Events 1 and 2 belonged to Unrelated Narratives (e.g. Johnny). For each participant, side-characters were randomly assigned to either the Coherent Narrative or Unrelated Narratives conditions (i.e. this is one example; see Methods). **(B)** Examples of narrative events: Synopses are provided for possible pairs of Coherent Narrative and Unrelated Narrative events. **(C)** Recall assessment during functional MRI: During scanning on Day 2, participants were asked to recall events involving each recurring character from the stories presented on Day 1 (Coherent Narratives = Blue; Unrelated Narratives = Red). The amount of time spent on recall, and the amount of details recalled, varied from character to character. Recall excerpts are presented within speech bubbles.

separately estimated hippocampal activity patterns immediately at the onset ("Boundary epoch") and during the body of each event ("Post-Boundary epoch"). We tested whether these patterns were more similar across Coherent versus Unrelated Narrative events, providing a basis for integration during memory encoding. We then investigated the extent to which activity patterns during these time periods were reinstated during recall of the events on the following day. We hypothesized that, if hippocampal activity supports the integration of distant events into a larger narrative, then activity patterns identified during encoding of Coherent Narrative events would be preferentially reinstated during recall of these events, as compared to reinstatement during recall of Unrelated Narrative events.

## Results

### Boundary and Post-Boundary hippocampal activity bridges Coherent Narrative events

As described above, our main hypothesis was that hippocampal activity might bridge distant events, if the events can form a coherent narrative. There was also reason to suspect that activity during event boundaries—which reliably engage the hippocampus—would provide support for memory integration (*36*). Each story was written to include *a priori* event boundaries (green bars in **Fig. 1A**), which reliably evoked the perception of event boundaries in an independent sample (point-biserial correlation: $r=0.85$, $p<0.0001$; **Fig. S1A;** see **Methods**). As in previous studies (*32–34*), *a priori* event boundaries evoked increased activation in the right and left hippocampus (**Fig. S1B**).

If event boundaries could evoke cognitive processes that integrate distant events, it was possible that boundary-evoked hippocampal activation would be higher for events that form a

Coherent Narrative (CN), compared with events that were taken from Unrelated Narratives (UN). To test this hypothesis, we first identified a Boundary epoch (TRs 5-11, ~6.1-14.64 seconds following the onset of an event) which corresponded to the onset and duration of boundary-evoked activation in the hippocampus (**Fig. S1B**). The Boundary epoch accounted for the lag between the real-time onset of an event, and an increase in activation on functional MRI (i.e. the hemodynamic response lag). Then, we modeled the unique activity corresponding to each story event (see **Methods**), and we tested whether boundary-evoked activation within story events would differ between Coherent and Unrelated Narrative events. Within the right and left hippocampus, paired two-tailed t-tests revealed that boundary-evoked activation was higher for CN than UN events (right: $t(24) = 2.46$, $p=0.02$, Cohen's $d=.49$; left: $t(24)=2.48$, $p=0.02$, Cohen's $d=0.50$). These findings preliminarily suggested that hippocampal activity during the Boundary epoch was modulated by whether temporally-distant events could be assimilated into a larger narrative.

However, the preceding analysis did not determine whether hippocampal activity provided any link between temporally-distant events. To assess this possibility, we identified the unique pattern of hippocampal activity (voxel pattern) corresponding to each story event (see **Methods**; **Fig. 2A**), and investigated the degree to which voxel patterns were shared between events which formed a Coherent Narrative (CN) versus events which were taken from Unrelated Narratives (UN). We hypothesized that narrative coherence would determine whether the hippocampus could integrate temporally-distant events, and if so, we would expect higher pattern similarity between CN than UN events. Alternatively, the hippocampus might support the integration of any overlapping events (i.e. via a shared character), and if so, we would not expect any difference in pattern similarity between CN and UN events.

Furthermore, we investigated whether voxel patterns within the Boundary epoch (TRs 5-11; **Fig. 2A**), versus other parts of an event, would play a unique role in integration. We therefore defined a Post-Boundary epoch (**Fig. 2A**), in order to examine patterns of activity during the remainder of an event. Because each event spanned 33 timepoints, the Post-Boundary epoch encompassed 26 timepoints following the Boundary epoch (TRs 12-37, ~ 14.64-46.36 seconds following the onset of an event). We computed and selectively averaged voxel pattern similarity between the Boundary and Post-Boundary epochs of paired events (**Fig. 2A**), and subsequently analyzed pattern similarity values within the right and left hippocampus using a multi-level mixed effects model (see **Methods**, Formula 1). This model tested whether Narrative

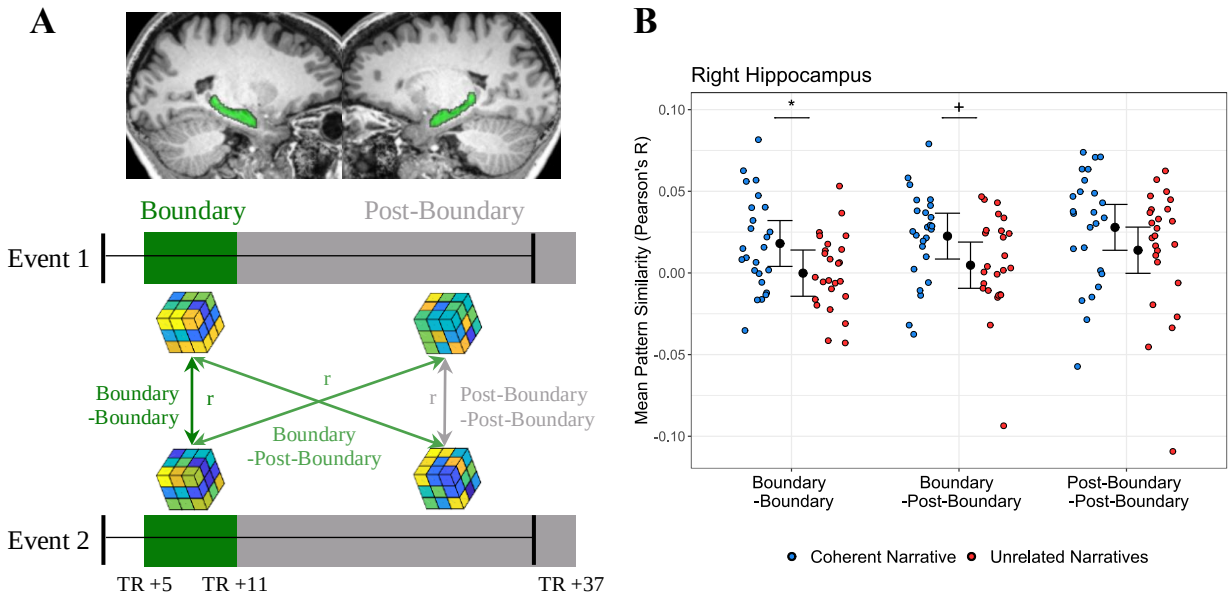


**Fig. 2: Right hippocampus activity patterns bridge events that form a coherent narrative.**
**(A)** Across-event pattern similarity analysis: within the left and right hippocampus (plotted on an individual subject's structural MRI, in green), the unique spatial pattern of activity ("voxel pattern", represented by cubes) was modeled for Boundary (dark green rectangles, TRs +5 to +11) and Post-Boundary epochs (grey rectangles, TRs +12 to +37). Black bars represent the real-time beginning and ending of each Coherent Narrative and Unrelated Narrative event during scanning—epoch definitions account for the lagged BOLD response. Pattern similarity (Pearson's r) was calculated between voxel patterns at Event 1 and Event 2, and selectively averaged to yield three epoch-by-epoch measures of similarity: Boundary-Boundary (dark green line), Boundary-Post-Boundary (light green lines), and Post-Boundary-Post-Boundary (grey line). **(B)** Right hippocampus pattern similarity: Pearson's correlations between Event 1 and 2 epochs were computed for Coherent Narrative and Unrelated Narrative events. Mean pattern similarity for individual subjects (colored dots) is plotted on the basis of which epoch was being examined, and whether events could form Coherent Narratives or not (Unrelated Narratives). Mixed model fits predicting pattern similarity are visualized as 95% confidence intervals (see text for model details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, +=p<0.10, *=p<0.05.

Coherence [CN vs UN] would modulate pattern similarity across events, and whether any effects of Narrative Coherence would be specifically driven by similarity between Boundary epochs (i.e. an interaction between Narrative Coherence and Epoch). Furthermore, the mixed model accounted for nuisance regressors which might confound effects of interest: particular subjects, specific events which were presented to subjects, and overall activation at boundaries.

As shown in **Fig. 2B,** within the right hippocampus, this model (Aikake Information Criterion: AIC = -1039) revealed a significant effect of Narrative Coherence ($F(1,11.33)=6.97$, $p=.02$). This finding reflected that, even after accounting for potential confounds, pattern similarity in the right hippocampus was higher for CN events, compared with UN events (see **Fig. 2B** for model estimates). However, there was no significant interaction between Narrative Coherence and Epoch ($F(2,257.71)=0.09$, $p=.92$; $ps \geq 0.10$ for other fixed effects). Because the interaction was not significant, we cannot conclude that the Narrative Coherence effect was driven by the Boundary epoch alone. These findings suggested that during encoding of an event (either Boundary or Post-Boundary epochs), activity patterns in the right hippocampus might bridge distant events that form a coherent narrative.

This pattern of findings was not evident in the left hippocampus (**Fig. S2**), wherein the mixed model (AIC = -1103) revealed only a significant effect of Overall Boundary Activation on pattern similarity ($F(1,289.69)=12.18$, $p<.001$; for other fixed effects, $ps>.36$). Interestingly, an exploratory analysis suggested that findings within the right hippocampus might be specific to the right posterior hippocampus, and not the right anterior hippocampus (**Fig. S3**); this analysis should be interpreted with caution.

**Hippocampal activity that bridges Coherent Narrative events is reinstated during recall**

The preceding findings supported the idea that, during memory encoding, patterns of activity in the right hippocampus are preferentially shared across distant events that form a coherent narrative. If activity patterns in the right hippocampus provided a basis for integrating Coherent Narrative events during memory encoding, we would expect that these activity patterns would play some role in subsequently retrieving these events from memory. Namely, we would expect that right hippocampal voxel patterns from encoding would be more highly reinstated during recall of CN events compared with UN events. Participants were scanned during recall of events involving CN and UN characters (**Fig. 1C**), providing an opportunity to compare voxel patterns between encoding and recall (i.e. "encoding-retrieval similarity"; **Fig. 3A**). Because preceding analyses revealed pattern similarity effects that spanned both Boundary and Post-Boundary epochs, we pooled timepoints from both epochs to measure reinstatement of whole-event activity patterns from encoding (**Fig. 3A**).

In order to accurately predict how reinstatement might work, it is important to consider how integration may have unfolded during encoding. During encoding, there were two events involving each CN and UN character (Event 1 and Event 2; **Fig. 1**). Because events were presented sequentially, Event 2 was the event which determined whether there was any relationship between the two events. In other words, in the CN condition, Event 2 may have served as a key period for integration. If so, we would expect that information from CN Event 2 would play a critical role in subsequent recall. In contrast, in the UN condition, it was unlikely that information from Event 2 would provide an opportunity to easily integrate events during encoding (i.e. aside from an overlapping character).

Based on this rationale, we hypothesized that activity patterns during encoding of Event 2 would be more highly reinstated during recall of CN events than UN events. To test this

hypothesis, for each CN and UN character, we analyzed the degree to which right hippocampal voxel patterns during Event 2 encoding were reinstated during recall (see **Fig. 3A**, **Methods**). We also analyzed reinstatement of activity patterns from Event 1—however, because we suspected that Event 2 served as the critical period for integration, we did not expect any difference in reinstatement of activity from Event 1 between CN and UN events. To test our hypothesis, we performed a multi-level mixed effects model (see **Methods**, Formula 2) which tested whether encoding-retrieval similarity would be modulated by Narrative Coherence [CN vs UN events], and whether an effect of Narrative Coherence would be limited to reinstatement of Event 2 (i.e. an interaction between Narrative Coherence and Event Number). The model statistically accounted for several nuisance variables: the amount of time spent on recall (28-299 TRs, ~ 0.5-6 minutes per recall cue), particular subjects, and particular event content.

