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

Authors: Cohn-Sheehy, Brendan I.[1-3]*, Delarazan, Angelique I.[5], Reagh, Zachariah M.[5], Crivelli-Decker, Jordan E.[3-4], Zacks, Jeffrey M.[5], Ranganath, C.[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 and Lead Contact: Brendan I. Cohn-Sheehy

Keywords: episodic memory, event cognition, fMRI, narratives, naturalistic stimuli, pattern similarity, hippocampus, medial temporal lobe

Summary

Recent studies have suggested that activity in the hippocampus is increased during divisions between events in memory, or "event boundaries." Although hippocampal activity at event boundaries may play a role in segregating adjacent events, it might also serve a counterintuitive purpose: to bridge the divide between distant events that form a larger narrative. We used functional neuroimaging to investigate whether patterns of hippocampal activity at event boundaries support the integration of distant events in memory, when these events can form a larger narrative. During encoding of fictional stories, boundary-evoked hippocampal activity patterns were more similar between distant events that formed one coherent narrative, than between distant events from unrelated narratives. Twenty-four hours later, detailed recall of coherent narrative events was supported by the reinstatement of boundary-evoked hippocampal activity from encoding. These findings reveal a key function of the hippocampus: the dynamic integration of events into a narrative structure for memory.

Introduction

The ability to remember past events, or episodic memory (Tulving, 1972), may depend on the processes that shape our initial experience of events. Substantial evidence suggests that throughout continuous experience, people reliably perceive event boundaries, key moments where one event ends, and another event begins (Newtson, 1973; Zacks, 2020). It is thought that the perception of event boundaries may reflect that people engage in event segmentation, a process by which continuous experience is segregated into discrete events in memory (Zacks, 2020; Zacks et al., 2007).

If event boundary perception is related to episodic memory, it is likely that this relation is supported by the hippocampus (Swallow et al., 2011). There is widespread agreement that the hippocampus plays a unique and necessary role in episodic memory (Addis et al., 2007; Eichenbaum et al., 2007; Scoville and Milner, 1957; Yonelinas et al., 2002). Furthermore, recent studies have demonstrated that during continuous viewing of naturalistic films, event boundaries evoke transient increases in activity in the hippocampus (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Reagh et al., 2020). Although these findings suggest that the hippocampus is recruited during event boundaries, the functional significance of boundary-evoked activity in the hippocampus remains unclear.

It is possible that hippocampal activity at event boundaries reflects the segregation of individual events in memory. Consistent with this idea, recent work has shown that people find it hard to retrieve information across event boundaries (Clewett et al., 2020; Ezzyat and Davachi, 2011; Pettijohn and Radvansky, 2016; Swallow et al., 2011), which supports the idea that individual events become dissociated in memory. Building on this finding, several studies have

isolated patterns of activity in the hippocampus at specific moments in time, and have found that these patterns of activity are more similar when they take place within the same event, versus across an event boundary (DuBrow and Davachi, 2014; Ezzyat and Davachi, 2014; Hsieh et al., 2014). These findings suggest that the hippocampus supports binding features within an event and discriminating that event from others in memory. However, hippocampal activity may also be important for integrating event information into larger temporal structures (Cohn-Sheehy and Ranganath, 2017).

In fact, people can often remember multiple, connected events, even when these events take place at very different times. Namely, people can construct narratives (van Dijk and Kintsch, 1983) that bridge distant events in memory. A recent series of experiments demonstrated that when two distant events can form one larger narrative, people recall these events in greater detail, compared with two distant events that are taken from unrelated narratives (Cohn-Sheehy et al., 2020). This reliable finding contrasts with the notion that memory is simply segmented into individual, discrete events. Moreover, it suggests that some neural mechanism can construct an integrated memory across distant events.

Counterintuitively, although event boundaries may support event segmentation, they may also evoke cognitive processes that support event integration. For instance, if your ex-romantic partner were to walk into the room right now, this would probably trigger the perception of an event boundary; but additionally, you might be reminded of previous interactions with your ex (Ross and Bradshaw, 1994; Stawarczyk et al., 2020). Event Segmentation Theory (Zacks, 2020; Zacks et al., 2007) proposes that when an event begins (i.e. during an event boundary), people may retrieve knowledge about past events in order to construct a representation of a new event. This entails that brain activity which coincides with an event boundary may support knowledge

about multiple, thematically-related events (e.g. events classified as a romantic relationship). Rather than simply acting as a wall between events, an event boundary may act as a bridge across distant, related events. If so, brain activity that is present during event boundaries may be shared by distant events which are part of the same narrative.

It is possible that, at event boundaries, the hippocampus activates knowledge about thematically-related events. For instance, hippocampal activity increases when people are asked to construct hypothetical events in response to a thematic prompt (Addis et al., 2009, 2007). Additionally, some studies have suggested that the hippocampus may support information about larger narrative structures in memory (Collin et al., 2015; Milivojevic et al., 2016, 2015), although it is not clear whether this information shapes memory for narrative events. We do not know whether the hippocampus integrates temporally-separated events into a narrative, nor whether this depends on information that is present at event boundaries.

In the current study, we used functional magnetic resonance imaging (fMRI) to test whether the hippocampus supports the ability to integrate two distant events into one unified narrative. Participants were scanned while they listened to fictional stories, in which particular characters appeared in two temporally-separated events (previously implemented in behavioral experiments; Cohn-Sheehy et al., 2020). Critically, these two distant events could either form parts of one overarching narrative (Coherent Narrative), or they belonged to two different narratives (Unrelated Narratives; **Figure 1A-B**). Furthermore, these events were unrelated to adjacent events—in other words, they could not easily form a narrative with any adjacent event, and could only form a coherent narrative with a distant, paired event in the Coherent Narrative condition.

We analyzed spatial patterns of hippocampal activity during each event ("voxel

patterns") to test whether and how the hippocampus supports the integration of distant events. Our primary hypothesis was that hippocampal activity at event boundaries supports retrieval of event-related knowledge, which can be used to integrate related events into a narrative. In this case, boundary-evoked voxel patterns would be more similar between Coherent Narrative events than between Unrelated Narrative events. Alternatively, the hippocampus might support the integration of events into a narrative, but this might not involve event boundaries—for instance, the hippocampus might support some kind of continuous semantic processing, and semantic features may be similar between related events (Duff et al., 2020; Piai et al., 2016; Solomon et al., 2019). If so, post-boundary voxel patterns would be more similar between Coherent Narrative events than between Unrelated Narrative events. Finally, the hippocampus might not support the integration of distant events into a narrative, in which case voxel patterns would not be more similar between Coherent Narrative events than Unrelated Narrative events.

We scanned the same participants a day later, while they recalled events involving each recurring character. This enabled us to compute the similarity between hippocampal voxel

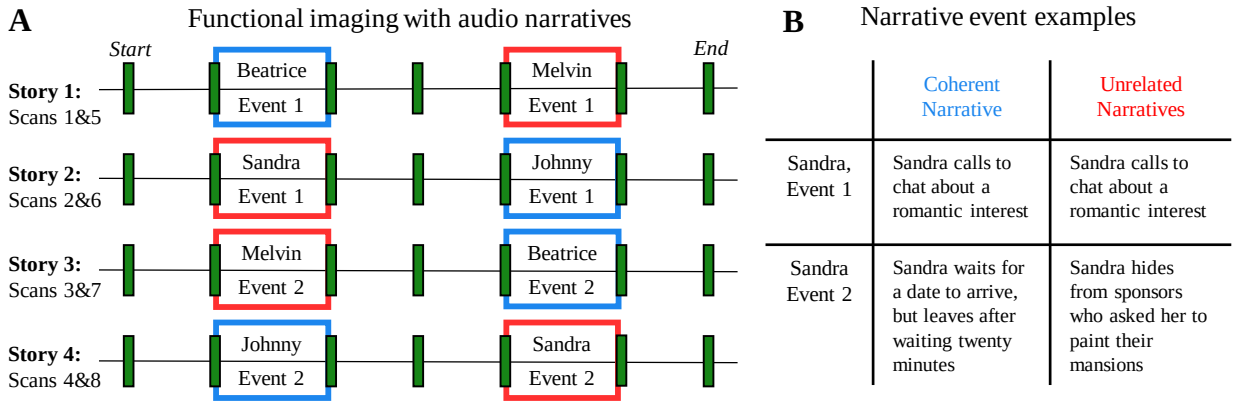

Figure 1: Experimental paradigm. (A) Stimulus design: Four fictional stories were presented serially during scanning, as audio clips (240s each), which were punctuated by discrete event boundaries (green bars). Events involving recurring characters (Beatrice, Melvin, Sandra, Johnny; 40s each) did not relate to the more continuous plot events surrounding them. For two of these recurring characters (Blue), two temporally-distant events could form one Coherent Narrative (e.g. Sandra Events 1 and 2). In contrast, for two other recurring characters (Red), 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 two possible pairs of events for Sandra, one which is Coherent, and one which is Unrelated.

patterns at encoding and at retrieval (encoding-retrieval similarity). As a corollary to our primary hypothesis, we hypothesized that, if the hippocampus integrates distant events into a larger narrative, then the hippocampus should also support the ability to retrieve these events from memory. If so, then boundary-evoked voxel patterns should exhibit higher encoding-retrieval similarity for Coherent Narratives than for Unrelated Narratives. Furthermore, in line with our previous behavioral work (Cohn-Sheehy et al., 2020), we hypothesized that this kind of reinstatement would predict the number of details recalled about Coherent Narrative events.

