Discovering Event Structure in Continuous Narrative Perception and Memory

Highlights

- Event boundaries during perception can be identified from cortical activity patterns

- Event timescales vary from seconds to minutes across the cortical hierarchy

- Hippocampal activity following an event predicts reactivation during recall

- Prior knowledge of a narrative enables anticipatory reinstatement of event patterns

Authors

Christopher Baldassano, Janice Chen, Asieh Zadbood, Jonathan W. Pillow, Uri Hasson, Kenneth A. Norman

Correspondence

chrisb@princeton.edu

In Brief

Using a new approach for identifying temporal structure in neuroimaging data, Baldassano et al. propose a theory of how continuous experience is divided into events that are represented in high-level cortex, are stored in long-term memory, and influence later perception.

CrossMark

CellPress

Article

Discovering Event Structure in Continuous Narrative Perception and Memory

Christopher Baldassano,[1,3,*] Janice Chen,[2] Asieh Zadbood,[1] Jonathan W. Pillow,[1] Uri Hasson,[1] and Kenneth A. Norman[1]
[1]Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, NJ 08544, USA
[2]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA
[3]Lead Contact
*Correspondence: chrisb@princeton.edu
http://dx.doi.org/10.1016/j.neuron.2017.06.041

SUMMARY

During realistic, continuous perception, humans automatically segment experiences into discrete events. Using a novel model of cortical event dynamics, we investigate how cortical structures generate event representations during narrative perception and how these events are stored to and retrieved from memory. Our data-driven approach allows us to detect event boundaries as shifts between stable patterns of brain activity without relying on stimulus annotations and reveals a nested hierarchy from short events in sensory regions to long events in high-order areas (including angular gyrus and posterior medial cortex), which represent abstract, multimodal situation models. High-order event boundaries are coupled to increases in hippocampal activity, which predict pattern reinstatement during later free recall. These areas also show evidence of anticipatory reinstatement as subjects listen to a familiar narrative. Based on these results, we propose that brain activity is naturally structured into nested events, which form the basis of long-term memory representations.

INTRODUCTION

Typically, perception and memory are studied in the context of discrete pictures or words. Real-life experience, however, consists of a continuous stream of perceptual stimuli. The brain therefore needs to structure experience into units that can be understood and remembered: "the meaningful segments of one's life, the coherent units of one's personal history" (Beal and Weiss, 2013). Although this question was first investigated decades ago (Newtson et al., 1977), a general "event segmentation theory" was proposed only recently (Zacks et al., 2007). These and other authors have argued that humans implicitly generate event boundaries when consecutive stimuli have distinct temporal associations (Schapiro et al., 2013), when the causal structure of the environment changes (Kurby and Zacks, 2008; Radvansky, 2012), or when our goals change (DuBrow and Davachi, 2016).

At what timescale are experiences segmented into events? When reading a story, we could chunk it into discrete units of in-

dividual words, sentences, paragraphs, or chapters, and we may need to chunk information on different timescales depending on our goals. Behavioral studies have shown that subjects can segment events into a nested hierarchy from coarse to fine timescales (Kurby and Zacks, 2008; Zacks et al., 2001b) and flexibly adjust their units of segmentation depending on their uncertainty about ongoing events (Newtson, 1973). The neural basis of this segmentation behavior is unclear; event perception could rely on a single unified system, which segments the continuous perceptual stream at different granularities depending on the current task (Zacks et al., 2007), or may rely on multiple brain areas that segment events at different timescales, as suggested by the selective deficits for coarse segmentations exhibited by some patient populations (Zalla et al., 2003, 2004).

A recent theory of cortical process-memory topography argues that information is integrated at different timescales throughout the cortex. Processing timescales increase from tens of milliseconds in early sensory regions (e.g., for detecting phonemes in early auditory areas), to a few seconds in mid-level sensory areas (e.g., for integrating words into sentences), up to hundreds of seconds in regions including the temporoparietal junction, angular gyrus, and posterior and frontal medial cortex (e.g., for integrating information from entire paragraphs) (Chen et al., 2016; Hasson et al., 2015). The relationship between the process-memory topography and event segmentation has not yet been investigated. On the one hand, it is possible that cortical representations are accumulated continuously, e.g., using a sliding window approach, at each level of the processing hierarchy (Stephens et al., 2013). On the other hand, a strong link between the timescale hierarchy and event segmentation theory would predict that each area chunks experience at its preferred timescale and integrates information within discretized units (e.g., phonemes, words, sentences, paragraphs) before providing its output to the next processing level (Nelson et al., 2017). In this view, "events" in low-level sensory cortex (e.g., a single phoneme; Giraud and Poeppel, 2012) are gradually integrated into minutes-long situation-level events, using a multistage nested temporal chunking. This chunking of continuous experience at multiple timescales along the cortical processing hierarchy has not been previously demonstrated in the dynamics of whole-brain neural activity.

A second critical question for understanding event perception is how real-life experiences are encoded into long-term memory. Behavioral experiments and mathematical models have argued that long-term memory reflects event structure during encoding
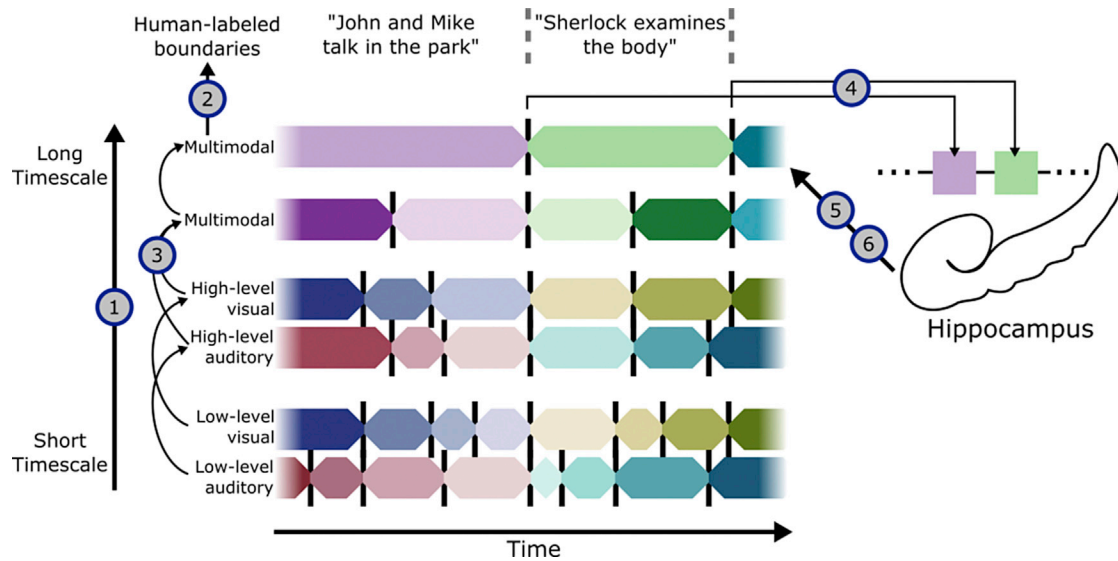
Figure 1. Theory of Event Segmentation and Memory

During perception, events are constructed at a hierarchy of timescales (1), with short events in primary sensory regions and long events in regions including the angular gyrus and posterior medial cortex. These high-level regions have event boundaries that correspond most closely to putative boundaries identified by human observers (2) and represent abstract narrative content that can be drawn from multiple input modalities (3). At the end of a high-level event, the situation model is stored into long-term memory (4) (resulting in post-boundary encoding activity in the hippocampus), and can be reinstated during recall back into these cortical regions (5). Prior event memories can also influence ongoing processing (6), facilitating prediction of upcoming events in related narratives. We test each of these hypotheses using a data-driven event segmentation model, which can automatically identify transitions in brain activity patterns and detect correspondences in activity patterns across datasets.

(Ezzyat and Davachi, 2011; Gershman et al., 2014; Sargent et al., 2013; Zacks et al., 2001b), suggesting that the event segments generated during perception may serve as the "episodes" of episodic memory. The hippocampus is thought to bind cortical representations into a memory trace (McClelland et al., 1995; Moscovitch et al., 2005; Norman and O'Reilly, 2003), a process that is typically studied using discrete memoranda (Danker et al., 2016). However, given a continuous stream of information in a real-life context, it is not clear at what timescale memories should be encoded and whether these memory traces should be continuously updated or encoded only after an event has completed. The hippocampus is connected to long-timescale regions, including the angular gyrus and posterior medial cortex (Kravitz et al., 2011; Ranganath and Ritchey, 2012; Rugg and Vilberg, 2013), suggesting that the main inputs to long-term memory come from these areas, which are thought to represent multimodal, abstract representations of the features of the current event ("situation models," Johnson-Laird, 1983; Van Dijk and Kintsch, 1983; Zwaan et al., 1995; Zwaan and Radvansky, 1998; or more generally, "event models," Radvansky and Zacks, 2011). Recent work has shown that hippocampal activity peaks at the *offset* of video clips (Ben-Yakov and Dudai, 2011; Ben-Yakov et al., 2013), suggesting that the end of a long-timescale event triggers memory encoding processes that occur *after* the event has ended.

