

<https://doi.org/10.1038/s42003-025-08377-1>

Hippocampal systems for event encoding and sequencing during ongoing narrative comprehension

Check for updates

Jiwoong Park ^{1,2,3}, Hayoung Song ⁴ & Won Mok Shim ^{1,2,3}

Narrative comprehension requires encoding individual events and sequencing them into coherent structures. This study demonstrates how the hippocampus contributes to these processes during ongoing narrative processing. Participants viewed a temporally scrambled movie and subsequently recounted its inferred original story during functional magnetic resonance imaging (fMRI) scans. Content encoding and event sequencing abilities were assessed by comparing semantic similarity and temporal order between movie annotations and recall. Functional connectivity between the hippocampus and ventromedial prefrontal cortex (vmPFC) predicted sequencing ability during moments when past and present information are integrated, identified through pre-defined narrative structures and data-driven language models. Conversely, hippocampus-posterior medial cortex (PMC) connectivity predicted content encoding abilities following event boundaries. These findings reveal two distinct hippocampus-centered memory systems in narrative processing: the hippocampus-PMC system for event encoding and the hippocampus-vmPFC system for their integration into coherent narratives.

Natural experiences unfold as interconnected events, each containing organized sensory, semantic, and social information. Understanding such narratives requires encoding detailed information from temporally discrete events while inferring their sequential relationships through temporal and causal connections^{1–6}. To construct coherent narrative representations in real time, the brain needs to integrate each new event with previously encoded context, analogous to solving an evolving puzzle where each piece contributes to an emerging sequential picture.

Neuroimaging studies using naturalistic stimuli, such as movies and verbal stories, have revealed the involvement of the hippocampus and default mode network (DMN) in encoding narrative events, particularly at event boundaries^{7–11}. Neural state dynamics in the posterior medial cortex (PMC) represent the event structure of narratives and show coupling with hippocampal activity following boundaries^{7,12,13}. The PMC exhibits response patterns to each event during movie viewing that closely resemble those during subsequent recall, suggesting its role in capturing narrative content shared across memory encoding and retrieval. Furthermore, hippocampal activity following event boundaries correlates with subsequent event memory^{14,15}.

Beyond encoding individual events, narrative comprehension requires the integration of events separated in time into coherent representations. The hippocampus^{16,17} and medial prefrontal cortex (mPFC)^{18,19} are known

to support these memory integration processes^{20–22}. Lesions to the hippocampus-mPFC system impair the ability to integrate information scattered across time and multiple episodes^{16,19,23}. Importantly, the hippocampus and mPFC interact dynamically to assess whether current events align with or diverge from existing schemas^{20,21}. Reduced hippocampus-mPFC synchronization may facilitate integration of schema-congruent information with prior memories, while increased synchronization may prioritize encoding of novel or incongruent events^{24–27}.

Recent computational studies have proposed optimal strategies for when memory encoding and sequencing should occur during ongoing processing of incoming stimuli^{3,25,28}. Neural network models trained to predict future states suggest that memory encoding occurs selectively at event boundaries to capture complete event representations²⁸. Additionally, when current events require additional contextual information from past experiences to resolve their temporal or causal relationships, the memory system dynamically sequences them with relevant prior events to construct a coherent schema^{28,29}. These computational principles align with recent fMRI studies showing that semantically related events trigger reactivation of past event representations in the hippocampus, facilitating integration of past and present information^{30–32}. These findings indicate two computationally efficient processes for constructing memories with predictable event sequences: encoding complete

¹Center for Neuroscience Imaging Research, Institute for Basic Science (IBS), Suwon, Republic of Korea. ²Department of Biomedical Engineering, Sungkyunkwan University (SKKU), Suwon, Republic of Korea. ³Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University (SKKU), Suwon, Republic of Korea. ⁴Center for Theoretical and Computational Neuroscience, Washington University in St. Louis, St. Louis, MO, USA. e-mail: wonmokshim@skku.edu

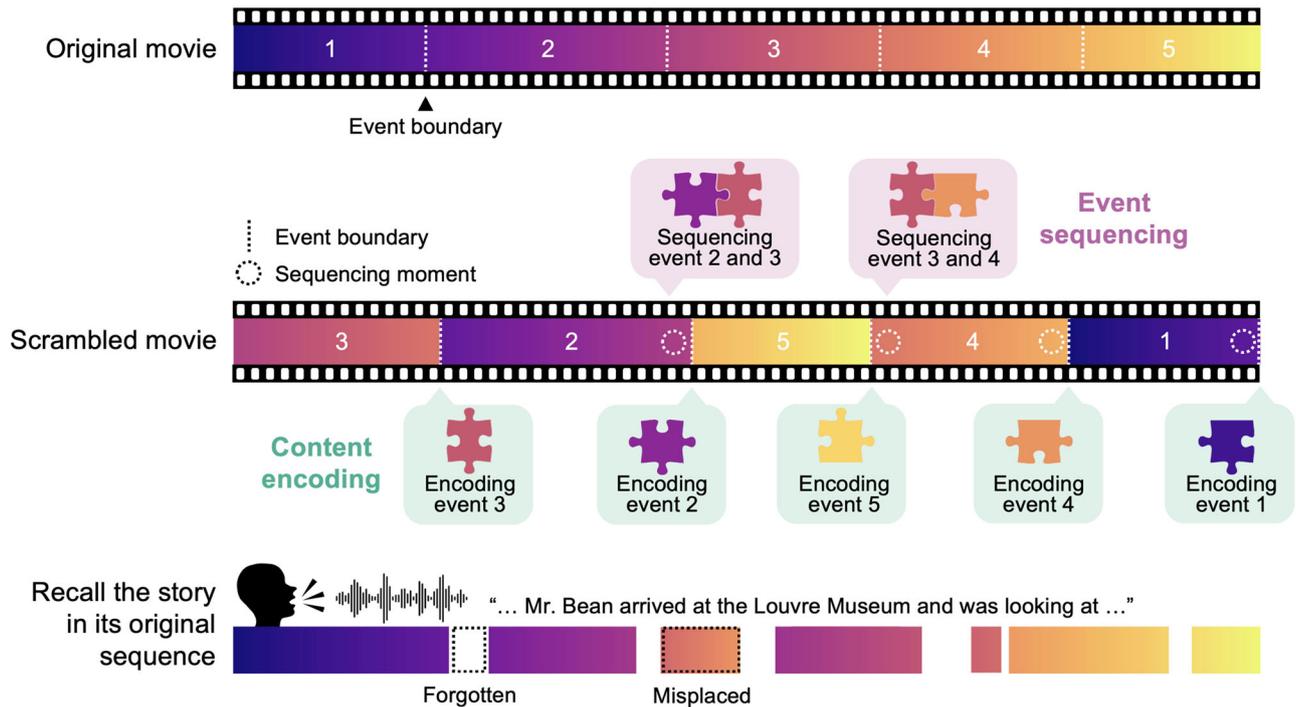


Fig. 1 | Experimental design and memory processes in narrative reconstruction. Participants viewed a temporally scrambled movie (middle) and later recounted the story in its inferred original sequence (bottom). We proposed two distinct memory processes essential for narrative comprehension: encoding individual event content (green) and sequencing events (purple). Event sequencing was hypothesized to occur at

moments where participants could link current events with previously viewed content (dotted circles), while content encoding was expected to follow each event boundary (dotted lines). The figure shows five event segments for illustration; the complete narrative contained seventeen segments. See Methods and Supplementary Fig. 1 for detailed scrambling procedures and sequencing moment specifications.

events at boundaries and sequencing them by reactivating related past events when additional context is needed to resolve uncertainty.

In this study, we investigated how the human brain leverages these distinct memory processes during ongoing narrative comprehension. Specifically, we examined how individual events are encoded and organized into coherent narratives through connections with accumulated context. We hypothesized that two distinct hippocampus-centered memory systems support key aspects of narrative processing: the hippocampus-PMC system encodes event content at boundaries, while the hippocampus-mPFC system sequences events by evaluating their congruency with ongoing narrative context.