Results

Empirically-defined event boundary epochs

In order to test hypotheses about the significance of hippocampal activation at event boundaries, it was first necessary to confirm that hippocampal activation increased at event boundaries as participants listened to each story. Each story was written so that it would naturally be segmented into six 40-second events within each story (green bars in **Figure 1A;** see **Methods**). We modeled BOLD activation time-courses in the left and right hippocampus (**Figure 2A**) to estimate boundary-related activity changes in the hippocampus at Boundary-Adjacent timepoints (i.e., -2 to +16TRs around an event boundary; 1 TR = 1.22s; see Methods), and compared them with hippocampal activation around Mid-Event timepoints (i.e. 20 seconds into each 40-second event).

Activation during story encoding was 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 * Timepoint interaction ($F(5.06, 121.53)=11.2$, $\eta_G^2=.02$, $p<.0001$). No other effects approached statistical significance (all $p>.15$). As shown in **Figure 2B**, these results reflect the fact that, as in previous studies (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Reagh et al., 2020), event boundaries evoked increases in hippocampal activation.

To break down the Condition * Timepoint interaction, we used cluster-based permutation testing (Maris and Oostenveld, 2007), to identify clusters of Boundary-Adjacent timepoints where activity was significantly higher than corresponding Mid-Event timepoints (see **Methods**). We incorporated both the left and right hippocampus into this analysis because we did not observe any effects of hemisphere in the previous ANOVA. This analysis revealed a cluster of significant activity increases ($p<0.05$) that began 5 timepoints after the onset of an event, and lasted for 7 timepoints (TRs 5-11, ~6.1-14.64 seconds following the onset of an event; **Figure 2B**). As described below, these timepoints were subsequently used to identify the period of boundary-evoked activity in the hippocampus.

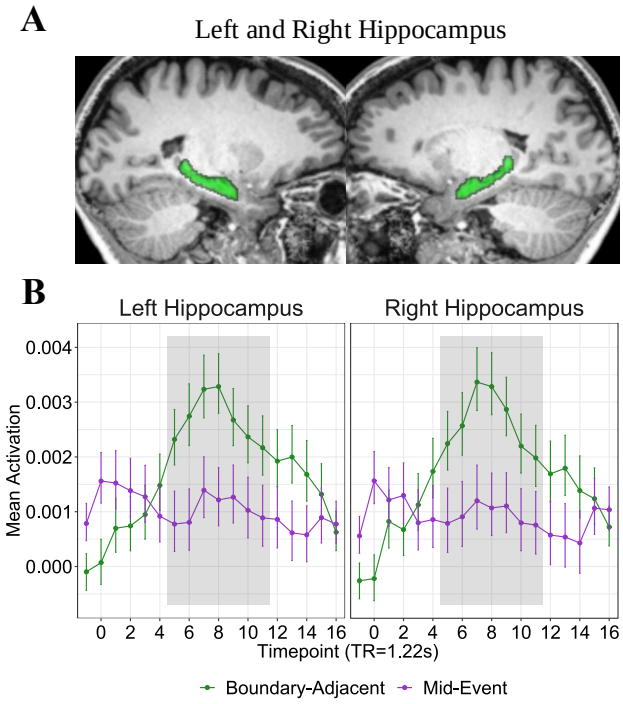

Figure 2: Hippocampal activation increases at event boundaries. (A) Regions of interest: the left and right hippocampus (green), are plotted on an individual subject's structural MRI. (B) Analysis of Boundary-Adjacent activation in the left and right hippocampus (green), vs activation at Mid-Event timepoints (purple). Means are represented by dots and lines, and error bars represent within-subjects estimates of standard error (see Methods for details). Grey shading represents a significant cluster of boundary activation (TRs +5 to +11; cluster-defining threshold is $p<.05$).

Boundary-evoked hippocampal activity bridges Coherent Narrative events

We hypothesized that hippocampal activation at event boundaries supports the integration of information from distant events. To test this hypothesis, we identified the unique pattern of hippocampal activity (voxel pattern) corresponding to each story event (see **Methods**), and investigated the degree to which voxel patterns around event boundaries were shared between events which formed a Coherent Narrative (CN) versus events which were taken from Unrelated Narratives (UN).

Voxel pattern similarity analyses were centered on two time-windows (**Figure 3A**). We identified a Boundary epoch, corresponding to TRs 5-11 following the onset of each CN and UN event, based on results from the previous analyses (**Figure 2B**). We also defined a Post-Boundary epoch, in order to examine patterns of activity during processing of 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). Voxel patterns were extracted from each time-window, and then compared across events.

We first tested whether boundary-evoked activity in the left and right hippocampus overlaps more across CN than UN events. Specifically, we analyzed voxel pattern similarity between the Event 1 Boundary and Event 2 Boundary (Boundary-Boundary similarity; **Figure 3A**). If hippocampal activation at event boundaries enables the integration of temporally-distant events, then we would expect Boundary-Boundary similarity to be higher for CN than UN events. In support of this hypothesis, a paired two-tailed t-test revealed that within the right hippocampus (**Figure 3B**), Boundary-Boundary similarity was higher for CN than UN events

($t(24) = 2.46$, p=0.02, Cohen's d=.49), reflecting that boundary-evoked activity patterns were disproportionately similar across events that formed a Coherent Narrative.

We next investigated similarity between boundary-evoked activity patterns and activity patterns maintained throughout the remainder of an event (i.e., the Post-Boundary epoch; **Figure 3A**). It has been theorized that, at event boundaries, people draw upon episodic memory to construct an initial representation of an event, and this representation is then maintained throughout the duration of an event (Franklin et al., 2020; Lu et al., 2020; Zacks, 2020). If so, we might expect that information from the event boundary would be maintained across the duration of an event, particularly for events that are integrated into a Coherent Narrative (i.e. higher Boundary-Post-Boundary similarity for CN events). In line with this hypothesis, a paired two-

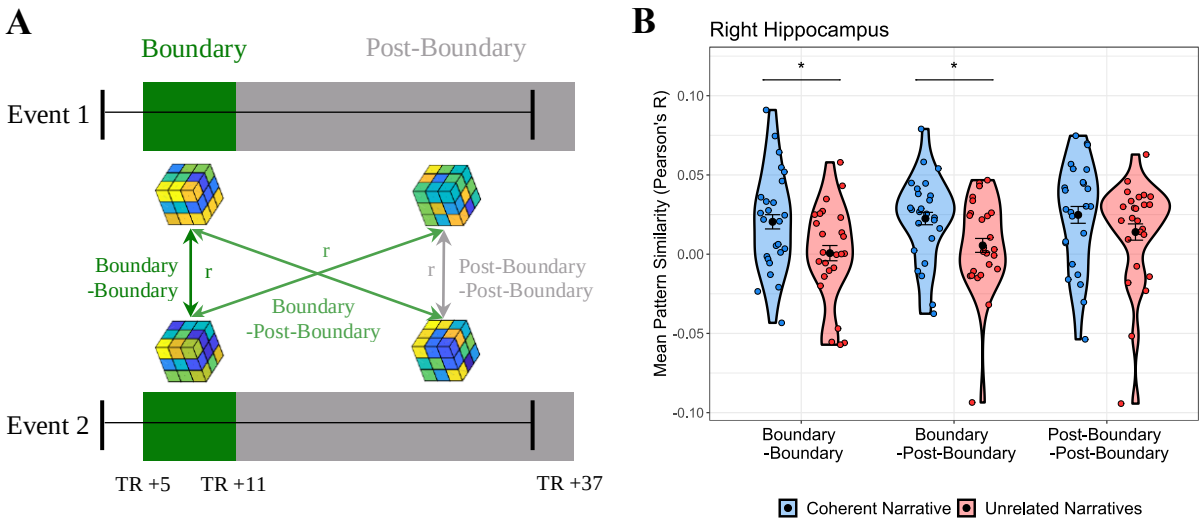

Figure 3: Boundary-evoked hippocampal activity bridges Coherent Narrative events.
(A) Across-event pattern similarity analysis: the unique spatial pattern of activity ("voxel pattern", represented by cubes) was modeled for Boundary (dark green rectangles, TRs +5 to +11) and Post-Boundary epochs (grey 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. Mean similarity (Pearson's r) was calculated between voxel patterns at Event 1 and Event 2, yielding 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. Violin plots display the range and density of pattern similarity for Coherent and Unrelated Narrative events, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, *=p<0.05.

tailed t-test revealed that within the right hippocampus (**Figure 3B**), Boundary-Post-Boundary similarity was higher for CN than UN events ($t(24) = 2.12$, $p=0.04$, Cohen's $d=.42$). This finding suggests that, at event boundaries, activity patterns in the right hippocampus were disproportionately similar to information maintained throughout the Post-Boundary epoch for events that formed a Coherent Narrative.

