The hippocampus constructs narrative memories across distant events

Brendan I. Cohn-Sheehy[1,2,3,*], **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**[2,3,4]

[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

Summary

Life's events are scattered throughout time, yet we often recall different events in the context of an integrated narrative. Prior research suggests that the hippocampus, which supports memory for past events, can support the integration of overlapping associations or separate events in memory. However, the conditions which lead to hippocampus-dependent memory integration are unclear. We used functional brain imaging to test whether the opportunity to form a larger narrative (narrative coherence) drives hippocampal memory integration. During encoding of fictional stories, patterns of hippocampal activity, including activity at boundaries between events, were more similar between distant events that formed one coherent narrative, compared with overlapping events taken from unrelated narratives. One day later, the hippocampus preferentially supported detailed recall of coherent narrative events, through reinstatement of hippocampal activity patterns from encoding. These findings demonstrate a key function of the hippocampus: the integration of events into a narrative structure for memory.

Keywords

---

[*]Corresponding Author and Lead Contact: Brendan I. Cohn-Sheehy, Ph.D. bcohnsheehy@ucdavis.edu.

Introduction

It is well-established that although experience unfolds over a continuous timeline, memory for one's experience, or episodic memory [1], can be segmented into individual units of time called "events" [2–8]. However, there is reason to think that real-life memory might be organized at a level above and beyond events. For example, in one event, you might encounter your neighbor Melvin, who tells you that he is locked out of his house while a pizza is baking in the oven. The next day, Melvin might call and tell you that his kitchen is covered in ash. Although Melvin 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].

Organizing events into a narrative can be beneficial for memory [12,15–18]. We recently demonstrated that, when temporally separated events could be assimilated into a larger narrative, they were recalled in greater detail than events involving an overlapping character in different narratives [10]. Furthermore, narratives are more than the semantic information conveyed in words and sentences alone [11–13,16]. Although some models have suggested that semantic associations can bind words or sentences in memory [19], we found that the memory advantage for events that formed a larger narrative was over and above any advantage at the word or sentence level [10]. Our findings suggested that episodic memory might be organized above and beyond words, sentences, or events, on the basis of *narrative coherence*: the degree to which individual units of information can be interrelated within a single narrative [9,11]. In the present study, we assessed whether narrative coherence shapes the way the 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 studies suggest that the hippocampus can build memories that integrate information across overlapping experiences [20–26], although this might not depend on narrative coherence. For instance, when one encounters any information that overlaps with a previous event (e.g. a recurring "B" item in separated "A-B" and "B-C" pairs), this overlap may trigger the hippocampus to form associations between temporally distant experiences [22,23,25,26]. Furthermore, some findings suggest that encountering a new event can trigger the hippocampus to reinstate activity patterns which correspond with a prior, overlapping event, thereby embedding information from the prior event into memory for the new event [20,25,27]. If the hippocampus can combine information across events during memory encoding, this might facilitate subsequent recall of multiple events.

However, overlapping information does not always result in memory integration. In fact, longstanding evidence suggests that events that share overlapping features can compete with each other during memory retrieval, thus making these events more difficult to remember [28,29]. Furthermore, some findings suggest that the hippocampus can hyper-differentiate representations of overlapping events [30,31], and results from another study suggest that the hippocampus can simultaneously support the ability to segregate and integrate overlapping associations [22]. Recent models have attempted to delineate which conditions lead to hippocampus-dependent memory integration versus differentiation [20,30,32–34]. These models

generally focus on simple associations between items and do not address any role for narrative coherence. Narrative coherence may play an important role in real-life memory —and, as our recent behavioral findings suggest [10], narrative coherence might determine whether or not events become integrated.

Recent evidence suggests that narrative coherence modulates hippocampal activity[35,36]. Milivojevic et al. [36] used functional magnetic resonance imaging (fMRI) to analyze patterns of hippocampal activity evoked by repeated pairs of temporally adjacent events. Although events were initially unrelated, some event pairs could form a narrative with a subsequently-presented, linker event—when events could become linked, hippocampal activity patterns became more similar between paired events [36]. This suggested that, after one initially encodes events, hippocampal activity patterns might become modified to support a larger narrative. Related findings [35] suggested that overlapping hippocampal activity patterns might only be observed in participants who consciously recognize links between events. However, overlapping activity patterns are not equivalent to memory integration per se. If pattern overlap reflects memory integration, then hippocampal pattern overlap should also support the ability to retrieve information from memory (i.e. during recall). Critically, no study to date has demonstrated whether narrative coherence modulates hippocampus-dependent memory integration across temporally distant events.

An additional possibility is that the hippocampus might integrate events into a narrative, but only at specific moments. Many recent findings [37–41] suggest that the hippocampus becomes particularly active when people perceive transitions between events, or *event boundaries* [2,42]. Although the significance of boundary-evoked activation is unclear, several models suggest that, at event boundaries, the hippocampus retrieves information from past events in order to construct a working model of a new event [27,43–45]. A key prediction from these models is that, at the moment when one perceives an event boundary, one might begin to integrate distant events into a narrative. For instance, re-encountering Melvin might not only mark the onset of a new event, but also lead you to retrieve Melvin's previous event and integrate it with the new event [2,16,27,46].

Building on prior work showing that narratives modulate activity in the hippocampus [35,36,47], we investigated whether the hippocampus supports memory integration for temporally distant events that form a larger, coherent narrative, by examining hippocampal activity patterns during encoding and recall of realistic events. Functional magnetic resonance imaging (fMRI) was used to image hippocampal activity while participants listened to fictional stories (Figure 1A), in which recurring characters appeared in pairs of temporally separated events. Critically, stories were written to include pairs of events that could form one Coherent Narrative (CN), and pairs involving an overlapping character in Unrelated Narratives (UN; examples in Figure 1B). These events were embedded in longer stories which did not involve these recurring characters. For example, one four-minute story (Figure 1A, Story 2) followed a protagonist who, by chance, encountered the characters Sandra and Johnny, and in a subsequent story (Figure 1A, Story 4), a different protagonist encountered Sandra and Johnny. In one version of these stories, events involving Sandra could be integrated into a coherent narrative (e.g., two events that describe a single dating experience) whereas events involving Johnny were parts of unrelated narratives (e.g., losing

a family recipe in Event 1, evading financial troubles in Event 2). However, both CN and UN events were distinct from the surrounding plot, such that events involving recurring characters could only form a larger narrative in the CN condition. Importantly, different versions of events were pseudo-randomly assigned to different subjects (STAR Methods), to minimize content-driven differences between CN and UN events. Participants returned one day later, and they were scanned during detailed recall of events involving each recurring character (Figure 1C).

We hypothesized that the hippocampus would play a disproportionate role in supporting memory for CN events, as compared with UN events. We tested this hypothesis by determining whether hippocampal activity patterns during memory encoding, including activity at event boundaries, carried shared information across CN events, and whether activity patterns during memory encoding were reinstated during recall of CN events, as compared to reinstatement during recall of UN events. As follows, our results suggest that the hippocampus supports the construction of coherent narratives that integrate distant events in memory.