To test these hypotheses, we conducted an fMRI experiment where participants viewed a temporally scrambled movie and later recounted the story in its inferred original sequence. Using topic modeling^{33,34}, we assessed participants’ narrative comprehension through free recall, measuring both content and temporal order memory. We then developed predictive models of content encoding and event sequencing abilities based on hippocampal functional connectivity (FC) with DMN regions, particularly the PMC and vmPFC. Lastly, we employed a pre-trained large language model (LLM)³⁵ to identify critical moments when event sequencing likely occurs based on narrative context. The LLM revealed periods of high narrative coherence associated with hippocampus-vmPFC interactions during memory sequencing. This novel data-driven approach demonstrates how LLMs can illuminate dynamic memory processes during naturalistic experiences of narratives without relying on experimentally manipulated narrative structure. Our findings reveal how human memory systems dynamically construct coherent narratives from temporally fragmented information during ongoing processing.

Results

Event encoding and sequencing during ongoing narrative comprehension

During fMRI scans, participants viewed a temporally scrambled movie and subsequently recalled the story in its original chronological sequence

($N = 65$). This temporal scrambling required them to actively infer the original narrative structure⁵. For successful story reconstruction, participants needed to both encode event content and determine temporal relationships between events. Previous studies suggest that encoding of individual events occurs optimally at event boundaries, where complete representations of current events can be captured with minimal interference from adjacent events^{28,36}. Supporting this boundary-specific encoding, prior work has shown selective hippocampal responses at event boundaries^{7,15}. Similarly, our study revealed elevated hippocampal activity following boundaries, enabling us to define moments of event encoding (4 s, from 4 to 7 s after boundaries, Supplementary Fig. 1a, b).

Beyond encoding individual events, constructing coherent narratives requires sequencing current events with relevant past information. The timing of this integration process varied according to the original temporal order of events. When current events preceded previously viewed events, their temporal relationships emerged toward the event’s end (e.g., sequencing events 2 and 3 in Fig. 1). However, when current events followed past events, these relationships became apparent at the event’s beginning (e.g., sequencing events 3 and 4 in Fig. 1). Based on these patterns, we identified expected sequencing moments (4 s at the beginning or end of each corresponding event) for event integration in the scrambled movie (Fig. 1; see Supplementary Fig. 1a for details).

To independently examine content encoding and event sequencing during narrative comprehension, we developed quantitative measures of each process using topic modeling. This approach embedded the semantic content of movie annotations and participants’ recalls in a common space^{33,34} (Fig. 2a), generating a movie-recall similarity matrix that captured semantic correspondences between the original narrative and each recall³³. From this matrix, we derived two independent measures of narrative memory: content scores and ordering scores (Fig. 2b). Content scores quantified semantic information retained from the original narratives, regardless of sequence^{12,32,33}, while ordering scores measured participants’

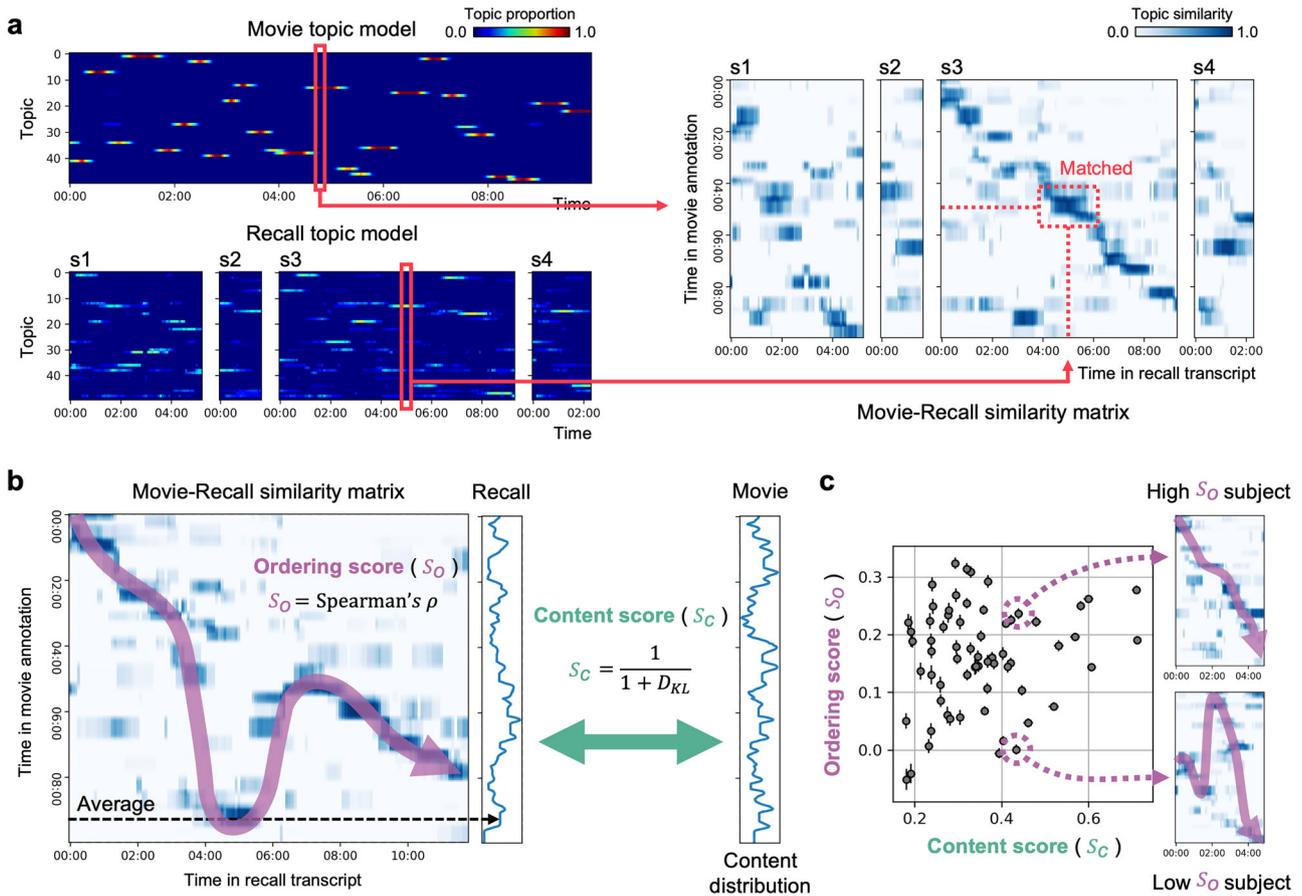


Fig. 2 | Assessment of content and sequence memory in ongoing narrative comprehension. **a** Topic modeling embedded movie annotations and participant recalls in a common topic space. Left panel: topic vectors of the movie (top) and sample recalls from four participants (bottom, s1–s4). Right panel: movie-recall similarity matrices computed from moment-by-moment cosine similarity between movie and recall topic vectors. The diagonal pattern emerges when narrative content is recalled in the correct sequence. **b** Analysis of two memory metrics from a movie-recall similarity matrix. Content scores quantified semantic information retained from the movie, regardless of sequence, based on similarity between movie and recall

content distributions. Ordering scores measured temporal sequence accuracy using rank correlation between movie and recall event orders. **c** Correlation between ordering and content scores. Each circle represents a participant. High ordering scores produced clear diagonal patterns in the movie-recall similarity matrix (top right). Some participants showed low content scores but high ordering scores, indicating accurate temporal ordering despite partial recall, but no participants showed high content scores with low ordering scores. Error bars indicate \pm one standard error per participant across scores, calculated from 50 topic model iterations with different random seeds. See the Methods and Supplementary Fig. 2 for details.

ability to reconstruct the chronological sequence of the original story through rank correlation of matched indices in the similarity matrix. These ordering scores reflect participants' ability to accurately determine event sequences by assessing how each event fits within the evolving narrative context. The independence of these measures was confirmed by their lack of significant correlation (Fig. 2c, $r = 0.173, p = 0.166$), indicating that they capture distinct aspects of narrative comprehension.

Hippocampo-cortical FC predicts event encoding and sequencing abilities

To examine the neural basis of individuals' event encoding and sequencing abilities in functional brain networks, we developed connectome-based models predicting memory performance from FC patterns^{37–39}. Previous studies have shown the involvement of the hippocampus and mPFC in memory sequencing^{20–23}, and the hippocampus and PMC in post-boundary memory encoding^{7,14,15,28}. Based on these findings, we hypothesized that hippocampus-vmPFC connectivity would predict ordering scores during sequencing moments, while hippocampus-PMC connectivity would predict content scores during post-event boundary periods. To test these hypotheses, we constructed a hippocampo-cortical FC model using the hippocampus as a seed and compared its predictive performance to a cortico-

cortical FC model excluding the hippocampus. We evaluated these models across three time periods: 1) expected sequencing moments when current and past events were integrated (4 s, Supplementary Fig. 1a), 2) post-event boundary moments, identified by peak hippocampal activity following event boundaries (4 s, from 4 to 7 s after boundaries, Supplementary Fig. 1a, b, and 3) all movie time points for comparison.