The above analyses suggested that hippocampal activity patterns carried information during the Boundary epoch that was preferentially shared across events that form a Coherent Narrative. We next considered the possibility that information which is processed by the hippocampus during the Post-Boundary epoch supports integration of events that form a Coherent Narrative (Post-Boundary-Post-Boundary, **Figure 3A**). If this was the case, Post-Boundary-Post-Boundary similarity would be higher for CN than UN events. However, a paired two-tailed t-test revealed that, within the right hippocampus (**Figure 3B**), there was no significant difference in Post-Boundary-Post-Boundary similarity between CN and UN events ($t(24) = 1.17$, $p=0.25$, Cohen's $d=.23$). As such, there was no evidence to suggest that integration of events was supported by Post-Boundary timepoints alone.

Taken together, this pattern of findings suggests that, in the right hippocampus, event boundaries played a key role in supporting the integration of distant events into a coherent narrative. No effects in the left hippocampus approached statistical significance (all $p>.12$, Cohen's $d<.32$; **Figure S1B**).

Boundary-evoked hippocampal patterns are not explained by sentence-level features

Having established that hippocampal activity patterns during the Boundary epoch are shared across Coherent Narrative events, we next investigated whether this effect could be

explained solely by simple features of the experimental stimuli (**Figure 1**). For instance, if certain semantic information, contained in words and sentences within the presented story events, tends to overlap more across CN than UN events, that might account for our finding of preferential hippocampal pattern similarity across CN events (Duff et al., 2020; Piai et al., 2016; Solomon et al., 2019). We therefore investigated whether boundary-evoked activity patterns in the right hippocampus were more similar across CN events even after accounting for the effects of word- and sentence-level similarity.

To account for this possibility, we quantified the degree of similarity between sentences in the four presented stories (see **Methods** for further details). We used the Universal Sentence Encoder (Cer et al., 2018) to embed each sentence into a multidimensional vector which represented the combination of words comprising each sentence, and calculated cosine similarity between each sentence vector (sentence similarity). We then regressed out sentence similarity values from voxel pattern similarity. This approach yielded residual voxel pattern similarity values for each time-window during encoding, which could not be explained by sentence similarity alone.

Using the residuals from this regression analysis, we tested whether boundary-evoked pattern similarity in the right hippocampus was higher for CN than UN events, even after accounting for word- and sentence-level semantic similarity. In support of this prediction, a paired two-tailed t-test revealed that in the right hippocampus, Boundary-Post-Boundary residuals significantly differed between Coherent Narrative and Unrelated Narrative events ($t(24) = 2.14$, $p=0.04$, Cohen's $d = 0.43$; **Figure S2A**). This suggests that semantic similarity could not completely account for the degree to which boundary-evoked hippocampal activity patterns were shared across events. For the other two epochs tested, residuals did not significantly differ

between CN and UN events (*t*s<1.85; **Figure S2A**).

Boundary-evoked hippocampal activity is preferentially reinstated during recall of Coherent Narrative events

The preceding findings supported the idea that, during encoding, patterns of boundary-evoked activity in the right hippocampus are shared across distant events that form a larger narrative. To the extent that boundary-evoked activity in the hippocampus helps to bridge CN events in memory, we would expect that hippocampal voxel patterns during the Boundary epoch should be preferentially reinstated during delayed recall of CN events (see **Figure 4A**).

However, 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; **Figure 1**). Because events were presented in order, the beginning of Event 2 (i.e. the Boundary epoch) provided the earliest opportunity for participants to determine any relationship between the two events. In the CN condition, the Event 2 Boundary may have served as the key moment when participants could integrate information from both events to form a coherent narrative. If so, we would expect that information from the CN Event 2 Boundary would play a critical role in subsequent recall. In contrast, in the UN condition, it was unlikely that information from the Event 2 Boundary could integrate events during encoding, nor that information from the UN Event 2 Boundary would be critical for subsequent recall.

Based on this rationale, we hypothesized that boundary-evoked activity 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 boundary-

evoked hippocampal voxel patterns during Event 2 encoding were reinstated during recall (encoding-retrieval similarity; see **Figure 4A, Methods**). We also analyzed reinstatement of boundary-evoked activity from Event 1—however, because we suspected that the Event 2 Boundary provided the critical moment for integration, we did not expect any difference in reinstatement of boundary-evoked activity from Event 1 between CN and UN events. Results

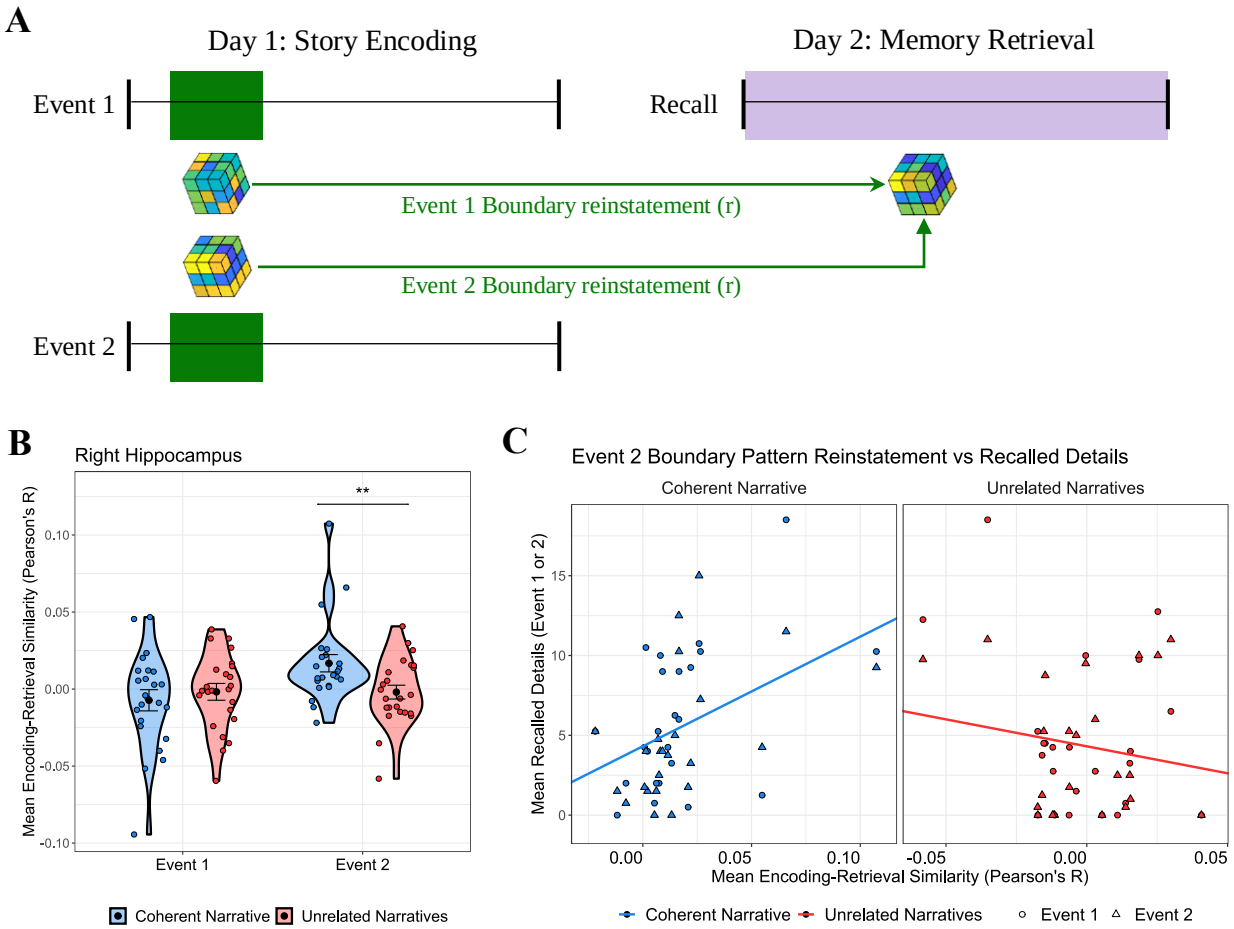

Figure 4: Recall is predicted by reinstatement of boundary-evoked hippocampal activity.
(A) Encoding-retrieval similarity analysis: Within the right hippocampus, encoding-retrieval similarity was calculated by computing correlations (Pearson's r) between boundary-evoked voxel patterns from each event involving a recurring character at encoding (Event 1 or Event 2) and delayed recall of that same character. (B) Violin plots display the range and density of encoding-retrieval similarity, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). (B) Reinstatement of boundary-evoked Event 2 activity patterns during recall (encoding-retrieval similarity, x-axis) is compared with the number of details recalled (y-axis), separately for Coherent Narrative events (left panels) and Unrelated Narrative events (right panels). Lines represent fits of multiple regressions (see text) predicting an interaction between the degree of reinstatement and whether or not events can form a coherent narrative. Key: Blue=Coherent Narrative, Red=Unrelated Narratives, triangles=Event 1 recall, circles=Event 2 recall, **=p<.01.

showed that, for Event 1, encoding-retrieval similarity did not significantly differ between the CN and UN conditions ($t(23) = 0.71$, p=0.49, Cohen's d=.14; **Figure 4B**). For Event 2, however, encoding-retrieval similarity was higher for CN Event 2 than UN Event 2 ($t(23) = 3.08$, p=0.005, Cohen's d=.63; **Figure 4B**).