Results

Hippocampal activity bridges coherent narrative events during memory encoding

Our overarching hypothesis was that hippocampal activity might bridge distant events in memory, if the events can form a coherent narrative. To assess this possibility, we first identified the unique pattern of hippocampal activity (voxel pattern) corresponding to each story event (STAR Methods; Figure 2A), and we tested whether voxel patterns were more highly shared between CN events than UN events (i.e. higher pattern similarity). Alternatively, if the hippocampus supported the integration of any overlapping events (i.e. via a shared character), we would not expect any difference in pattern similarity between CN and UN events.

Additionally, drawing from models of event cognition and narrative processing [2,16,43,45,46], we predicted that hippocampal activity which is evoked by the onset of an event—at an event boundary—would support memory integration. Each story was written to include *a priori* event boundaries (Figure 1A), which reliably evoked the perception of event boundaries in an independent sample (point-biserial r=0.85, p<0.0001; Figure S1A; STAR Methods). As in previous studies [37,39,40], *a priori* boundaries evoked increased activation in the right and left hippocampus (Figure S1B). The time period which corresponded to the onset and duration of boundary-evoked activation in the hippocampus (TRs 5-11, 6.1-14.64 seconds following an event's onset; Figure S1B) is subsequently referred to as the Boundary epoch (Figure 2A). The Boundary epoch accounted for the lag between the real-time onset of an event, and increased fMRI activation (i.e. the hemodynamic response lag). For comparison, we also defined a Post-Boundary epoch (Figure 2A), in order to examine 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 an event's onset).

Within the right and left hippocampus, we computed voxel pattern similarity between the Boundary and Post-Boundary epochs of paired events (Figure 2A), and subsequently analyzed pattern similarity values using a multi-level mixed effects model (STAR Methods, Formula 1). This model tested whether Narrative Coherence [CN vs UN] modulated pattern similarity across events, and whether any effects of Narrative Coherence were specifically driven by similarity between Boundary epochs (i.e. an interaction between Narrative Coherence and Epoch). Furthermore, the mixed model statistically accounted for potential confounds: particular subjects, specific events which were presented to subjects, and overall activation at boundaries.

As shown in Figure 2B, within the right hippocampus, this model (Akaike Information Criterion: AIC = −1039) revealed a significant effect of Narrative Coherence (F(1,11.33)=6.97, p=.02), indicating that pattern similarity in the right hippocampus was higher for CN events, compared with UN events. However, there was no significant interaction between Narrative Coherence and Epoch (F(2,257.71)=0.09, p=.92; ps 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, throughout 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. A follow-up analysis of timepoint-by-timepoint correlations between events revealed similar findings (Figure S2).

This pattern of findings was not evident in the left hippocampus (Figures S2–3), 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 narrative coherence effects within the right hippocampus were significant within the right posterior hippocampus, but not the right anterior hippocampus (Figure S3B–C). Although the implications of these findings are unclear, a related study suggested that anterior and posterior hippocampus provide differential support for memory integration. [22]

Additionally, we assessed whether effects of narrative coherence were driven by low-level text characteristics. If the hippocampus supports information about words and sentences [48–50], then hippocampal pattern similarity might simply reflect the degree to which word- or sentence-level semantic information overlaps across events. We therefore quantified textual similarity between events using the Universal Sentence Encoder [51], and we statistically accounted for textual similarity within models of hippocampal pattern similarity (STAR Methods, Formula 2). Within the right hippocampus, this model (AIC = −1037) revealed that, even after accounting for lower-level semantic similarity between events, there was a significant effect of Narrative Coherence (F(1,12.06)=5.56, p=.036; other ps>0.14). This suggested that, within the right hippocampus, the degree to which narrative coherence modulated pattern similarity across events was not purely driven by lower-level semantic overlap between events. This pattern of findings was not evident within the left hippocampus (AIC = −1101; effect of overall boundary activation, F(1,288.70)=12.17, p<.001; all other effects, ps>.58).

Hippocampal activity that bridges coherent narrative events is reinstated during recall

One day after memory encoding, participants were scanned while they freely recalled information about each pair of CN and UN events (Figure 1C). Participants were oriented to recall as many details as possible from any of the events involving each CN or UN character, in order to encourage participants to recall events in an integrated manner. If activity patterns in the right hippocampus provided a basis for integrating CN events during memory encoding, we would expect that these activity patterns would facilitate subsequent memory retrieval for CN events.

We therefore investigated the degree to which right hippocampal voxel patterns from encoding were reinstated during recall (i.e. "encoding-retrieval similarity"; Figure 3A). Because preceding analyses revealed pattern similarity effects across both Boundary and Post-Boundary epochs, we pooled timepoints from both epochs to measure reinstatement of whole-event activity patterns from encoding (Figure 3A). We hypothesized that pattern reinstatement would be disproportionately higher for CN events compared with UN events.

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 1A–B). Because events were presented sequentially, Event 2 was the event which determined whether there was any relationship between the two events – that is, whether or not the two events could form a coherent narrative. In other words, if narratives spanning Events 1 and 2 were bound together, this binding should occur during Event 2. We therefore predicted that Event 2 served as a key period for memory integration, specifically in the CN condition, and that activity patterns which were present during this interval would be beneficial for recall of multiple events. In contrast, UN Event 2 would not provide an opportunity to form a single coherent narrative. Thus, we predicted disproportionately higher reinstatement for CN Event 2 relative to UN Event 2. Alternatively, if overlapping associations (e.g. a recurring character) can drive the hippocampus to reinstate activity patterns across events[25] regardless of narrative coherence, we might expect no difference in reinstatement between CN Event 2 and UN Event 2.

To test our hypothesis, we analyzed the degree to which right hippocampal voxel patterns during Event 2 encoding were reinstated during recall (Figure 3A). We also analyzed reinstatement of activity patterns from Event 1—however, because we theorized that Event 2 served as the critical period for integration, we did not expect any difference in Event 1 pattern reinstatement between CN and UN events during the recall task. To assess the overall pattern of reinstatement findings, we performed a multi-level mixed effects model (STAR Methods, Formula 3) which tested whether encoding-retrieval similarity would be modulated by Narrative Coherence [CN vs UN], 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: recall duration (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 effects were significant (ps>.48). As shown in Figure 3B, model estimates revealed that the significant interaction was driven by higher reinstatement for CN than UN Events, and that this effect was specific to Event 2. Additionally, model estimates revealed significantly higher reinstatement of CN Event 2 than CN Event 1. This does not signify that there was zero reinstatement for other events, but that reinstatement was disproportionately higher for CN Event 2. This pattern of findings remained even after accounting for pattern similarity effects during memory encoding (see STAR Methods). Interestingly, an exploratory analysis revealed similar findings when applying an identical model to predict reinstatement from Boundary epochs alone (AIC=−1427.5; significant interaction of Narrative Coherence and Event Number, F(1,343.00)=7.58, p=.006).