Based on our experiments (described below), we propose that the full life cycle of an event can be described in a unified theory, illustrated in Figure 1. During perception, each brain region along the processing hierarchy segments information at its preferred timescale, beginning with short events in primary visual and auditory cortex and building to multimodal situation models in long-timescale areas, including the angular gyrus and posterior medial cortex. This model of processing requires that (1) all regions driven by audio-visual stimuli should exhibit event-structured activity, with segmentation into short events in early sensory areas and longer events in high-order areas; (2) events throughout the hierarchy should have a nested structure, with coarse event boundaries annotated by human observers most strongly related to long events at the top of the hierarchy; and (3) event representations in long-timescale regions, which build a coarse model of the situation, should be invariant across different descriptions of the same situation (e.g., when the same situation is described visually in a movie or a verbally in a story). We also argue that event structure is reflected in how experiences are stored into episodic memory. At event boundaries in long-timescale areas, the situation model is transmitted to the hippocampus, which can later reinstate the situation model in long-timescale regions during recall. This implies that (4) the end of an event in long-timescale cortical regions should trigger the hippocampus to encode information about the just-concluded event into episodic memory, and (5) stored event memories can be reinstated in long-timescale cortical regions during recall, with stronger reinstatement for more strongly encoded events. Finally, this process can come full circle, with prior event memories influencing ongoing processing, such that (6) prior memory for a narrative should lead to anticipatory reinstatement in long-timescale regions.
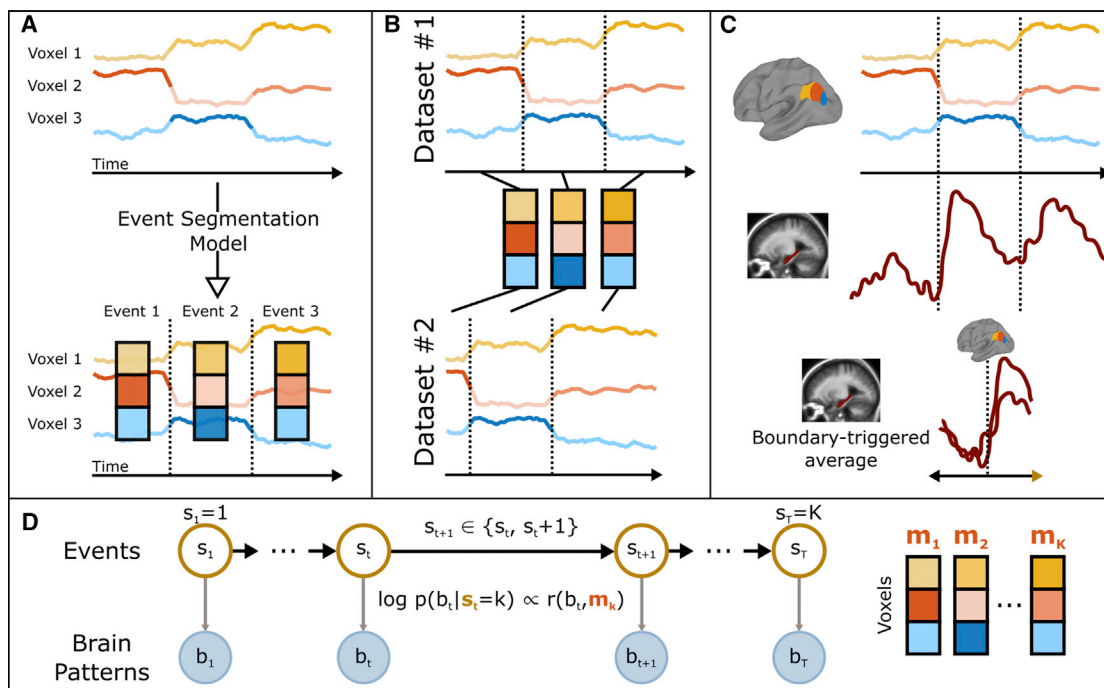
Figure 2. Event Segmentation Model

(A) Given a set of (unlabeled) time courses from a region of interest, the goal of the event segmentation model is to temporally divide the data into "events" with stable activity patterns, punctuated by "event boundaries" at which activity patterns rapidly transition to a new stable pattern. The number and locations of these event boundaries can then be compared across brain regions or to stimulus annotations.

(B) The model can identify event correspondences between datasets (e.g., responses to movie and audio versions of the same narrative) that share the same sequence of event activity patterns, even if the duration of the events is different.

(C) The model-identified boundaries can also be used to study processing evoked by event transitions, such as changes in hippocampal activity coupled to event transitions in the cortex.

(D) The event segmentation model is implemented as a modified Hidden Markov Model (HMM) in which the latent state s_t for each time point denotes the event to which that time point belongs, starting in event 1 and ending in event K. All datapoints during event K are assumed to be exhibit high similarity with an event-specific pattern m_k. See also Figures S1, S2, and S3.

To test these hypotheses, we need the ability to identify how different brain areas segment events (at different timescales), align events across different datasets with different timings (e.g., to see whether the same situation model is being elicited by a movie versus a verbal narrative, or a movie versus recall), and track differences in event segmentations in different observers of the same stimulus (e.g., depending on prior experience). Thus, to search for the neural correlates of event segmentation, we have developed a new data-driven method that allows us to identify events directly from fMRI activity patterns across multiple timescales and datasets (Figure 2).

Our analysis approach (described in detail in STAR Methods) starts with two simple assumptions: that while processing a narrative stimulus, observers progress through a sequence of discrete event representations (hidden states), and that each event has a distinct (observable) signature (a multi-voxel fMRI pattern) that is present throughout the event. We implement these assumptions using a data-driven event segmentation model, based on a Hidden Markov Model (HMM). Fitting the model to fMRI data (e.g., evoked by viewing a movie) entails estimating the optimal number of events, the mean activity pattern for each event, and when event transitions occur. When applying the model to multiple datasets evoked by the same narrative (e.g., during movie viewing and during later verbal recall), the model is constrained to find the same sequence of patterns (because the events are the same), but the timing of the transitions between the patterns can vary (e.g., since the spoken description of the events might not take as long as the original events).

In prior studies, using human-based segmentation of coarse event structure, we demonstrated that event-related representations generalize across modalities and between encoding and recall (thereby supporting hypotheses 3 and 5, Chen et al., 2017; Zadbood et al., 2016). In the current study, we extend these findings by testing whether we can use the data-driven HMM to detect stable and abstract event boundaries in high-order areas without relying on human annotations. This new analysis approach also allows us to test for the first time whether the brain segments information, hierarchically, at multiple timescales (hypotheses 1 and 2); how segmentation of information interacts with the storage and retrieval of this information by the hippocampus (hypothesis 4); and how prior exposure to a sequence of events can later lead to anticipatory reinstatement of those events (hypothesis 6). Taken together, our results
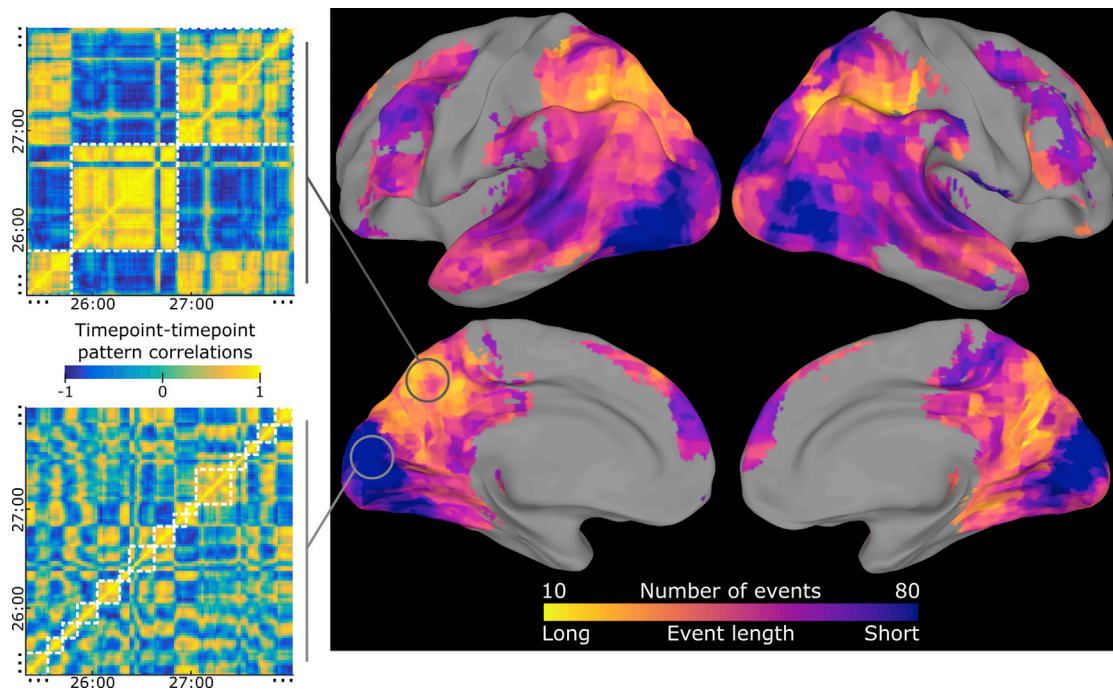
Figure 3. Event Segmentation Model for Movie-Watching Data Reveals Event Timescales
The event segmentation model identifies temporally clustered structure in movie-watching data throughout all regions of cortex with high intersubject correlation. The optimal number of events varied by an order of magnitude across different regions, with a large number of short events in sensory cortex and a small number of long events in high-level cortex. For example, the time point correlation matrix for a region in the precuneus exhibited coarse blocks of correlated patterns, leading to model fits with a small number of events (white squares), while a region in visual cortex was best modeled with a larger number of short events (note that only ∼3 min of the 50-min stimulus are shown and that the highlighted searchlights were selected post hoc for illustration). The searchlight was masked to include only regions with intersubject correlation > 0.25 and voxelwise thresholded for greater within- than across-event similarity, q < 0.001. See also Figures S3, S4, and S8.

provide the first direct evidence that realistic experiences are discretely and hierarchically chunked at multiple timescales in the brain, with chunks at the top of the processing hierarchy playing a special role in cross-modal situation representation and episodic memory.

RESULTS

All of our analyses are carried out using our new HMM-based event segmentation model (summarized above, and described in detail in the Event Segmentation Model subsection in STAR Methods), which can automatically discover the fMRI signatures of each event and its temporal boundaries in a particular dataset. We validated this model using both synthetic data (Figure S1) and narrative data with clear event breaks between stories (Figure S2), confirming that we could accurately recover the number of event boundaries and their locations (see STAR Methods). We then applied the model to test six predictions of our theory of event perception and memory.

Timescales of Cortical Event Segmentation
We first tested the hypothesis that all regions driven by audio-visual stimuli should exhibit event-structured activity, with segmentation into short events in early sensory areas and longer events in high-order areas. We measured the extent to which continuous stimuli evoked the event structure hypothesized by

our model (periods with stable event patterns punctuated by shifts between events) and whether the timescales of these events varied along the cortical hierarchy. We tested the model by fitting it to fMRI data collected while subjects watched a 50-min movie (Chen et al., 2017) and then assessing how well the learned event structure explained the activity patterns of a held-out subject. Note that previous analyses of this dataset have shown that the evoked activity is similar across subjects, justifying an across-subjects design (Chen et al., 2017). We found that essentially all brain regions that responded consistently to the movie (across subjects) showed evidence for event-like structure, and that the optimal number of events varied across the cortex (Figure 3). Sensory regions like visual and auditory cortex showed faster transitions between stable activity patterns, while higher-level regions like the posterior medial cortex, angular gyrus, and intraparietal sulcus had activity patterns that often remained constant for over a minute before transitioning to a new stable pattern (see Figure 3 insets and Figure S3). This topography of event timescales is broadly consistent with that found in previous work (Hasson et al., 2015) measuring sensitivity to temporal scrambling of a movie stimulus (see Figure S8).