Although the hippocampo-cortical FC model included far fewer initial edges than the cortico-cortical FC model (200 vs. 19,900), a much higher proportion of its edges were significantly correlated with memory performance during training (~ 10 vs. 4%, all $ps < 0.001$), indicating the robust behavioral relevance of hippocampal connections. The hippocampo-cortical FC model successfully predicted both memory measures: using sequencing moments, it showed significant predictive performance for ordering scores, (Fig. 3a, cross-validated $r = 0.314, p = 0.023$, one-tailed for this and all subsequent predictive modeling analyses), while using post-event boundary moments, it showed significant predictive performance for content scores (Fig. 3a, Post-event boundary moments: cross-validated $r = 0.299, p = 0.034$; All time points: cross-validated $r = 0.247, p = 0.067$). In contrast, the cortico-cortical FC model did not yield significant prediction performance for either memory score (Fig. 3a), even when the number of edges was matched to the hippocampal model (Supplementary Fig. 3a).

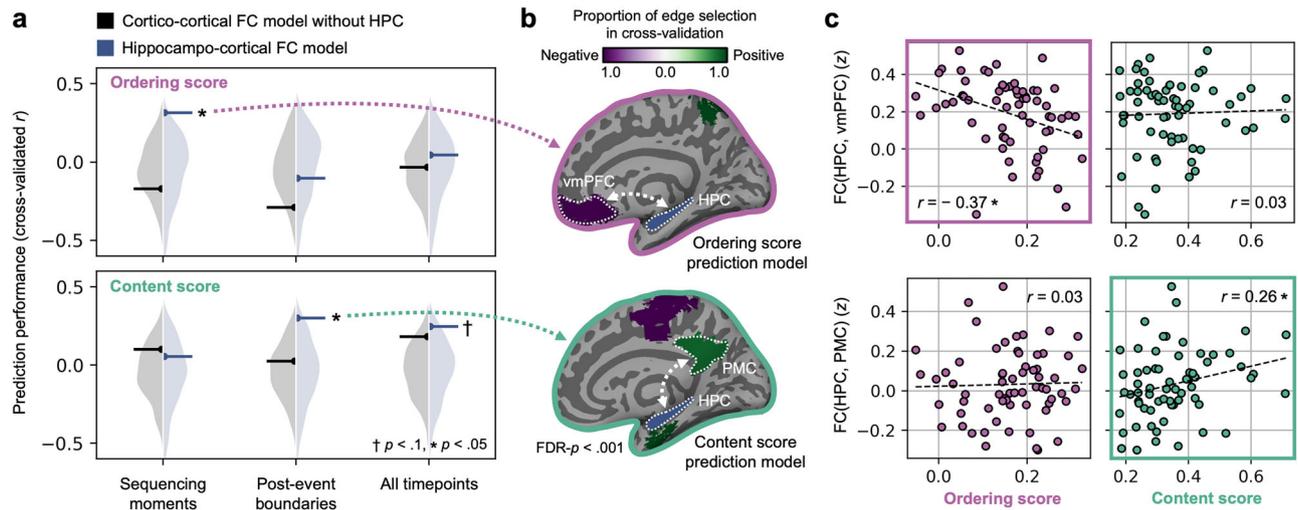


Fig. 3 | Predictive modeling of content and ordering scores using functional connectivity patterns. **a** Hippocampo-cortical FC models (blue) successfully predicted both memory scores, whereas cortical FC models excluding the hippocampus (black) did not. The model predicted ordering scores during sequencing moments and content scores during post-event boundaries, with marginal prediction when using all time points. Statistical significance was determined by comparing actual model performance against a null distribution generated by randomly permuting narrative memory scores across participants ($n = 1000$, one-tailed). Results from a control analysis matching the number of edges between the hippocampo-cortical and cortico-cortical FC models are shown in Supplementary Fig. 3a. **b** FC edges consistently selected for predicting each memory score were identified. For ordering

score prediction, the FC between the hippocampus and vmPFC was reliably selected (top), while for content score prediction, the FC between the hippocampus and PMC was consistently selected (bottom). Additional selected ROIs are presented in Supplementary Fig. 3b. **c** Hippocampus-vmPFC connectivity negatively correlated with ordering scores during sequencing moments (top left), while hippocampus-PMC connectivity positively correlated with content scores during post-event boundaries (bottom right). Neither hippocampus-PMC FC correlated with ordering scores (top right), nor did hippocampus-vmPFC FC correlate with content scores (bottom left), supporting distinct hippocampal-cortical systems for sequence and content memory.

These findings support our hypothesis about distinct roles of the hippocampus in narrative memory: integrating past and present events for sequence reconstruction and encoding event content at boundaries (Fig. 3a). The results remained consistent across multiple cortical parcellations with varying numbers of ROIs, confirming their robustness (Supplementary Fig. 4).

To further evaluate the impact of the hippocampus on narrative memory prediction, we first analyzed how the number of ROIs included in the model affects performance. We hypothesized that if the hippocampus-centered system is essential, incorporating additional irrelevant cortical ROIs would reduce performance, potentially due to overfitting or underfitting. Indeed, models including the hippocampus showed significantly decreased predictive performance with more cortical ROIs (Supplementary Fig. 6, ordering score: $r = -0.921, p < 0.001$; content score: $r = -0.902, p < 0.001$), highlighting the importance of functionally relevant edges for accurate prediction. In contrast, cortico-cortical FC models without the hippocampus consistently underperformed in predicting both memory scores, regardless of the number of ROIs (Supplementary Fig. 6, all $ps < 0.001$). To validate these findings, we compared the prediction performance of FC models using different seed regions. The hippocampus-centered FC models outperformed most models based on other cortical regions in predicting both memory scores (Supplementary Fig. 7, all $ps < 0.05$). Notably, the hippocampus-based FC model achieved the highest accuracy in predicting ordering scores, providing further evidence for its critical role in integrating narrative information over time.

Next, to identify critical hippocampal functional connections for predicting narrative memory, we examined the consistently selected edges of the hippocampo-cortical FC model across cross-validation folds^{40,41} and cortical parcellations. The model revealed two key connections: the FC between the hippocampus and vmPFC significantly contributed to predicting ordering scores during sequencing moments, while the FC between the hippocampus and PMC was crucial for predicting content scores during post-event boundary moments (Fig. 3b, see Supplementary Fig. 3b for an extended view). Statistical analysis confirmed the specificity of these

connections. The hippocampus-vmPFC edge was selected exclusively for ordering score prediction (100% of CV folds) and not for content score prediction (0%; $\chi^2(1, N = 65) = 130.0, p < 0.001$). Conversely, the hippocampus-PMC edge was predominantly selected for content score prediction (87.6% of CV folds), and not for ordering score prediction (0%; $\chi^2(1, N = 65) = 101.5, p < 0.001$). These selective contributions remained consistent across different cortical parcellations (Supplementary Fig. 5), supporting distinct roles of hippocampus-vmPFC connections in event sequencing and hippocampus-PMC connections in content encoding.

Finally, we observed distinct patterns in the relationships between FC and each memory score. Hippocampus-vmPFC connectivity negatively correlated with ordering scores during sequencing moments (Fig. 3c top left, $r = -0.369, p = 0.002$), while hippocampus-PMC connectivity positively correlated with content scores during post-event boundaries (Fig. 3c bottom right, $r = 0.258, p = 0.037$). This functional decoupling between the hippocampus and vmPFC in participants with high ordering scores aligns with previous studies showing enhanced hippocampus-vmPFC interactions during encoding of schema-incongruent information^{24–27}. In our task, constructing accurate narrative sequences requires recognizing relationships with existing event schemas rather than encoding novel information. Thus, higher ordering scores may reflect successful integration of current events into previously structured schemas through hippocampus-vmPFC desynchronization, rather than treating them as novel, schema-incongruent events. In contrast, the positive correlations between hippocampus-PMC connectivity and content scores suggest that enhanced hippocampus-PMC coupling supports event encoding at boundaries^{7,14,15,28}. Supporting the specificity of hippocampal-cortical systems for each memory process, hippocampus-vmPFC FC showed no correlation with content scores (Fig. 3c, $r = 0.036, p = 0.77$), nor did hippocampus-PMC FC with ordering scores (Fig. 3c, $r = 0.029, p = 0.812$). Statistical comparison using Steiger's Z-test⁴² further revealed stronger hippocampus-vmPFC FC correlations with ordering than content scores ($z(62) = 2.614, p = 0.008$), while hippocampus-PMC FC correlations showed no significant difference ($z(62) = 1.436, p = 0.15$).