These findings suggested that, in the right hippocampus, activity patterns from Event 2 encoding were preferentially reinstated during recall of CN events. In line with our hypothesis, this suggests that Event 2 provided an opportunity for memory integration, and that hippocampal activity patterns during Event 2 may have provided a basis for integrating CN events in memory.

Narrative coherence modulates the relationship between hippocampal pattern reinstatement and detailed recall

Having established that information from Event 2 was preferentially reinstated during recall of CN events, our next analyses assessed whether this reinstatement effect was behaviorally relevant. In previous work, we found that, across three experiments, participants were able to recall more details about CN than UN events [10]. 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 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 that this variance might be better explained by accounting for inter-subject differences in hippocampal activity. If so, we would expect that recall of CN events should be predicted by reinstatement of right hippocampus activity patterns during recall.

Following up on our previous analyses (Figure 3B), we tested whether reinstatement of right hippocampus activity patterns from CN Event 2 was predictive of how many details participants recalled about CN events. In line with our prediction, there was a significant positive correlation between mean CN Event 2 reinstatement and mean recalled details for CN events (r=0.41, p=0.048). That is, the degree to which Event 2 patterns were reinstated during recall predicted the degree to which participants recalled CN events in detail. For comparison, we also tested whether UN Event 2 reinstatement would predict detailed recall. In contrast, there was no significant correlation between mean UN Event 2 reinstatement and mean recall of UN events (r=−0.23, p=0.28). This initially suggested that narrative coherence might modulate the relationship between right hippocampal pattern reinstatement

and detailed recall. To formally assess this possibility, we designed a rigorous mixed-effects model which could discern between several theoretical predictions at the level of individual events.

In line with our pattern reinstatement findings (Figure 3B), we reasoned that Event 2 provided an opportunity to integrate CN events, but not UN events, in memory. If so, detailed recall would be predicted by a statistical interaction between Event 2 pattern reinstatement and narrative coherence. Furthermore, the recall task was designed to elicit recall of multiple events—as described previously [10,29], the ability to recall details from overlapping events depends on whether or not they can be easily integrated. As such, if memory integration took place during encoding of CN Event 2, then pattern reinstatement from CN Event 2 should not only predict how well Event 2 was recalled, but also how well Event 1 was recalled, reflecting that these events became integrated. In contrast, we did not expect that UN Event 2 provided a clear opportunity for memory integration, and as such, we predicted that pattern reinstatement from UN Event 2 would be less predictive of recall from multiple overlapping events. Alternatively, it was also possible that no memory integration took place during encoding of either CN or UN Event 2—if so, we might expect that CN or UN Event 2 pattern reinstatement would predict recall of details from CN or UN Event 2, respectively, but not details from Event 1 (i.e. an interaction between Event 2 reinstatement and which event was recalled).

With these hypotheses in mind, we designed a mixed effects model (STAR Methods, Formula 5) to test whether and how recalled details would be predicted by Event 2 pattern reinstatement. Critically, this model tested whether the relationship between recall and pattern reinstatement would be modulated by Narrative Coherence. Furthermore, 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). Interestingly, this suggested that, after accounting for a full pattern of mixed effects, there was significantly higher recall for CN than UN events.

Moreover, the main effect of Narrative Coherence was qualified by a significant interaction between Narrative Coherence and Event 2 Reinstatement ($F(1,156.42)$=5.20, p=0.024). As shown in Figure 3C, this interaction reflected that in the right hippocampus, the predicted relationship between Event 2 pattern reinstatement and recalled details (from either event) was significantly more positive for CN events than for UN events ($t(156)$=2.28, p=0.024). That is, reinstatement of hippocampal activity patterns from CN Event 2 was associated with a greater behavioral benefit for recalling multiple events than reinstatement from UN Event 2. Consistent with our hypothesis, this suggests that hippocampal activity from Event 2 preferentially supported memory integration for CN events. Similar findings were revealed after removing potential outliers (see STAR Methods). Interestingly, similar findings were also revealed when examining reinstatement from Boundary epochs alone (AIC=1114.7; significant interaction of Event 2 Reinstatement by Narrative Coherence, $F(1,160.15)$=10.56, p=.001).

No other fixed effects or interactions were significant (ps>.34). For one, there were no overall differences in recall of Event 1 versus Event 2. Furthermore, Event 2 Reinstatement did not differentially predict recalled details for Event 1 versus Event 2, even in the case of UN events (i.e. UN Event 2 reinstatement did not predict UN Event 2 recall). This null effect likely reflects the nature of the recall task (see Discussion).

Post-hoc analyses of cortical networks

Our study was specifically designed to test *a priori* hypotheses about hippocampal contributions to memory. It is also possible to generate testable hypotheses about many other brain areas that might be involved in this task. For completeness, we have included post-hoc analyses (see STAR Methods) which test for narrative coherence effects, within two key cortical networks that interact with the hippocampus[52–55]: the Posterior Medial (PM) network, which supports information about event contexts (e.g. types of situations); and the Anterior Temporal (AT) network, which supports information about particular entities within events (e.g. people or objects). Because stories were presented aurally, we also investigated auditory cortex.

If, for instance, narrative coherence findings within the hippocampus were driven by contextual similarity between events, one might also expect narrative coherence effects within the PM network. However, there were no significant effects of narrative coherence within the PM network. Interestingly, there were significant effects of narrative coherence within the AT network and auditory cortex. These findings are inconclusive, and further work is needed to understand how cortical networks contribute to memory integration for coherent narrative events.

Discussion

In the current study, we investigated whether the hippocampus supports a narrative-level organization for episodic memories, by integrating distant events into a larger narrative. Consistent with this hypothesis, activity patterns in the right hippocampus during memory encoding were more similar across two temporally separated events that formed a coherent narrative, compared with unrelated events that shared an overlapping character. Activity patterns elicited during encoding of coherent narrative events were preferentially reinstated when these events were recalled one day later. 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 both events within a narrative.

Taken together, these findings suggest that, rather than simply integrating memories of overlapping experiences, the hippocampus supports the construction of narratives that integrate distant events in memory. A considerable body of evidence suggests that hippocampus-dependent processes can support memory integration[21–23,25,26] — the ability to put together information across two simple associations with an overlapping element —though it is also the case that the hippocampus can assign highly differentiated

representations to overlapping associations [30,31]. In fact, some findings have suggested that the hippocampus keeps event representations distinct during memory encoding, instead supporting integration of overlapping associations on the fly during memory retrieval [33,56].

Although important questions remain about dynamics of memory integration in these studies, it is important to clarify differences between the present study and previous work on memory integration. Studies of memory integration typically focus on overlapping associations that share the same element—for instance, two houses or scenes that are associated with the same face [24,57]. In the present study, both CN and UN events shared overlapping characters, yet hippocampal pattern similarity was significantly higher across CN events, and CN event patterns were preferentially reinstated during subsequent recall. Thus, overlap across events, while potentially necessary, was not sufficient to integrate information across distant events. Furthermore, in contrast with prior work, this study investigated memory integration in the context of complex, realistic events and narratives.