Comparison of Event Boundaries across Regions and to Human Annotations
The second implication of our theory is that events throughout the hierarchy should have a nested structure, with coarse event
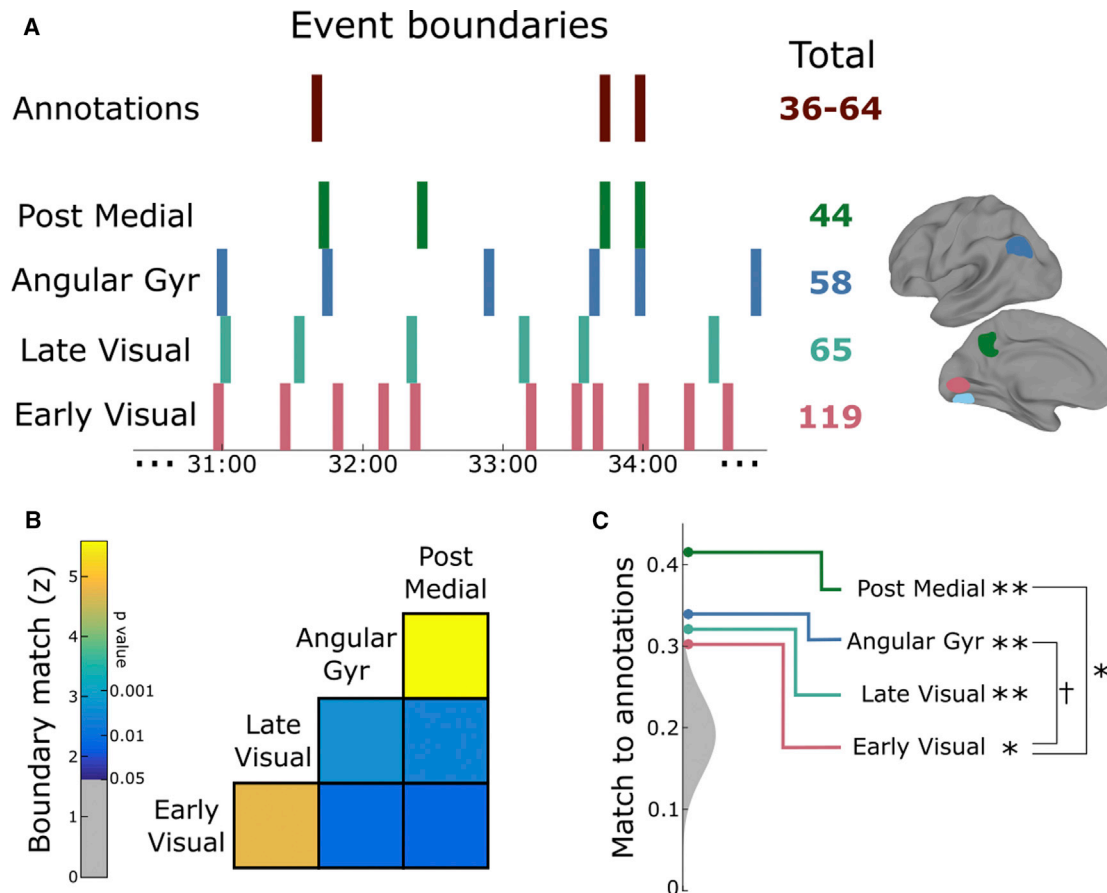
Figure 4. Cortical Event Boundaries Are Hierarchically Structured and Are Related to Human-Labeled Event Boundaries, Especially in Posterior Medial Cortex

(A) An example of boundaries evoked by the movie over a 4-min period shows how the number of boundaries decreases as we proceed up the hierarchy, with boundaries in posterior medial cortex most closely related to human annotations of event transitions.

(B) Event boundaries in higher regions are present in lower regions at above-chance levels (especially pairs of regions that are close in the hierarchy), suggesting that event segmentation is in part hierarchical, with lower regions subdividing events in higher regions.

(C) All four levels of the hierarchy show an above-chance match to human annotations (the null distribution is shown in gray), but the match increases significantly from lower to higher levels (†p = 0.058, *p < 0.05, **p < 0.01).

boundaries annotated by human observers most strongly related to long events at the top of the hierarchy. To examine how event boundaries changed throughout the cortical hierarchy, we created four regions of interest, each with 300 voxels, centered on ROIs from prior work (see STAR Methods): early visual cortex, late visual cortex, angular gyrus, and posterior medial cortex. We identified the optimal timescale for each region as in the previous analysis and then fit the event segmentation model at this optimal timescale, as illustrated in Figure 4A. We found that a significant portion of the boundaries in a given layer were also present in lower layers (Figure 4B), especially for adjacent layers in the hierarchy. This suggests that event segmentation is at least partially hierarchical, with finer event boundaries nested within coarser boundaries.

We asked four independent raters to divide the movie into "scenes" based on major shifts in the narrative (such as in location, topic, or time). The number of event boundaries identified by the observers varied between 36 and 64, but the boundaries had

a significant amount of overlap, with an average pairwise Dice's coefficient of 0.63 and 20 event boundaries that were labeled by all four raters. We constructed a "consensus" annotation containing boundaries marked by at least two raters, which split the narrative into 54 events, similar to the mean timescale for individual annotators (49.5). We then measured, for each region, what fraction of its fMRI-defined boundaries were close to (within three time points of) a consensus event boundary. As shown in Figure 4C, all regions showed an above-chance match to human annotations (early visual, p = 0.0135; late visual, p = 0.0065; angular gyrus, p = 0.0011; posterior medial, p < 0.001), but this match increased across the layers of the hierarchy and was largest in angular gyrus and posterior medial cortex (angular gyrus > early visual, p = 0.0580; posterior medial > early visual, p = 0.033).

Shared Event Structure across Modalities

The third requirement of our theory is that activity patterns in long-timescale regions should be invariant across different
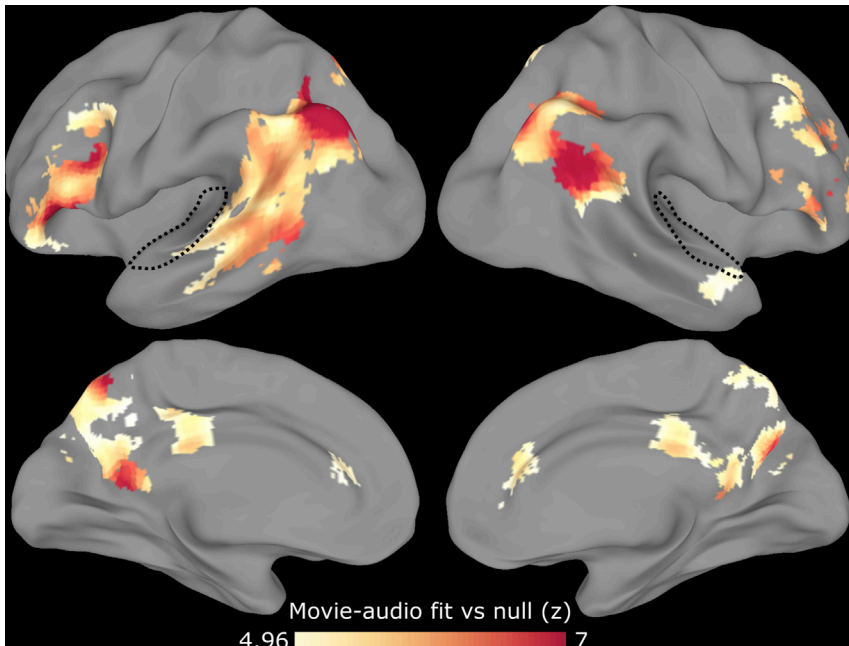
Figure 5. Movie-Watching Model Generalizes to Audio Narration in High-Level Cortex

After identifying a series of event patterns in a group of subjects who watched a movie, we tested whether this same series of events occurred in a separate group of subjects who heard an audio narration of the same story. The movie and audio stimuli were not synchronized and differed in their duration. We restricted our searchlight to voxels that responded to both the movie and audio stimuli (having high intersubject correlation within each group). Movie-watching event patterns in early auditory cortex (dotted line) did not generalize to the activity evoked by audio narration, while regions including the angular gyrus, temporoparietal junction, posterior medial cortex, and inferior frontal cortex exhibited shared event structure across the two stimulus modalities. This analysis, conducted using our data-driven model, replicates and extends the previous analysis of this dataset (Zadbood et al., 2016) in which the event correspondence between the movie and audio narration was specified by hand. The searchlight is masked to include only regions with intersubject correlation > 0.1 in all conditions and voxelwise thresholded for above-chance movie-audio fit, q < 10^{-5}. See also Figure S8.

descriptions of the same situation. This hypothesis is based on a prior study (Zadbood et al., 2016) in which one group of subjects watched a 24-min movie while the other group listened to an 18-min audio narration describing the events that occurred in the movie. The prior study used a hand-labeled event correspondence between these two stimuli to show that activity patterns evoked by corresponding events in the movie and narration were correlated in a network of regions including angular gyrus, precuneus, retrosplenial cortex, posterior cingulate cortex, and mPFC (Zadbood et al., 2016). Here, we use this dataset to ask whether our event segmentation model can replicate these results in a purely unsupervised manner, without using any prior event labeling.

For each cortical searchlight, we first segmented the movie data into events and then tested whether this same sequence of events from the movie-watching subjects was present in the audio-narration subjects. Regions including the angular gyrus, temporoparietal junction, posterior medial cortex, and inferior frontal cortex showed a strongly significant correspondence between the two modalities (Figure 5), indicating that a similar sequence of event patterns was evoked by the movie and audio narration irrespective of the modality used to describe the events. In contrast, though low-level auditory cortex was reliably activated by both of these stimuli, there was no above-chance similarity between the series of activity patterns evoked by the two stimuli (movie versus verbal description), presumably because the low-level auditory features of the two stimuli were markedly different.

Relationship between Cortical Event Boundaries and Hippocampal Encoding

We tested a fourth hypothesis from our theory, that the end of an event in long-timescale cortical regions should trigger the hippo-campus to encode information about the just-concluded event into episodic memory. Prior work has shown that the end of a video clip is associated with increased hippocampal activity, and the magnitude of the activity predicts later memory (Ben-Yakov and Dudai, 2011; Ben-Yakov et al., 2013). These experiments, however, have used only isolated short video clips with clear transitions between events. Do neurally defined event boundaries in a continuous movie, evoked by subtler transitions between related scenes, generate the same kind of hippocampal signature? Using a searchlight procedure, we identified event boundaries with the HMM segmentation model for each cortical area across the timescale hierarchy, using the 50-min movie dataset (Chen et al., 2017). We then computed the average *hippocampal* activity around the event boundaries of each cortical area to determine whether a cortical boundary tended to trigger a hippocampal response. We found that event boundaries in a distributed set of regions including angular gyrus, posterior medial cortex, and parahippocampal cortex all showed a strong relationship to hippocampal activity, with the hippocampal response typically peaking within several time points after the event boundary (Figure 6). This network of regions closely overlaps with the posterior medial memory system (Ranganath and Ritchey, 2012). Note that both the event boundaries and the hippocampal response are hemodynamic signals, so there is no hemodynamic offset between these two measures. The hippocampal response does start slightly before the event boundary, which could be due to uncertainty in the model estimation of the exact boundary time point and/or anticipation that the event is about to end.