These findings suggest that in the right hippocampus, boundary-evoked activity at Event 2 was preferentially reinstated during recall of events which formed a Coherent Narrative. In other words, hippocampal activity during the Event 2 Boundary may have provided a bridge between distant events in memory. As we describe below, we investigated this possibility further by comparing encoding-retrieval similarity with a behavioral measure of memory.

Recall of Coherent Narrative events scales with reinstatement of boundary-evoked hippocampal activity

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 (Cohn-Sheehy et al., 2020). In the present study, the number of details recalled from CN events (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 of CN events. If so, we would expect that recall of CN events should be predicted by reinstatement of boundary-evoked activity patterns in the hippocampus during recall.

As described above (**Figure 4B**), hippocampal activity during the Event 2 Boundary was more highly reinstated during recall of CN than UN events. One possibility was that reinstatement of boundary-evoked activity from Event 2 was related to recall success for individual events (i.e. recall of Event 2, not Event 1), regardless of whether events were taken from Coherent or Unrelated Narratives—this would be in line with the fact that the hippocampus generally supports recollection. If so, reinstatement of boundary-evoked Event 2 patterns during retrieval might predict recall of either CN or UN events. A second possibility was that the ability to form a coherent narrative facilitated encoding of CN Event 2. If so, we might expect that reinstatement of boundary-evoked activity from CN Event 2, but not UN Event 2, would predict the number of recalled details from Event 2. A third, more intriguing possibility was that higher reinstatement of CN Event 2 reflected that CN Event 2 was integrated with CN Event 1 – if so, reinstatement of CN Event 2 would predict recall of both CN Events 1 and 2, reflecting their integration in memory.

To adjudicate between these possibilities, we performed a multiple regression model (see **Methods**), predicting recall performance based on the following regressors: (1) the degree of reinstatement of boundary-evoked hippocampal activity from Event 2 (Encoding-Retrieval Similarity); (2) whether events were from Coherent Narratives versus Unrelated Narratives (Narrative Coherence); and (3) whether details were recalled from Event 1 versus Event 2 (Event Number). In addition, the model also included Recall Duration (i.e. number of TRs for recall of each cue) as a nuisance regressor, in order to control for the possibility that the amount of time spent on recall might correspond with more successful recall. This model significantly accounted for 15% of the variance in recalled details ($F(8,87)=3.16$, $p=.003$, $R^2 = 0.15$). Significant predictors included a main effect of Encoding-Retrieval Similarity (B=70.2, t=2.22, p=0.03)

which was qualified by an interaction between Encoding-Retrieval Similarity and Narrative Coherence (B=-129.1, t=2.55, p=0.01). No other predictors were statistically significant (all t's <1). As shown in **Figure 4C**, the degree to which right hippocampal activity patterns at the Event 2 Boundary were reinstated during recall predicted individual differences in recall of CN events, and not UN events.

In other words, reinstatement of boundary-evoked hippocampal activity from Event 2 specifically predicted recall of events which formed a Coherent Narrative. Furthermore, this reinstatement predicted recall of both CN Events 1 and 2. These findings suggest that, in the right hippocampus, information that is present during the critical moment of integration at encoding can bridge distant events in memory.

Boundary and post-boundary activity in medial temporal cortex bridges Coherent Narrative events

The parahippocampal cortex (**Figure 5A**) and perirhinal cortex (**Figure 5C**) are known to provide important input to the hippocampal formation (Eichenbaum et al., 2007; Reagh and Ranganath, 2018). It was possible that these areas would also play a role in integrating events into a Coherent Narrative. We therefore investigated whether boundary-evoked activity patterns in these regions exhibited higher similarity between CN than UN events.

We first tested whether boundary-evoked activity overlaps more across CN than UN events (Boundary-Boundary similarity). Within the left parahippocampal cortex (**Figure 5B**), a paired two-tailed t-test revealed that Boundary-Boundary similarity was significantly higher between CN than UN events ($t(24) = 2.11$, p=0.045, Cohen's d=0.42). In other words, boundary-evoked activity patterns in the left parahippocampal cortex were more shared between events

that formed a Coherent Narrative than between events which did not. This suggests that the left parahippocampal cortex may support some kind of boundary-dependent integration of distant events into a larger narrative. It was possible that this boundary-evoked activity might also overlap with other event timepoints, therefore we also tested whether Boundary-Post-Boundary similarity in the left parahippocampal cortex was higher for CN than UN events. However, a paired two-tailed t-test revealed that within the left parahippocampal cortex (**Figure 5B**), Boundary-Post-Boundary similarity did not significantly differ between CN and UN events ($t(24) = 1.71$, p=0.10, Cohen's d=.34). No other medial temporal cortical regions exhibited significant differences in either Boundary-Boundary or Boundary-Post-Boundary similarity (all *t*s<1.5; **Figure 5D, Figure S1C-F**).

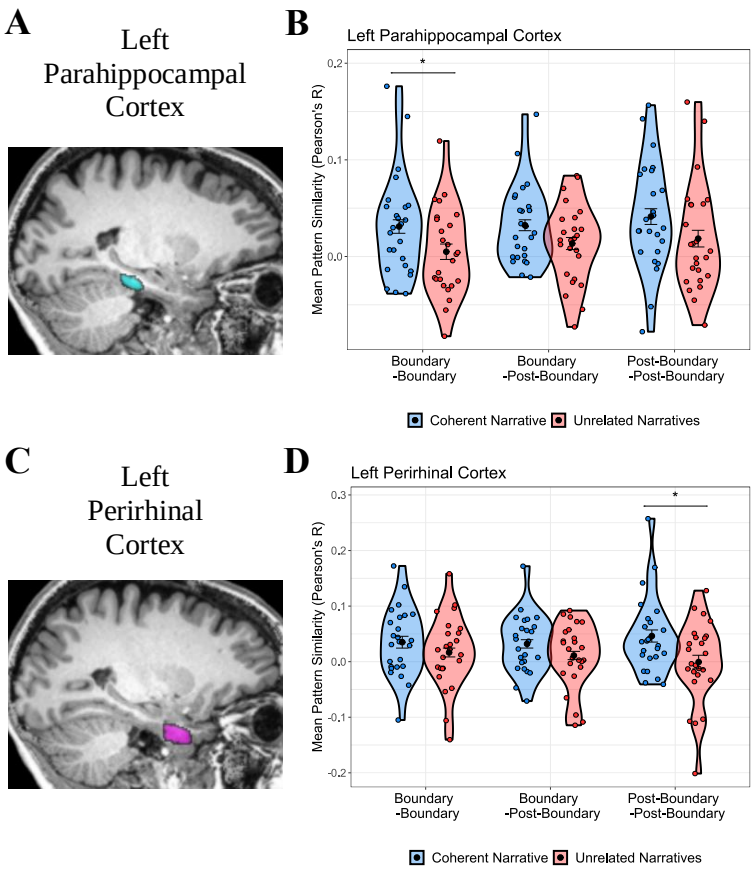

Figure 5: Medial temporal cortex activity bridges Coherent Narrative events. (A,C) Regions of interest: the left parahippocampal cortex (A, in cyan) and left perirhinal cortex (C, in magenta) are plotted on an individual subject's structural MRI. (B, D) Cortical pattern similarity: Pearson's correlations between Event 1 and 2 timepoints were computed separately for Coherent Narrative and Unrelated Narrative events. Violin plots in B (left parahippocampal cortex) and D (left perirhinal cortex) display the range and density of pattern similarity for Coherent and Unrelated Narrative events, averaged within each event window, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, *=p<0.05.