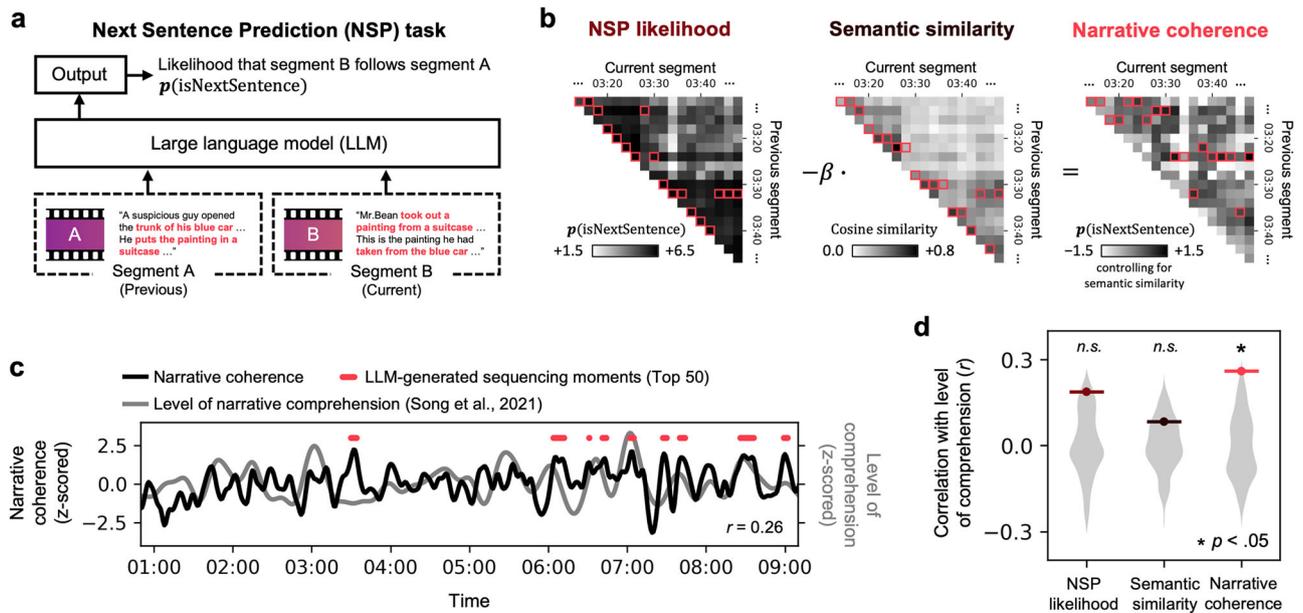


Fig. 4 | Identification of event sequencing moments using LLMs based on moment-by-moment narrative coherence. **a** An LLM pre-trained on the NSP task assessed logical and causal relationships between narrative segments by measuring the likelihood that one sentence follows another. **b** Initial pairwise NSP likelihoods showed bias toward adjacent segments due to shared semantics. After controlling for semantic similarities, we identified narrative relationships between distant events, enabling detection of sequencing moments (red box) **c** LLM-derived narrative

coherence correlated significantly with group-level narrative comprehension during movie viewing, as previously reported by participants⁵. High-coherence moments (red dots, top 50) were identified as likely sequencing moments. **d** LLM-generated narrative coherence showed robust correlation with human-rated narrative comprehension, while NSP likelihood without semantic control and semantic similarity alone showed no correlation. Null distributions, generated from phase-randomized time series, are indicated in gray.

LLM reveal sequencing moments in narrative memory formation

While our previous analysis identified sequencing moments using pre-defined narrative structures, recent research suggests that the human brain reactivates related memories based on perceived narrative coherence during ongoing comprehension^{30,31}. Approaches based solely on pre-defined narrative structures may fail to capture a full range of contextually relevant moments of coherence during naturally unfolding narrative experiences. To address this, we utilized pre-trained LLMs to identify sequencing moments in a data-driven manner.

Using bidirectional encoder representations from transformers (BERT), we quantified the narrative coherence of each moment within the movie. We first fed detailed movie annotations into BERT as input and computed next sentence prediction (NSP) likelihoods³⁵—the likelihood that one sentence logically and temporally follows another—between each 2-second segment and its preceding segments (Fig. 4a). Each segment’s highest NSP likelihood among its preceding segments (Fig. 4b) served as its narrative coherence measure, indicating its degree of connection with preceding segments. To isolate narrative connections from semantic effects, we regressed out semantic similarity using the universal sentence encoder (USE)^{5,43}. The resulting LLM-generated narrative coherence index significantly correlated with human-rated moment-by-moment narrative comprehension from our previous study⁵ (Fig. 4c, $r = 0.259, p = 0.016$), while raw NSP likelihood and semantic similarity alone showed no correlation with the reported narrative comprehension levels (Fig. 4d, NSP likelihood: $r = 0.187, p = 0.06$; semantic similarity: $r = 0.08, p = 0.212$). These findings suggest that LLM-generated coherence effectively captures the cognitive dynamics of narrative processing, highlighting the importance of sequential coherence in ongoing comprehension. Since high narrative understanding moments reported by humans involve integrating past and present events based on their causal relationships⁵, we identified high-coherence moments generated by the LLM as likely moments of event sequencing within narrative context (LLM-generated sequencing moments, Fig. 4c). For a more detailed comparison of the temporal characteristics of LLM-

generated and pre-defined sequencing moments, see Supplementary Fig. 8a.

Our hippocampo-cortical predictive model based on the LLM-generated sequencing moments demonstrated significant prediction accuracy for ordering scores, comparable to the original model based on pre-defined sequencing pairs (64 time points from 16 pairs with 4 s each) from the scrambled narrative structure (Fig. 5a, model using top 50 moments, $r = 0.305, p = 0.034$). However, models using either moments of high raw NSP likelihood, or moments of high semantic similarity alone, failed to predict ordering scores (NSP likelihood: $r = -0.026, p = 0.51$; semantic similarity: $r = 0.036, p = 0.42$). Notably, model performance declined as the number of sequencing moments used for training is increased, indicating that event sequencing likely occurs primarily at specific moments of high narrative coherence (Fig. 5a). Consistent with previous results, hippocampus-vmPFC connectivity remained crucial for predicting ordering scores (Fig. 5b, c, $r = -0.301, p = 0.014$), highlighting its role in integrating previously acquired information.

Our model, based on LLM-generated sequencing moments, both validated and extended the findings from the model based on pre-defined sequencing moments. While pre-defined sequencing moments exhibited higher narrative coherence than other movie segments ($t_{(608)} = 2.91, p = 0.003$, Supplementary Fig. 8b), the LLM-based approach identified additional high coherence moments not captured by the pre-defined structure. With ~15% overlap between the two sets of moments (Supplementary Fig. 8b), these findings demonstrate how LLMs can reveal previously unidentified moments critical for event sequencing. These converging results across independent methods further validate the role of hippocampus-vmPFC connectivity in event sequencing.