This model (AIC = -1334.7) revealed a trending main effect of Narrative Coherence ($F(1,303.98)=3.60$, $p=0.059$), which was qualified by a significant interaction between Narrative Coherence and Event Number ($F(1,343.46)=7.72$, $p=0.006$); no other fixed effects were significant (ps>.48). As shown in **Fig. 3B**, model estimates revealed that the interaction was driven by an effect of higher reinstatement for CN than UN Events, and that this effect was specific to Event 2. An exploratory analysis revealed a similar pattern of findings when examining pattern reinstatement from Boundary epochs alone (**Fig. S4A-B**).

These findings suggested that in the right hippocampus, patterns of activity from encoding of Event 2 were preferentially reinstated during recall of events which formed a Coherent Narrative. In line with our predictions, this suggested that Event 2 provided an opportunity for memory integration, and that hippocampal activity patterns during Event 2 may have bridged distant events which formed a coherent narrative. We investigated this possibility

further by comparing encoding-retrieval similarity with a behavioral measure of memory.

## Narrative coherence modulates whether hippocampal pattern reinstatement predicts detailed recall

Our next analyses focused on recall of CN and UN events. In a previous behavioral study, we found that, across three experiments, participants were able to recall more details about CN than UN events (*9*). In the present study, the number of details recalled from CN events

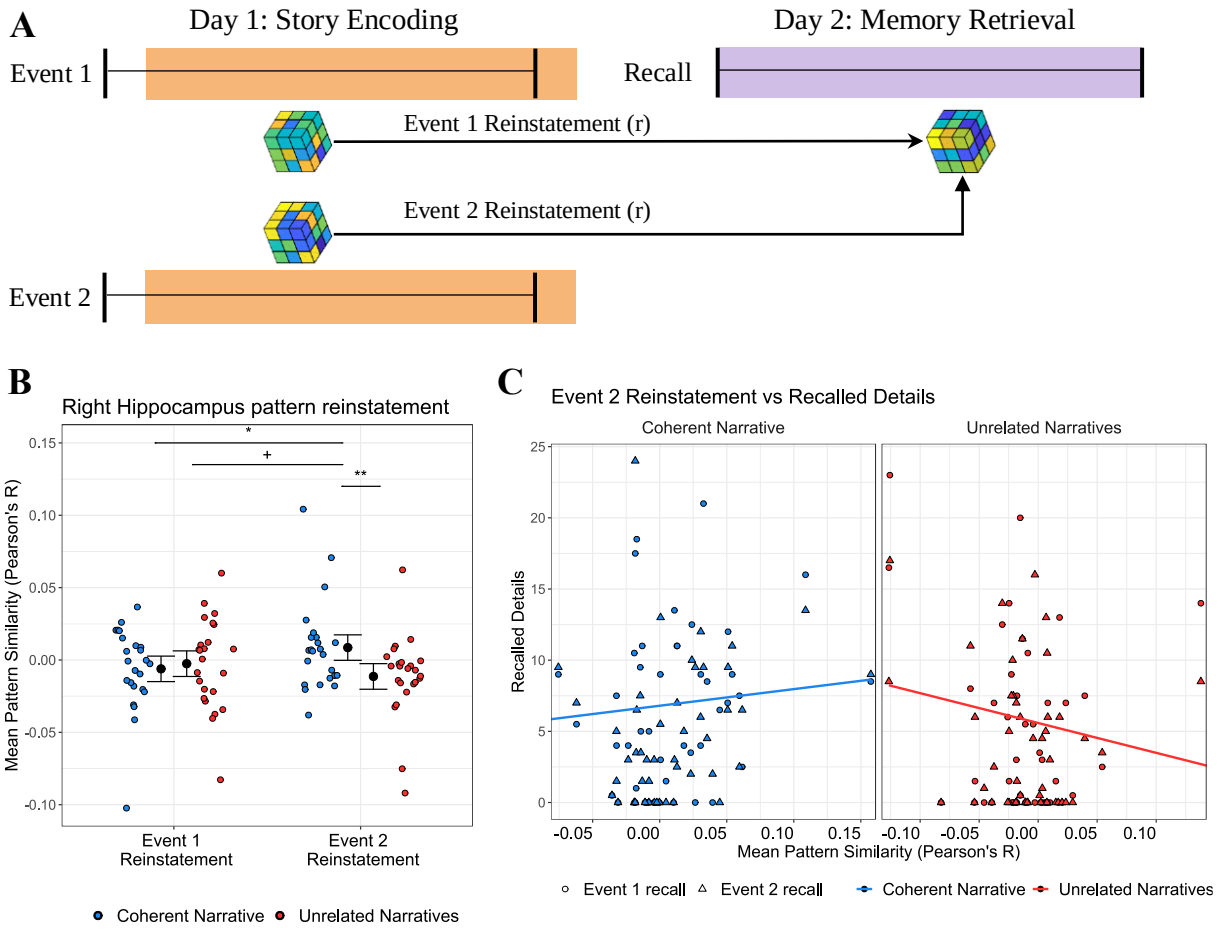


**Fig. 3: Reinstatement of hippocampal activity from memory encoding supports recall of Coherent Narrative events. (A)** Encoding-retrieval similarity: Pattern similarity (Pearson's r) was calculated between activity from each event involving a recurring character at encoding (Event 1 or 2), and activity during delayed recall of events involving that character. **(B)** Event 2 pattern reinstatement is higher for Coherent Narrative recall: Encoding-retrieval similarity for each subject (colored dots) is plotted for each epoch of each event during recall. 95% confidence intervals represent mixed model estimates (see text). **(C)** Event recall predicted by Event 2 pattern reinstatement: Recalled details from Event 1 (circles) and Event 2 (triangles) are plotted for each subject, predicted by Event 2 pattern reinstatement. Colored lines represent mixed model estimates (see text). Key: Blue=Coherent Narrative, Red=Unrelated Narratives, +=p<.10, *=p<.05, **=p<.01.

(mean=10.71 details per cue, SD=9.4, max=41.5) was numerically higher than recall of details from UN events (mean=8.65 details per cue, SD=9.1, max=40.0), though this effect did not meet the threshold for statistical significance ($t(24) = 1.98$, $p = .059$, Cohen's $d = 0.22$). Examination of the data revealed a large degree of inter-subject variance in recall performance, and we considered the possibility that this variance might be explained by inter-subject differences in hippocampal activity that supports recall. If so, we would expect that recall of CN events should be predicted by reinstatement of right hippocampus activity patterns during recall.

As described above (**Fig. 3B**), evidence from analyses of encoding-retrieval similarity suggested that Event 2 provided an opportunity for memory integration during encoding, enabling CN events, and not UN events, to become integrated in memory. If so, then we might expect that reinstatement of right hippocampus activity patterns from CN Event 2 would not only predict how well Event 2 was recalled, but would also predict how well Event 1 was recalled, reflecting that these events became integrated in memory. In contrast, if UN events were not integrated in memory, we would not expect reinstatement of UN Event 2 activity patterns to predict detailed recall of both UN Events. Alternatively, if our pattern similarity findings did not reflect memory integration, we might expect that the right hippocampus would generally support recollection of individual events. If so, we might expect that reinstatement of activity from any event (i.e. either CN or UN Event 2) would predict only the number of details recalled about that specific event (i.e. recall of CN Event 2, recall of UN Event 2).

To assess whether and how memory integration occurred, we performed a multi-level mixed effects model (see **Methods**, Formula 3), to test whether and how recalled details would be predicted by Event 2 pattern reinstatement. Moreover, the model tested whether the relation between recall and pattern reinstatement would be modulated by Narrative Coherence [CN vs

UN events]. The model also tested whether Event 2 pattern reinstatement would differentially predict the number of details recalled about Event 1 versus Event 2. After accounting for nuisance regressors (recall duration, individual subjects, specific events), this model (AIC=1120.5) revealed a significant main effect of Narrative Coherence ($F(1,160.17)=4.40$, $p=0.037$), which was qualified by a significant interaction of Narrative Coherence and Event 2 Reinstatement ($F(1,156.42)=5.20$, $p=0.024$). No other fixed effects were significant (all ps>.34), suggesting that Event 2 Reinstatement did not differentially predict recalled details for Event 1 versus Event 2, and that there were no overall differences in recall of Event 1 versus Event 2.

As shown in **Fig. 3C**, the significant interaction reflected that in the right hippocampus, the relationship between Event 2 pattern reinstatement and recalled details (from either event) was modified by whether events could form a Coherent Narrative. To break down this interaction further, we performed a pairwise comparison of linear trends that were predicted by the mixed model (shown in **Fig. 3C**). This comparison revealed that the relationship between recalled details and Event 2 Reinstatement was significantly more positive for CN events than UN events ($t(156)=2.28$, $p=0.024$). This pattern of findings suggested that hippocampal activity patterns which coincided with the second of two overlapping events did not always confer a benefit for subsequent recall. Instead, consistent with our hypothesis, this suggested that narrative coherence modulated the degree to which hippocampal pattern reinstatement from the second of two overlapping events predicted any benefit for subsequent recall. Moreover, this suggested that when participants encoded the second overlapping event, the hippocampus may have supported information that could integrate events in memory, but only if events could form a coherent narrative. A similar pattern of findings was revealed when limiting the model to reinstatement from the Boundary epoch (**Fig. S4C**).

**Post-hoc analyses: narrative coherence modulates pattern similarity in cortical networks**

Although our main hypotheses centered on hippocampus-dependent memory integration, there is reason to think that the hippocampus interacts with broader networks of cortical regions – particularly within the default mode network (*42*, *43*) – to process information in narratives (*44–50*). We performed post-hoc analyses to determine whether two sub-networks of the default mode network, the Posterior Medial and Anterior Temporal networks (*42*, *43*), would support the ability to integrate Coherent Narrative events during memory encoding. To that end, we examined the multivoxel pattern similarity within these networks as participants encoded CN and UN events. Because stories were presented aurally, we also tested for narrative coherence effects within bilateral auditory cortex. Additionally, in line with analyses of hippocampal responses to event boundaries, we tested whether these networks exhibited similar activation increases at event boundaries.

**Figs. S5-6** display cortical network findings which survived correction for multiple comparisons. Surprisingly, the Posterior Medial Network did not exhibit a significant effect of Narrative Coherence on pattern similarity (**Fig. S5B**). However, in line with previous studies (*34*, *50*), the Posterior Medial Network exhibited a cluster of significant boundary-driven activation (**Fig. S6A**), which overlapped with timepoints where the hippocampus responded to event boundaries (in **Fig. 2B**). Interestingly, pattern similarity in the Anterior Temporal Network exhibited an interaction between Narrative Coherence and Epoch (**Fig. S5D**), which reflected that pattern similarity was higher for CN than UN events, but only at timepoints within the Post-Boundary epoch. In bilateral auditory cortex, there was a main effect of Narrative Coherence on pattern similarity (**Fig. S5F**), reflecting that pattern similarity was higher for CN than UN events,

irrespective of epoch. However, neither the Anterior Temporal Network nor bilateral auditory cortex exhibited any significant clusters of boundary-evoked activation increases (**Figs. S6B-C**). Taken together, this set of post-hoc findings suggests that cortical networks might provide additional support for the representation of distant events that form a coherent narrative. Subsequent analyses did not reveal any network-wide effects of narrative coherence on cortical pattern reinstatement during memory recall.