A few studies have shown that hippocampal activity during encoding of temporally adjacent events can be sensitive to narrative coherence [35,36]. Building on these studies, the present findings directly support a relationship between narrative coherence effects in the hippocampus and detailed recall of temporally distant events. In other words, the present findings suggest that narrative coherence does more than modulate hippocampal activity across repeatedly presented events—rather, they directly support the hypothesis that narrative coherence drives hippocampal memory integration, and that this integration can even take place across events that are encountered at disparate times.

Aside from the present results, only a few studies [35,36,47,58] have implicated the hippocampus in narrative construction, and most neuroimaging studies of narrative processing have focused on cortical areas, particularly within the PM network [37,58–62]. Notably, there are important methodological differences between prior work, and the present study, in which post-hoc analyses did not reveal significant PM network effects. Most fMRI studies of narrative structure have analyzed activity patterns averaged over the course of an entire event [37,54,61,63]. Although PM network activity patterns are often stable throughout an event [37,61], hippocampal activity tends to be more dynamic [39]. For instance, Reagh & Ranganath[54] compared activity patterns across repetitions of a 40-second film, and results showed that hippocampal patterns primarily carried information about the time period immediately after the onset of the film[54]. This finding is interesting, because some of the key hippocampal findings in the present study were seen during the Boundary epoch.

To be clear, we can neither conclude, nor rule out, that event boundaries played a unique or disproportionate role in integrating distant events. Our hypothesis-driven analyses revealed that pattern similarity was higher across CN events than across UN events during the Boundary epoch, and between the Boundary epoch and Post-Boundary epoch. In other words, hippocampal activity during the Boundary epoch carried similar information across coherent narrative events. Notably, several models suggest that when a new event begins (i.e. at an event boundary), people can update their current representation of an event, by retrieving information about relevant past events [7,27,43–45]. In one recent model, Lu et al.[45] demonstrated that it is computationally advantageous to restrict hippocampal encoding

and retrieval of cortical states to periods around an event boundary, and this fits with the established idea that people often rely on episodic retrieval to build mental representations of an ongoing event [7,27,43,44,46,64]. That is, even if event boundaries serve as "walls" between temporally adjacent events, hippocampal retrieval at boundaries might construct "bridges" across temporally distant events.

Interestingly, other recent findings have suggested that boundary-evoked hippocampal activity supports storage of events that precede an event boundary [37,39,41]. These findings are not 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. Our single-trial modeling approach (STAR Methods) enabled us to statistically separate the components of hippocampal activity which correspond to an ongoing event, from the signal corresponding to temporally adjacent events, enabling us to specifically test hypotheses about boundary-evoked hippocampal activity patterns at the start of an event. Future work 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[41,44,45,65]) or multiple processes (e.g. memory storage vs retrieval).

Another key finding is that hippocampal activity patterns during CN events were preferentially reinstated during memory retrieval, and this reinstatement was more predictive of individual differences in memory for CN than UN events. In our previous behavioral study[10], memory performance was significantly higher for events that formed a coherent narrative than for overlapping, but unrelated events. In order to adapt this paradigm for fMRI pattern analyses, we increased the number of trials per condition by presenting stories twice during memory encoding. It is likely that repeated presentation enabled participants to encode more information about UN events, which would explain why detailed recall was numerically, but not significantly higher for CN than UN events. Nonetheless, participants showed preferential reinstatement of information from Event 2—precisely when integration would be expected to occur—during recall of CN events. If hippocampal activity simply reflected successful encoding of Event 2, we would expect to see that reinstatement would be equally predictive of memory performance for CN and UN events, since memory performance was not significantly different across these conditions. In fact, reinstatement of information during this key period when participants could build a coherent narrative was predictive of memory for CN events, whereas reinstatement of information from an independent event did not significantly predict memory performance.

Importantly, the recall task cued memory for multiple events, rather than one event at a time, in order to assess effects of memory integration. Notably, previous work suggests that re-encountering a feature of a prior event (i.e. during Event 2) can act as a reminder which enables online retrieval of memory for the prior event, and therefore embeds information about the prior event within memory for the new event [46,66]. This kind of reminder-evoked integration can be useful for narrative comprehension[16,67], and it may have provided a mechanism for integrating CN events, explaining the correspondence between Event 2 pattern reinstatement and recall of both CN events. However, reminders do not always lead to integration, and can also lead to differentiation or competition between events [46,64,68].

This may explain why reinstatement from UN Event 2, in which reminders would not easily lead to integration, did not provide a clear benefit for recall of multiple overlapping events. We also speculate that adjusting the recall task to cue recall of one event at a time might reveal a more specific, timepoint-by-timepoint correspondence between neural pattern reinstatement and recall of details from specific events.

Finally, it might be of interest that significant effects of narrative coherence were observed within the right hippocampus, but not the left hippocampus. It is unclear whether the right hippocampus plays a disproportionately important role in narrative construction. Notably, other studies have implicated the right hippocampus in the integration of information across overlapping events [20,47], but it is also the case that fMRI studies of hippocampal function often report significant effects in only one hemisphere [69]. Further hypothesis-driven research is needed to understand the relative contributions of the right and left hippocampus to episodic memory.

In closing, even though life's events occur at disparate times, the hippocampus can form memories that integrate events into a larger, coherent narrative. By bridging the divide between distant events, the hippocampus may support a narrative-level architecture for episodic memory.

STAR Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Brendan Cohn-Sheehy (bcohnsheehy@ucdavis.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability—Upon publication, behavioral and neuroimaging data, analysis code, and other materials will be available via the Open Science Framework: https://osf.io/7zsxd/. Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants—Human subjects protocols were approved by the UC Davis Institutional Review Board.

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).

METHOD DETAILS

Experimental Stimuli—The stimulus design is depicted in Figure 1A–B, and has been described elsewhere [10]. 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") [4,70]. 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; confirmed by human raters, Figure 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 (Figure 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 Figure 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 versions were 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 [10], 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 [71].

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). Stories were repeated in order to increase the number of available trials and timepoints for pattern similarity analyses. 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 [10]. 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 explicitly instruct participants to integrate information between events. Within one extended scanning run (Figure 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 (e.g. "Melvin Doyle (Charles Bort's neighbor, Karen Joyce's friend)"). This cueing approach was designed to encourage recall of both Events 1 and 2, thus providing an opportunity for participants to naturally integrate CN events during recall (or not, as hypothesized for UN events). 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 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 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 [10], and which adapted a well-characterized scoring method from the Autobiographical Memory Interview [72]. 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 ($\overline{X} = 0.06$ 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 ($\overline{X} = 0.14$ 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). To ensure proper alignment between functional images from 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/spml2/), 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 [73]. 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 cumulatively across the scanning session, 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/) [74] to define the 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 (per 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 [40], for exploratory analyses of right hippocampal subregions (Figure 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. Other cortical ROIs which were implemented in post-hoc analyses (see Ancillary analyses, below) were derived from individual subject parcellations in Freesurfer (https://surfer.nmr.mgh.harvard.edu), using the Destrieux atlas [75]. ROIs were then co-registered and resliced to functional images.