The hippocampus-vmPFC FC during LLM-generated sequencing moments showed selective prediction for predicting ordering scores, being selected in every cross-validation fold (100% for ordering scores and never selected (0%) for content scores across models with varying numbers of moments (all $ps < 0.001$, Supplementary Fig. 8c), consistent with our findings using pre-defined narrative structures. While hippocampus-

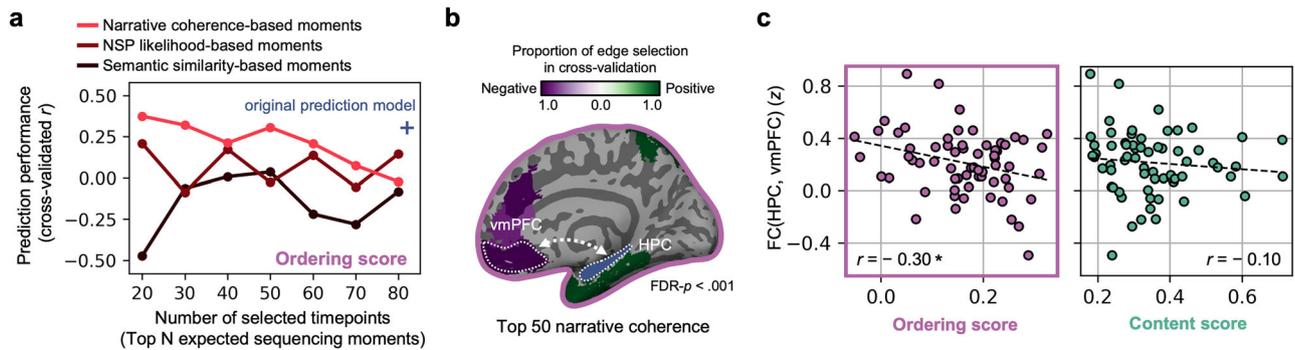


Fig. 5 | Predictive performance of the hippocampo-cortical model using LLM-generated sequencing moments. **a** The hippocampo-cortical model using LLM-generated sequencing moments (red) achieved prediction accuracy comparable to the original model based on pre-defined movie structure (blue cross). In contrast, models using either high NSP likelihood without semantic control (brown) or high semantic similarities alone (black) failed to predict ordering scores. **b** The predictive model using LLM-generated sequencing moments consistently selected

hippocampus-vmPFC FC as a significant predictor of ordering scores.

c Hippocampus-vmPFC connectivity showed a significant negative correlation with ordering scores during LLM-generated sequencing moments (left), but no correlation with content scores (right), confirming its specific role in event sequencing (as in Fig. 3c). See Supplementary Fig. 5 for analyses with varying numbers of top moments.

vmPFC connectivity significantly correlated with ordering but not with content scores (Fig. 5c, ordering score: $r = -0.301$, $p = 0.014$; content score: $r = -0.104$, $p = 0.409$), Steiger's Z-test revealed no significant differences between these correlations. The results remained robust across different cortical parcellations, including both prediction performance and the importance of vmPFC connections (Supplementary Figs. 4 and 5). These findings highlight the significance of narrative coherence in sequencing related events and the value of our data-driven approach for identifying critical event integration moments during ongoing comprehension.

Discussion

This study demonstrated how distinct hippocampus and DMN memory systems support content encoding and temporal sequencing during ongoing narrative comprehension. Using natural language processing techniques, we developed quantitative measures that selectively captured the processes of encoding event content and reconstructing their temporal sequence from participants' free recall data. Our findings revealed that these distinct memory processes rely on separate hippocampus-centered systems operating at specific moments during narrative processing. The hippocampus-vmPFC FC during expected event sequencing moments predicted participants' ability to reconstruct the temporal order of narratives, while hippocampus-PMC connectivity at post-event boundaries predicted content memory. Furthermore, our LLM-based approach identified periods of high narrative coherence where hippocampus-vmPFC interactions support event sequencing.

Using FC-based predictive modeling^{37–39} across different time points and brain regions, we identified key neural features that predict memory performance during narrative processing. These findings align with and extend previous lesion^{19,23}, electrophysiology^{10,27}, and neuroimaging studies^{7,13,30,31}. First, recent work has demonstrated that the hippocampus and DMN regions are engaged during narrative processing, particularly at event boundaries^{12,13} and during encounters with semantically related past events^{30,31}. In line with these findings, our study found that hippocampus-PMC connectivity at event boundaries was positively correlated with participants' ability to encode detailed narrative content. These findings not only align with recent human neuroimaging studies showing content encoding around event boundaries but also deepen our understanding of individual differences in narrative memory encoding. Second, the negative correlation between hippocampus-vmPFC connectivity and narrative sequencing ability (Fig. 5c) extends previous human and animal studies implicating these regions in memory sequencing. For instance, hippocampal or mPFC damage disrupted memory sequencing of odor experiences in rodents^{19,23}, and hippocampal-mPFC desynchronization has been observed when animals explored contextually coherent objects while

discriminating between familiar and novel events²⁷. Human neuroimaging studies have also reported that the FC between the hippocampus and vmPFC was strengthened when processing novel information inconsistent with prior event schemas^{24,26}. This suggests that elevated hippocampus-vmPFC connectivity during predicted sequencing moments in our study may reflect difficulty integrating current events into existing narrative schemas, resulting in lower ordering scores.

Despite its crucial importance, demonstrating memory sequencing during ongoing narrative processing presents two key challenges: limited sequencing demands in typical linear narratives, and the difficulty of identifying precise moments when sequencing processes occur during continuous experiences. To address these challenges, we experimentally induced enhanced demands for event sequencing⁵ by presenting a temporally scrambled movie. This scrambled narrative structure enabled us to infer likely moments when current events could be linked to prior ones as part of a coherent narrative. Our results demonstrate that neural signals centered on the hippocampus and vmPFC at these moments are predictive of participants' ability to reconstruct narrative structure through event sequencing. While natural experiences rarely contain such salient temporal discontinuities, our LLM-based approach provides a data-driven method for identifying critical moments of narrative integration, when ongoing narrative events are processed in the context of related past information. This approach offers broad utility for analyzing existing datasets without requiring additional experimental manipulations or a priori assumptions, though further validation across diverse narrative contexts is needed.

One limitation of the present study is that we did not acquire direct measurements of event sequencing performance, as we intentionally avoided including an explicit sequencing task during narrative viewing to preserve the naturalistic experience of ongoing narrative comprehension. Instead, we inferred the likely occurrence of sequencing based on the scrambled narrative structure. The event sequencing process likely involves multiple cognitive operations, such as retrieval⁴⁴, reactivation³¹, and integration³⁰, which were not independently dissociated or explicitly measured in our study. Future research combining neuroimaging with concurrent behavioral probes during narrative viewing or computational modeling approaches may allow for a more precise characterization of these component processes and their neural correlates during narrative comprehension. Also, the fixed scrambling order and the limited number of narrative events in this study potentially constrain the generalizability of our findings. While these methodological limitations are common challenges associated with using naturalistic stimuli, future research could systematically address these constraints by using multiple randomized scrambling schemes or utilizing more complex and longer narratives with diverse

temporal structures. Such extensions would help establish the robustness of the observed findings across varied narrative contexts.

Natural experiences unfold as open-ended and uncertain sequences requiring continuous updating and integration of newly encountered information into coherent representations. To understand how the brain solves this sequential puzzle, it is essential to examine how narrative models are constructed to weave our experiences together. By combining novel behavioral metrics, FC-based predictive modeling, and LLM-based analyses, we have demonstrated the crucial role of the hippocampus and DMN in both encoding and sequencing events to comprehend ongoing narratives. These findings illuminate how distinct memory systems dynamically support the integration of complex, naturalistic information into meaningful narratives, providing insights into the neural mechanisms underlying real-world memory processes.

Methods

Participants

A total of 71 participants (26 females, mean age = 22.78 ± 2.28 years) were recruited for the study. Six participants were excluded from the data analysis: one due to a global artifact in the functional images and five due to excessive head motion during the experiment, with a framewise displacement (FD) exceeding 0.5 mm for more than 5% of the total images acquired. All participants received monetary compensation for their participation and provided informed consent before the experiment, which had been approved by the Institutional Review Board of Sungkyunkwan University. All ethical regulations relevant to human research participants were followed. It was confirmed that none of the participants had viewed the movie used in the study prior to their participation.

Stimulus

A 10 min audiovisual animated movie, “*Mr. Bean: The Animated Series, Art Thief*” (season 2, episode 13; 2003, Fehrenbach), was used for the fMRI experiment. This movie was composed of 17 events, primarily defined by the director’s cut, each lasting ~36 s. These events were temporally scrambled in an identical pseudorandom order across all participants (Supplementary Fig. 1a). The chronological event order of the movie was scrambled to engage participants in inferring the narrative structure of the original story while viewing the scrambled movie. This scrambling resulted in twelve event boundaries, each preceded or followed by an event paired with its related event. In these pairs, one event occurring either before (e.g., sequencing events 2 and 3 in Fig. 1) or after the other event (e.g., sequencing events 3 and 4 in Fig. 1) in its original sequence, appeared earlier in the scrambled order (see Supplementary Fig. 1a for details). The movie stimulus was presented using the Psychophysics Toolbox 3⁴⁵ and an MR-compatible video (PRO-Pixx projector, VPixx Technologies) and audio (OptoActive ANC headphones, Opto-acoustics) system. To prevent abrupt transitions in audiovisual features and provide a buffer period, a 30 s video of a nature scene, such as a waterfall, with moderate audiovisual features similar to those of the movie, was inserted before the start of the movie (for a detailed description of the movie stimulus, see our previous study⁵).