## Discussion

In the current study, we investigated whether the hippocampus supports a narrative-level organization for episodic memories, by integrating temporally-distant events into a larger narrative. In support of this hypothesis, activity patterns in the right hippocampus during memory encoding were more similar across two temporally separated events that formed a Coherent Narrative, than across Unrelated Narrative events with an overlapping character. Twenty-four hours later, during memory retrieval, activity patterns in the right hippocampus were more highly reinstated for Coherent Narratives than for Unrelated Narratives. In particular, activity reinstatement was disproportionately higher for the second of two events that formed a Coherent Narrative—that is, the event which provided the clearest opportunity to meaningfully integrate information across events. Furthermore, narrative coherence determined the degree to which reinstatement of hippocampal activity patterns from the second event predicted the ability to recall details about events.

Taken together, these findings suggest that the hippocampus supports the construction of a narrative that integrates distant events in memory, and which can be used to facilitate memory

retrieval. Furthermore, these findings suggest that integration specifically takes place at key periods during memory encoding, when one can draw upon information about a prior event that informs one's understanding of a current event.

When and how does the hippocampus integrate distant events in memory? Previous work has suggested that hippocampal activity during events can reflect the existence of larger narratives (*30*, *31*, *41*). However, previous work has not investigated whether narrative coherence is a determining factor for hippocampus-dependent memory integration across distant events. The present findings suggest that the second of two Coherent Narrative events provided an opportunity for hippocampus-dependent memory integration. To some degree, this is consistent with prior evidence that, when new events share overlapping features with a prior event, the hippocampus can incorporate information about the prior event when encoding the new event in memory, leading to memory integration (*16–18*, *20*, *21*). However, in the present study, both Coherent and Unrelated Narrative events shared overlapping characters, suggesting that overlap, while potentially necessary, was not sufficient to promote memory integration across distant events.

The present findings are less consistent with models which suggest that the hippocampus keeps event representations distinct during memory encoding, and can only integrate them during memory retrieval (*28*). However, there is reason to suspect that memory integration could be supported by some mechanism which interleaves hippocampus-dependent memory retrieval with memory encoding (*28*). Behavioral evidence suggests that when overlapping information reminds people of distant events, people often retrieve information about prior events (*22*, *38*, *51–53*). This kind of reminder-evoked retrieval, or an analogous process (e.g. hippocampal pattern completion), may evoke overlapping hippocampal activity patterns which integrate

events during memory encoding (*15*, *17*, *20*).

However, reminder-evoked retrieval does not always lead to integration of events, and can also lead to differentiation and/or competition between events (*38*, *52*, *53*). In fact, in the current study, it was possible that re-encountering either a CN or UN character could act as a reminder of a previous event—yet, only CN events appeared to become integrated in memory. One possibility is that reminder-evoked retrieval can lead to interference between events that share weak associations or feature overlap, but can lead to integration when events can be understood as part of a larger structure (*24*, *54*, *55*). Narratives can provide a coherent structure for information that spans multiple events (*9*, *10*, *12*, *13*, *56*, *57*). Here, our findings suggest that narrative coherence may determine whether or not hippocampal activity patterns support the integration of distant events during memory encoding.

Additionally, we investigated whether hippocampal activity patterns at event boundaries, versus post-boundary timepoints within an event, support the integration of temporally-distant events. Recent models suggest that when a new event begins (i.e. at an event boundary), people might update their current representation of an event, by retrieving information about relevant past events (*7*, *22*, *36*, *58*, *59*). Similar to the idea of reminder-evoked retrieval (described above), it is possible that some kind of boundary-evoked memory retrieval could support the integration of temporally-distant events into a larger narrative. However, it is also possible that boundary-evoked hippocampal activity supports event segmentation (*32–34*, *60*), as well as memory encoding for a preceding event (*60*, *61*). Our findings suggest that, even if boundary-evoked activity in the hippocampus can reflect a “wall” between temporally adjacent events, it might also reflect the construction of a “bridge” between temporally distant events.

The anatomical distribution of our findings warrants further investigation. We are not

certain as to why narrative coherence effects within the hippocampus were right-lateralized. Some previous studies have also implicated a role for the right hippocampus in the integration of overlapping events (*15*, *41*). One possibility is that activity in the right hippocampus reflects its interactions with broader networks of cortical regions (*62*, *63*). Several studies have suggested that hippocampal-cortical interactions support the ability to continuously track information across multiple events within a narrative, as well as event boundaries (*32*, *33*, *44*, *46*, *64*). Recent findings from Heusser et al. (*65*) suggest that the Posterior Medial and Anterior Temporal subnetworks of the default mode network (*42*, *43*) provide complementary support for the representation of events within larger narratives. Future studies should elucidate the link between cortical support for processing temporally extended narratives, and hippocampal support for integrating temporally distant events in memory.

In closing, the current findings suggest that even though life's events can occur at disparate times, the hippocampus can form memories that integrate these events into a larger, coherent narrative. This integration might even take place at event boundaries. Surprisingly, even the moments that divide events in memory may hold the key to putting them back together.

## Materials and Methods

### Participants

**Functional MRI study:** Thirty-six right-handed participants aged 18-32 years-old (mean age= 24.4, SD= 3.6; 24 female, 12 male) were recruited to participate in the imaging study. Among other criteria, participants were screened for native English language fluency. After informed consent was obtained, additional screening at the time of scanning revealed that two of these participants had contraindications for MRI scanning, and thus could not be included. Of the thirty-four remaining participants scanned on Day 1, seven participants were excluded because they did not return for the recall session on Day 2 (due to unanticipated issues with scanner or peripherals, or substantial motion or imaging artifacts on Day 1). For one additional subject, a scanner error precluded imaging during two of the eight story presentation scans on Day 1, however this participant was able to complete the story listening task (without concurrent scanning during the final two blocks), and was therefore retained for participation on Day 2. As a result, twenty-seven participants were scanned on Day 2. Two of these subjects were excluded because they were unable to successfully recall any details during the recall task. In all, twenty-five participants (mean age = 25.2, SD = 3.7; 15 female, 10 male) were included in the analyses reported in this paper, twenty-four of whom had sufficient data quality during recall to be included in encoding-retrieval similarity analyses.

**Event boundary ratings**: An independent sample of eighteen participants aged 18-26 years-old (mean age=20.4 , SD=2.0; 12 female), screened for native English language fluency, provided informed consent and were recruited to provide annotations of event boundaries (i.e. no imaging data was collected).

**Experimental Stimuli**

The stimulus design is depicted in **Fig. 1A-B,** and has been described elsewhere (*9*). We constructed four fictional stories in which we manipulated whether temporally-distant events could form a coherent narrative (**Data S1**). Story audio was scripted, recorded, and thoroughly edited by the first author (BICS) to ensure equivalent lengths for all sentences (5s each), events (8 sentences/40s each), and stories (6 events/240s each), and to avoid any differences in perceptual distinctiveness (i.e. similar audio quality and smooth transitions between sentences). Importantly, most 40s events were initiated and terminated by shifts in information that are known to trigger the perception of event boundaries (e.g. characters entering or exiting, temporal shifts like "20 minutes later") (*4*, *66*). Two additional boundaries marked the start and end of each story, respectively. These seven *a priori* event boundaries for each story (green bars in **Fig. 1A**; confirmed by human raters, **Fig. S1A**) provided the basis for subsequent fMRI analysis.

Two of the stories are centered on a character named Charles, who is attempting to get a big scoop for a newspaper, and two are centered on a character named Karen, who is attempting to find employment as a chef. To examine the effects of narrative-level organization on memory, we incorporated events involving key "side-characters" into each story (**Fig. 1A**). Two stories involved "Beatrice" and "Melvin" as side-characters, and two involved "Sandra" and "Johnny." Each side-character appeared in two temporally-distant, distinct events that occurred 6-12 minutes apart (lags matched between CN and UN conditions), in disparate contexts across two unrelated stories. For instance, Sandra's first event (Event 1) occurs in Story 2, when she calls Charles, who is sitting at a park at noon on a Tuesday. Sandra's second event (Event 2) occurs in Story 4, when Karen runs into her at a French restaurant, at early evening the next day. Events involving the side-characters were tangential to adjacent main plot events which centered on

Charles and Karen (i.e. they were “sideplots”).

Critically, for two recurring side-characters, these sideplot event pairs could form a Coherent Narrative (CN) about one particular situation involving that side-character. In contrast, for the other two recurring side-characters, each sideplot event described a different situation involving that side-character, such that the two sideplot events could not easily form a singular coherent narrative (Unrelated Narratives, UN). Each story contained one CN event and one UN event (see **Fig. 1A**). Because the stimulus design controlled for several other features that could support integration of sideplot events (temporal proximity, contextual similarity, attention to intervening main plot events), only CN events, and not UN events, could be easily integrated into a larger narrative.

We sought to control for any effects of specific event content or character identity that could confound the coherence manipulation, by randomizing CN and UN event content across subjects. For each side-character, we created two alternate pairs of CN events (e.g. Sandra Events 1 and 2, version A; Sandra Events 1 and 2, version B) which had similar syntax. For a given subject, two side-characters were randomly selected to be CN, and two side-characters were randomly selected to be UN. If a side-character was selected to be CN, one of the two possible CN event pairs was selected (e.g. Sandra Events 1 and 2, version A). If a side-character was selected to be UN, the two events were drawn from different possible CN event pairs, such that they belonged to unrelated narratives (e.g. Sandra Event 1, version A, and Event 2, version B). For instance, in one CN version for Sandra, she calls to discuss someone she is dating in Event 1, and the aftermath of that date occurs in Event 2. In one UN version for Sandra, she calls to discuss someone she is dating in Event 1, and she is seen hiding from art sponsors in Event 2.

This approach resulted in 32 possible arrangements of CN and UN events, which were

pseudo-randomly assigned across subjects. For most subjects, 1 of 32 versions was randomly assigned by a seeded random number generator in MATLAB. After two-thirds of recruited subjects were scanned, it was determined that some versions had been under-sampled, therefore we generated a randomly-selected subset of the under-sampled versions to determine which version was assigned to the remaining subjects.

**Behavioral Tasks**

For the fMRI study, on Day 1, subjects first completed consent forms, demographics questionnaires, and MRI screening. Prior to entering the scanning facility, participants completed brief familiarization tasks that were aimed at facilitating their attention to the story presentation task. Then, structural and functional imaging was performed, including the story presentation task. On Day 2, additional structural and functional imaging was performed, including the recall task. Familiarization, story presentation, and recall tasks were administered in MATLAB (https://www.mathworks.com/products/matlab.html) using Psychophysics Toolbox 3 (https://www.psychtoolbox.org). For event boundary ratings, an independent subject sample completed a one-day experiment in which they provided annotations of event boundaries.

**Familiarization tasks (Day 1).** As described previously (*9*), three brief tasks were administered to familiarize participants with character names and relationships, to orient participants to the upcoming stories, and to increase the likelihood that character names and relationships could be used as successful recall cues. Participants were first presented with character names, then character relationships, and then a test on character relationships (feedback provided). To encourage familiarity, these tasks incorporated face pictures for each character that were selected for high memorability (*67*).

**Story presentation (Day 1).** Stories were presented aurally using MRI-compatible

Sensimetrics model S14 earbuds for binaural audio (https://www.sens.com/products/model-s14/). Stories 1-4 were, respectively, presented in scanning runs 1-4, and were presented again in the same order (Stories 1-4 in runs 5-8; **Fig. 1A).** This repetition of the stories was designed to increase the number of available trials and timepoints for pattern similarity analyses, which may benefit from large numbers of available trials. Prior to the task, participants were instructed verbally and onscreen that they would hear fictional clips involving characters from the familiarization tasks, and that they should devote their full attention and imagination to the clips as if they were listening to a book they enjoyed. Furthermore, they would later be asked to remember these stories in detail. Each story run began with 7 empty scanning timepoints, after which stories were presented binaurally through over-ear headphones, followed by 7 additional empty scanning timepoints. Only a white fixation cross was present onscreen during each story.