Reinstatement of Event Patterns during Free Recall

Our theory further implies that stored event memories can be reinstated in long-timescale cortical regions during recall, with
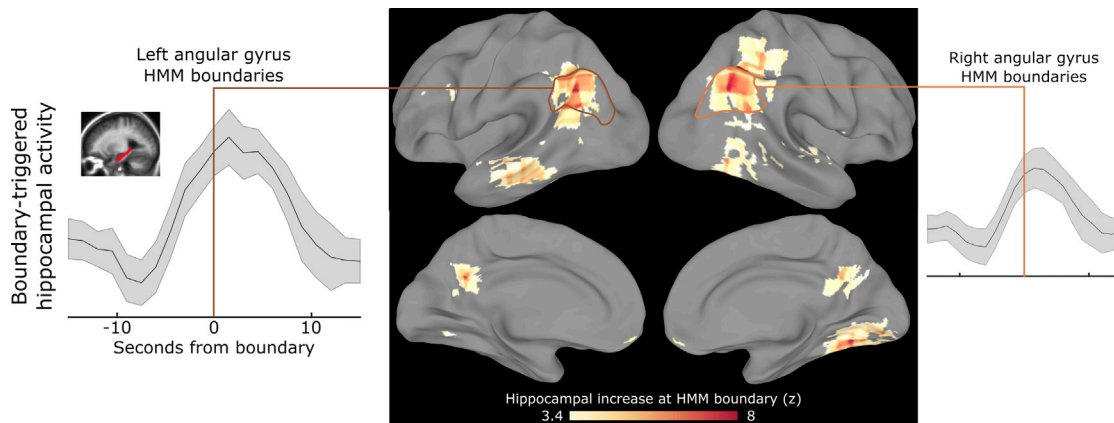
Figure 6. Hippocampal Activity Increases at Cortically Defined Event Boundaries
To determine whether event boundaries may be related to long-term memory encoding, we identify event boundaries based on a *cortical* region and then measure *hippocampal* activity around those boundaries. In a set of regions including angular gyrus, posterior medial cortex, and parahippocampal cortex, we find that event boundaries robustly predict increases in hippocampal activity, which tends to peak just after the event boundary (shaded region indicates 95% confidence interval). The searchlight is masked to include only regions with intersubject correlation > 0.25 and voxelwise thresholded for post-boundary hippocampal activity greater than pre-boundary activity, q < 0.001. See also Figures S5 and S8.

stronger reinstatement for more strongly encoded events. After watching the movie, all subjects in this dataset were asked to retell the story they had just watched, without any cues or stimulus (see Chen et al., 2017 for full details). We focused our analyses on the high-level regions that showed a strong relationship with hippocampal activity in the previous analysis (posterior cingulate and angular gyrus), as well as early auditory cortex for comparison.

Using the event segmentation model, we first estimated the (group-average) series of event-specific activity patterns evoked by the movie and then attempted to segment each subject's recall data into corresponding events. When fitting the model to the recall data, we assumed that the same event-specific activity patterns seen during the movie-viewing will be reinstated during the spoken recall. Analyzing the spoken recall transcriptions revealed that subjects generally recalled the events in the same order as they appeared in the movie (see Table S1 in Chen et al., 2017). Therefore, the model was constrained to use the same order of multi-voxel event patterns for recall that it had learned from the movie-watching data. However, crucially, the model was allowed to learn different event timings for the recall data compared to the movie data—this allowed us to accommodate the fact that event durations differed for free recall versus movie-watching.

For each subject, the model attempted to find a sequence of latent event patterns that was shared between the movie and recall, as shown in the example with 25 events in Figure 7A (see Figure S6 for examples from all subjects). Green shading indicates the probability that a movie and a recall time point belong to the same latent event (with darker shading indicating higher probability), and boxes indicate segments of the movie and recall that were labeled as corresponding to the same event by human annotators. Compared to the null hypothesis that there was no shared event order between the movie and recall, we found significant model fits in both the posterior cingulate

(p = 0.015) and the angular gyrus (p = 0.002), but not in low-level auditory cortex (p = 0.277) (Figure 7B). This result demonstrates that we can identify shared temporal structure between perception and recall without any human annotations. A similar pattern of results can be found regardless of the number of latent events used (see Figure S6). The identified correspondences for each subject (using both posterior cingulate and angular gyrus) were also significantly similar to the human-labeled correspondences (probability mass inside annotations = 17.8%, significantly greater than null model [11.9%], p < 0.001) (see Figure S6).

We then assessed whether the hippocampal response evoked by the end of an event during the encoding of the movie to memory was predictive of the length of time for which the event was strongly reactivated during recall. As shown in Figures 7C and 7D, we found that encoding activity and event reactivation were positively correlated in both angular gyrus (r = 0.362, p = 0.002) and the posterior cingulate (r = 0.312, p = 0.042), but not early auditory cortex (r = 0.080, p = 0.333). Note that there was no relationship between the hippocampal activity at the *starting* boundary of an event and that event's later reinstatement in the angular gyrus (r = −0.119, p = 0.867; difference from ending boundary correlation p = 0.004) and only a weak, nonsignificant relationship in posterior cingulate (r = 0.189, p = 0.113; difference from ending boundary correlation p = 0.274). The relationship between the *average* hippocampal activity throughout an event and later cortical reinstatement was actually negative (angular gyrus, r = −0.247, p = 0.017; posterior cingulate, r = −0.092, p = 0.175), suggesting that encoding is strongest when hippocampal activity is relatively low during an event and high at its offset.

Anticipatory Reinstatement for a Familiar Narrative
Finally, we tested a sixth hypothesis, that prior memory for a narrative should lead to anticipatory reinstatement in long-timescale regions. Our ongoing interpretation of events can be
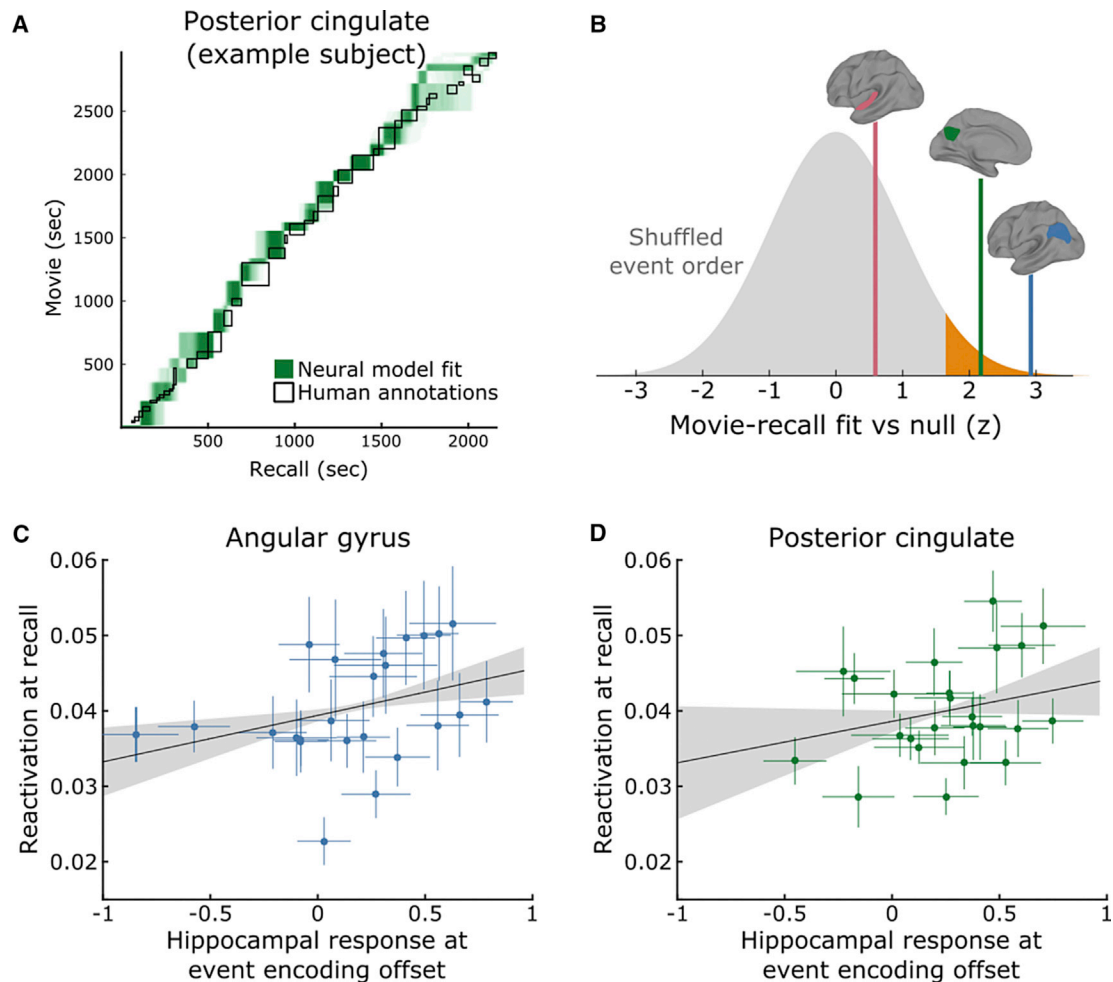
Figure 7. Movie-Watching Events Are Reactivated during Individual Free Recall, and Reactivation Is Related to Hippocampal Activation at Encoding Event Boundaries

(A) We can obtain an estimated correspondence between movie-watching data and free-recall data in individual subjects by identifying a shared sequence of event patterns, shown here for an example subject using data from posterior cingulate cortex.

(B) For each region of interest, we tested whether the movie and recall data shared an ordered sequence of latent events (relative to a null model in which the order of events was shuffled between movie and recall). We found that both angular gyrus (blue) and posterior cingulate cortex (green) showed significant reactivation of event patterns, while early auditory cortex (red) did not.

(C and D) Events whose offset drove a strong hippocampal response during encoding (movie-watching) were strongly reactivated for longer fractions of the recall period, both in the angular gyrus and the posterior cingulate. Error bars for event points denote SEM across subjects, and error bars on the best-fit line indicate 95% confidence intervals from bootstrapped best-fit lines. See also Figure S6.

influenced by prior knowledge; specifically, if subjects listening to the audio version of a narrative had already seen the movie version, they may anticipate upcoming events compared to subjects experiencing the narrative for the first time. Detecting this kind of anticipation has not been possible with previous approaches that rely on stimulus annotations, since the difference between the two groups is not in the stimulus (which is identical) but rather in the temporal dynamics of their cognitive processes.

Using data from Zadbood et al. (2016), we simultaneously fit our event segmentation model to three conditions—watching the movie, listening to the narration after watching the movie ("memory"), and listening to the narration without having previously seen the movie ("no memory")—looking for a shared

sequence of event patterns across conditions. By analyzing which time points were assigned to the same event, we can generate a time point correspondence indicating—for each time point during the audio narration datasets—which time points of the movie are most strongly evoked (on average) in the mind of the listeners.