Experimental procedure

The fMRI session comprised four consecutive functional runs and one anatomical run. In the first functional run, participants viewed a 10 min scrambled movie, followed by a free recall task where they recounted the story they believed to be the most likely original version. During this task, participants were encouraged to construct a coherent story based on their recollections, even if they could not perfectly reconstruct the original story and were given unlimited time for recall. Participants verbally indicated the conclusion of their recall by stating, “I am finished,” at which point the corresponding run was terminated. In the subsequent run, participants viewed the intact version of the same movie to fully comprehend the original story. Following this, they viewed the scrambled movie again, followed by a second recall task. The last run was a resting-state fMRI run without a task. For the present study, only the fMRI data acquired during the first run,

involving the initial viewing of the scrambled movie and subsequent free recall, were analyzed. The duration of the functional run was contingent upon the length of participants’ free recall (mean = 3.59 min, SD = 2.03 min). Prior to the fMRI experiment, participants completed a practice session involving a different scrambled movie, “*Oggy and the Cockroaches: The Animated Series, Panic Room*” (season 4, episode 8; 2013, Jean-Marie), and recounted its story in the inferred original sequence. The free recall data from the practice session were qualitatively evaluated to ensure that participants understood the task instructions and recalled the story with an appropriate level of detail.

Measure of narrative memory: content and ordering scores

Four independent annotators generated detailed annotations of the original movie content at two-second intervals. Free recall data were transcribed at 5 s intervals. Merging sentences from all annotations and transcripts, the text data were tokenized using the KoNLPy Python package, designed for Korean natural language processing⁴⁶. Only nouns and verbs were extracted from the annotations and transcripts through part-of-speech tagging. A user-defined word dictionary, containing 67 words, was utilized to match synonymous words between the recall transcripts and movie annotations. For example, “wrench” and “spanner,” and “artworks” and “paintings” were considered equivalent. Next, sentences were aggregated based on a specified window size parameter and converted into sentence vectors using the bag-of-words model and the scikit-learn Python package⁴⁷. This preprocessed dataset was used for topic modeling to assess the similarity between the movie annotations and participants’ free recalls, enabling the evaluation of narrative memory performance³³.

We initially trained a topic model using the annotated movie data. The latent Dirichlet allocation model was employed to discern the abstract semantic content within the movie annotation. Both the annotated movie data and participants’ recall were then mapped onto semantic vectors in the topic space using this trained model. Similarities between these vectors were determined based on cosine similarity. Unlike prior studies that used events identified through a hidden Markov model³³, our study calculated the movie-recall similarity for each sentence, since the event structure in free recall of a scrambled movie was less discernible than that of intact movies.

Using the topic model, we quantified two types of narrative memory scores: ordering scores evaluated how accurately participants reconstructed the temporal sequence of events during recall, while content scores measured the degree of semantic overlap between recalled narrative elements and the movie annotation. To compute ordering scores, we converted the movie-recall cosine similarity matrix into a binary matrix by applying a threshold, indicating the matched indices between the sentences in recall and the movie annotation. Using this binarized matrix, we computed Spearman’s rank correlation to compare the recall order and the original chronological order to assess the similarity in temporal sequencing of matched sentences. For content scores, we generated a recall content distribution by averaging topic similarities across all recalled elements for each time segment of the original movie in the movie-recall similarity matrix, independent of recall order. This distribution represents the amount of content recalled from the movie across time. We then derived a movie content distribution from the movie-movie similarity matrix. Content scores, which reflect the similarity between these two distributions, were obtained as the inverse of the Kullback-Leibler divergence, $1/(1 + D_{KL})$, ranging from 0 to 1 (Supplementary Fig. 2a).

To optimize the hyperparameters for the topic model and validate our narrative memory scores it produced, we recruited two independent human raters to manually evaluate participants’ free recall data. These raters segmented the original annotation into distinct events based on perceived narrative transitions, such as character changes, shifts in time or location, and changes in topic. This segmentation resulted in 46 and 25 events for the two raters, respectively. The raters then matched events from each participant’s recall with corresponding events in the annotation, arranging the recalled events chronologically. Ordering scores were computed using Spearman’s rank correlation based on the recalled events identified by the

raters. Content scores were determined by computing the ratio of recalled events identified by the raters against the total number of events in the annotation. The ordering and content scores from both raters were averaged to provide a comparative human-rated measure against the model-derived scores. To optimize the topic model, we tested a range of hyperparameters including the number of topics (from 10 to 80), the window size for both movie annotation (0–10) and free recall (0–6), and the threshold hyperparameter for binarizing the movie-recall similarity matrix (0.3–0.9). The optimal settings selected were 80 topics, a window size of 0 for both movie annotation and free recall, and a threshold of 0.3 for the similarity matrix. These parameters were chosen to maximize correlations between scores generated by the topic model and those assessed by human raters (Supplementary Fig. 2c). Notably, even across diverse hyperparameter settings, model-derived scores closely matched human-rated scores (Supplementary Fig. 2d), validating the topic model's effectiveness in accurately quantifying narrative memory performance. Building on this validation, we further evaluated the robustness of these narrative memory scores by assessing their predictive power using FC-based modeling across the full range of tested hyperparameter configurations. Most parameter combinations yielded positive prediction performance, with several even outperforming the selected model (Supplementary Fig. 2e). This comprehensive evaluation demonstrates that our narrative memory scores are stable across a broad set of modeling choices and are not simply the result of narrowly optimized parameters.

Data acquisition

The data were acquired using a 3 T Siemens Prisma MRI scanner with a 64-channel head coil located at Sungkyunkwan University and the Institute of Basic Science, Center for Neuroscience Imaging Research. T2*-weighted functional images sensitive to blood oxygenation level-dependent contrast were obtained using an echo-planar imaging sequence (voxel size: 3 mm isotropic; TR: 1000 ms; TE: 30 ms; FOV: 240 x 240 mm; 48 slices covering the whole brain; flip angle: 90°; multi-band factor: 3). High-resolution anatomical data were also acquired using a T1-weighted magnetization-prepared rapid gradient echo sequence (voxel size: 1 mm isotropic; TR: 2200 ms; TE: 2.44 ms; FOV: 256 x 256 mm; 256 slices; flip angle: 8°).

Preprocessing

The functional and anatomical data were preprocessed using the fMRIprep⁴⁸ pipeline. The anatomical data underwent intensity non-uniformity correction, skull-stripping, brain segmentation, and surface reconstruction. All functional data were motion corrected and registered to the MNI152 standard space for further analyses. Additional denoising was performed using noise components obtained during preprocessing⁴⁹, including six motion parameters (x, y, z, roll, pitch, and yaw), their derivatives, global signals extracted from whole-brain masks, FD, and six physiological regressors extracted from cerebrospinal fluid and white matter provided by aCompCor⁵⁰. Following denoising, spatial smoothing with a full-width half-maximum of 5 mm and intensity normalization were applied to the functional images using AFNI⁵¹.

FC-based predictive modeling

To predict participants' narrative memory scores, we employed FC-based predictive modeling^{37–39}. We identified the hippocampus from the Brainnetome atlas⁵² and cortical ROIs from the Schaefer atlas⁵³, which offers parcellations with varying numbers of cortical ROIs, to assess the reliability of prediction performance across different parcellations used for training the model. After averaging each ROI's time course across its voxels, we computed the FC using Pearson's correlation, and then transformed it to a Fisher's *z*-value.

Connectivity patterns were calculated during three distinct periods: expected sequencing moments, post-event boundaries, and all time points. For all time points, the full fMRI data of 610 TRs, including the 10 min movie (600 TRs) and ten additional TRs from a blank screen, were used. Among these time points, post-event boundary moments were identified

based on hippocampal activity^{7,14,15} around event boundaries (21 time points, 10 s before and after boundaries). A *t* test compared activity at each time point to baseline, with false discovery rate (FDR) correction applied ($q < 0.001$). Four seconds showing significantly elevated hippocampal activity were selected for each of the 17 scrambled events, resulting in 68 time points (Supplementary Fig. 1b). Sequencing moments were defined by aligning four seconds per sequencing pair with the temporal structure of post-event boundaries. These moments were selected either three seconds before or after event boundaries, depending on the relative position of current events within the original narrative, yielding a total of 64 time points (Supplementary Fig. 1a).