**Recall task (Day 2).** Prior to performing the scanned recall task, participants were instructed that, after seeing the name of each character, they were to recall everything they could remember involving the particular character, from all stories, in as much detail as possible (*9*). They were encouraged to attempt to recall as many details as possible for at least five minutes, and, if they were able to remember additional information, they were allowed to continue beyond five minutes. We did not instruct participants to recall events in any particular order, nor did we instruct participants to integrate information between events. Within one extended scanning run (**Fig. 1C**), participants were cued with the four CN and UN side-characters, in a randomized order. Each cue was presented onscreen, and included their relations to main protagonists, in order to encourage recall of both Events 1 and 2 (e.g. “Melvin Doyle (Charles Bort’s neighbor, Karen Joyce’s friend)”). After cueing CN and UN characters, cues were shown for the two main protagonists, in a randomized order. Each cue was preceded and followed by 7 empty scanning

timepoints, with a white fixation cross onscreen. Spoken recall was recorded using an Optoacoustics FOMRI-III MRI-compatible microphone (http://www.optoacoustics.com/medical/fomri-iii).

**Event boundary ratings:** An independent sample completed a version of the Story Presentation task at a desktop computer (i.e. without brain imaging), with an additional instruction. Participants were presented with one of two versions of the story, and during story listening, were asked to press the spacebar whenever they perceived that one event had ended, and another event had begun.

**Recall scoring**

Free recall performance was scored using a procedure developed in our previous behavioral experiments with a similar paradigm (*9*), and which adapted a well-characterized scoring method from the Autobiographical Memory Interview (*68*). Recall data were scored in a blinded fashion by segmenting each participant's typed recall into meaningful detail units, and then determining how many of these details could be verified within specific story events (**Data S2-3**). Briefly, each recall transcript was segmented into the smallest meaningful unit possible ("details"), and details were assigned labels that describe their content. We primarily sought to label details that could be verified within the story text ("verifiable details"). Verifiable details are details that specifically refer to events that center on the cued character (i.e. neither inferences about cued events, nor "external details" about other characters or events), which are recalled with some degree of certainty (e.g. not preceded by "I think" or "Maybe"), and which do not merely restate recall cues or other previously recalled details for a given cue.

We were also interested to discern recall performance for each CN or UN event (Events 1 or 2). For each CN or UN cue, verifiable details that referred to one particular event were labeled

as “Event 1” or “Event 2.” If a verifiable detail could have originated from either Event 1 or Event 2, it was scored as “Either”; these details were rare ($\bar{X}$=0.06 details per cue, SD=0.24 details/cue, max=1.5 details/cue). Finally, if a verifiable detail merged information from both Event 1 and Event 2, it was marked as “Integrated”; these details were both rare and only observed for CN events ($\bar{X}$=0.14 details/cue, SD=0.56 details/cue, max=4.5 details/cue). To ensure that analyzed details were event-specific, Either and Integrated details were excluded from all reported analyses.

Audio transcription (AG, EM, JD, MD) and recall scoring (EM, JD, MD) was performed by individuals who were blinded to the experimental hypotheses and coherence conditions. Each participant’s recall transcript was first segmented by two raters, who were required to agree on the particular start and end times for each segment. Each rater subsequently scored the recall segments for verifiable details. Interrater reliability (IRR) for CN and UN events was high (mean Pearson’s $r$=0.95), and took into account how many verifiable details were scored per event label (1, 2, Either, Integrated), per character, and per participant. All reported comparisons between imaging and recalled details are based on counts of verifiable details for CN or UN events, averaged across raters.

**MRI acquisition**

Scans were acquired at the UC Davis Imaging Research Center on a Siemens Skyra 3T scanner with a 32-channel head coil. Functional magnetic resonance imaging (fMRI) was performed using multi-band gradient-echo echo planar imaging (TR=1220 ms, TE=24ms, FA=67, MB Factor=2, FOV=19.2cm, matrix: 64 x 64, voxel size: 3.0mm isotropic, 38 slices). In order to isolate functional activity within particular structures of the brain, we also collected high-resolution structural imaging using a T1-weighted magnetization prepared rapid acquisition

gradient echo (MPRAGE) pulse sequence (1 $mm^3$ isotropic voxels). In order to ensure proper alignment between functional images between encoding and recall, we collected structural scans on both days.

On both Day 1 and Day 2, we first collected a brief localizer, and then the MPRAGE. We then adjusted the field of view for subsequent functional imaging to prioritize full inclusion of the temporal lobes (i.e. excluding some dorsal cortex, if necessary) and to align to the AC-PC axis. Initial shimming was then performed during a brief, task-free functional scan, after which we checked for artifacts (e.g. ghosting, aliasing) and determined whether the field of view needed to be adjusted (if so, this scan was repeated, with re-shimming). These preliminary scans were followed by functional imaging—story presentation scans on Day 1, and recall scans on Day 2. As time permitted, we additionally collected task-free functional imaging (Day 1), diffusion tensor imaging (Day 1 or Day 2), or additional task-concurrent imaging (Day 2, following the recall task). These additional scans are beyond the scope of the current experiment, and will be reported elsewhere.

**MRI preprocessing**

Images were processed in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), including slice-time correction and realignment. After realignment, we checked for data quality using the MemoLab QA Toolbox (https://github.com/memobc/memolab-fmri-qa), which identifies excess motion and suspect timepoints or voxels within scanning runs using base functions in MATLAB, SPM, and ArtRepair (https://www.nitrc.org/projects/art_repair/), as well as detailed plots of signal intensity across all brain voxels and timepoints for each scanning run (*69*). Furthermore, the QA Toolbox uses realignment parameters to create 6-directional motion regressors for each scan, which we implemented in subsequent models.

We excluded scanning runs if they exceeded 3mm motion in any direction, or if >20% of timepoints were identified as suspect based on framewise displacement >0.5mm or large deviations in signal intensity. We then merged realigned images into 4-dimensional files and used "film mode" in FslView (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslView/UserGuide) to inspect whether realignment failed – that is, whether there were any visible "jumps" between volumes after realignment. As previously mentioned, these approaches led us to exclude some subjects from participation in Day 2 (see Participants, above). When checking data quality for recall scans, we identified poor realignment due to large framewise displacement in two subjects. We attempted to salvage all data that preceded these moments of framewise displacement, by excluding subsequent volumes and re-performing realignment. This approach enabled us to salvage most of the recall scan for one subject (including all CN and UN recall), however, realignment still failed for the other subject. As such, one out of the twenty-five subjects could not be incorporated into encoding-retrieval similarity analyses.

Functional EPIs were co-registered with structural MPRAGE scans from the same day (Day 1 or Day 2). In order to compare functional volumes across days, we aligned images from Day 1 and Day 2, using the following approach. We co-registered Day 2 MPRAGEs to Day 1 MPRAGEs, and we applied these co-registration parameters to the realigned Day 2 images. This enabled the Day 2 functional images to directly overlap with Day 1 functional images. Additionally, we performed tissue segmentation on coregistered Day 1 MPRAGE scans, which we used to create brainmasks out of grey matter, white matter, and cerebrospinal fluid segmentations. For multi-voxel fMRI analyses (see below), these brainmasks were used to mask out extraneous voxels from first-level models of encoding and retrieval.

Furthermore, this enabled us to use the same structurally-defined regions of interest

(ROIs) for both encoding and recall. We used a published probabilistic atlas of the medial temporal lobe (https://neurovault.org/collections/3731/) (*70*) to define the left and right hippocampus, perirhinal cortex, and parahippocampal cortex, within each subject's structural MPRAGE, by reverse-transforming these ROIs from template space (in MNI space) to native space (for each subject) using Advanced Normalization Tools (ANTs; http://stnava.github.io/ANTs/). Each ROI was visually inspected against the native space scan. Because this parcellation separately defines the head, body, and tail of the hippocampus, we obtained whole-hippocampus ROIs by merging the head, body, and tail into one ROI on each side of the brain (left and right). As previously described (*34*), for exploratory analyses of right hippocampal subregions (**Fig. S3**), we merged the right hippocampal body and tail ROIs to create a right posterior hippocampus ROI, and used the right hippocampal head as the right anterior hippocampus ROI; these ROIs had similar counts of voxels. Finally, other cortical regions which were implemented in post-hoc analyses (**Figs. S5-6**) were derived from individual subject parcellations in Freesurfer (https://surfer.nmr.mgh.harvard.edu/), using the Destrieux atlas (*71*). All ROIs were then co-registered and resliced to functional images.

**Univariate analysis of boundary-evoked activation**

For univariate analysis of fMRI activation at event boundaries, structurally aligned BOLD data were smoothed with a 6mm-FWHM smoothing kernel and filtered with a 128s high-pass temporal filter prior to analysis. We modeled hippocampal activation at event boundaries using subject-level generalized linear models (GLMs) in SPM12. In order to control movement related confounds, all models include 6 rigid-body motion regressors (3 translational, 3 rotational). In line with previously reported analyses of event boundary activation (*33*), these models incorporated finite impulse response functions (FIRs) to model activation at event

boundaries in a timepoint-by-timepoint fashion. The FIR approach enabled us to not only replicate the previous finding that hippocampal activation increases at event boundaries. It also enabled us to determine the lag between the onset of an event and the fMRI response to an event boundary. Furthermore, this enabled us to determine how long the hippocampus exhibits activation at event boundaries, such that we could use this epoch to constrain subsequent multivoxel fMRI analyses (modeled separately).

Event boundaries were defined *a priori*—they were intrinsic to the stimulus design (starts and ends of 40s events; **Fig. 1A**). Within the GLMs, we modeled activation at event boundaries (Boundary-Adjacent timepoints) versus Mid-Event timepoints (20s into each event), which did not contain any shifts in information that were expected to trigger event boundaries (see story text, **Data S1**). Importantly, we did not distinguish between specific types of events in this model (e.g. CN, UN, main plot). FIRs were modeled from -2 to +16 TRs around each timepoint—we selected this window in order to model all available event timepoints with minimal overlap between Boundary-Adjacent and Mid-Event windows. As such, the GLM resulted in beta images for each regressor, for each subject, which we then masked using the left and right hippocampus ROIs, averaging betas within each ROI, for each condition and each timepoint (e.g. Boundary-Adjacent TR +3, Mid-Event TR +14, etc.).

These averaged betas were then imported into RStudio for visualization and statistical analysis. Analyses of variance were performed using the Afex package in R (https://github.com/singmann/afex), including Greenhouse-Geisser corrections for any non-sphericity. For bilateral hippocampus, permutation tests were performed using Permuco (https://cran.r-project.org/web/packages/permuco/), testing for main effects at every timepoint using 100,000 permutations and a $p<0.05$ threshold for selecting clusters of significant

timepoints, and then corrected for significance based on the mass of each cluster (*72*). For post-hoc analyses of cortical networks (**Figs. S5-6**), the cluster threshold was adjusted for the number of comparisons performed (3 comparisons, therefore $p<0.017$). This approach enabled us to account for interdependency between adjacent timepoints, in line with well-documented approaches for time-series analysis of neuroimaging data (*73*).

**Multivoxel fMRI analysis**

**Single trial models.** We isolated the unique patterns of activity evoked by each CN or UN event during encoding, and each CN or UN cue during recall, using single-trial modeling (LS-S approach; Mumford et al., 2012; Turner et al., 2012). Specifically, we used finite impulse response functions (FIRs) to extract timepoint-by-timepoint activity in each voxel, for each modeled event for a given subject. These models were performed on unsmoothed data, in order to optimally characterize spatially-distributed patterns of activity (*76*). For each story event (CN or UN), we modeled each timepoint (-2 TRs to + 45 TRs around the event onset) with a separate regressor, and used additional regressors to model each timepoint for all other story events within a given scanning run. This approach enabled the activity within each timepoint of a given event to be distinguished from activity from corresponding timepoints for all other events within a scanning run. The -2 to +45 TR window was designated to include extra available timepoints around the starts and ends of an event in order to: (1) differentiate activity corresponding with the event of interest from activity corresponding to adjacent events; and (2) capture the hemodynamic lag in the BOLD response around event boundaries (seen in the Univariate analysis of boundary-evoked activation, above).