We searched for cortical regions along the hierarchy of timescale showing anticipation, in which this correspondence for the memory group was consistently *ahead* of the correspondence for the no-memory group (relative to chance). As shown in Figure 8, we found anticipatory event reinstatement in the angular gyrus, posterior medial cortex, and medial frontal cortex. Examining the movie-audio correspondences in these regions,
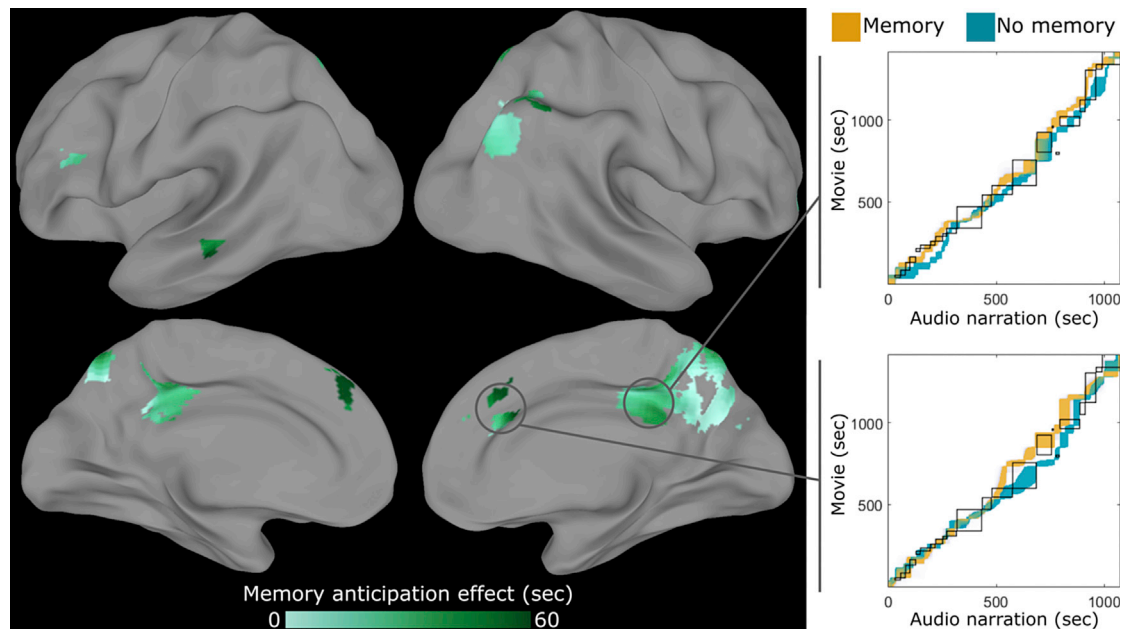
Figure 8. Prior Memory Shifts Movie-Audio Correspondence
The event segmentation model was fit simultaneously to data from a group watching the movie, the same group listening to the audio narration after having seen the movie ("memory"), and a separate group listening to the audio narration for the first time ("no memory"). By examining which time points were estimated to fall within the same latent event, we obtained a correspondence between time points in the audio data (for both groups) and time points in the movie data. We found regions in which the correspondence in both groups was close to the human-labeled correspondence between the movie and audio stimuli (black boxes), but the memory correspondence (orange) significantly led the non-memory correspondence (blue) (indicated by an upward shift on the correspondence plots; note that the highlighted searchlights were selected post hoc for illustration). This suggests that cortical regions of the memory group were anticipating events in the narration based on knowledge of the movie. The searchlight is masked to include only regions with intersubject correlation > 0.1 in all conditions and voxelwise thresholded for above-chance differences between memory and no memory groups, q < 0.05. See also Figures S7 and S8.

the memory group was consistently ahead of the no-memory group, indicating that for a given time point of the audio narration the memory group had event representations that corresponded to later time points in the movie. A similar result can be obtained by directly aligning the two listening conditions without reference to the movie-watching condition (see Figure S7).

DISCUSSION

We found that narrative stimuli evoke event-structured activity throughout the cortex, with pattern dynamics consisting of relatively stable periods punctuated by rapid event transitions. Furthermore, the angular gyrus and posterior medial cortex exhibit a set of overlapping properties associated with high-level situation model representations: long event timescales, event boundaries closely related to human annotations, generalization across modalities, hippocampal response at event boundaries, reactivation during free recall, and anticipatory coding for familiar narratives. Identifying all of these properties was made possible by using naturalistic stimuli with extended temporal structure, paired with a data-driven model for identifying activity patterns shared across time points and across datasets.

Event Segmentation Theory
Our results are the first to validate a number of key predictions of event segmentation theory (Zacks et al., 2007) directly from

fMRI data of naturalistic narratives, without using specially constructed stimuli or subjective labeling of where events should start and end. Previous work has shown that hand-labeled event boundaries are associated with univariate activity increases in a network of regions overlapping our high-level areas (Ezzyat and Davachi, 2011; Speer et al., 2007; Swallow et al., 2011; Whitney et al., 2009; Zacks et al., 2001a, 2010), but by modeling fine-scale spatial activity patterns we were able to detect these event changes without an external reference. This allowed us to identify regions with temporal event structure at many different timescales, only some of which matched human-labeled boundaries. Other analyses of these datasets also found reactivation during recall (Chen et al., 2017) and shared event structure across modalities (Zadbood et al., 2016); however, because these other analyses defined events based on the narrative rather than brain activity, they were unable to identify differences in event segmentation across brain areas or across groups with different prior knowledge.

Processing Timescales and Event Segmentation
The topography of event timescales revealed by our analysis provides converging evidence for an emerging view of how information is processed during real-life experience (Hasson et al., 2015). The "process memory framework" argues that perceptual stimuli are integrated across longer and longer timescales along a hierarchy from early sensory regions to regions in

the default-mode network. Using a variety of experimental approaches, including fMRI, electrocorticography (ECoG), and single-unit recording, this topography has previously been mapped either by temporally scrambling the stimulus at different timescales to see which regions' responses are disrupted (Hasson et al., 2008; Honey et al., 2012; Lerner et al., 2011) or by examining the power spectrum of intrinsic dynamics within each region (Honey et al., 2012; Murray et al., 2014; Stephens et al., 2013). Our model and results yield a new perspective on these findings, suggesting that *all* processing regions can exhibit rapid activity shifts, but that these fast changes are much less frequent in long-timescale regions. The power spectrum is therefore an incomplete description of voxel dynamics, since the correlation timescale changes dynamically, with faster changes at event boundaries and slower changes within boundaries (see also Figure S4B). We also found evidence for nested hierarchical structure, suggesting that chunks of information are transmitted from lower to higher levels primarily at event boundaries, as in recent multiscale recurrent neural network models (Chung et al., 2017).

The specific features encoded in the event representations of long-timescale regions like the angular gyrus and posterior cingulate cortex during naturalistic perception are still an open question. These areas are involved in high-level, multimodal scene processing tasks including memory and navigation (Baldassano et al., 2016; Kumar et al., 2017), are part of the "general recollection network" with strong anatomical and functional connectivity to the hippocampus (Rugg and Vilberg, 2013), and are the core components of the posterior medial memory system (Ranganath and Ritchey, 2012), which is thought to represent and update a high-level situation model (Johnson-Laird, 1983; Van Dijk and Kintsch, 1983; Zwaan et al., 1995; Zwaan and Radvansky, 1998). Since event representations in these regions generalized across modalities and between perception and recall, our results provide further evidence that they encode high-level situation descriptions. We also found that event patterns could be partially predicted by key characters and locations from the narrative (see Figure S4D, and Figure S6 in Chen et al., 2017), but future work (with richer descriptions of narrative events, Vodrahalli et al., 2017) will be required to understand how event patterns are evoked by semantic content.

Events in Episodic Memory

Behavioral experiments have shown that long-term memory reflects event structure during encoding (Ezzyat and Davachi, 2011; Sargent et al., 2013; Zacks et al., 2001b). Here, we were able to identify the reinstatement of events that were automatically discovered during perception, extending previous work demonstrating reinstatement of individual items or scenes in angular gyrus and posterior medial cortex (Chen et al., 2017; Johnson et al., 2009; Kuhl and Chun, 2014; Ritchey et al., 2013; Wing et al., 2015) to continuous perception without any stimulus annotations.

We demonstrated that the hippocampal encoding activity previously shown to be present at the end of movie clips (Ben-Yakov and Dudai, 2011; Ben-Yakov et al., 2013) and at abrupt switches between stimulus category and task (DuBrow and Davachi, 2016) also occurs at the much more subtle transitions between events (defined by pattern shifts in high-level regions), providing evidence that event boundaries trigger the storage of the current situation representation into long-term memory. We have also shown that this post-event hippocampal activity is related to pattern reinstatement during recall, as has been recently demonstrated for the encoding of discrete items (Danker et al., 2016), thereby supporting the view that events are the natural units of episodic memory during everyday life. Changes in cortical activity patterns may drive encoding through a comparator operation in the hippocampus (Lisman and Grace, 2005; Vinogradova, 2001), or the prediction error associated with event boundaries may potentiate the dopamine pathway (Zacks et al., 2011), thereby leading to improved hippocampal encoding (Kempadoo et al., 2016; Takeuchi et al., 2016). Notably, the positive relationship between hippocampal activity and subsequent cortical reinstatement was specific to hippocampal activity at the end of an event; there was no significant relationship between hippocampal activity at the start of an event and subsequent reinstatement, and higher hippocampal activity during an event was associated with worse reinstatement (a similar relationship was observed in parietal cortex by Lee et al., 2017). In this respect, our results differ from other results showing that the hippocampal response to novel events drives memory for the novel events *themselves* (for a review, see Ranganath and Rainer, 2003)—here, we show that the hippocampal response to a new event is linked to subsequent memory for the *previous* event.

Most extant models of memory consolidation (McClelland et al., 1995) and recall (Polyn et al., 2009) have been formulated under the assumption that the input to the memory system is a series of discrete items to be remembered. Although this is true for experimental paradigms that use lists of words or pictures, it was not clear how these models could function for realistic autobiographical memory. Our model connects naturalistic perception with theories about discrete memory traces, by proposing that cortex chunks continuous experiences into discrete events; these events are integrated along the processing hierarchy into meaningful, temporally extended, episodic structures, to be later encoded into memory via interaction with the hippocampus. The fact that the hippocampus is coupled to both angular gyrus and posterior medial cortex, which have slightly different timescales, raises the interesting possibility that these event memories could be stored at multiple temporal resolutions.