We next investigated whether perirhinal or parahippocampal cortex might support integration outside of

event boundaries. We therefore analyzed Post-Boundary-Post-Boundary similarity in these cortical areas. A two-tailed t-test revealed that within the left perirhinal cortex (**Figure 5D**), there was a significant difference in Post-Boundary-Post-Boundary pattern similarity between CN and UN events ($t(24) = 2.31$, $p=0.03$, Cohen's d=46). Interestingly, this finding suggests that left perirhinal activity patterns might support the integration of distant events into a Coherent Narrative, however, this information might be supported by Post-Boundary timepoints alone. No other regions exhibited a significant difference in Post-Boundary-Post-Boundary similarity (all *t*s<1.65; **Figure 5B, Figure S1C-F**).

Finally, we investigated whether pattern similarity findings in the left parahippocampal and left perirhinal cortex could be solely explained by word- or sentence-level features of the stimuli. Although we initially found that these regions supported preferential pattern similarity across CN events (**Figure 5**), these effects were no longer significant after controlling for semantic similarity (all *t*s<1.8; **Figure S2B-C**). This suggests that in the medial temporal cortex, disproportionate pattern similarity across CN events might be explained by the degree to which word- or sentence-level features overlap across story events.

Discussion

Event boundaries have previously been shown to play a role in segregating contiguous events (Zacks, 2020). In the current study, we investigated whether boundary-evoked activity in the hippocampus supports the integration of distant events into a larger narrative in memory. Results showed that, consistent with prior studies using film stimuli, hippocampal activity increased at event boundaries. During memory encoding, activity patterns in the right

hippocampus during event boundaries were more similar across two temporally-separated events that formed a Coherent Narrative, than across events from Unrelated Narratives. Twenty-four hours later, during memory retrieval, boundary-evoked activity from encoding was more highly reinstated for Coherent Narratives than for Unrelated Narratives. Specifically, boundary-evoked activity from the second Coherent Narrative event–the initial point at which a narrative could be formed across events – was more highly reinstated during retrieval than boundary-evoked activity from the second Unrelated Narrative event. This reinstatement specifically predicted the ability to recall details about either of two events that formed a Coherent Narrative. Taken together, our findings suggest that, during event boundaries, the hippocampus supports the construction of a narrative that bridges distant events in memory, and which can be used to facilitate memory retrieval.

The present study builds on previous work suggesting that the hippocampus might support representations of complex, temporally extended episodes (Collin et al., 2015; Milivojevic et al., 2016, 2015). For instance, one project examined hippocampal activity while participants viewed seemingly unrelated scenes, and then after viewing "linker" scenes that enabled participants to build a narrative out of the initially-unrelated events (Collin et al., 2015; Milivojevic et al., 2015). In a separate study, participants watched a commercial film in which alternating scenes corresponded to one of two narratives, which diverged in meaning as the film progressed (Milivojevic et al., 2016). In all three studies, the hippocampus showed higher voxel pattern similarity across events which belonged to the same narrative versus events which could not be linked into a narrative. As in the present study, Milivojevic et al. (2016) found that narrative effects were limited to the right hippocampus.

The studies described above (Collin et al., 2015; Milivojevic et al., 2015; Milivojevic et

al., 2016) support the idea that the hippocampus supports information about narratives during encoding, but notably, they did not investigate the role of event boundaries, and they did not examine the role of the hippocampus in retrieval of narrative events. The present study expands upon prior work by demonstrating: (1) that hippocampal activity at event boundaries, reflecting processes engaged at the onset of a new event, supports the integration of temporally-separated events; (2) that hippocampal representations of event content during the critical moment of integration are reactivated during recall of the integrated narrative, and (3) that hippocampal reinstatement of boundary-evoked activity is predictive of detailed recall of these narratives.

Research on event boundaries and memory has typically focused on how event segmentation breaks up continuous experiences. It follows that hippocampal activity increases at event boundaries (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Reagh et al., 2020) might reflect the process of segmenting experiences into distinct event representations. Although the present results do not rule out that possibility, our results suggest that boundary-evoked activity in the hippocampus may not only reflect a "wall" between events, but also a "bridge."

Consistent with this idea, Event Segmentation Theory (Zacks, 2020; Zacks et al., 2007) proposes that, when people perceive that a new event has begun, they draw upon prior knowledge about similar events in order to better predict the new event (see also, Lu et al., 2020; Ross and Bradshaw, 1994; Stawarczyk et al., 2020). This proposal has been formalized in the Structured Event Memory computational model (SEM; Franklin et al., 2020), which is built on the assumption that the brain generates predictions of what will happen in an ongoing event. In this model, moments of high prediction error (i.e. event boundaries) trigger an updating mechanism, in which prior memories are retrieved in order to better predict what will happen next in an event. Although SEM has not yet been applied to model the current findings, it is

plausible that the model's boundary-driven updating mechanism during memory encoding can explain how boundary-evoked activity patterns in the hippocampus support the comprehension and subsequent recall of events that can be integrated into a larger narrative.

In contrast with the present study, other studies have suggested that boundary-evoked hippocampal activity supports storage of events that precede an event boundary (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Sinclair et al., 2020). These findings are not necessarily incompatible with findings of the present study. In fact, event boundaries mark both the onset of an event and the offset of a preceding event. In contrast with previous studies, our single-trial modeling approach (see **Methods**) enabled us to separate overlapping events, and to specifically investigate the patterns of boundary-evoked hippocampal activity at the start of an event. Future investigations should determine whether the correspondence of hippocampal activity at event boundaries with memory for preceding and subsequent events reflects one common, underlying process (e.g. updating representations in response to high prediction error; Antony et al., 2020; Bein et al., 2020; Franklin et al., 2020; Lu et al., 2020; Sinclair et al., 2020) or two overlapping processes (e.g. memory storage vs retrieval).

Looking beyond the hippocampus, our results suggest that medial temporal lobe cortical areas might also contribute to the integration of distant events in memory. Prior work has suggested that the parahippocampal cortex may support representations of spatial and semantic context which are shared with the hippocampus, and which contrast with representations (e.g. specific items or concepts) that are supported by the perirhinal cortex (Davachi, 2006; Diana et al., 2007; Eacott and Gaffan, 2005; Eichenbaum et al., 2007; Mishkin et al., 1997; Montaldi and Mayes, 2010; Ranganath, 2010; Ranganath and Ritchey, 2012). Mishkin et al. (1997), for instance, proposed that neocortical areas in the medial temporal lobe may represent semantic

information which is bound by the hippocampus to support representations of specific episodes. Consistent with this idea, activity patterns in the left parahippocampal and perirhinal cortex overlapped to a greater degree across CN than UN events, but this effect was not significant after accounting for overlap in word- and sentence-level semantic information. There is reason to believe that the medial temporal lobe interacts with broader networks of regions – particularly within the default mode network (Cohn-Sheehy and Ranganath, 2017; Ranganath and Ritchey, 2012) – to process information in narratives (Aly et al., 2018; Chen et al., 2017, 2016; Speer et al., 2009; Zacks, 2010). Although a more detailed investigation of cortical networks is outside the scope of the current study, future work will investigate whether these networks support the ability to integrate distant events in memory.

In closing, the current findings demonstrate that the hippocampus supports the ability to bridge events in memory to form a narrative, even when these events take place at different times. Surprisingly, the moments that divide events in memory may also hold the key to putting them back together.

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 staff for their technical contributions. We thank Alex Barnett, Halle Dimsdale-Zucker, Anna Leshinskaya, Kamin Kim, Andy Yonelinas, and the Dynamic Memory Lab (http://dml.ucdavis.edu) for consultation on experimental design and analysis, and James Antony for feedback on the manuscript. This research was supported by

Multi-University Research Initiative grant N00014-17-1-2961 from the United States Office of Naval Research/Department of Defense (C.R., J.M.Z), Ruth L. Kirchstein National Research Service Award F30AG062053 from the National Institute on Aging (B.I.C.S.), and a Floyd and Mary Schwall Medical Research Fellowship from the University of California, Davis (B.I.C.S.). 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.

Author Contributions

Conceptualization, B.I.C.S. and C.R.; Methodology, B.I.C.S., C.R., and Z.M.R.; Investigation, B.I.C.S. and A.I.D.; Analysis, B.I.C.S. and A.I.D.; Software, B.I.C.S., J.E.C.D., A.I.D., Z.M.R.; Writing—Original Draft, B.I.C.S. and C.R.; Writing – Review and Editing, all authors; Funding Acquisition, C.R., J.M.Z., and B.I.C.S.; Supervision, C.R., Z.M.R., and J.M.Z.

Declaration of Interests

The authors declare no competing interests.

Methods

Participants

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 study. 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.