To assess the model's prediction performance, we implemented a leave-one-subject-out cross-validation method. During the training phase, FC edges significantly correlated with the narrative memory score ($p < 0.05$) were identified from the training set. The Fisher's *z*-transformed correlation coefficients of these selected edges were then summed separately for both positive and negative correlations, establishing two predictors³⁷. Using these predictors, a linear regression model was created to estimate the narrative memory score, and this model was subsequently applied to the left-out subject for prediction. Finally, the model's cross-validated performance was determined using Pearson's correlation between the actual and predicted narrative memory scores. The statistical significance of the model's performance was assessed using a permutation test, where the narrative memory scores were shuffled across participants to create a null distribution with 1000 iterations. *P*-values were calculated based on the fraction of sampled permutations that were greater than or equal to the actual prediction performance.

To evaluate the predictive role of hippocampal connectivity, we constructed a hippocampo-cortical FC model using the hippocampus as a seed (200 edges) and compared its performance to a whole-brain cortico-cortical FC model excluding the hippocampus (19,900 edges). To control for differences in feature dimensionality, we also conducted a control analysis in which the number of edges was matched between the two models. For each training fold, we recorded the number of functional edges selected in the hippocampo-cortical model and selected the same number of top-ranking cortico-cortical edges based on their absolute correlation with memory performance (Supplementary Fig. 3a).

To determine whether specific FC edges were consistently selected during predictive model building at a rate significantly higher than expected by chance, we conducted a one-tailed binomial test for each edge. For each model and participant, we recorded the number of cross-validation folds in which a given edge was selected during feature selection. Assuming a null hypothesis where edges are selected randomly with a probability of 0.05 (reflecting the nominal *p* value threshold for initial selection), we computed the probability of observing the actual or greater selection count using the binomial distribution. The resulting *p*-values were corrected for multiple comparisons across all 200 edges using FDR correction. Edges that survived FDR correction at $q < 0.001$ were considered significantly and consistently selected, and only these edges were visualized in the brain figures.

Hippocampal contribution to FC-based predictive model

To evaluate the role of the hippocampus in predicting narrative memory performance, we utilized two models: a hippocampo-cortical FC model and a cortico-cortical FC model excluding the hippocampus. The cortico-cortical FC model incorporated all interconnecting edges among cortical parcellations, excluding subcortical ROIs, including the hippocampus. For *n* cortical ROIs, the model included $n(n - 1)/2$ edges. The hippocampo-cortical FC model employed the hippocampus as a seed in a seed-based FC model and focused on the FC between the hippocampus and each cortical ROI, excluding connections among cortical ROIs. We derived the hippocampus' time course by averaging across all voxels in its four subregions (left anterior, left posterior, right anterior, and right posterior hippocampus) from the Brainnetome atlas⁵¹. Consequently, the number of FC edges in this model was equal to the number of cortical ROIs. We assessed the robustness of model performance using various cortical parcellations, ranging from 100

to 1000 parcels (Supplementary Fig. 4). Note that our main analysis employed the model with 200 parcels.

To further assess the specific contribution of the hippocampus in predicting narrative memory scores, we conducted two complementary analyses. First, we tested whether including the hippocampus improved predictive performance in our FC-based models. For this comparison, we constructed two models: one that included the hippocampus and one that excluded it. Both models were constructed using identical sets of randomly selected cortical ROIs (ranging from 10 to 200), generated using a fixed random seed. This controlled design ensured that any differences in performance could be attributed specifically to the unique contribution of hippocampal connectivity rather than to differences in model complexity (Supplementary Fig. 6).

In the second analysis, we evaluated the predictive power of each brain region by using it as a seed in an FC-based model, encompassing both cortical and subcortical regions. To test whether the hippocampus's predictive performance was significantly better than expected by chance, we conducted a permutation-based rank analysis. Specifically, we ranked all ROIs based on their prediction accuracy for ordering and content scores, then generated a null distribution of hippocampal ranks by randomly shuffling the behavioral scores 500 times and recalculating the ranks for each iteration. The observed hippocampus rank was compared to this null distribution to calculate a *p* value, assessing the statistical significance of its predictive advantage (Supplementary Fig. 7).

To test whether each FC (hippocampus-vmPFC and hippocampus-PMC) differentially contributes to predicting distinct types of narrative memory scores (ordering and content), we employed a chi-square test and Steiger's *Z*-test⁴². We selected Steiger's *Z*-test to evaluate the association between FC and narrative memory scores since this test addresses differences between two correlations that share a common variable (i.e., FC) within the same sample. For the chi-square test, we compared observed frequencies of functional edge selection against expected frequencies calculated under the null hypothesis of no association between FC selection and memory score type in our predictive modeling.

LLM-generated narrative coherence

We proposed a quantitative measure to assess the likelihood of memory sequencing during ongoing narratives by leveraging the LLM, BERT³⁵, pre-trained on the NSP task. The NSP task is one of the two pre-training objectives employed by BERT to understand sequential contextual relationships between sentences in a text corpus. During this task, BERT was trained to predict whether a given sentence logically or causally follows another sentence, enabling the model to capture narrative relationships that extend beyond individual sentences.

Three additional independent annotators provided detailed descriptions of each segment of the scrambled movie at two-second intervals, which were then fed into the BERT model. We calculated the NSP likelihood values between each segment and all of its preceding segments in the movie. For each segment, we assigned the highest likelihood among all of its past segments to serve as its NSP likelihood. This provided an estimate of potential memory sequencing likelihood for each segment, similar to predefined sequencing moments of the scrambled narrative. To account for elevated NSP likelihoods between two segments due to shared semantics rather than logical or causal relationships, we used the USE⁴³ to create semantic vectors for each segment. We then computed semantic similarities between segments using cosine similarity. By regressing out semantic similarities from NSP likelihoods and applying a canonical hemodynamic response function, we derived an LLM-generated narrative coherence index for each movie segment. All metrics were averaged across multiple annotators.

To validate our approach, moment-by-moment human-rated narrative comprehension was adapted from our previous study⁵, where 20 participants responded when they felt they had understood the narrative while viewing the same scrambled movie used in the current study. We computed the time course similarity between this human-rated narrative

comprehension and the LLM-generated narrative coherence using Pearson's correlation. We determined statistical significance by comparing the actual correlation with a null distribution, generated by randomly permuting the phase of the time series for each dataset ($n = 1000$ permutations) (for details of the behavioral experiments, see our previous study⁵).

To examine the effectiveness of LLM-generated narrative coherence in assessing temporal memory sequencing during ongoing comprehension, we constructed a predictive model for ordering scores using LLM-generated sequencing moments. These sequencing moments were determined by selecting the top *N* time points ranked by LLM-generated narrative coherence. We varied the number of top time points used to compute the hippocampo-cortical FC pattern, ranging from 20 to 80. For the results shown in Fig. 5b and 5c, we selected the top 50 time points with the highest NSP likelihood scores. The predictive modeling procedure was identical to that of the main analysis. For comparison, we also extracted sequencing moments using raw NSP likelihood values from an LLM, which did not account for semantic similarity effects, and sequencing moments based on semantic similarity alone, calculated from USE. All other procedures, including the process of assigning the highest value obtained from all past pairs to each moment, remained identical to the original analysis.

Statistics and reproducibility

FC-based predictive performance was evaluated using leave-one-subject-out cross-validation, with statistical significance assessed via permutation testing. To assess the consistency of FC edge selection across cross-validation folds, one-tailed binomial tests were conducted, and results were corrected for multiple comparisons using FDR. Permutation-based rank analyses were performed to determine whether hippocampus-seeded predictive models ranked significantly higher than expected by chance in predicting memory scores. Differences between correlation coefficients were evaluated using Steiger's *Z*-test, and chi-square tests assessed the association between FC edge selection and memory score type. For LLM-based analyses, predictive performance and the association between LLM-derived features and human-rated narrative comprehension were tested using the same cross-validation and permutation procedures. Sample sizes were informed by prior naturalistic fMRI studies. No data were excluded beyond predefined quality control criteria. Results were reproducible across participants and robust to variations in model parameters and brain parcellations. All analysis code was verified to run successfully in an independent environment, and all results are fully reproducible with the provided code, dependencies, and data resources.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The fMRI data are openly available at OpenNeuro⁵⁴ with the following link <https://openneuro.org/datasets/ds005215>.