For recall, we used a similar approach to model activity at each timepoint during a specific character cue, however, the number of available timepoints varied from cue to cue. For

each recall cue, we therefore modeled timepoint regressors from -2 TRs before the onset of a recall cue to 7 TRs after the offset of that recall cue (corresponding with inter-cue times during scanning), and an equivalent number of regressors was designated to model activity from any other recall cue in the extended scanning run. For all models, we incorporated a 128s high-pass filter as well as 6 rigid-body motion regressors (3 translational, 3 rotational). This approach enabled us to distinguish patterns of activity that were evoked by each recall cue.

**Representational similarity analysis:** Once we isolated the unique patterns of activity associated with each story event and recall cue, we sought to test our hypotheses by comparing these patterns of activity across events. This type of analysis is referred to as Representational Similarity Analysis (RSA) (*77*). In brief, RSA involves extracting the pattern of activity that is spread across a group of voxels in a given part of the brain (a "voxel pattern"), and computing correlations between voxel patterns that are evoked by different trials in an experiment. If voxel patterns are more highly correlated between trials in one experimental condition versus another, this provides evidence to suggest that whichever information distinguishes between those conditions is supported by that particular area of the brain.

To this end, we adapted several functions from the RSA Toolbox (http://www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/) and developed custom code in MATLAB. Voxel patterns were extracted within all six ROIs for a given subject, for every modeled timepoint for every trial (story event or recall cue). For each subject, and within each ROI, we then constructed a representational similarity matrix (RSM), by correlating voxel patterns (Pearson's r) from every modeled timepoint with every modeled timepoint.

To test our hypotheses about how activity patterns are shared between CN versus UN events, and between these events and recall, we selectively averaged correlations from the

overall RSM. For analyses of encoding, we selected all correlations between timepoints from two events involving the same CN or UN character (e.g. Beatrice Event 1 vs Beatrice Event 2). For encoding-retrieval similarity, we selected all correlations between timepoints from a CN or UN encoding event and timepoints corresponding to the recall cue for the same CN or UN character (e.g. Beatrice Event 1 vs Beatrice recall cue). Importantly, all of these correlations were computed across scanning runs (at least two scans apart), such that any effects of temporal autocorrelation would be minimized. After selecting these specific sets of correlations, we averaged them to derive one condition-specific, timepoint-by-timepoint correlation matrix for each ROI, for each subject**.** As described below, within each condition, we then averaged pattern similarity values (mean Pearson's r) within empirically-defined epochs (e.g. Boundary epoch).

**Across-event similarity epochs:** To test our specific hypotheses about representational similarity involving event boundaries, we computed selective averages of each condition-specific RSA matrix (described above) based on empirically-derived Boundary and Post-Boundary epochs, which were then imported into RStudio for statistical analysis. Within each across-event similarity matrix, for each subject and ROI, we selectively averaged timepoint-by-timepoint correlations between events (**Fig. 2**): Boundary-Boundary similarity, between TRs +5 to +11 from one event, and the same TRs from the other event; Boundary-Post-Boundary similarity, between TRs +5 to +11 from one event, and TRs +12 to +37 from the other event; and Post-Boundary-Post-Boundary similarity, between TRs +12 to +37 from one event, and the same TRs from the other event. These three measures of epoch-by-epoch pattern similarity constitute the three levels of the "Epoch" factor within mixed models of across-event pattern similarity (see **Multi-level mixed effects models**, below).

**Encoding-retrieval similarity:** Within the right hippocampus, for each encoding event

(Event 1 or 2), we selectively averaged timepoint-by-timepoint correlations between timepoints spanning both Boundary and Post-Boundary epochs at encoding (i.e. the whole event), and any timepoints during recall of the same CN or UN character (**Fig. 3**). This yielded two measures of encoding-retrieval similarity: Event 1 Reinstatement and Event 2 Reinstatement. These measures were specified by a factor of Event Number [Event 1 Reinstatement, Event 2 Reinstatement] within mixed models (see **Multi-level mixed effects models**, below). For ancillary analyses, we also selectively averaged correlations between recall and the Boundary epoch alone (**Fig. S4**).

**Univariate analysis of narrative coherence**

To supplement the pattern similarity analyses described above, we analyzed whether hippocampal activation in the Boundary epoch was modulated by narrative coherence. This analysis was conducted using beta volumes which were initially modeled for representational similarity analysis (see **Single trial models**, above), and then smoothed with a 6mm-FWHM smoothing kernel for use in univariate analysis. To directly compare univariate activity with pattern similarity findings in the left and right hippocampus, we extracted beta values during the Boundary epoch, and averaged betas within left and right hippocampus ROIs. Overall boundary activation within each ROI was incorporated into mixed models of hippocampal pattern similarity during memory encoding (see **Multi-level mixed effects models**, below).

**Statistical Analysis**

Statistical tests were performed in R using standard functions t-tests, as well as Afex (https://github.com/singmann/afex) for mixed models and ANOVAs, and Permuco (https://cran.r-project.org/web/packages/permuco/) for permutation tests. For ANOVAs, Greenhouse-Geisser corrections were implemented for any non-sphericity. For visualization (**Figs. S1** and **S6**), within-subjects standard error was calculated using Rmisc (https://cran.r-

project.org/web/packages/Rmisc/index.html), by normalizing data to remove between-subject variability, and computing variance from this normalized data (*78*).

**Multi-level mixed effects models:** Mixed model formulas are reported within the Results text. Fixed effects specifications were based on experimental hypotheses, random effects structures which accounted for effects of individual subjects ("Subject") and the content of particular events ("Event Content"). For models of across-event similarity at encoding, effects of Event Content (i.e. "item" effects) were operationalized as the particular pair of events which were presented for a particular CN or UN character (e.g. Beatrice Events 1A+2B, Melvin Events 1B+2B). For models of encoding-retrieval similarity or recalled details, Event Content was operationalized as which particular event was being recalled (e.g. Beatrice Event 1A, Beatrice Event 2B, Melvin Event 1B, or Melvin Event 2B). Random effects structures were specified by first attempting to fit a maximal random effects structure justified by the experimental design (including random slopes) (*79*), followed by systematically pruning the random effects structure until the model converged without a singular solution (i.e. without overfitting; Singmann & Kellen, 2019). Because models which included random slopes either did not converge or reached a singular solution, reported models only included intercepts for random effects. This approach resulted in the following model formulas:

Formula 1 (**Fig. 2B**): Pattern similarity ~ Narrative Coherence * Epoch + Overall Boundary Activation + (1 | Subject) + (1 | Event Content)

Formula 2 (**Fig. 3B**): Encoding-retrieval similarity ~ Narrative Coherence * Event Number + Recall Duration + (1 | Subject) + (1 | Event Content)

Formula 3 (**Fig. 3C**): Recalled details ~ Event 2 Reinstatement * Narrative Coherence * Event Recalled + Recall duration + (1 | Subject) + (1 | Event Content)

Model fits were quantified using the Aikake Information Criterion (AIC), and significance for fixed effects was calculated using the Kenward-Roger approximation for degrees of freedom, which is useful for unbalanced mixed designs (*80*). To follow up on significant fixed effects and interactions, we used the emmeans package (https://cran.r-project.org/web/packages/emmeans/index.html) to conduct pairwise comparisons and visualization of factors (emmeans; **Figs. 2-3, S2-5**) and interactions between factors and continuous variables (emtrends and emmip; **Fig. 3C**, **Fig. S4C**).

## References


1. E. Tulving, *Episodic and semantic memory. From E. Tulving & W. Donaldson,(Eds.) Organization of memory (pp. 381-403)* (New York. Academic Press, 1972).

2. J. M. Zacks, Event Perception and Memory. *Annu. Rev. Psychol.* **71**, 165–191 (2020).

3. D. Clewett, C. Gasser, L. Davachi, Pupil-linked arousal signals track the temporal organization of events in memory. *Nat. Commun.* **11**, 4007 (2020).

4. Y. Ezzyat, L. Davachi, What constitutes an episode in episodic memory? *Psychol. Sci.* **22**, 243–252 (2011).

5. A. J. Horner, J. A. Bisby, A. Wang, K. Bogus, N. Burgess, The role of spatial boundaries in shaping long-term event representations. *Cognition*. **154**, 151–164 (2016).

6. K. M. Swallow, D. M. Barch, D. Head, C. J. Maley, D. Holder, J. M. Zacks, Changes in events alter how people remember recent information. *J. Cogn. Neurosci.* **23**, 1052–1064 (2011).

7. J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, J. R. Reynolds, Event perception: a mind-brain perspective. *Psychol. Bull.* **133**, 273–293 (2007).

8. G. A. Radvansky, D. E. Copeland, Walking through doorways causes forgetting: Situation models and experienced space. *Mem. Cognit.* **34**, 1150–1156 (2006).

9. B. Cohn-Sheehy, A. Delarazan, Z. Reagh, N. Mundada, A. P. Yonelinas, J. M. Zacks, C. Ranganath, "Narratives bridge the divide between distant events in episodic memory" (preprint, PsyArXiv, 2020), , doi:10.31234/osf.io/bmt4s.

10. F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press, New York, 1932).

11. A. C. Graesser, M. Singer, T. Trabasso, Constructing inferences during narrative text comprehension. *Psychol. Rev.* **101**, 371–395 (1994).

12. T. Trabasso, T. Secco, P. Van Den Broek, in *Learning and Comprehension of Text*, H. Mandl, N. L. Stein, T. Trabasso, Eds. (Lawrence Erlbaum Associates, Inc., Hillsdale, N.J., U.S., 1984).

13. T. A. van Dijk, W. Kintsch, *Strategies of discourse comprehension* (Academic Press, New York, 1983).

14. R. M. Willems, S. A. Nastase, B. Milivojevic, Narratives for Neuroscience. *Trends Neurosci.* **43**, 271–273 (2020).

15. A. J. Horner, J. A. Bisby, D. Bush, W.-J. Lin, N. Burgess, Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat. Commun.* **6**, 7462 (2015).

16. B. A. Kuhl, A. T. Shah, S. DuBrow, A. D. Wagner, Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501–506 (2010).

17. M. L. Schlichting, J. A. Mumford, A. R. Preston, Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun.* **6**, 8151 (2015).

18. M. L. Schlichting, D. Zeithamova, A. R. Preston, CA1 subfield contributions to memory integration and inference: Memory Integration in CA1. *Hippocampus.* **24**, 1248–1260 (2014).

19. D. Shohamy, A. D. Wagner, Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*. **60**, 378–389 (2008).

20. D. Zeithamova, A. L. Dominick, A. R. Preston, Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*. **75**, 168–179 (2012).

21. D. Zeithamova, A. R. Preston, Flexible Memories: Differential Roles for Medial Temporal Lobe and Prefrontal Cortex in Cross-Episode Binding. *J. Neurosci.* **30**, 14676–14684 (2010).

22. D. Stawarczyk, C. N. Wahlheim, J. A. Etzel, A. Z. Snyder, J. M. Zacks, Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. *Proc. Natl. Acad. Sci.* **117**, 29346–29353 (2020).

23. G. A. Radvansky, R. T. Zacks, Mental models and the fan effect. *J. Exp. Psychol. Learn. Mem. Cogn.* **17**, 940–953 (1991).

24. G. A. Radvansky, Across the Event Horizon. *Curr. Dir. Psychol. Sci.* **21**, 269–272 (2012).

25. A. J. H. Chanales, A. Oza, S. E. Favila, B. A. Kuhl, Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Curr. Biol.* **27**, 2307-2317.e5 (2017).