Univariate measures of boundary-evoked activation

Boundary vs. Mid-Event timepoints: 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 published analyses of event boundary activation [39], 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; 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, 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 [76]. For post-hoc analyses of cortical networks (see Ancillary analyses, below), the cluster threshold was adjusted for the number of models performed (3 models, 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 [77].

Activation during CN and UN events: To supplement the multivoxel fMRI analyses described below, we separately quantified univariate activation for CN and UN events, and incorporated this measure ("overall boundary activation") as a fixed effect in mixed models of hippocampal pattern similarity during memory encoding (see Multi-level mixed effects models, below). In contrast to the univariate analysis reported above, this was performed using beta volumes which were initially modeled for multivoxel fMRI analysis (see Single trial models, below), and then smoothed with a 6mm-FWHM smoothing kernel for use in univariate analysis. We then extracted beta values during the Boundary epoch, and averaged these beta values within left and right hippocampus ROIs.

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) [78,79]. 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 [80]. 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) [81]. 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 each ROI 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 (Figure 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).

Follow-up timepoint-by-timepoint analysis: As shown in Figure S2, we also analyzed timepoint-by-timepoint pattern similarity between pairs of events at encoding, to follow-up on epoch-based analyses (i.e. to see which timepoints had CN>UN similarity). To assess statistical significance for narrative coherence effects, and to correct for multiple comparisons, we used cluster-based permutation tests[76] with 1000 permutations, with a cluster defining threshold of p=0.05 (one-tailed, to test for CN>UN similarity) and a cluster mass threshold of p=0.05. Each pixel of a statistical comparison (t-value) was converted into a Z value by normalizing it to the mean and standard error generated from our permutation distributions. Above-threshold (Z>1.96) pixels and above-threshold clusters are plotted in Figure S2C–D.

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 (Figure 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.

Textual similarity analysis—We modeled textual similarity between sentences from different story events using the freely available, "transformer" version of the Universal Sentence Encoder (USE), a text embedding model designed to convert text into numerical vectors [51] (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, and the respective positions of words, within each sentence. Then, cosine similarity is calculated between each sentence vector, yielding pairwise measures of textual similarity for all sentences. As such, USE similarity quantifies 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 time-course 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. Finally, USE similarity was selectively averaged within each condition of interest (CN vs UN) and time-window (Boundary-Boundary, Boundary-Post-Boundary, Post-Boundary-Post-Boundary).

These averaged USE similarity values were subsequently incorporated into mixed-effects models (see Statistical Analysis, Formula 2, below).

Ancillary analyses

Accounting for pattern similarity during encoding within pattern reinstatement findings: As an additional control, we evaluated whether findings from reinstatement analyses (Figure 3B) were biased by the degree to which activity patterns overlapped across events during memory encoding (Figure 2B). Even after accounting for pattern similarity at encoding within the mixed model formula (see Statistical Analysis, Formula 4, below; AIC = −1349.1), reinstatement was predicted by a significant main effect of Narrative Coherence ($F_{(1,291.90)}$=6.28, p=.013) and a significant interaction of Narrative Coherence and Event Number ($F_{(1,342.25)}$=8.15, p=.005; also a significant main effect of encoding-only pattern similarity, $F_{(1,322.83)}$=16.68, p<0.001).

Accounting for potential outliers within reinstatement-versus-recall findings: As an additional control, we evaluated whether the relationship between CN Event 2 reinstatement and recall of CN events (Figure 3C) was driven by potential outliers. A similar pattern of findings was revealed when excluding events with zero recalled details from the model (see Statistical Analysis, Formula 5, below; AIC=796.91; significant interaction of Narrative Coherence and Event 2 Reinstatement, $F_{(1,99.01)}$=4.45, p=0.037). An additional follow-up model excluded potential outliers based on Event 2 reinstatement or recalled details (+/− 3 standard deviations from the mean), and applied a square-root transform to recalled details to better approximate normality – this approach also revealed a similar pattern of findings (AIC=598.65; significant effect of Narrative Coherence, $F_{(1,154.17)}$=6.43, p=0.012; significant interaction of Narrative Coherence and Event 2 Reinstatement, $F_{(1,154.63)}$=5.31, p=0.023).

Post-hoc analyses of cortical networks: Post-hoc models (see Statistical Analysis, Formula 6, below) accounted for specific epochs and regions of interest within the PM network (parahippocampal cortex, angular gyrus, posterior medial cortex, medial prefrontal cortex), AT network (perirhinal cortex, temporal pole, orbitofrontal cortex), and bilateral auditory cortex (left and right transverse temporal gyrus). Within the PM network, post-hoc modeling which accounted for statistical interactions with specific epochs and regions of interest (AIC= −4372.2) did not reveal a significant effect of Narrative Coherence ($F_{(1,12.35)}$=2.20, p=0.16). Within the AT network, post-hoc modeling (AIC=−5159.5) revealed a significant interaction between Narrative Coherence and Epoch ($F_{(2,1727.42)}$=6.95, p<.001), and this interaction was driven by significantly higher pattern similarity across CN than UN events within the Post-Boundary-Post-Boundary epoch alone (Holm-corrected $t_{(24.2)}$=3.48, p=.02; other ts>.065). Within bilateral auditory cortex, post-hoc modeling (AIC=−608.1) revealed a significant main effect of Narrative Coherence ($F_{(1,8.81)}$=13.21, p=0.006).

Additionally, in line with analyses of boundary-evoked hippocampal activation (see Figure S1B), we assessed whether these cortical networks exhibited activation increases at event boundaries (see Univariate measures of boundary-evoked activation, above). Only the PM

network exhibited significant boundary-evoked activation (TRs +4 to +14; cluster-defining threshold corrected for multiple comparisons, p<0.017).

QUANTIFICATION AND 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 (Figure S1), 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 [82].

Multi-level mixed effects models:

Mixed model formulas are described within the Results text. Fixed effects specifications were based on experimental hypotheses, and random effects structures 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 Event 1A and Event 2B, Melvin Event 1B and Event 2B; "A" and "B" refer to different possible event versions, based on randomization – see Experimental Stimuli, above). 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) [83], followed by systematically pruning the random effects structure until the model converged while avoiding a singular solution (i.e. overfitting) [84]. 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: Pattern similarity ~ Narrative Coherence * Epoch + Overall Boundary Activation + USE similarity + (1 | Subject) + (1 | Event Content)

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

Formula 4 (Supplemental Analysis 1): Encoding-retrieval similarity ~ Narrative Coherence * Event Number + Encoding-only pattern similarity + Recall Duration + (1 | Subject) + (1 | Event Content)

Formula 5 (Fig. 3C): Recalled details ~ Event 2 Reinstatement * Narrative Coherence * Event Recalled + Recall duration + (1 | Subject) + (1 | Event Content)

Formula 6 (post-hoc cortical analyses; Supplemental Analysis 3): Pattern similarity ~ Narrative Coherence * Epoch * Region of interest (ROI) + (1 | Subject) + (1 | Event Content)

Model fits were quantified using the Akaike 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 [84]. 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; Figures 2–3, Figure S3) and interactions between factors and continuous variables (emtrends and emmip; Figure 3C).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