Our Event Segmentation Model

Temporal latent variable models have been largely absent from the field of human neuroscience, since the vast majority of experiments have a temporal structure that is defined ahead of time by the experimenter. One notable exception is the recent work of Anderson and colleagues, which has used HMM-based models to discover temporal structure in brain activity responses during mathematical problem solving (Anderson and Fincham, 2014; Anderson et al., 2014, 2016). These models are used to segment problem-solving operations (performed in less than 30 s) into a small number of cognitively distinct stages such as encoding, planning, solving, and responding. Our work is the first to show that (using a modified HMM and an annealed fitting procedure) this latent-state approach can be extended to much longer

experimental paradigms with a much larger number of latent states.

For finding correspondences between continuous datasets, as in our analyses of shared structure between perception and recall or perception under different modalities, several other types of approaches (not based on HMMs) have been proposed in psychology and machine learning. Dynamic time warping (Kang and Wheatley, 2015; Silbert et al., 2014) locally stretches or compresses two time series to find the best match, and more complex methods such as conditional random fields (Zhu et al., 2015) allow for parts of the match to be out of order. However, these methods do not explicitly model event boundaries.

Future work will be required to investigate what types of neural correspondences are well modeled by continuous warping versus event-structured models. Logically, a strictly event-structured model (with static event patterns) cannot be a complete description of brain activity during narrative perception, since subjects are actively accumulating information during each event, and extensions of our model could additionally model these within-event dynamics (see Figure S4B).

Perception and Memory in the Wild

Our results provide a bridge between the large literature on long-term encoding of individual items (such as words or pictures) and studies of memory for real-life experience (Nielson et al., 2015; Rissman et al., 2016). Since our approach does not require an experimental design with rigid timing, it opens the possibility of having subjects be more actively and realistically engaged in a task, allowing for the study of events generated during virtual reality navigation (such as spatial boundaries, Horner et al., 2016) or while holding dialogs with a simultaneously scanned subject (Hasson et al., 2012). The model also is not fMRI specific and could be applied to other types of neuroimaging time series such as ECoG, electroencephalography (EEG), or functional near-infrared spectroscopy (fNIRS), including portable systems that could allow experiments to be run outside the lab (McKendrick et al., 2016).

Conclusion

Using a novel event segmentation model that can be fit directly to neuroimaging data, we showed that cortical responses to naturalistic stimuli are temporally organized into discrete events at varying timescales. In a network of high-level association regions, we found that these events were related to subjective event annotations by human observers, predicted hippocampal encoding, generalized across modalities and between perception and recall, and showed anticipatory coding of familiar narratives. Our results provide a new framework for understanding how continuous experience is accumulated, stored, and recalled.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Interleaved Stories dataset

- METHOD DETAILS
  - Interleaved Stories dataset
  - Sherlock Recall dataset
  - Sherlock Narrative dataset
  - Event annotations by human observers
  - Event Segmentation Model
  - Finding event structure in narratives
  - Timescales of cortical event segmentation
  - Comparison of event boundaries across regions and to human annotations
  - Shared event structure across modalities
  - Relationship between cortical event boundaries and hippocampal encoding
  - Reinstatement of event patterns during free recall
  - Anticipatory reinstatement for a familiar narrative
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Timescales of cortical event segmentation
  - Comparison of event boundaries across regions and to human annotations
  - Shared event structure across modalities
  - Relationship between cortical event boundaries and hippocampal encoding
  - Reinstatement of event patterns during free recall
  - Anticipatory reinstatement for a familiar narrative
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2017.06.041.

REFERENCES

Anderson, J.R., and Fincham, J.M. (2014). Discovering the sequential structure of thought. Cogn. Sci. *38*, 322–352.

Anderson, J.R., Lee, H.S., and Fincham, J.M. (2014). Discovering the structure of mathematical problem solving. Neuroimage *97*, 163–177.

Anderson, J.R., Pyke, A.A., and Fincham, J.M. (2016). Hidden Stages of Cognition Revealed in Patterns of Brain Activation. Psychol. Sci. 27, 1215–1226.

Baldassano, C., Esteva, A., Beck, D.M., and Fei-Fei, L. (2016). Two distinct scene processing networks connecting vision and memory. eNeuro 3, http://dx.doi.org/10.1523/ENEURO.0178-16.2016.

Beal, D.J., and Weiss, H.M. (2013). The Episodic Structure of Life at Work. In A Day in the Life of a Happy Worker, A.B. Bakker and K. Daniels, eds. (Psychology Press), pp. 8–24.

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.

Ben-Yakov, A., Eshel, N., and Dudai, Y. (2013). Hippocampal immediate post-stimulus activity in the encoding of consecutive naturalistic episodes. J. Exp. Psychol. Gen. 142, 1255–1263.

Chen, J., Chow, M., Norman, K.A., and Hasson, U. (2015a). Differentiation of neural representations during processing of multiple information streams (Society for Neuroscience).

Chen, P.-H. (Cameron), Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., and Ramadge, P.J. (2015b). A Reduced-Dimension fMRI Shared Response Model. In Advances in Neural Information Processing Systems 28 (NIPS 2015). http://papers.nips.cc/paper/5855-a-reduced-dimension-fmri-shared-response-model.

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

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

Chung, J., Ahn, S., and Bengio, Y. (2017). Hierarchical Multiscale Recurrent Neural Networks. In 5th International Conference on Learning Representations (ICLR 2017).

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. http://www.ncbi.nlm.nih.gov/pubmed/8812068.

Danker, J.F., Tompary, A., and Davachi, L. (2016). Trial-by-Trial Hippocampal Encoding Activation Predicts the Fidelity of Cortical Reinstatement During Subsequent Retrieval. Cereb. Cortex. Published online June 10, 2016. http://dx.doi.org/10.1093/cercor/bhw146.

DuBrow, S., and Davachi, L. (2016). Temporal binding within and across events. Neurobiol. Learn. Mem. 134 (Pt A), 107–114.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., and Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.

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

Gershman, S.J., Radulescu, A., Norman, K.A., and Niv, Y. (2014). Statistical computations underlying the dynamics of memory updating. PLoS Comput. Biol. 10, e1003939.

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 28, 2539–2550.

Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends Cogn. Sci. 16, 114–121.

Hasson, U., Chen, J., and Honey, C.J. (2015). Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. 19, 304–313.

Honey, C.J., Thesen, T., Donner, T.H., Silbert, L.J., Carlson, C.E., Devinsky, O., Doyle, W.K., Rubin, N., Heeger, D.J., and Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76, 423–434.

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

Johnson, J.D., McDuff, S.G.R., Rugg, M.D., and Norman, K.A. (2009). Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron 63, 697–708.

Johnson-Laird, P.N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (Harvard University Press).

Kang, O., and Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. Conscious. Cogn. 35, 128–135.

Kempadoo, K.A., Mosharov, E.V., Choi, S.J., Sulzer, D., and Kandel, E.R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. Proc. Natl. Acad. Sci. USA 113, 14835–14840.

Kravitz, D.J., Saleem, K.S., Baker, C.I., and Mishkin, M. (2011). A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.

Kuhl, B.A., and Chun, M.M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. J. Neurosci. 34, 8051–8060.

Kumar, M., Federmeier, K.D., Fei-Fei, L., and Beck, D.M. (2017). Evidence For Similar Patterns of Neural Activity Elicited by Picture- and Word-based Representations of Natural Scenes. Neuroimage. Published online March 24, 2017. http://dx.doi.org/10.1016/j.neuroimage.2017.03.037.

Kurby, C.A., and Zacks, J.M. (2008). Segmentation in the perception and memory of events. Trends Cogn. Sci. 12, 72–79.

Lee, H., Chun, M.M., and Kuhl, B.A. (2017). Lower Parietal Encoding Activation Is Associated with Sharper Information and Better Memory. Cereb. Cortex 27, 2486–2499.

Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 31, 2906–2915.

Lisman, J.E., and Grace, A.A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46, 703–713.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102, 419–457.

McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., and Ayaz, H. (2016). Into the Wild: Neuroergonomic Differentiation of Hand-Held and Augmented Reality Wearable Displays during Outdoor Navigation with Functional Near Infrared Spectroscopy. Front. Hum. Neurosci. 10, 216.

Moscovitch, M., Rosenbaum, R.S., Gilboa, A., Addis, D.R., Westmacott, R., Grady, C., McAndrews, M.P., Levine, B., Black, S., Winocur, G., and Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. J. Anat. 207, 35–66.

Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., and Wang, X.J. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci. 17, 1661–1663.

Nelson, M.J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S.S., Naccache, L., Hale, J.T., Pallier, C., and Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. Proc. Natl. Acad. Sci. USA 114, E3669–E3678.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. Journal of Personality and Social Psychology 28, 28–38.

Newtson, D., Engquist, G.A., and Bois, J. (1977). The objective basis of behavior units. Journal of Personality and Social Psychology 35, 847–862.

Nielson, D.M., Smith, T.A., Sreekumar, V., Dennis, S., and Sederberg, P.B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. Proc. Natl. Acad. Sci. USA 112, 11078–11083.

Norman, K.A., and O'Reilly, R.C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110, 611–646.

Polyn, S.M., Norman, K.A., and Kahana, M.J. (2009). A context maintenance and retrieval model of organizational processes in free recall. Psychol. Rev. 116, 129–156.

Radvansky, G.A. (2012). Across the Event Horizon. Current Directions in Psychological Science 21, 269–272.

Radvansky, G., and Zacks, J.M. (2011). Event Perception. Wiley Interdiscip. Rev. Cogn. Sci. 2, 608–620.

Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4, 193–202.

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

Rissman, J., Chow, T.E., Reggente, N., and Wagner, A.D. (2016). Decoding fMRI Signatures of Real-world Autobiographical Memory Retrieval. J. Cogn. Neurosci. 28, 604–620.

Ritchey, M., Wing, E.A., LaBar, K.S., and Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb. Cortex 23, 2818–2828.

Rugg, M.D., and Vilberg, K.L. (2013). Brain networks underlying episodic memory retrieval. Curr. Opin. Neurobiol. 23, 255–260.

Sargent, J.Q., Zacks, J.M., Hambrick, D.Z., Zacks, R.T., Kurby, C.A., Bailey, H.R., Eisenberg, M.L., and Beck, T.M. (2013). Event segmentation ability uniquely predicts event memory. Cognition 129, 241–255.

Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., and Botvinick, M.M. (2013). Neural representations of events arise from temporal community structure. Nat. Neurosci. 16, 486–492.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M.D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex 22, 158–165.

Silbert, L.J., Honey, C.J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. Proc. Natl. Acad. Sci. USA 111, E4687–E4696.

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

Speer, N.K., Zacks, J.M., and Reynolds, J.R. (2007). Human brain activity time-locked to narrative event boundaries. Psychol. Sci. 18, 449–455.

Stephens, G.J., Honey, C.J., and Hasson, U. (2013). A place for time: the spatiotemporal structure of neural dynamics during natural audition. J. Neurophysiol. 110, 2019–2026.