26. S. E. Favila, A. J. H. Chanales, B. A. Kuhl, Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun.* **7**, 11066 (2016).

27. A. J. H. Chanales, N. M. Dudukovic, F. R. Richter, B. A. Kuhl, Interference between overlapping memories is predicted by neural states during learning. *Nat. Commun.* **10**, 5363 (2019).

28. D. Kumaran, What representations and computations underpin the contribution of the hippocampus to generalization and inference? *Front. Hum. Neurosci.* **6** (2012), doi:10.3389/fnhum.2012.00157.

29. V. J. H. Ritvo, N. B. Turk-Browne, K. A. Norman, Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends Cogn. Sci.* **23**, 726–742 (2019).

30. S. H. P. Collin, B. Milivojevic, C. F. Doeller, Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci.* **18**, 1562–1564 (2015).

31. B. Milivojevic, A. Vicente-Grabovetsky, C. F. Doeller, Insight Reconfigures Hippocampal-Prefrontal Memories. *Curr. Biol.* **25**, 821–830 (2015).

32. C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, K. A. Norman, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. **95**, 709-721.e5 (2017).

33. A. Ben-Yakov, R. N. Henson, The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J. Neurosci.* **38**, 10057–10068 (2018).

34. Z. M. Reagh, A. I. Delarazan, A. Garber, C. Ranganath, Aging alters neural activity at event boundaries in the hippocampus and Posterior Medial network. *Nat. Commun.* **11**, 3980 (2020).

35. D. Newtson, Attribution and the unit of perception of ongoing behavior. *J. Pers. Soc. Psychol.* **28**, 28–38 (1973).

36. B. J. Griffiths, L. Fuentemilla, Event conjunction: How the hippocampus integrates episodic memories across event boundaries. *Hippocampus*. **30**, 162–171 (2020).

37. W. Kintsch, in *Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes.*, A. F. Healy, S. M. Kosslyn, R. M. Shiffrin, Eds. (Lawrence Erlbaum Associates, Inc., Hillsdale, N.J., U.S., 1992), vol. 2, pp. 261–278.

38. B. H. Ross, G. L. Bradshaw, Encoding effects of remindings. *Mem. Cognit.* **22**, 591–605 (1994).

39. M. Silva, C. Baldassano, L. Fuentemilla, Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. *J. Neurosci.* **39**, 8538–8548 (2019).

40. I. Sols, S. DuBrow, L. Davachi, L. Fuentemilla, Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Curr. Biol.* **27**, 3499-3504.e4 (2017).

41. B. Milivojevic, M. Varadinov, A. Vicente Grabovetsky, S. H. P. Collin, C. F. Doeller, Coding of Event Nodes and Narrative Context in the Hippocampus. *J. Neurosci. Off. J. Soc. Neurosci.* **36**, 12412–12424 (2016).

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

43. Z. M. Reagh, C. Ranganath, What does the functional organization of cortico-hippocampal networks tell us about the functional organization of memory? *Neurosci. Lett.* **680**, 69–76 (2018).

44. M. Aly, J. Chen, N. B. Turk-Browne, U. Hasson, Learning Naturalistic Temporal Structure in the Posterior Medial Network. *J. Cogn. Neurosci.* **30**, 1345–1365 (2018).

45. J. Chen, Y. C. Leong, C. J. Honey, C. H. Yong, K. A. Norman, U. Hasson, Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).

46. J. Chen, C. J. Honey, E. Simony, M. J. Arcaro, K. A. Norman, U. Hasson, Accessing Real-Life Episodic Information from Minutes versus Hours Earlier Modulates Hippocampal and High-Order Cortical Dynamics. *Cereb. Cortex N. Y. N 1991*. **26**, 3428–3441 (2016).

47. E. Simony, C. J. Honey, J. Chen, O. Lositsky, Y. Yeshurun, A. Wiesel, U. Hasson, Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* **7**, 12141 (2016).

48. N. K. Speer, J. R. Reynolds, K. M. Swallow, J. M. Zacks, Reading Stories Activates Neural Representations of Visual and Motor Experiences. *Psychol. Sci.* **20**, 989–999 (2009).

49. K. Tylén, P. Christensen, A. Roepstorff, T. Lund, S. Østergaard, M. Donald, Brains striving for coherence: Long-term cumulative plot formation in the default mode network. *NeuroImage*. **121**, 106–114 (2015).

50. J. M. Zacks, The brain's cutting-room floor: segmentation of narrative cinema. *Front. Hum. Neurosci.* **4** (2010), doi:10.3389/fnhum.2010.00168.

51. L. L. Jacoby, The role of mental contiguity in memory: registration and retrieval effects. *J. Verbal Learn. Verbal Behav.* **13**, 483–496 (1974).

52. C. N. Wahlheim, L. L. Jacoby, Remembering change: The critical role of recursive remindings in proactive effects of memory. *Mem. Cognit.* **41**, 1–15 (2013).

53. C. N. Wahlheim, J. M. Zacks, Memory guides the processing of event changes for older and younger adults. *J. Exp. Psychol. Gen.* **148**, 30–50 (2019).

54. A. J. Horner, N. Burgess, Pattern Completion in Multielement Event Engrams. *Curr. Biol.* **24**, 988–992 (2014).

55. A. J. Horner, N. Burgess, The associative structure of memory for multi-element events. *J. Exp. Psychol. Gen.* **142**, 1370–1383 (2013).

56. D. E. Rumelhart, in *Basic processes in reading: Perception and comprehension*, D. Laberge, S. J. Samuels, Eds. (Lawrence Erlbaum Associates, Hillsdale, N.J., 1977), pp. 265–303.

57. P. W. Thorndyke, Cognitive structures in comprehension and memory of narrative discourse. *Cognit. Psychol.* **9**, 77–110 (1977).

58. N. T. Franklin, K. A. Norman, C. Ranganath, J. M. Zacks, S. J. Gershman, Structured Event Memory: A neuro-symbolic model of event cognition. *Psychol. Rev.* **127**, 327–361 (2020).

59. Q. Lu, U. Hasson, K. A. Norman, "Learning to use episodic memory for event prediction" (preprint, bioRxiv, 2020).

60. A. Ben-Yakov, Y. Dudai, Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *J. Neurosci.* **31**, 9032–9042 (2011).

61. A. H. Sinclair, G. M. Manalili, I. K. Brunec, R. A. Adcock, M. D. Barense, "Prediction errors during naturalistic events modulate hippocampal representations and drive episodic memory updating" (preprint, Neuroscience, 2020), , doi:10.1101/2020.09.29.319418.

62. A. J. Barnett, W. Reilly, H. R. Dimsdale-Zucker, E. Mizrak, Z. Reagh, C. Ranganath, "Organization of cortico-hippocampal networks in the human brain" (preprint, Neuroscience, 2020), , doi:10.1101/2020.06.09.142166.

63. B. R. Geib, M. L. Stanley, E. A. Wing, P. J. Laurienti, R. Cabeza, Hippocampal Contributions to the Large-Scale Episodic Memory Network Predict Vivid Visual Memories. *Cereb. Cortex.* **27**, 680–693 (2017).

64. S. Michelmann, A. R. Price, B. Aubrey, W. K. Doyle, D. Friedman, P. C. Dugan, O. Devinsky, S. Devore, A. Flinker, U. Hasson, K. A. Norman, "Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions" (preprint, Neuroscience, 2020), , doi:10.1101/2020.12.09.416438.

65. A. C. Heusser, P. C. Fitzpatrick, J. R. Manning, Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nat. Hum. Behav.* (2021), doi:10.1038/s41562-021-01051-6.

66. J. M. Zacks, N. K. Speer, J. R. Reynolds, Segmentation in reading and film comprehension. *J. Exp. Psychol. Gen.* **138**, 307–327 (2009).

67. W. A. Bainbridge, P. Isola, A. Oliva, The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen.* **142**, 1323–1334 (2013).

68. B. Levine, E. Svoboda, J. F. Hay, G. Winocur, M. Moscovitch, Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol. Aging.* **17**, 677–689 (2002).

69. J. D. Power, A simple but useful way to assess fMRI scan qualities. *NeuroImage*. **154**, 150–158 (2017).

70. M. Ritchey, M. E. Montchal, A. P. Yonelinas, C. Ranganath, Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. *eLife*. **4** (2015), doi:10.7554/eLife.05025.

71. C. Destrieux, B. Fischl, A. Dale, E. Halgren, Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*. **53**, 1–15 (2010).

72. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods*. **164**, 177–190 (2007).

73. M. X. Cohen, *Analyzing neural time series data: theory and practice* (2014).

74. J. A. Mumford, B. O. Turner, F. G. Ashby, R. A. Poldrack, Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*. **59**, 2636–2643 (2012).

75. B. O. Turner, J. A. Mumford, R. A. Poldrack, F. G. Ashby, Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*. **62**, 1429–1438 (2012).

76. N. Kriegeskorte, R. Goebel, P. Bandettini, Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3863–3868 (2006).

77. N. Kriegeskorte, M. Mur, P. Bandettini, Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

78. R. D. Morey, Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* **4**, 61–64 (2008).

79. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).

80. H. Singmann, D. Kellen, in *New Methods in Cognitive Psychology*, D. H. Spieler, E. Schumacher, Eds. (Psychology Press, 2019), pp. 4–31.

## Acknowledgments

We thank Alexander Garber, Elena Markantonakis, Mehar Durah, and June Dy for transcribing and scoring recall. We thank Jesse Doty, Sarah Morse, Costin Tanase, Dennis Thompson, and the Imaging Research Center for their technical contributions. We thank Halle Dimsdale-Zucker, Derek Huffman, Anna Leshinskaya, Walter Reilly, Andy Yonelinas, and the Dynamic Memory Lab (http://dml.ucdavis.edu) for consultation on experimental design and analysis, and James Antony and Cameron Riddell for feedback on the manuscript. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views of the Office of Naval Research, U.S. Department of Defense, or National Institutes of Health.

**Funding:** Multi-University Research Initiative grant N00014-17-1-2961 from the United States Office of Naval Research/Department of Defense (CR, JMZ); Ruth L. Kirchstein National Research Service Award F30AG062053 from the National Institute on Aging (BICS); Floyd and Mary Schwall Medical Research Fellowship from the University of California, Davis (BICS)

**Author contributions:** Conceptualization, BICS and CR; Methodology, BICS, CR, ZMR, KK, AJB; Investigation, BICS, AID.; Analysis, BICS, AID; Software, BICS, JECD, KK, AID, ZMR, AJB; Writing—Original Draft, BICS, CR; Writing – Review and Editing, all authors; Funding Acquisition, CR, JMZ, BICS; Supervision, CR, ZMR, JMZ.

**Competing interests:** The authors declare no competing interests.

**Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and Supplementary Materials. Upon publication, additional data, code, and materials will be available in the following repository: https://osf.io/7zsxd/.

## Supplementary Figures

**Figure S1: A priori event boundaries evoke boundary perception and increased hippocampal activation.** Each story was written to include *a priori* event boundaries, such that it would naturally be segmented into six 40-second events (see Methods). (A) Perceived boundaries correspond with *a priori* boundaries: An independent sample (N=18, not scanned) annotated perceived event boundaries during story presentation. The total number of observers who perceived a boundary during each 5-second story sentence (N(Observors), green; dots/line = mean across stories, vertical lines = range) was compared with the locations of *a priori* event boundaries (black dots/line; not including story onset and offset). (B) Boundary-evoked hippocampal activation: BOLD activation time-courses were modeled in the left and right hippocampus, to estimate activity changes at Boundary-Adjacent (green; -2 to +16 TRs around an a priori event boundary) versus Mid-Event timepoints (purple; 20 seconds into each 40-second event). These activation time-courses were modeled using an analysis of variance (ANOVA) with three within-subjects factors: Condition [Boundary-Adjacent vs Mid-Event], Timepoint, and Hemisphere [left vs right hippocampus]. This analysis revealed significant main effects of Condition ($F(1,24)=29.0$, $\eta_G^2=.007$, $p<.0001$) and Timepoint ($F(2.56, 61.37)=4.92$, $\eta_G^2=.02$, $p=.006$) that were qualified by a significant Condition by Timepoint interaction ($F(5.06, 121.53)=11.2$, $\eta_G^2=.02$, $p<.0001$; for other effects, ps>.15). Following up on the significant Condition by Timepoint interaction, cluster-based permutation testing revealed a cluster of Boundary-Adjacent timepoints where activity was significantly higher than corresponding Mid-Event timepoints (grey shading; TRs +5 to +11; cluster-defining threshold is $p<.05$). We incorporated both the left and right hippocampus into this analysis because we did not observe

any effects of Hemisphere in the preceding ANOVA. Means are represented by dots and lines, and error bars represent within-subjects estimates of standard error (see Methods for details).