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

2. Zacks JM. (2020). Event Perception and Memory. Annu. Rev. Psychol 71, 165–191. [PubMed: 31905113]

3. Clewett D, Gasser C, and Davachi L (2020). Pupil-linked arousal signals track the temporal organization of events in memory. Nat. Commun 11, 4007. [PubMed: 32782282]

4. Ezzyat Y, and Davachi L (2011). What constitutes an episode in episodic memory? Psychol. Sci 22, 243–252. [PubMed: 21178116]

5. Horner AJ, Bisby JA, Wang A, Bogus K, and Burgess N (2016). The role of spatial boundaries in shaping long-term event representations. Cognition 154, 151–164. [PubMed: 27295330]

6. Swallow KM, Barch DM, Head D, Maley CJ, Holder D, and Zacks JM (2011). Changes in events alter how people remember recent information. J. Cogn. Neurosci 23, 1052–1064. [PubMed: 20521850]

7. Zacks JM, Speer NK, Swallow KM, Braver TS, and Reynolds JR (2007). Event perception: a mind-brain perspective. Psychol. Bull. 133, 273–293. [PubMed: 17338600]

8. Radvansky GA, and Copeland DE (2006). Walking through doorways causes forgetting: Situation models and experienced space. Mem. Cognit 34, 1150–1156.

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

10. Cohn-Sheehy BI, Delarazan AI, Crivelli-Decker JE, Reagh ZM, Mundada NS, Yonelinas AP, Zacks JM, and Ranganath C (2021). Narratives bridge the divide between distant events in episodic memory. Mem. Cognit

11. Graesser AC, Singer M, and Trabasso T (1994). Constructing inferences during narrative text comprehension. Psychol. Rev 101, 371–395. [PubMed: 7938337]

12. Trabasso T, Secco T, and Van Den Broek P (1984). Causal Cohesion and Story Coherence. In Learning and Comprehension of Text, Mandl H, Stein NL, and Trabasso T, eds. (Lawrence Erlbaum Associates, Inc.).

13. van Dijk TA, and Kintsch W (1983). Strategies of discourse comprehension (Academic Press).

14. Willems RM, Nastase SA, and Milivojevic B (2020). Narratives for Neuroscience. Trends Neurosci. 43, 271–273. [PubMed: 32353331]

15. Bransford JD, and Johnson MK (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. J. Verbal Learn. Verbal Behav 11, 717–726.

16. Kintsch W (1992). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes, Healy AF, Kosslyn SM, and Shiffrin RM, eds. (Lawrence Erlbaum Associates, Inc.), pp. 261–278.

17. Rumelhart DE (1977). Understanding and Summarizing Brief Stories. In Basic processes in reading: Perception and comprehension, Laberge D and Samuels SJ, eds. (Lawrence Erlbaum Associates), pp. 265–303.

18. Thorndyke PW (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognit. Psychol 9, 77–110.

19. Reder LM, and Anderson JR (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. Cognit. Psychol 12, 447–472. [PubMed: 7418365]

20. Horner AJ, Bisby JA, Bush D, Lin W-J, and Burgess N (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. Nat. Commun 6, 7462. [PubMed: 26136141]

21. Kuhl BA, Shah AT, DuBrow S, and Wagner AD (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat. Neurosci 13, 501–506. [PubMed: 20190745]

22. Schlichting ML, Mumford JA, and Preston AR (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun 6, 8151. [PubMed: 26303198]

23. Schlichting ML, Zeithamova D, and Preston AR (2014). CA1 subfield contributions to memory integration and inference: Memory Integration in CA1. Hippocampus 24, 1248–1260. [PubMed: 24888442]

24. Shohamy D, and Wagner AD (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. Neuron 60, 378–389. [PubMed: 18957228]

25. Zeithamova D, Dominick AL, and Preston AR (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. Neuron 75, 168–179. [PubMed: 22794270]

26. Zeithamova D, and Preston AR (2010). Flexible Memories: Differential Roles for Medial Temporal Lobe and Prefrontal Cortex in Cross-Episode Binding. J. Neurosci 30, 14676–14684. [PubMed: 21048124]

27. Stawarczyk D, Wahlheim CN, Etzel JA, Snyder AZ, and Zacks JM (2020). Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. Proc. Natl. Acad. Sci 117, 29346–29353. [PubMed: 33229530]

28. Radvansky GA, and Zacks RT (1991). Mental models and the fan effect. J. Exp. Psychol. Learn. Mem. Cogn 17, 940–953. [PubMed: 1834775]

29. Radvansky GA (2012). Across the Event Horizon. Curr. Dir. Psychol. Sci 21, 269–272.

30. Chanales AJH, Oza A, Favila SE, and Kuhl BA (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. Curr. Biol 27, 2307–2317.e5. [PubMed: 28736170]

31. Favila SE, Chanales AJH, and Kuhl BA (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat. Commun 7, 11066. [PubMed: 27925613]

32. Chanales AJH, Dudukovic NM, Richter FR, and Kuhl BA (2019). Interference between overlapping memories is predicted by neural states during learning. Nat. Commun 10, 5363. [PubMed: 31767880]

33. Kumaran D. (2012). What representations and computations underpin the contribution of the hippocampus to generalization and inference? Front. Hum. Neurosci 6.

34. Ritvo VJH, Turk-Browne NB, and Norman KA (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. Trends Cogn. Sci 23, 726–742. [PubMed: 31358438]

35. Collin SHP, Milivojevic B, and Doeller CF (2015). Memory hierarchies map onto the hippocampal long axis in humans. Nat. Neurosci. 18, 1562–1564. [PubMed: 26479587]

36. Milivojevic B, Vicente-Grabovetsky A, and Doeller CF (2015). Insight Reconfigures Hippocampal-Prefrontal Memories. Curr. Biol 25, 821–830. [PubMed: 25728693]

37. Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, and Norman KA (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron 95, 709–721.e5. [PubMed: 28772125]

38. Ben-Yakov A, and Dudai Y (2011). Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. J. Neurosci 31, 9032–9042. [PubMed: 21677186]

39. Ben-Yakov A, and Henson RN (2018). The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. J. Neurosci 38, 10057–10068. [PubMed: 30301758]

40. Reagh ZM, Delarazan AI, Garber A, and Ranganath C (2020). Aging alters neural activity at event boundaries in the hippocampus and Posterior Medial network. Nat. Commun 11, 3980. [PubMed: 32769969]

41. Sinclair AH, Manalili GM, Brunec IK, Alison Adcock R, and Barense MD (2020). Prediction Errors Disrupt Hippocampal Representations and Update Episodic Memories (Neuroscience).