Experimental Stimuli

The stimulus design is depicted in **Figure 1A-B,** and has been described elsewhere (Cohn-Sheehy et al., 2020). We constructed four fictional stories in which we manipulated whether temporally-distant events could form a coherent narrative (**Supplementary Data 1**). 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 (i.e. characters entering or exiting, temporal shifts like “20 minutes later”; Ezzyat and Davachi, 2011; Zacks et al., 2009). Two additional boundaries marked the start and end of each story, respectively. These seven *a priori* event boundaries for each story (green bars in **Figure 1A**) 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. Two stories involved “Beatrice” and “Melvin” as side-characters, and two involved “Sandra” and “Johnny.” Each side-character appeared in two distinct events that occurred 6-12 minutes apart (i.e. they were temporally distant), in disparate contexts across two unrelated stories. For instance, Sandra’s first event occurs in Story 2, when she calls Charles, who is sitting at a park at noon on a Tuesday. Sandra’s second event 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). 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; **Figure 1B**). 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; **Figure 1B**). However, each story always contained one CN and one UN event (e.g. in **Figure 1A**, Sandra is CN and Johnny is UN). This approach resulted in 32 possible arrangements of CN and UN events, which were pseudo-randomly assigned. 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

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).

Familiarization tasks (Day 1). As described previously (Cohn-Sheehy et al., 2020), three brief tasks aimed to familiarize participants with character names and relationships, in order 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 (Bainbridge et al., 2013).

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; **Figure 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 (see Cohn-Sheehy et al., 2020). 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, 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).

Recall scoring

Free recall performance was scored using a procedure developed in our previous behavioral experiments with a similar paradigm (Cohn-Sheehy et al., 2020), and which adapted a well-characterized scoring method from the Autobiographical Memory Interview (Diamond et al., 2020; Levine et al., 2002). 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 (**Supplementary Data 2-3**). 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 (Power, 2017). 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/; Ritchey et al., 2015) 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). These ROIs were then co-registered and resliced to functional images (either univariate activation or multivoxel patterns; see below).

Univariate fMRI analysis

For univariate fMRI analyses, 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 (Ben-Yakov and Henson, 2018), 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.

Event boundaries were defined *a priori*—they were intrinsic to the stimulus design (starts and ends of 40s events; **Figure 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, **Supplementary Data 1**). 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. This was followed by permutation tests in Permuco (https://cran.r-project.org/web/packages/permuco/), which tested 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 (Maris and Oostenveld, 2007). This approach enabled us to account for interdependency between adjacent timepoints, in line with well-documented approaches for time-series analysis of neuroimaging data (Cohen, 2014).

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. 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 fMRI Analysis, 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 event or 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; Kriegeskorte et al., 2008). 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. This whole matrix was later utilized for text-based models (see Textual similarity analysis, below).

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 (i.e. Beatrice Event 1 vs Beatrice Event 2, Melvin Event 2 vs Melvin Event 1). For encoding-retrieval similarity, we selected all correlations between timepoints from a CN or UN encoding event (e.g. Beatrice Event 1) and timepoints corresponding to the recall cue for the same CN or UN character (e.g. 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, we then used empirically-defined epochs to derive mean similarity values (mean Pearson's r) for each condition. Finally, for both encoding and encoding-retrieval similarity analysis, we tested our *a priori* hypotheses for each type of similarity using paired two-tailed t-tests.

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 (**Figure 3A**): 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.

Encoding-retrieval similarity: To follow up on across-event similarity findings within the right hippocampus, for each encoding event (Event 1 or 2), we selectively averaged timepoint-by-timepoint correlations between the Boundary epoch at encoding (TRs +5 to +11), and any timepoints during recall of the same CN or UN character (**Figure 4A**). However, participants spent varying amounts of time recalling details for each cue (28-299 TRs, ~ 0.5-6 minutes per recall cue). In order to analyze an equivalent amount of data for each CN and UN character, our analysis of encoding-retrieval similarity was drawn from the maximum number of timepoints that were shared across all recall cues (28 TRs; results presented in **Figure 4** and corresponding text). We then used a multiple linear regression to analyze the correspondence between encoding-retrieval and recalled details, with the following formula: Details recalled ~ Encoding-Retrieval Similarity * Coherence * Event Number + Recall Duration. We also followed up on these analyses by computing encoding-retrieval similarity using all available recall timepoints (widely varying across recall cues); this approach revealed a similar pattern of results (see **Figure S3**).

Textual similarity analysis

We modeled textual similarity between sentences from different story events using the freely available, "deep averaging network" version of the Universal Sentence Encoder (USE), a text embedding model designed to convert text into numerical vectors (Cer et al., 2018; see also, https://ai.googleblog.com/2018/05/advances-in-semantic-textual-similarity.html). Briefly, the USE uses pre-weighted layers, pre-trained on an expansive textual database, to transform inputted sentences into 512-dimensional embedding vectors that account for the combination of words in each sentence. Then, cosine similarity is calculated between each sentence vector, yielding pairwise measures of textual similarity for all inputted sentences. As such, USE

similarity serves as a proxy of word- and sentence-level semantic relatedness for text.

For each subject, we created a textual-similarity matrix for cosine similarity between all sentences from all stories. In order to compare textual similarity with fMRI voxel pattern similarity, we used custom MATLAB code to align the whole USE similarity matrix with the whole RSM for each subject. Because all sentences were of equivalent lengths (5s each), we were able to align the time-course of sentences with the timecourse of fMRI scans (1.22s for each timepoint). However, in order to align the 8-sentence window with the 40s window for each event, it was necessary to only include fMRI timepoints that corresponded with that 40s window. Therefore, we excluded any timepoints that did not correspond with Boundary or Post-Boundary event epochs (i.e. including only TRs +5 to +37, 40s total). The whole USE similarity matrix was upsampled and interpolated from a 5s/sentence timescale to match the 1.22s/TR timescale of the whole RSM. We then vectorized the upsampled USE matrix and the whole RSMs (for right hippocampus, left parahippocampal cortex, and left perirhinal cortex). These vectors were imported into RStudio, z-scored, and then entered into linear regression models that predicted voxel pattern similarity from sentence similarity for every possible pair of timepoints during story encoding (within each subject and ROI): pattern similarity ~ USE similarity.

From each model (for each subject and ROI), residuals were extracted for every timepoint-by-timepoint correlation. Residuals were imported into MATLAB, and selectively averaged within each condition of interest (CN vs UN) and time-window (Boundary-Boundary, Boundary-Post-Boundary, Post-Boundary-Post-Boundary). Finally, these averaged residuals were imported into RStudio for visualization and statistical analyses.

Statistical analysis

Statistical tests were performed in R using standard functions for linear models and t-

tests, as well as Afex (https://github.com/singmann/afex) for ANOVAs and Permuco (https://cran.r-project.org/web/packages/permuco/) for permutation tests that followed up on ANOVAs. For ANOVAs, Greenhouse-Geisser corrections were implemented for any non-sphericity. For visualization, 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 (Morey, 2008).

References

Addis, D.R., Pan, L., Vu, M.-A., Laiser, N., Schacter, D.L., 2009. Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. Neuropsychologia 47, 2222–2238. https://doi.org/10.1016/j.neuropsychologia.2008.10.026

Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45, 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Aly, M., Chen, J., Turk-Browne, N.B., Hasson, U., 2018. Learning Naturalistic Temporal Structure in the Posterior Medial Network. J. Cogn. Neurosci. 30, 1345–1365. https://doi.org/10.1162/jocn_a_01308

Antony, J.W., Hartshorne, T.H., Pomeroy, K., Gureckis, T.M., Hasson, U., McDougle, S.D., Norman, K.A., 2020. Behavioral, Physiological, and Neural Signatures of Surprise during Naturalistic Sports Viewing. Neuron S0896627320308539. https://doi.org/10.1016/j.neuron.2020.10.029

Bainbridge, W.A., Isola, P., Oliva, A., 2013. The intrinsic memorability of face photographs. J. Exp. Psychol. Gen. 142, 1323–1334. https://doi.org/10.1037/a0033872

Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A., 2017. Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron 95, 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Bein, O., Duncan, K., Davachi, L., 2020. Mnemonic prediction errors bias hippocampal states. Nat. Commun. 11, 3451. https://doi.org/10.1038/s41467-020-17287-1

Ben-Yakov, A., Henson, R.N., 2018. The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. J. Neurosci. 38, 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., Kurzweil, R., 2018. Universal Sentence Encoder. ArXiv180311175 Cs.

Chen, J., Honey, C.J., Simony, E., Arcaro, M.J., Norman, K.A., Hasson, U., 2016. 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. https://doi.org/10.1093/cercor/bhv155

Chen, J., Leong, Y.C., Honey, C.J., Yong, C.H., Norman, K.A., Hasson, U., 2017. Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci. 20, 115–125. https://doi.org/10.1038/nn.4450

Clewett, D., Gasser, C., Davachi, L., 2020. Pupil-linked arousal signals track the temporal organization of events in memory. Nat. Commun. 11, 4007. https://doi.org/10.1038/s41467-020-17851-9

Cohen, M.X., 2014. Analyzing neural time series data: theory and practice.