**Figure S2: No effect of narrative coherence on left hippocampus pattern similarity.** Left hippocampus pattern similarity between Event 1 and 2 epochs was computed for Coherent Narrative and Unrelated Narrative events. Pattern similarity for individual subjects (colored dots) is plotted on the basis of which epoch was being examined, and whether events could form Coherent Narratives or not (Unrelated Narratives). Mixed model estimates predicting pattern similarity are visualized as 95% confidence intervals (see text for model details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives.

**Figure S3: Exploratory pattern similarity analysis in right anterior and posterior hippocampus.** Following up on findings in the right hippocampus (Figure 3), pattern similarity between Event 1 and 2 epochs was computed for Coherent Narrative and Unrelated Narrative events, within the right anterior (A) and posterior hippocampus (B). Pattern similarity for individual subjects (colored dots) is plotted on the basis of which epoch was being examined, and whether events could form Coherent Narratives or not (Unrelated Narratives). Mixed model estimates predicting pattern similarity are visualized as 95% confidence intervals (similar to Figure 3); pairwise differences in right posterior hippocampus reflect a significant main effect of Narrative Coherence on pattern similarity ($F(1,11.66)=6.85$, $p=0.023$; model AIC=-868.56). This main effect was not significant within the right anterior hippocampus ($F(1,11.73)=2.16$, $p=0.17$; model AIC=-998.62). Key: Blue=Coherent Narratives, Red=Unrelated Narratives; *=p<.05, +=p<.10, uncorrected for multiple comparisons.

**Figure S4: Reinstatement of boundary-evoked hippocampal activity from memory encoding supports recall of Coherent Narrative events.** (A) Encoding-retrieval similarity: Pattern similarity (Pearson's r) was calculated between boundary-evoked activity from each Event (Event 1 Boundary, Event 2 Boundary) involving a recurring character at encoding, and delayed recall of events involving that character. (B) Event 2 boundary pattern reinstatement is higher for Coherent Narrative recall: Encoding-retrieval similarity for each subject (colored dots) is plotted for each epoch of each event during recall. 95% confidence intervals represent mixed model estimates (AIC=-1427.5; interaction of Narrative Coherence by Event Number: $F(1,343.00)=7.58$, $p=.006$). (C) Event recall predicted by Event 2 boundary pattern reinstatement: Recalled details from Event 1 (circles) and Event 2 (triangles) are plotted for each subject, predicted by Event 2 boundary pattern reinstatement during recall. Colored lines represent mixed model estimates (AIC=1114.7; interaction of Event 2 Reinstatement by Narrative Coherence: $F(1,160.15)=10.56$, $p=.001$), which were significantly different ($t(160)=3.25$, $p=.001$, uncorrected for multiple comparisons). Key: Blue=Coherent Narrative, Red=Unrelated Narratives; **=p<.01, ***=p<.001, uncorrected for multiple comparisons.

**Figure S5: Post-hoc analysis of narrative coherence effects in cortical networks.** (A,C,E) Regions of interest within three cortical networks are overlaid on a single-subject brain: the Posterior Medial Network (A: medial prefrontal cortex, posterior medial cortex, angular gyrus, parahippocampal cortex), Anterior Temporal Network (C: perirhinal cortex, orbitofrontal cortex, temporal pole), and Bilateral Auditory Cortex (transverse temporal gyrus). (B,D,F) Similar to Figure 3, Pearson's correlations between Event 1 and 2 epochs were computed for Coherent

Narrative and Unrelated Narrative events. Pattern similarity for individual subjects (colored dots) is plotted on the basis of which epoch was being examined, and whether events could form Coherent Narratives or not (Unrelated Narratives). 95% confidence intervals (black) represent the predictions of mixed models with the following formula: Pattern similarity ~ Narrative Coherence * Epoch * Region of interest (ROI) + (1 | Subject) + (1 | Event Content). Pairwise comparisons following up on significant fixed effects were corrected for multiple comparisons using the Holm method. (B) Within the Posterior Medial Network (AIC=-4372.2), this model revealed significant effects of Epoch ($F(2,2315.22)=3.38$, $p=.034$) and ROI ($F(7,2315.22) = 19.05$, $p<0.001$; not shown here), however, the effect of Narrative Coherence was not significant ($F(1,12.35)=2.20$, $p=0.16$; for interactions, ps>.15). (D) Within the Anterior Temporal Network (AIC=-5159.5), there was a significant effect of ROI ($F(5,1727.42)=3.04$,$p=.01$; not shown here) and a significant interaction between Narrative Coherence and Epoch ($F(2,1727.42)=6.95$, $p<.001$; all other effects, p>.07). (F) Within bilateral auditory cortex (AIC=-608.1), there were significant effects of ROI ($F(1,552.88)=18.61$, $p<0.001$; not shown here), Epoch ($F(2,552.88)=5.10$, $p=0.006$), and Narrative Coherence ($F(1,8.81)=13.21$, $p=0.006$). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, *=p<0.05.

**Figure S6: Post-hoc analysis of boundary-evoked activation in cortical networks.** Similar to Figure 2B, Boundary-Adjacent activation (green) is compared with Mid-Event activation (purple) within the Posterior Medial Network (A), Anterior Temporal Network (B), and Bilateral Auditory Cortex (C). Means are represented by dots and lines, and error bars represent within-subjects estimates of standard error (see Methods for details). Grey shading (A) represents a significant cluster of boundary activation in the Posterior Medial Network (TRs +4 to +14;

cluster-defining threshold corrected for multiple comparisons, $p<0.017$).