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

43. Griffiths BJ, and Fuentemilla L (2020). Event conjunction: How the hippocampus integrates episodic memories across event boundaries. Hippocampus 30, 162–171. [PubMed: 31566860]

44. Franklin NT, Norman KA, Ranganath C, Zacks JM, and Gershman SJ (2020). Structured Event Memory: A neuro-symbolic model of event cognition. Psychol. Rev 127, 327–361. [PubMed: 32223284]

45. Lu Q, Hasson U, and Norman KA (2020). Learning to use episodic memory for event prediction (bioRxiv).

46. Ross BH, and Bradshaw GL (1994). Encoding effects of remindings. Mem. Cognit 22, 591–605.

47. Milivojevic B, Varadinov M, Vicente Grabovetsky A, Collin SHP, and Doeller CF (2016). Coding of Event Nodes and Narrative Context in the Hippocampus. J. Neurosci. Off. J. Soc. Neurosci 36, 12412–12424.

48. Duff MC, Covington NV, Hilverman C, and Cohen NJ (2020). Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship. Front. Hum. Neurosci 13, 471. [PubMed: 32038203]

49. Solomon EA, Lega BC, Sperling MR, and Kahana MJ (2019). Hippocampal theta codes for distances in semantic and temporal spaces. Proc. Natl. Acad. Sci 116, 24343–24352. [PubMed: 31723043]

50. Piai V, Anderson KL, Lin JJ, Dewar C, Parvizi J, Dronkers NF, and Knight RT (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. Proc. Natl. Acad. Sci 113, 11366–11371. [PubMed: 27647880]

51. Cer D, Yang Y, Kong S, Hua N, Limtiaco N, John RS, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, et al. (2018). Universal Sentence Encoder. ArXiv180311175 Cs.

52. Ranganath C, and Ritchey M (2012). Two cortical systems for memory-guided behaviour. Nat. Rev. Neurosci 13, 713–726. [PubMed: 22992647]

53. Reagh ZM, and 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. [PubMed: 29704572]

54. Reagh ZM, and Ranganath C (2021). A cortico-hippocampal scaffold for representing and recalling lifelike events (Neuroscience).

55. Baldassano C, Hasson U, and Norman KA (2018). Representation of Real-World Event Schemas during Narrative Perception. J. Neurosci 38, 9689–9699. [PubMed: 30249790]

56. Koster R, Chadwick MJ, Chen Y, Berron D, Banino A, Düzel E, Hassabis D, and Kumaran D (2018). Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. Neuron 99, 1342–1354.e6. [PubMed: 30236285]

57. Preston AR, Shrager Y, Dudukovic NM, and Gabrieli JDE (2004). Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 14, 148–152. [PubMed: 15098720]

58. Chen J, Honey CJ, Simony E, Arcaro MJ, Norman KA, and 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.

59. Aly M, Chen J, Turk-Browne NB, and Hasson U (2018). Learning Naturalistic Temporal Structure in the Posterior Medial Network. J. Cogn. Neurosci 30, 1345–1365. [PubMed: 30004848]

60. Chang CHC, Lazaridi C, Yeshurun Y, Norman KA, and Hasson U (2021). Relating the Past with the Present: Information Integration and Segregation during Ongoing Narrative Processing. J. Cogn. Neurosci, 1–23.

61. Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, and Hasson U (2017). Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci 20, 115–125. [PubMed: 27918531]

62. Heusser AC, Fitzpatrick PC, and Manning JR (2021). Geometric models reveal behavioural and neural signatures of transforming experiences into memories. Nat. Hum. Behav

63. Oedekoven CSH, Keidel JL, Berens SC, and Bird CM (2017). Reinstatement of memory representations for lifelike events over the course of a week. Sci. Rep 7.

64. Wahlheim CN, and Jacoby LL (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. Mem. Cognit 41, 1–15.

65. Antony JW, Hartshorne TH, Pomeroy K, Gureckis TM, Hasson U, McDougle SD, and Norman KA (2020). Behavioral, Physiological, and Neural Signatures of Surprise during Naturalistic Sports Viewing. Neuron, S0896627320308539.

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

67. Kintsch W (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychol. Rev 95, 163–182. [PubMed: 3375398]

68. Wahlheim CN, and Zacks JM (2019). Memory guides the processing of event changes for older and younger adults. J. Exp. Psychol. Gen 148, 30–50. [PubMed: 29985021]

69. Diana RA, Yonelinas AP, and Ranganath C (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci 11, 379–386. [PubMed: 17707683]

70. Zacks JM, Speer NK, and Reynolds JR (2009). Segmentation in reading and film comprehension. J. Exp. Psychol. Gen 138, 307–327. [PubMed: 19397386]

71. Bainbridge WA, Isola P, and Oliva A (2013). The intrinsic memorability of face photographs. J. Exp. Psychol. Gen 142, 1323–1334. [PubMed: 24246059]

72. Levine B, Svoboda E, Hay JF, Winocur G, and Moscovitch M (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. Psychol. Aging 17, 677–689. [PubMed: 12507363]

73. Power JD. (2017). A simple but useful way to assess fMRI scan qualities. NeuroImage 154, 150–158. [PubMed: 27510328]

74. Ritchey M, Montchal ME, Yonelinas AP, and Ranganath C (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. eLife 4.

75. Destrieux C, Fischl B, Dale A, and Halgren E (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53, 1–15. [PubMed: 20547229]

76. Maris E, and Oostenveld R (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. [PubMed: 17517438]

77. Cohen MX (2014). Analyzing neural time series data: theory and practice.

78. Mumford JA, Turner BO, Ashby FG, and Poldrack RA (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59, 2636–2643. [PubMed: 21924359]

79. Turner BO, Mumford JA, Poldrack RA, and Ashby FG (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. NeuroImage 62, 1429–1438. [PubMed: 22659443]

80. Kriegeskorte N, Goebel R, and Bandettini P (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A 103, 3863–3868. [PubMed: 16537458]

81. Kriegeskorte N, Mur M, and Bandettini P (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci 2, 4. [PubMed: 19104670]