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

Cohn-Sheehy, B.I., Ranganath, C., 2017. Time regained: how the human brain constructs memory for time. Curr. Opin. Behav. Sci. 17, 169–177. https://doi.org/10.1016/j.cobeha.2017.08.005

Collin, S.H.P., Milivojevic, B., Doeller, C.F., 2015. Memory hierarchies map onto the hippocampal long axis in humans. Nat. Neurosci. 18, 1562–1564. https://doi.org/10.1038/nn.4138

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16, 693–700. https://doi.org/10.1016/j.conb.2006.10.012

Diamond, N.B., Armson, M.J., Levine, B., 2020. The Truth Is Out There: Accuracy in Recall of Verifiable Real-World Events. Psychol. Sci. 095679762095481. https://doi.org/10.1177/0956797620954812

Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci. 11, 379–386. https://doi.org/10.1016/j.tics.2007.08.001

DuBrow, S., Davachi, L., 2014. Temporal memory is shaped by encoding stability and intervening item reactivation. J. Neurosci. Off. J. Soc. Neurosci. 34, 13998–14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014

Duff, M.C., Covington, N.V., Hilverman, C., Cohen, N.J., 2020. Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship. Front. Hum. Neurosci. 13, 471. https://doi.org/10.3389/fnhum.2019.00471

Eacott, M.J., Gaffan, E.A., 2005. The Roles of Perirhinal Cortex, Postrhinal Cortex, and the Fornix in Memory for Objects, Contexts, and Events in the Rat. Q. J. Exp. Psychol. Sect. B 58, 202–217. https://doi.org/10.1080/02724990444000203

Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The Medial Temporal Lobe and Recognition Memory. Annu. Rev. Neurosci. 30, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Ezzyat, Y., Davachi, L., 2014. Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. Neuron 81, 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Ezzyat, Y., Davachi, L., 2011. What constitutes an episode in episodic memory? Psychol. Sci. 22, 243–252. https://doi.org/10.1177/0956797610393742

Franklin, N.T., Norman, K.A., Ranganath, C., Zacks, J.M., Gershman, S.J., 2020. Structured Event Memory: A neuro-symbolic model of event cognition. Psychol. Rev. 127, 327–361. https://doi.org/10.1037/rev0000177

Hsieh, L.-T., Gruber, M.J., Jenkins, L.J., Ranganath, C., 2014. Hippocampal activity patterns carry information about objects in temporal context. Neuron 81, 1165–1178. https://doi.org/10.1016/j.neuron.2014.01.015

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. https://doi.org/10.3389/neuro.06.004.2008

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

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

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Milivojevic, B., Varadinov, M., Vicente Grabovetsky, A., Collin, S.H.P., Doeller, C.F., 2016. Coding of Event Nodes and Narrative Context in the Hippocampus. J. Neurosci. Off. J. Soc. Neurosci. 36, 12412–12424. https://doi.org/10.1523/JNEUROSCI.2889-15.2016

Milivojevic, B., Vicente-Grabovetsky, A., Doeller, C.F., 2015. Insight Reconfigures Hippocampal-Prefrontal Memories. Curr. Biol. 25, 821–830. https://doi.org/10.1016/j.cub.2015.01.033

Mishkin, M., Suzuki, W.A., Gadian, D.G., Vargha–Khadem, F., 1997. Hierarchical organization of cognitive memory. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 352, 1461–1467. https://doi.org/10.1098/rstb.1997.0132

Montaldi, D., Mayes, A.R., 2010. The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. Hippocampus 20, 1291–1314. https://doi.org/10.1002/hipo.20853

Morey, R.D., 2008. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutor. Quant. Methods Psychol. 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59, 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

Newtson, D., 1973. Attribution and the unit of perception of ongoing behavior. J. Pers. Soc. Psychol. 28, 28–38. https://doi.org/10.1037/h0035584

Pettijohn, K.A., Radvansky, G.A., 2016. Walking through doorways causes forgetting: Event structure or updating disruption? Q. J. Exp. Psychol. 69, 2119–2129. https://doi.org/10.1080/17470218.2015.1101478

Piai, V., Anderson, K.L., Lin, J.J., Dewar, C., Parvizi, J., Dronkers, N.F., Knight, R.T., 2016. Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. Proc. Natl. Acad. Sci. 113, 11366–11371. https://doi.org/10.1073/pnas.1603312113

Power, J.D., 2017. A simple but useful way to assess fMRI scan qualities. NeuroImage 154, 150–158. https://doi.org/10.1016/j.neuroimage.2016.08.009

Ranganath, C., 2010. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus 20, 1263–1290. https://doi.org/10.1002/hipo.20852

Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. Nat. Rev. Neurosci. 13, 713–726. https://doi.org/10.1038/nrn3338

Reagh, Z.M., Delarazan, A.I., Garber, A., Ranganath, C., 2020. Aging alters neural activity at event boundaries in the hippocampus and Posterior Medial network. Nat. Commun. 11, 3980. https://doi.org/10.1038/s41467-020-17713-4

Reagh, Z.M., Ranganath, C., 2018. What does the functional organization of cortico-hippocampal networks tell us about the functional organization of memory? Neurosci. Lett. 680, 69–76. https://doi.org/10.1016/j.neulet.2018.04.050

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

Ross, B.H., Bradshaw, G.L., 1994. Encoding effects of remindings. Mem. Cognit. 22, 591–605. https://doi.org/10.3758/BF03198398

Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry 20, 11–21.

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

Solomon, E.A., Lega, B.C., Sperling, M.R., Kahana, M.J., 2019. Hippocampal theta codes for distances in semantic and temporal spaces. Proc. Natl. Acad. Sci. 116, 24343–24352. https://doi.org/10.1073/pnas.1906729116

Speer, N.K., Reynolds, J.R., Swallow, K.M., Zacks, J.M., 2009. Reading Stories Activates Neural Representations of Visual and Motor Experiences. Psychol. Sci. 20, 989–999. https://doi.org/10.1111/j.1467-9280.2009.02397.x

Stawarczyk, D., Wahlheim, C.N., Etzel, J.A., Snyder, A.Z., Zacks, J.M., 2020. Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. Proc. Natl. Acad. Sci. 117, 29346–29353. https://doi.org/10.1073/pnas.1918063117

Swallow, K.M., Barch, D.M., Head, D., Maley, C.J., Holder, D., Zacks, J.M., 2011. Changes in events alter how people remember recent information. J. Cogn. Neurosci. 23, 1052–1064. https://doi.org/10.1162/jocn.2010.21524

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

Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. NeuroImage 62, 1429–1438. https://doi.org/10.1016/j.neuroimage.2012.05.057

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

Yonelinas, A.P., Kroll, N.E.A., Quamme, J.R., Lazzara, M.M., Sauvé, M.-J., Widaman, K.F., Knight, R.T., 2002. Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. Nat. Neurosci. 5, 1236–1241. https://doi.org/10.1038/nn961

Zacks, J.M., 2020. Event Perception and Memory. Annu. Rev. Psychol. 71, 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

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

Zacks, J.M., Speer, N.K., Reynolds, J.R., 2009. Segmentation in reading and film comprehension. J. Exp. Psychol. Gen. 138, 307–327. https://doi.org/10.1037/a0015305

Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., Reynolds, J.R., 2007. Event perception: a mind-brain perspective. Psychol. Bull. 133, 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Supplemental Figures

Figure S1: Across-event pattern similarity at encoding. (A,C,E) Regions of interest: the left hippocampus (A, in green), right parahippocampal cortex (C, in cyan) and right perirhinal cortex (E, in magenta) are plotted on an individual subject's structural MRI. (B,D,F) Violin plots display the range and density of pattern similarity within the left hippocampus (B), right parahippocampal cortex (D), and right perirhinal cortex (F) for Coherent and Unrelated Narrative events, averaged within each event window, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives.

Figure S2: Accounting for semantic similarity in neural activity patterns. Violin plots display the range and density of model residuals within the right hippocampus (A), left parahippocampal cortex (B), and left perirhinal cortex (C) for Coherent and Unrelated Narrative events, averaged within each event window, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, *=$p<0.05$.

Figure S3: All-timepoints analysis of right hippocampal encoding-retrieval similarity and recall. In contrast with **Figure 5**, encoding-retrieval similarity analysis here included all available timepoints, which widely varied across recall cues and subjects. Within the right hippocampus, encoding-retrieval similarity was calculated by computing correlations (Pearson's r) between boundary-evoked voxel patterns from each event involving a recurring character at

encoding (Event 1 or Event 2) and delayed recall of that same character. (A) Violin plots display the range and density of encoding-retrieval similarity, along with individual subject means (colored dots), group means (black dots), and estimated within-subject standard error (black error-bars; see Methods for details). (B) Reinstatement of boundary-evoked Event 2 activity patterns during recall (encoding-retrieval similarity, x-axis) is compared with the number of details recalled (y-axis), separately for Coherent Narrative events (left panels) and Unrelated Narrative events (right panels). Lines represent fits of multiple regressions predicting an interaction between the degree of reinstatement and whether or not events can form a coherent narrative. Key: Blue=Coherent Narrative, Red=Unrelated Narratives, triangles=Event 1 recall, circles=Event 2 recall, **=p<.01.