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

Takeuchi, T., Duszkiewicz, A.J., Sonneborn, A., Spooner, P.A., Yamasaki, M., Watanabe, M., Smith, C.C., Fernández, G., Deisseroth, K., Greene, R.W., and

Morris, R.G. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. Nature 537, 357–362.

Van Dijk, T.A., and Kintsch, W. (1983). Strategies of discourse comprehension (Academic Press).

Vinogradova, O.S. (2001). Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. Hippocampus 11, 578–598.

Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., Honey, C., Hasson, U., Ramadge, P., Norman, K., and Arora, S. (2017). Mapping Between fMRI Responses to Movies and their Natural Language Annotations. NeuroImage. Published online June 23, 2017. http://dx.doi.org/10.1016/j.neuroimage.2017.06.042.

Wang, L., Mruczek, R.E., Arcaro, M.J., and Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. Cereb. Cortex 25, 3911–3931.

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., and Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. Neuroimage 47, 360–366.

Wing, E.A., Ritchey, M., and Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. J. Cogn. Neurosci. 27, 679–691.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L., and Raichle, M.E. (2001a). Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 651–655.

Zacks, J.M., Tversky, B., and Iyer, G. (2001b). Perceiving, remembering, and communicating structure in events. J. Exp. Psychol. Gen. 130, 29–58.

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

Zacks, J.M., Speer, N.K., Swallow, K.M., and Maley, C.J. (2010). The Brain's Cutting-Room Floor: Segmentation of Narrative Cinema. Front. Hum. Neurosci. 4, 1–15.

Zacks, J.M., Kurby, C.A., Eisenberg, M.L., and Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. J. Cogn. Neurosci. 23, 4057–4066.

Zadbood, A., Chen, J., Leong, Y.C., Norman, K.A., and Hasson, U. (2016). How we transmit memories to other brains: constructing shared neural representations via communication. bioRxiv. http://dx.doi.org/10.1101/081208.

Zalla, T., Pradat-Diehl, P., and Sirigu, A. (2003). Perception of action boundaries in patients with frontal lobe damage. Neuropsychologia 41, 1619–1627.

Zalla, T., Verlut, I., Franck, N., Puzenat, D., and Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia. Psychiatry Res. 128, 39–51.

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., and Fidler, S. (2015). Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books. arXiv, arXiv:1506.06724, https://arxiv.org/abs/1506.06724.

Zwaan, R.A., and Radvansky, G.A. (1998). Situation models in language comprehension and memory. Psychol. Bull. 123, 162–185.

Zwaan, R.A., Langston, M.C., and Graesser, A.C. (1995). The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. Psychol. Sci. 6, 292–297.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Event Segmentation Model	Custom Software	https://github.com/intelpni/brainiak, https://github.com/cbaldassano/Event-Segmentation
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
Python	Python	https://www.python.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Christopher Baldassano (chrisb@princeton.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Interleaved Stories dataset

Twenty-two subjects (all native English speakers) were recruited from the Princeton community (9 male, 13 female, ages 18-26). All subjects provided informed written consent prior to the start of the study in accordance with experimental procedures approved by the Princeton University Institutional Review Board. The study was approximately 2 hr long and subjects received $20 per hour as compensation for their time. Data from 3 subjects were discarded due to falling asleep during the scan, and 1 due to problems with audio delivery.

METHOD DETAILS

Interleaved Stories dataset

To test our model in a dataset with clear, unambiguous event boundaries, we used data from subjects who listened to two unrelated audio narratives (Chen et al, 2015a). We used data from 18 subjects who listened to the two audio narratives in an interleaved fashion, with the audio stimulus switching between the two narratives approximately every 60 s at natural paragraph breaks. The total stimulus length was approximately 29 min, during which there were 32 story switches. The audio was delivered via in-ear headphones.

Imaging data were acquired on a 3T full-body scanner (Siemens Skyra) with a 20-channel head coil using a T2*-weighted echo planar imaging (EPI) pulse sequence (TR 1500 ms, TE 28 ms, flip angle 64, whole-brain coverage 27 slices of 4 mm thickness, in-plane resolution 3 × 3 mm, FOV 192 × 192 mm). Preprocessing was performed in FSL, including slice time correction, motion correction, linear detrending, high-pass filtering (140 s cutoff), and coregistration and affine transformation of the functional volumes to a template brain (MNI). Functional images were resampled to 3 mm isotropic voxels for all analyses.

The analyses in this paper were carried out using data from a posterior cingulate region of interest, the posterior medial cluster in the "dorsal default mode network" defined by whole-brain resting state connectivity clustering (Shirer et al., 2012).

Sherlock Recall dataset

Our primary dataset consisted of 17 subjects who watched the first 50 min of the first episode of BBC's *Sherlock*, and were then asked to freely recall the episode in the scanner without cues (Chen et al., 2017). Subjects varied in the length and richness of their recall, with total recall times ranging from 11 min to 46 min (and a mean of 22 min). Imaging data was acquired using a T2*-weighted echo planar imaging (EPI) pulse sequence (TR 1500 ms, TE 28 ms, flip angle 64, whole-brain coverage 27 slices of 4 mm thickness, in-plane resolution 3 × 3 mm, FOV 192 × 192 mm). A standard preprocessing pipeline was performed using FSL, including motion correction. Since acoustic output was not correlated across subjects (Chen et al., 2017), shared activity patterns at recall are unlikely to be driven by correlated motion artifacts. All subjects were aligned to a common MNI template, and analyses were carried out in this common volume space. We also conducted an alternate version of the segmentation analysis that does not rely on precise MNI alignment (Chen et al., 2015b) and obtained similar results (see Figure S4C).

We restricted our searchlight analyses to voxels that were reliably driven by the stimuli, measured using intersubject correlation (Hasson et al., 2004). Voxels with a correlation less than r = 0.25 during movie-watching were removed before running the searchlight analysis.

We defined five regions of interest based on prior work. In addition to the posterior cingulate region defined above, we defined the angular gyrus as area PG (both PGa and PGp) using the maximum probability maps from a cytoarchitectonic atlas (Eickhoff et al.,

2005), early auditory cortex as voxels within the Heschl's gyrus region (Harvard-Oxford cortical atlas) with high intersubject correlation during an audio narrative ("Pieman," Simony et al., 2016), early visual cortex as voxels near the calcarine sulcus with high intersubject correlation during an audio-visual movie ("The Twilight Zone," Chen et al., 2016), and hV4 based on a group maximum-probability atlas (Wang et al., 2015).

Sherlock Narrative dataset

To investigate cross-modal event representations and the impact of prior memory, we used a separate dataset in which subjects experienced multiple versions of a narrative. One group of 17 subjects watched the first 24 min of the first episode of *Sherlock* (a portion of the same episode used in the Sherlock Recall dataset), while another group of 17 subjects (who had never seen the episode before) listened to an 18 min audio description of the events during this part of the episode (taken from the audio recording of one subject's recall in the Sherlock Recall dataset). The subjects who watched the episode then listened to the same 18 min audio description. This yielded three sets of data, all based on the same story: watching a movie of the events, listening to an audio narration of the events *without* prior memory, and listening to an audio narration of the events *with* prior memory. Imaging data was acquired using the same sequence as in Sherlock Recall dataset; see Zadbood et al. (2016) for full acquisition and preprocessing details.

As in the Sherlock Recall experiment, we removed all voxels that were not reliably driven by the stimuli. Only voxels with an intersubject correlation of at least r = 0.1 across all three conditions were included in searchlight analyses.

Event annotations by human observers

Four human observers were given the video file for the 50 min *Sherlock* stimulus, and given the following directions: "Write down the times at which you feel like a new scene is starting; these are points in the movie when there is a major change in topic, location, time, etc. Each 'scene' should be between 10 seconds and 3 minutes long. Also, give each scene a short title." The similarity among observers was measured using Dice's coefficient (number of matching boundaries divided by mean number of boundaries, considering boundaries within three time points of one another to match).

Event Segmentation Model

Our model is built on two hypotheses: (1) while processing narrative stimuli, observers experience a sequence of discrete events, and (2) each event has a distinct neural signature. Mathematically, the model assumes that a given subject (or averaged group of subjects) starts in event $s_1 = 1$ and ends in event $s_T = K$, where T is the total number of time points and K is the total number of events. In each time point the subject either remains in the same state or advances to the next state, i.e., $s_{t+1} \in \{s_t, s_t+1\}$ for all time points t. Each event has a signature mean activity pattern m_k across all V voxels in a region of interest, and the observed brain activity b_t at any time point t is assumed to be highly correlated with m_k, as illustrated in Figure 2.

Given the sequence of observed brain activities b_t, our goal is to infer both the event signatures m_k and the event structure s_t. To accomplish this, we cast our model as a variant of a Hidden Markov Model (HMM). The latent states are the events s_t that evolve according to a simple transition matrix, in which all elements are zero except for the diagonal (corresponding to $s_{t+1} = s_t$) and the adjacent off-diagonal (corresponding to $s_{t+1} = s_t+1$), and the observation model is an isotropic Gaussian $p(b_t|s_t=k) = (1/\sqrt{2\pi\sigma^2})e^{-(1/2\sigma^2)\|z(b_t)-z(m_k)\|_2^2}$, where $z(x)$ denotes z-scoring an input vector x to have zero mean and unit variance. Note that, due to this z-scoring, the log probability of observing brain state b_t in an event with signature m_k is simply proportional to the Pearson correlation between b_t and m_k plus a constant offset.

The HMM is fit to the fMRI data by using an annealed version of the Baum-Welch algorithm, which iterates between estimating the fMRI signatures m_k and the latent event structure s_t. Given the signature estimates m_k, the distributions over latent events $p(s_t = k)$ can be computed using the forward-backward algorithm. Given the distributions $p(s_t = k)$, the updated estimates for the signatures m_k can be computed as the weighted average $m_k = \sum_t p(s_t=k)b_t / \sum_t p(s_t=k)$. To encourage convergence to a high-likelihood solution, we anneal the observation variance σ^2 as $4 \cdot 0.98^i$ where i is the number of loops of Baum-Welch completed so far. We stop the fitting procedure when the log-likelihood begins to decrease, indicating that the observation variance has begun to drop below the actual event activity variance. We can also fit the model simultaneously to multiple datasets; on each round of Baum-Welch, we run the forward-backward algorithm on each dataset separately, and then average across all datasets to compute a single set of shared signatures m_k.

The end state requirement of our model – that all states should be visited, and the end state should be symmetrical to all other states – requires extending the traditional HMM by modifying the observation probabilities $p(b_t|s_t=k)$. First, we enforce $s_T=K$ by requiring that, on the final timestep, only the final state K could have generated the data, by setting $p(b_T|s_T=k)=0$ for all $k \neq K$. Equivalently, we can view this as a modification of the backward pass, by initializing the backward message $\beta(s_T=k)$ to 1 for $k=K$ and 0 otherwise. Second, we must modify the transition matrix to ensure that all valid event segmentations (which start at event 1 and end at event K, and proceed monotonically through all events) have the same prior probability. Formally, we introduce a dummy absorbing state K+1 to which state K can transition, ensuring that the transition probabilities for state K are identical to those for previous states, and then set $p(b_t|s_t=K+1)=0$ to ensure that this state is never actually used.