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

83. Barr DJ, Levy R, Scheepers C, and Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. J. Mem. Lang 68, 255–278.

84. Singmann H, and Kellen D (2019). An Introduction to Mixed Models for Experimental Psychology. In New Methods in Cognitive Psychology, Spieler DH and Schumacher E eds. (Psychology Press), pp. 4–31.
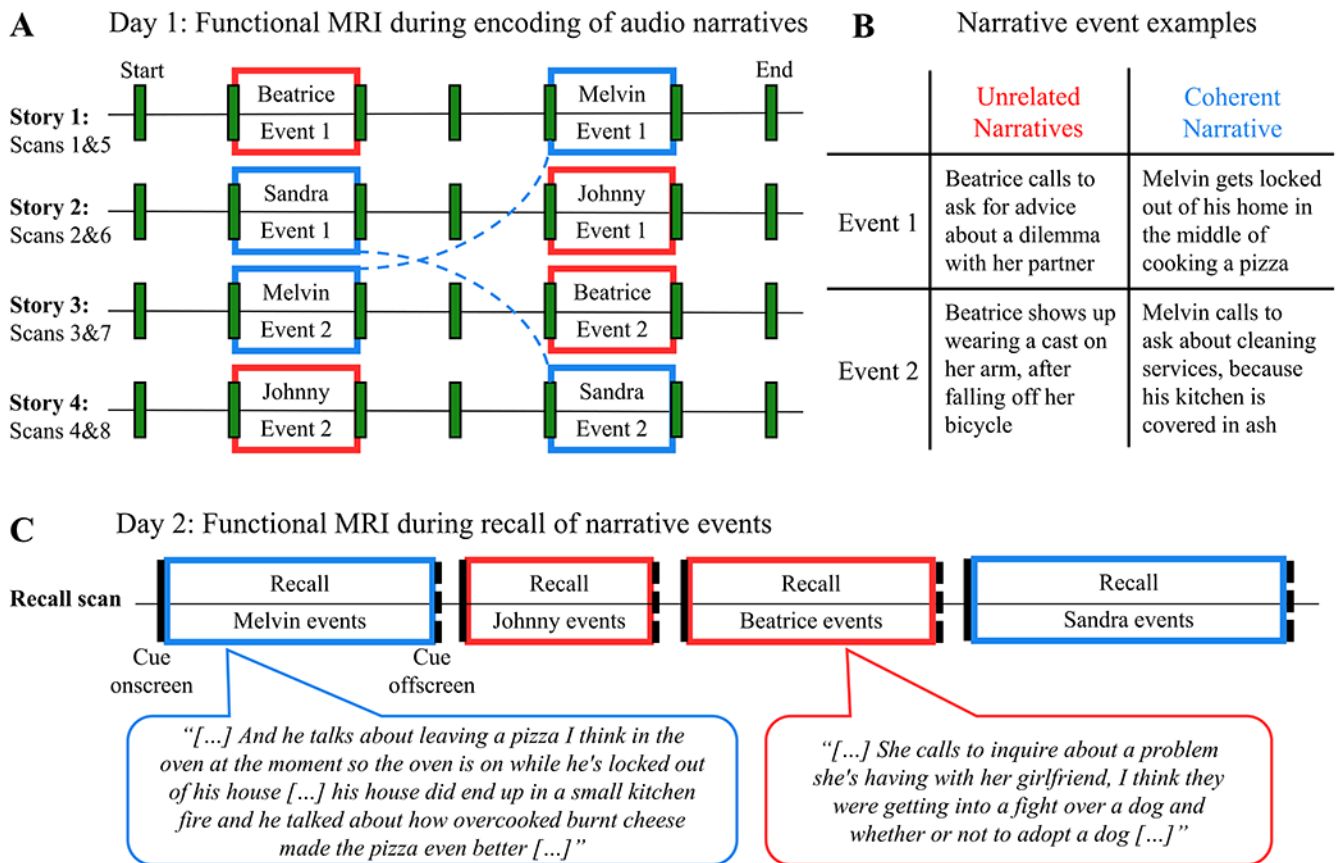
Figure 1: Experimental paradigm.
(A) Narrative stimuli presented during functional MRI: On Day 1, four fictional stories were presented serially during scanning, as audio clips (240s each), which were punctuated by discrete event boundaries (green bars). Events involving recurring characters (Beatrice, 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 STAR 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.
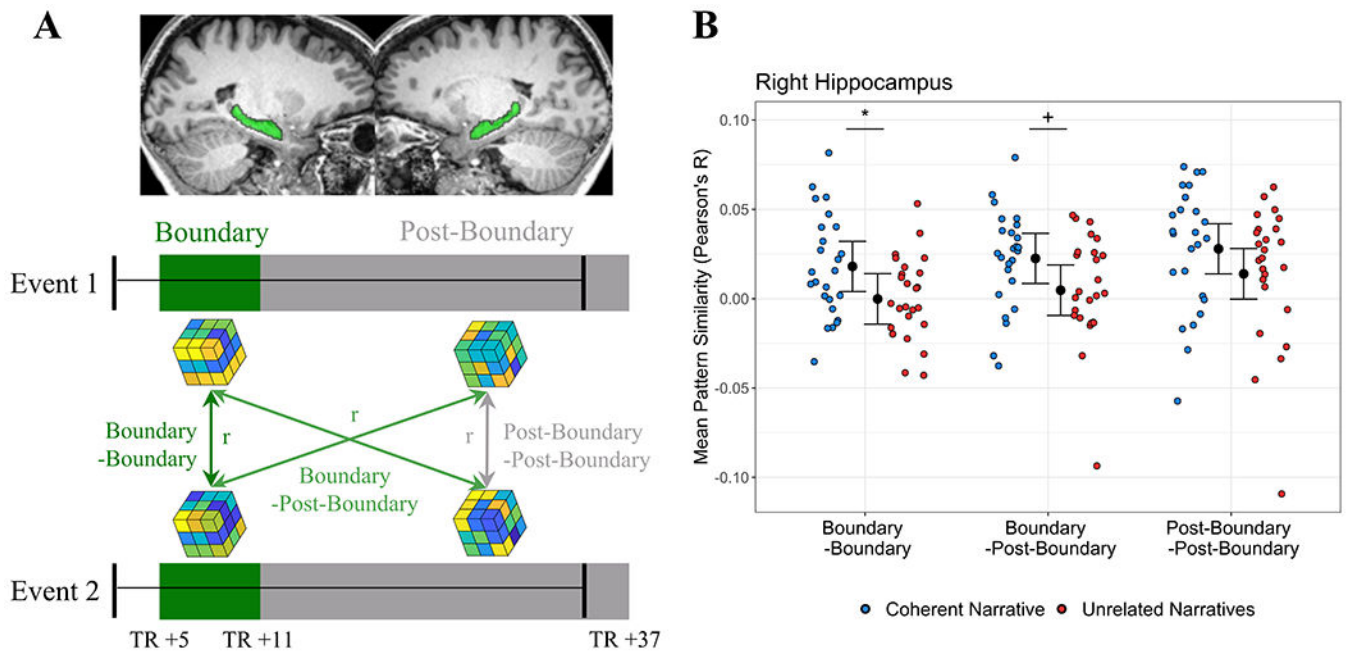
Figure 2: Right hippocampus activity patterns bridge events that form a coherent narrative.
(A) Across-event pattern similarity analysis: within the left and right hippocampus (plotted on an individual subject's structural MRI, in green), the unique spatial pattern of activity ("voxel pattern", represented 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). See also Figure S1. **(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). Note: pattern similarity is on a continuous scale, and zero or negative values simply reflect relatively lower similarity than positive values (i.e. not "no similarity" or "negative similarity"). Key: Blue=Coherent Narratives, Red=Unrelated Narratives, *=$p<0.05$, + = $p<0.10$. See also Figures S2–3.

Figure 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: Mean 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 for model details). Note: pattern similarity is on a continuous scale, and zero or negative values simply reflect relatively lower similarity than positive values (i.e. not "no similarity" or "negative similarity"). **(C)** Event recall predicted by Event 2 pattern reinstatement: Recalled details from Event 1 (circles) and Event 2 (triangles) are plotted for each specific event for each subject, predicted by Event 2 pattern reinstatement. Colored lines represent mixed model estimates (see text for model details). Key: Blue=Coherent Narrative, Red=Unrelated Narratives, +=p<.10 *=p<.05, **=p<.01.