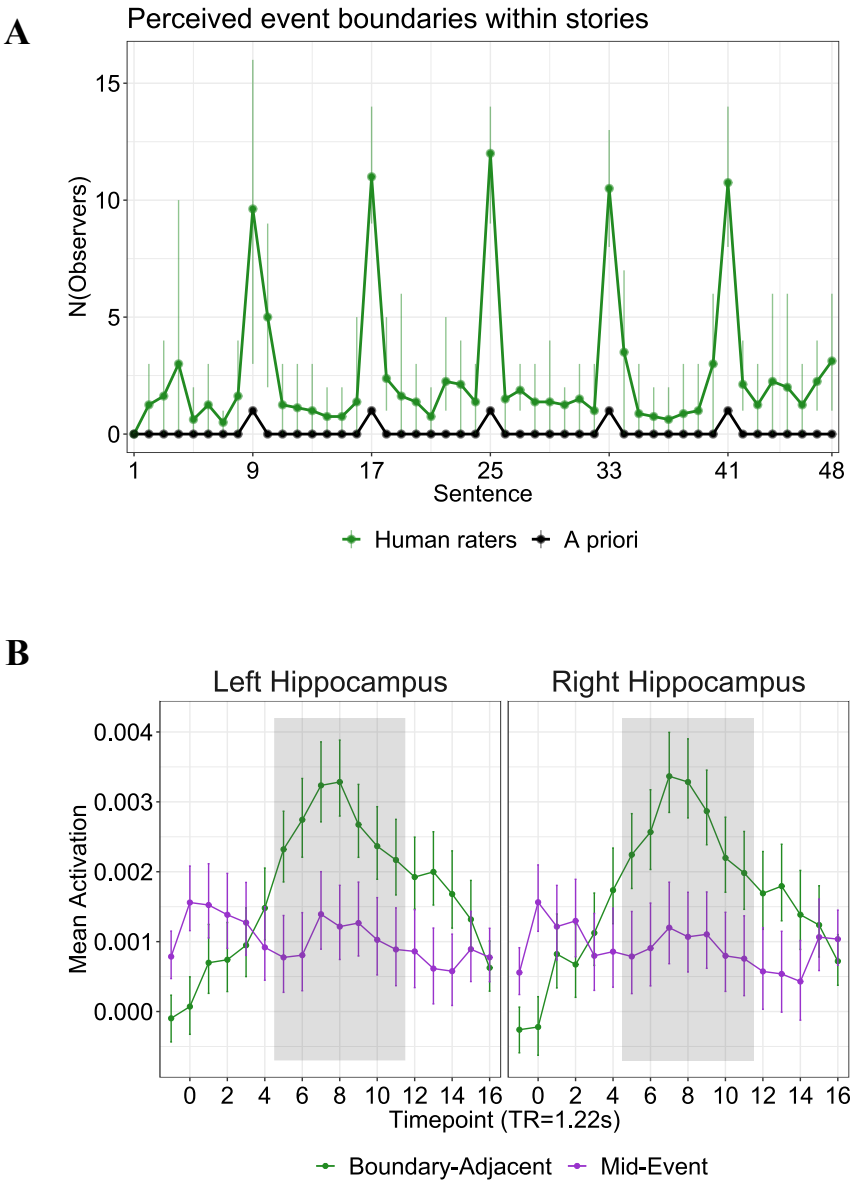


**Figure S1: A priori event boundaries evoke boundary perception and increased hippocampal activation.**

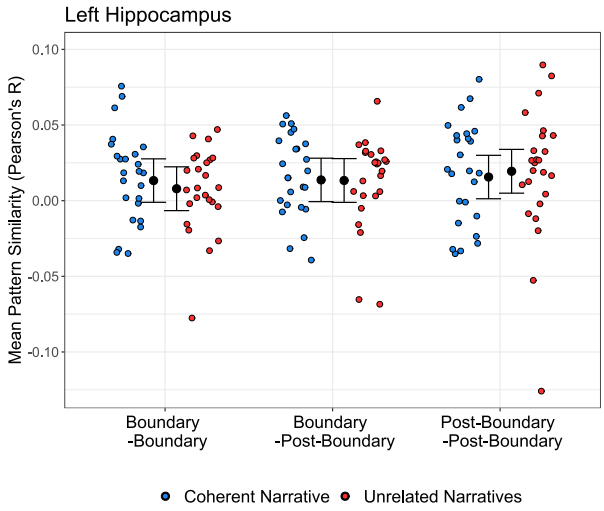


**Figure S2: No effect of narrative coherence on left hippocampus pattern similarity.**

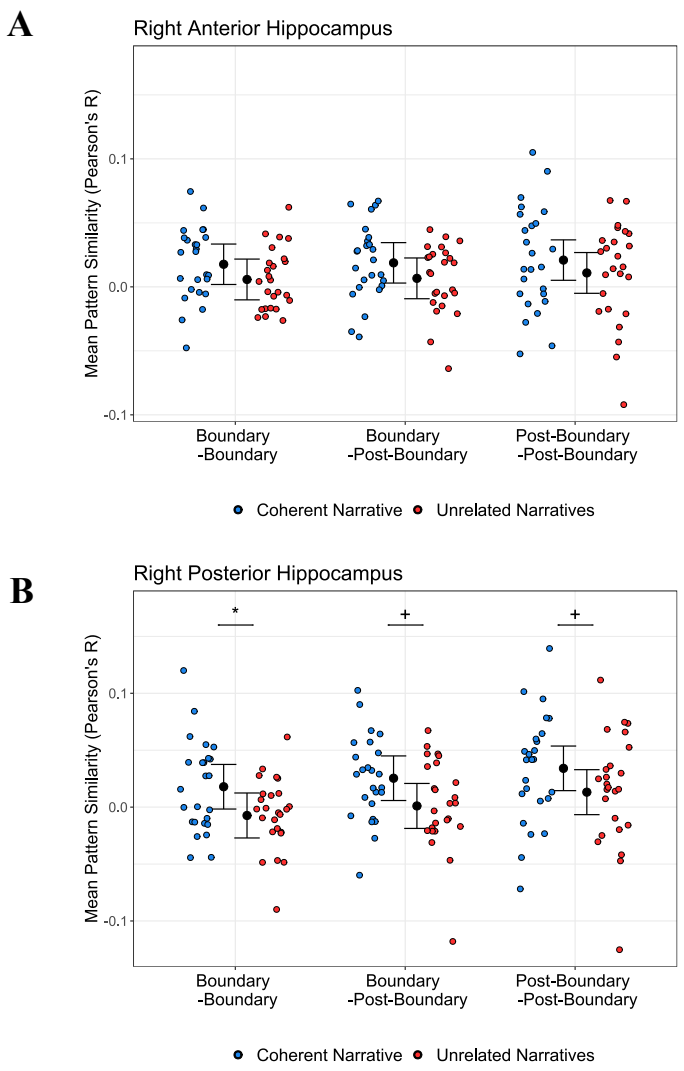


**Figure S3: Exploratory pattern similarity analysis in right anterior and posterior hippocampus.**

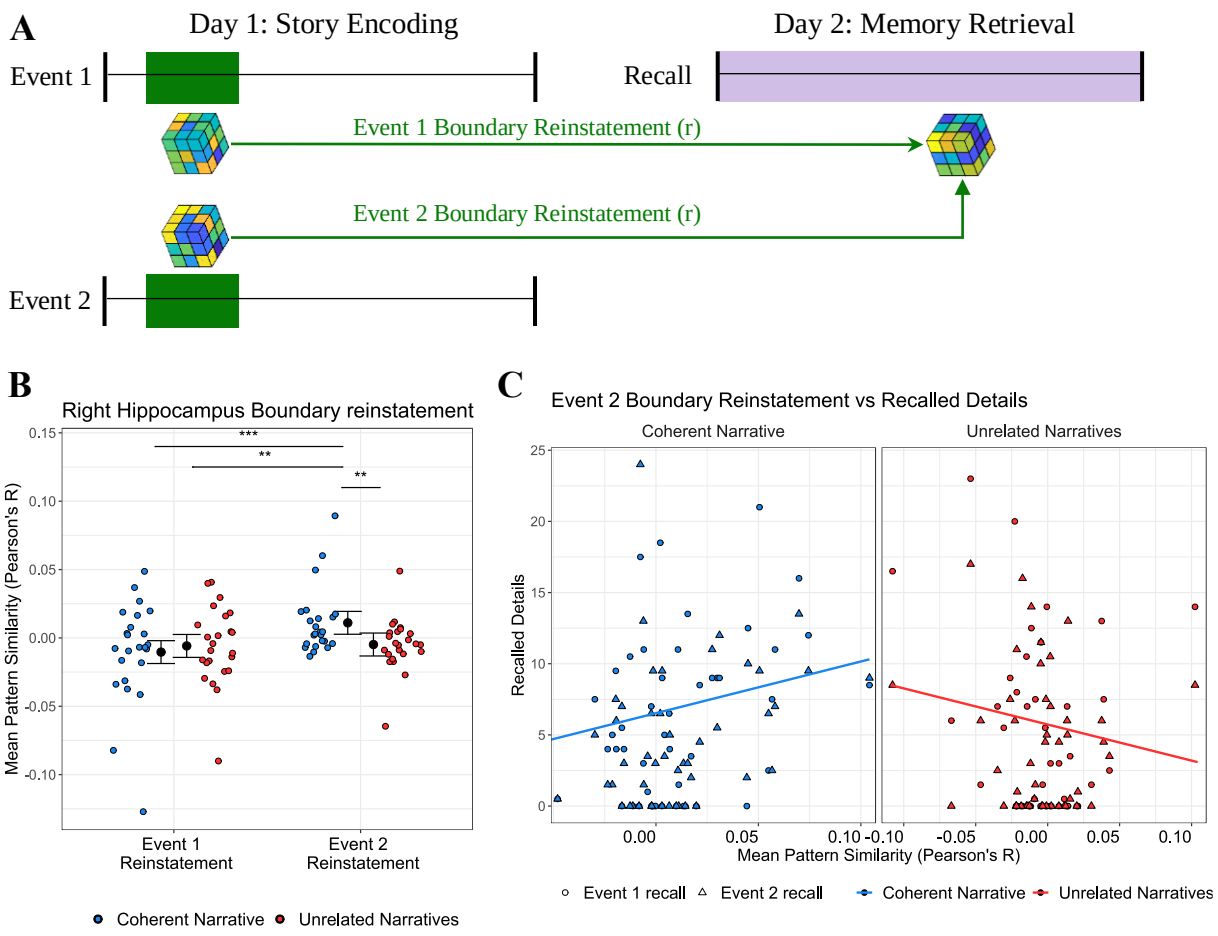


**Figure S4: Reinstatement of boundary-evoked hippocampal activity from memory encoding supports recall of Coherent Narrative events.**

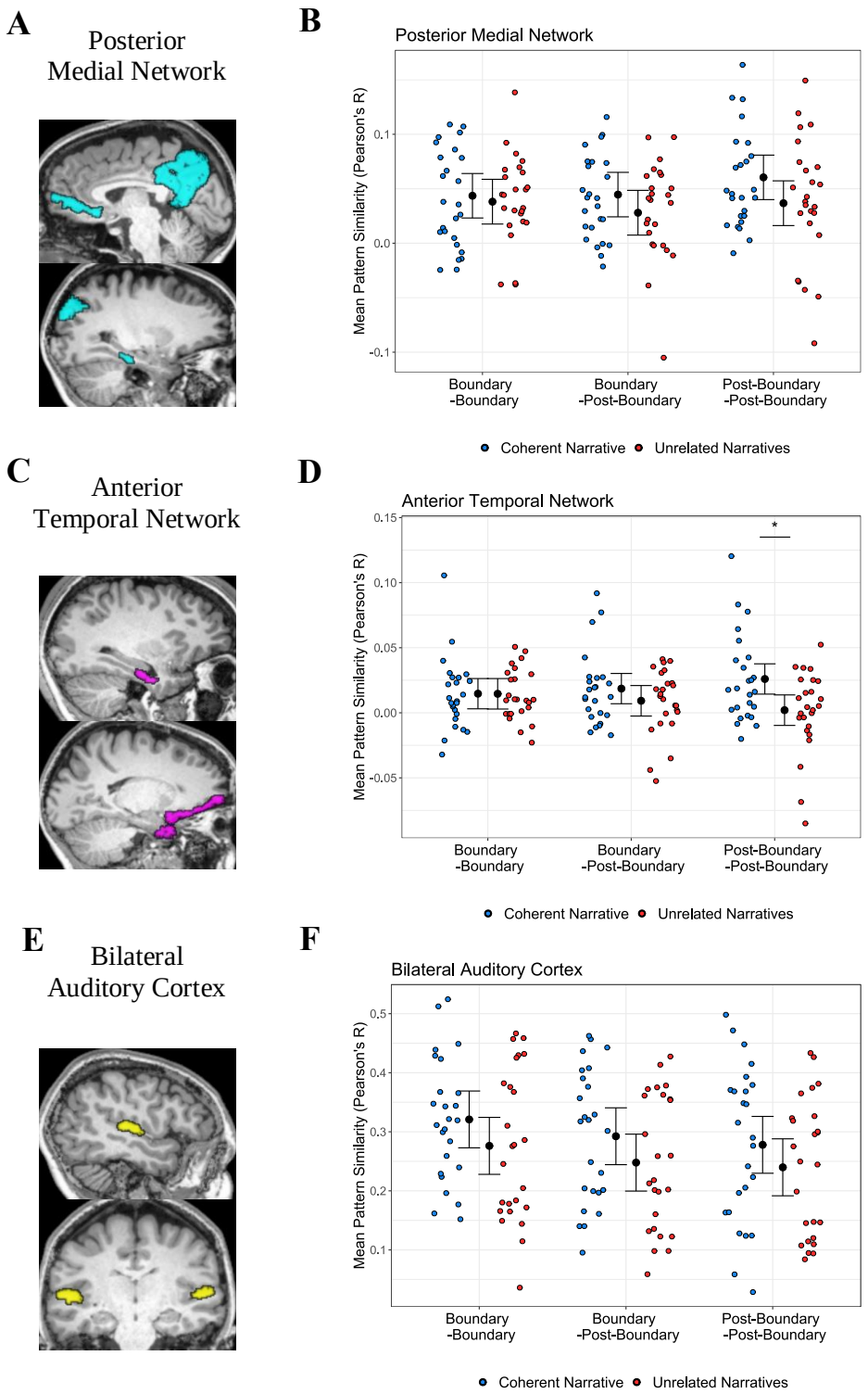


**Figure S5: Post-hoc analysis of narrative coherence effects in cortical networks.**

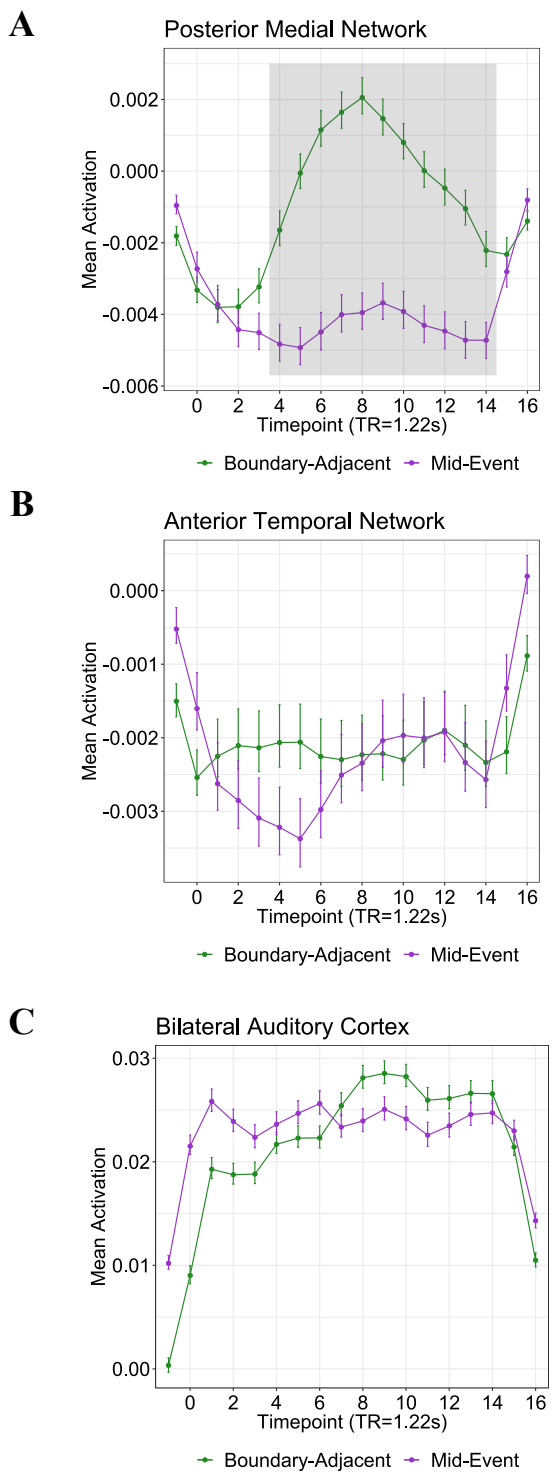


**Figure S6: Post-hoc analysis of boundary-evoked activation in cortical networks.**