Code availability

The behavioral data and Python code are available on GitHub (<https://github.com/jwparks/NarrativePuzzle>), with a fixed version archived on Zenodo⁵⁵. The bert-base-uncased language model is publicly available on Hugging Face⁵⁶.

Received: 5 March 2025; Accepted: 11 June 2025;

Published online: 23 June 2025

References

1. Mandler, J. M. & Johnson, N. S. Remembrance of things parsed: story structure and recall. *Cogn. Psychol.* **9**, 111–151 (1977).

2. Bower, G. H. & Morrow, D. G. Mental models in narrative comprehension. *Science* **247**, 44–48 (1990).
3. Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M. & Gershman, S. J. Structured event memory: a neuro-symbolic model of event cognition. *Psychol. Rev.* **127**, 327–361 (2020).
4. PARIS, A. H. & PARIS, S. G. Assessing narrative comprehension in young children. *Read. Res. Quart.* **38**, 36–76 (2003).
5. Song, H., Park, B.-Y., Park, H. & Shim, W. M. Cognitive and neural state dynamics of narrative comprehension. *J. Soc. Neurosci.* **41**, 8972–8990 (2021).
6. Grall, C., Equita, J. & Finn, E. S. Neural unscrambling of temporal information during a nonlinear narrative. *Cereb. Cortex* **33**, 7001–7014 (2023).
7. Baldassano, C., et al. Discovering event structure in continuous narrative perception and memory. *Neuron* **95**, 709–721 (2017).
8. Kurby, C. A. & Zacks, J. M. Segmentation in the perception and memory of events. *Trends Cogn. Sci.* **12**, 72–79 (2008).
9. Zacks, J. M. et al. Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* **4**, 651–655 (2001).
10. Zheng, J. et al. Neurons detect cognitive boundaries to structure episodic memories in humans. *Nat. Neurosci.* **25**, 358–368 (2022).
11. Ben-Yakov, A. & Henson, R. N. The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. *J. Neurosci.* **38**, 10057–10068 (2018).
12. Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).
13. Barnett, A. J. et al. Hippocampal-cortical interactions during event boundaries support retention of complex narrative events. *Neuron* **112**, 319–330.e7 (2024).
14. Ben-Yakov, A., Eshel, N. & Dudai, Y. Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.* **142**, 1255–1263 (2013).
15. Ben-Yakov, A. & Dudai, Y. Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J. Neurosci.* **31**, 9032–9042 (2011).
16. Bunsey, M. & Eichenbaum, H. Conservation of hippocampal memory function in rats and humans. *Nature* **379**, 255–257 (1996).
17. Kolibius, L. D. et al. Hippocampal neurons code individual episodic memories in humans. *Nat. Hum. Behav.* **11**, 1968–1979 (2023).
18. Kosciak, T. R. & Tranel, D. The human ventromedial prefrontal cortex is critical for transitive inference. *J. Cogn. Neurosci.* **24**, 1191–1204 (2012).
19. DeVito, L. M., Lykken, C., Kanter, B. R. & Eichenbaum, H. Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. *Learn. Mem.* **17**, 161–167 (2010).
20. Preston, A. R. & Eichenbaum, H. Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* **23**, R764–R773 (2013).
21. Schlichting, M. L. & Preston, A. R. Memory integration: neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci.* **1**, 1–8 (2015).
22. Zeithamova, D., Dominick, A. L. & Preston, A. R. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* **75**, 168–179 (2012).
23. DeVito, L. M. & Eichenbaum, H. Memory for the order of events in specific sequences: contributions of the hippocampus and medial prefrontal cortex. *J. Neurosci.* **31**, 3169–3175 (2011).
24. van Kesteren, M. T. R., Fernández, G., Norris, D. G. & Hermans, E. J. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc. Natl. Acad. Sci. USA* **107**, 7550–7555 (2010).
25. van Kesteren, M. T. R., Ruiter, D. J., Fernández, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
26. Gerraty, R. T., Davidow, J. Y., Wimmer, G. E., Kahn, I. & Shohamy, D. Transfer of learning relates to intrinsic connectivity between hippocampus, ventromedial prefrontal cortex, and large-scale networks. *J. Neurosci.* **34**, 11297–11303 (2014).
27. Morici, J. F., Weisstaub, N. V. & Zold, C. L. Hippocampal-medial prefrontal cortex network dynamics predict performance during retrieval in a context-guided object memory task. *Proc. Natl. Acad. Sci. USA* **119**, e2203024119 (2022).
28. Lu, Q., Hasson, U. & Norman, K. A. A neural network model of when to retrieve and encode episodic memories. *Elife* **11**, e74445 (2022).
29. Ritter, S. et al. Been there, done that: meta-learning with episodic recall. *Arxiv* (2018).
30. Cohn-Sheehy, B. I. et al. The hippocampus constructs narrative memories across distant events. *Curr. Biol.* **31**, 4935–4945 (2021).
31. Hahamy, A., Dubossarsky, H. & Behrens, T. E. J. The human brain reactivates context-specific past information at event boundaries of naturalistic experiences. *Nat. Neurosci.* **26**, 1080–1089 (2023).
32. Lee, H. & Chen, J. Predicting memory from the network structure of naturalistic events. *Nat. Commun.* **13**, 4235 (2022).
33. Heusser, A. C., Fitzpatrick, P. C. & Manning, J. R. Geometric models reveal behavioural and neural signatures of transforming experiences into memories. *Nat. Hum. Behav.* **5**, 905–919 (2021).
34. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
35. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *Arxiv* <https://doi.org/10.48550/arxiv.1810.04805> (2018).
36. Richmond, L. L. & Zacks, J. M. Constructing experience: event models from perception to action. *Trends Cogn. Sci.* **21**, 962–980 (2017).
37. Shen, X. et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* **12**, 506–518 (2017).
38. Finn, E. S. et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
39. Rosenberg, M. D. et al. A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).
40. Finn, E. S. & Bandettini, P. A. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *Neuroimage* **235**, 117963 (2021).
41. Song, H., Finn, E. S. & Rosenberg, M. D. Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proc. Natl. Acad. Sci. USA* **118**, e2021905118 (2021).
42. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).
43. Cer, D. et al. Universal sentence encoder. *Arxiv* <https://doi.org/10.48550/arxiv.1803.11175> (2018).
44. Chen, J. et al. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cereb. Cortex* **26**, 3428–3441 (2016).
45. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
46. Park, E. L. & Cho, S. KoNLPy: Korean natural language processing in Python. in *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology (NLP, 2014)*.
47. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012).
48. Esteban, O. et al. FMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
49. Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G. & Gobbini, M. I. An fMRI dataset in response to “The Grand Budapest Hotel”, a socially-rich, naturalistic movie. *Sci. Data* **7**, 383 (2020).
50. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

51. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
52. Fan, L. et al. The human brainnetome atlas: a new brain atlas based on connectonal architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
53. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2017).
54. Park, J. et al. Hippocampal systems for event encoding and sequencing during ongoing narrative comprehension. *OpenNeuro*. [Dataset] <https://doi.org/10.18112/openneuro.ds005215.v1.0.2> (2025).
55. Park, J. et al. Hippocampal systems for event encoding and sequencing during ongoing narrative comprehension. *Zenodo* <https://doi.org/10.5281/zenodo.15597906> (2025).
56. Wolf, T. et al. Transformers: state-of-the-art natural language processing. In: *Proc. 2020 Conf. Empir. Methods Nat. Lang. Process. Syst. Demonstr.* 38–45 (NLP, 2020).

Acknowledgements

We would like to thank Min-Suk Kang and Seok-Jun Hong for their valuable feedback on the manuscript; Boohee Choi for technical support in fMRI data collection. This work was supported by the National Research Foundation of Korea (RS-2024-00348130, RS-2025-02304581), and the Fourth Stage of Brain Korea 21 Project (S-2023-0794-000).

Author contributions

J.P., H.S., and W.M.S. designed the research. J.P., H.S., and W.M.S. performed the research. J.P. analyzed the data. J.P. wrote the first draft of the paper. J.P., H.S., and W.M.S. edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08377-1>.

Correspondence and requests for materials should be addressed to Won Mok Shim.

Peer review information *Communications Biology* thanks Christopher Baldassano and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jasmine Pan. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025