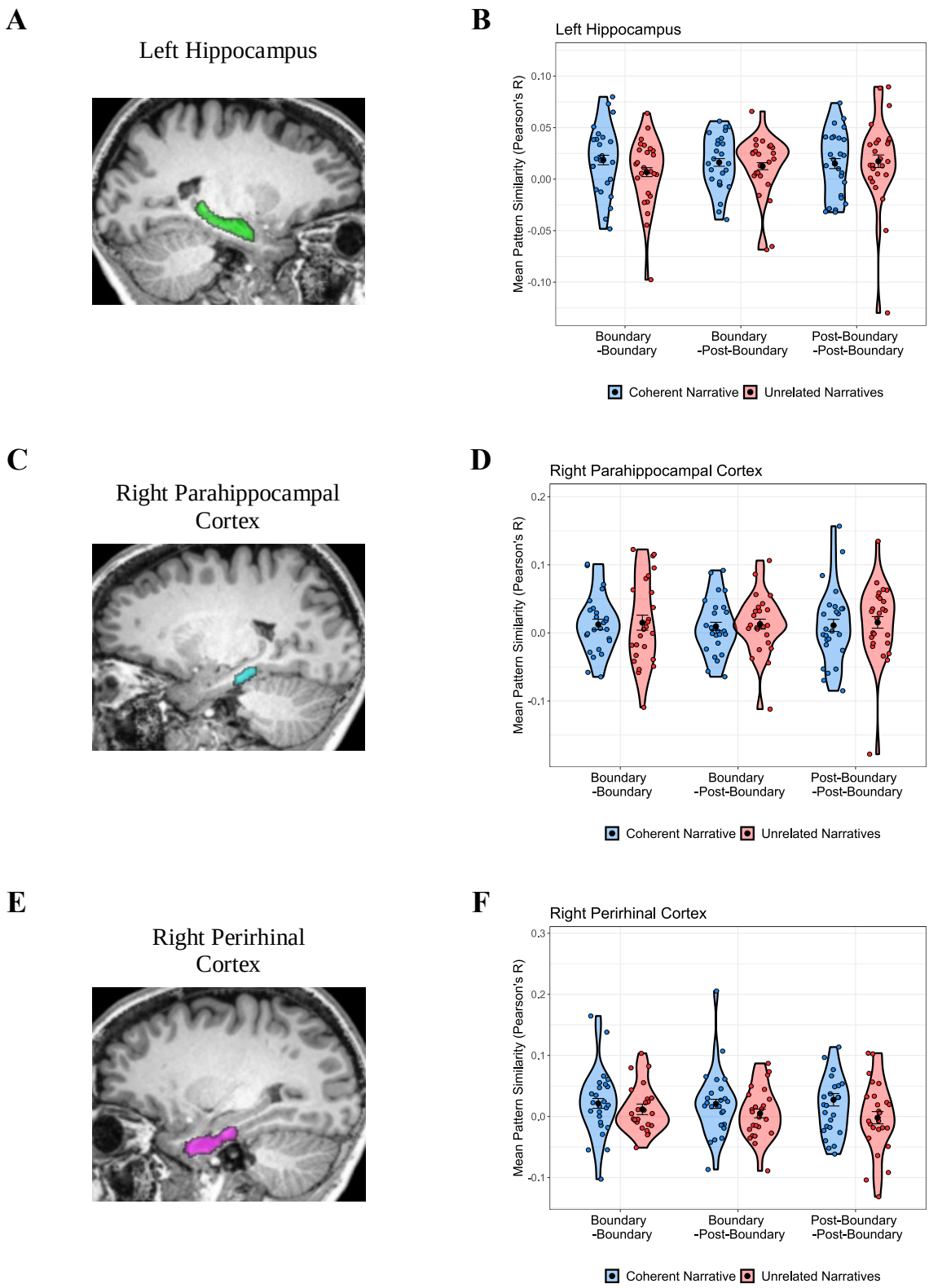

Figure S1: Across-event pattern similarity at encoding

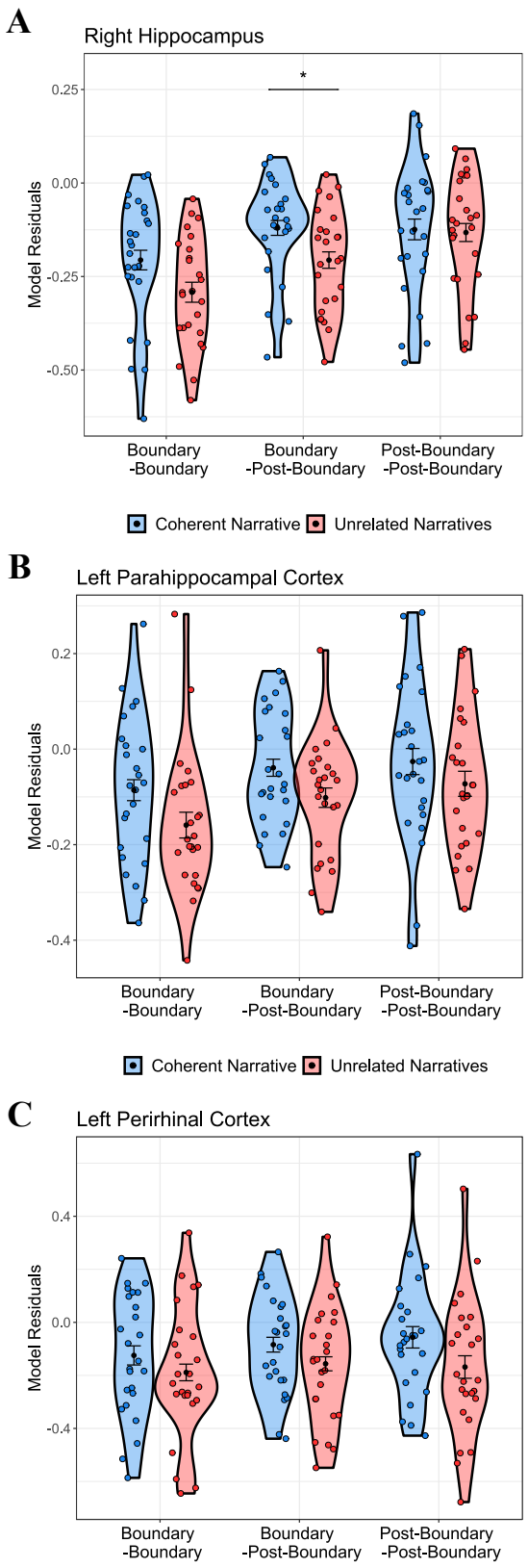

Figure S2: Accounting for semantic similarity in neural activity patterns

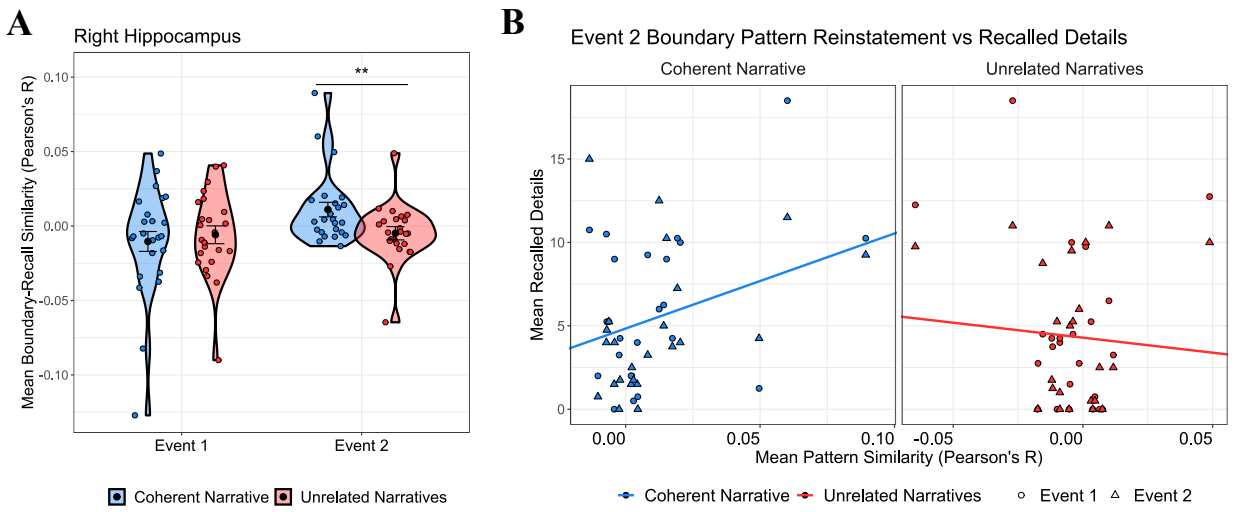

Figure S3: All-timepoints analysis of right hippocampal encoding-retrieval similarity