Since we do not want to assume that events will have the same relative lengths across different datasets (such as a movie and audio-narration version of the same narrative), we fix all states to have the same probability of staying in the same state ($s_{t+1} = s_t$) versus jumping to the next state ($s_{t+1} = s_t+1$). Note that the shared probability of jumping to the next state can take any value between

0 and 1 with no effect on the results (up to a normalization constant in the log-likelihood), since every valid event segmentation will contain exactly the same number of jumps (K-1).

Our model induces a prior over the locations of the event boundaries. There are a total of $\binom{T}{K-1}$ equally likely placements of the K-1 event boundaries, and the number of ways to have event boundary k fall on time point t is the number of ways that k-1 boundaries can be placed in t-1 time points times the number of ways that (K-1)-(k-1)-1 boundaries can be placed in T-t time points. Therefore $p(s_t = k \ \& \ s_{t+1} = k+1) = \binom{t-1}{k-1}\binom{T-t}{K-k-1} \Big/ \binom{T}{K-1}$. An example of this distribution is shown in Figure S1. During the annealing process, the distribution over boundary locations starts at this prior, and slowly adjusts to match the event structure of the data.

After fitting the model on one set of data, we can then look for the same sequence of events in another dataset. Using the signatures m_k learned from the first dataset, we simply perform a single round of the forward-backward algorithm to obtain event estimates $p(s_t = k)$ on the second dataset. If we expect the datasets to have similar noise properties (e.g., both datasets are group-averaged data from the same number of subjects), we set the observation variance to the final σ^2 obtained while fitting the first dataset. When transferring events learned on group-averaged data to individual subjects, we estimate the variance for each event across the individual subjects of the first dataset.

The model implementation was first verified using simulated data. An event-structured dataset was constructed with V = 10 voxels, K = 10 events, and T = 500 time points. The event structure was chosen to be either uniform (with 50 time points per event), or the length of each event was sampled (from first to last) from $N(1,0.25)$*(time points remaining)/(events remaining). A mean pattern was drawn for each event from a standard normal distribution, and the simulated data for each time point was the sum of the event pattern for that time point plus randomly distributed noise with zero mean and varying standard deviation. The noisy data were then input to the event segmentation model, and we measured the fraction of the event boundaries that were exactly recovered from the true underlying event structure. As shown in Figure S1, were able to recover a majority of the event boundaries even when the noise level was as large as the signature patterns themselves.

Finding event structure in narratives

To validate our event segmentation model on real fMRI data, we first fit the model to group-averaged PCC data from the Interleaved Stories experiment. In this experiment, we expect that an event boundary should be generated every time the stimulus switches stories, giving a ground truth against which to compare the model's segmentations. As shown in Figure S2, our method was highly effective at identifying events, with the majority of the identified boundaries falling close to a story switch.

The following subsections describe how the model was used to obtain each of the experimental results, with subsection titles corresponding to subsections of the Results.

Timescales of cortical event segmentation

We applied the model in a searchlight to the whole-brain movie-watching data from the Sherlock Recall study. Cubical searchlights were scanned throughout the volume at a step size of 3 voxels and with a side length of 7 voxels. For each searchlight, the event segmentation model was applied to group-averaged data from all but one subject. We measured the robustness of the identified boundaries by testing whether these boundaries explained the data in the held-out subject. We measured the spatial correlation between all pairs of time points that were separated by four time points, and then binned these correlations according to whether the pair of time points fell within the same event or crossed over an event boundary. The average difference between the within- versus across-event correlations was used as an index of how well the learned boundaries captured the temporal structure of the held-out subject. The analysis was repeated for every possible held-out subject, and with a varying number of events from K = 10 to K = 120. After averaging the results across subjects, the number of events with the best within- versus across-event correlations was chosen as the optimal number of events for this searchlight.

Since the topography of the results was similar to previous work on temporal receptive windows, we compared the map of the optimal number of events with the short and medium/long timescale maps derived by measuring intersubject correlation for intact versus scrambled movies (Chen et al., 2016). The histogram of the optimal number of events for voxels was computed within each of the timescale maps.

Comparison of event boundaries across regions and to human annotations

We defined four equally-sized regions along the cortical hierarchy by taking the centers of mass of the early visual cortex, hV4, angular gyrus and PCC ROIs in each hemisphere, and finding the nearest 150 voxels to these centers (yielding 300 bilateral voxels for each region). For each region, we calculated its optimal number of events using the same within- versus across-event correlation procedure described in the previous section, and then fit a final event segmentation model with the optimal number of events (using the Sherlock Recall data). We measured the match between levels of the hierarchy by computing the fraction of boundaries in the "upper" (slower) level that were close to a boundary in the "lower" (faster) level. We defined "close to" as "within three time points," since the typical uncertainty in the model about exactly where an event switch occurred was approximately three time points.

To measure similarity to the human annotations, we first constructed a "consensus" annotation from the four observers, consisting of boundary locations that were within three time points of boundaries marked by at least two observers. We then measured the match between the consensus boundaries and the boundaries from each region, treating the consensus boundaries as the "upper" level. To ensure that differences between regions were not driven by timescale differences, for this comparison we refit the event segmentation model to each cortical region using the same number of events as in the consensus annotation (rather than using each region's optimal timescale).

Shared event structure across modalities

After fitting the event segmentation model to a searchlight of movie-watching data from the Sherlock Narration experiment, we took the learned event signatures m_k and used them to run the forward-backward algorithm on the audio narration data, to test whether audio narration of a story elicited the same sequence of events as a movie of that story. Since both the movie and audio data were averaged at the group level, they should have similar levels of noise, and therefore we simply used the fit movie variance σ^2 for the observation variance.

Relationship between cortical event boundaries and hippocampal encoding

After applying the event segmentation model throughout the cortex to the Sherlock Recall study as described above, we measured whether the data-driven event boundaries were related to activity in the hippocampus. For a given cortical searchlight, we extracted a window of mean hippocampal activity around each of the searchlight's event boundaries. We then averaged these windows together, yielding a profile of boundary-triggered hippocampal response according to this region's boundaries. To assess whether the hippocampus showed a significant increase in activity related to these event boundaries, we measured the mean hippocampal activity for the 10 time points following the event boundary minus the mean activity for the 10 time points preceding the event boundary.

Reinstatement of event patterns during free recall

For each region of interest, we fit the event segmentation model as described above (on the group-averaged Sherlock Recall data). We then took the learned sequence of event signatures m_k and ran the forward-backward algorithm on each individual subject's recall data. We set the variance of each event's observation model by computing the variance within each event in the movie-watching data of individual subjects, pooling across both time points and subjects. The analysis was run for 10 events to 60 events in steps of 5.

We operationalized the overall reinstatement of an event k, as $\sum_t p(s_t = k)$; that is, the sum across all recall time points of the probability that the subject was recalling perceptual event k at that time point. We measured whether this per-event re-activation during recall could be predicted during movie-watching, based on the hippocampal response at the end of the event. For each subject, we computed the difference between hippocampal activity after versus before the event boundary as above. We then averaged the event re-activation and hippocampal offset response across subjects, and measured their correlation. For comparison purposes, we also performed the same analysis but with hippocampal differences at the *beginning* of each event, rather than the end, and with the mean hippocampal activity throughout the event.

Anticipatory reinstatement for a familiar narrative

To determine whether memory changed the event correspondence between the movie and narration, we then fit the segmentation model simultaneously to group-averaged data from the movie-watching condition, audio narration no-memory condition, and audio narration with memory condition, yielding a sequence of events in each condition with the same activity signatures. We computed the correspondence between the movie states $s_{m,t}$ and the audio no-memory states $s_{anm,t}$ as as $p(s_{m,t_1} = s_{anm,t_2}) = \sum_k p(s_{m,t_1} = k) \cdot p(s_{anm,t_2} = k)$, and similarly for the audio memory states $s_{am,t}$. We computed the differences between the group correspondences as $\sum_{t1} \sum_{t2} (p(s_{m,t_1} = s_{anm,t_2}) - p(s_{m,t_1} = s_{am,t_2}))^2$. For visualization, we also computed how far the memory correspondence was ahead of the no-memory correspondence as the mean over t_2 of the difference in the expected values $\sum_{t_1} t_1 p(s_{m,t_1} = s_{anm,t_2}) - \sum_{t_1} t_1 p(s_{m,t_1} = s_{am,t_2})$. We also performed the same analysis but with only the two narration conditions, computing the correspondence between the audio memory and audio no-memory states as $p(s_{am,t_1} = s_{anm,t_2}) = \sum_k p(s_{am,t_1} = k) \cdot p(s_{anm,t_2} = k)$. Since deviation from a diagonal correspondence would indicate anticipation in the memory group, we measured the expected deviation from the diagonal as $\sum_{t_1} \sum_{t_2} (|t_1 - t_2| / \sqrt{2}) p(s_{am,t_1} = s_{anm,t_2})$, and for visualization calculated the amount of anticipation as $\sum_t t p(s_{am,t_1} = s_{anm,t_2}) - T/2$.

QUANTIFICATION AND STATISTICAL ANALYSIS

Permutation or resampling analyses were used to statistically evaluate all of the results. As above, the analyses for each subsection are presented under a corresponding heading.

Timescales of cortical event segmentation

To generate a null distribution, the same analysis was performed except that the event boundaries were scrambled before computing the within- versus across-event correlation. This scrambling was performed by reordering the events with their durations held

constant, to ensure that the null events had the same distribution of event lengths as the real events. The within versus across difference for the real events compared to 1000 null events was used to compute a z value, which was converted to a p value using the normal distribution. The p values were Bonferroni corrected for the 12 choices of the number of events, and then the false discovery rate q was computed using the same calculation as in AFNI (Cox, 1996).

Comparison of event boundaries across regions and to human annotations

For each pairwise comparison between regions or between a region and the human annotations, we scrambled the event boundaries using the same duration-preserving procedure described above to produce 1000 null match values. The true match value was compared to this distribution to compute a z value, which was converted to a p value. To assess whether two cortical regions had significantly *different* matches to human annotations, we scrambled the boundaries from the regions (keeping the human annotations intact), and computed the fraction of scrambles for which the difference in the match to human annotations was larger in the null data than the original data.

Shared event structure across modalities

We compared the log-likelihood of the fit to the narration data against a null model in which the movie event signatures m_k were randomly re-ordered, and computed the z value of the true log-likelihood compared to 100 null shuffles, then converted to a p value. This null hypothesis test therefore assessed whether the narration exhibited *ordered* reactivation of the events identified during movie-watching.

Relationship between cortical event boundaries and hippocampal encoding

For each searchlight, we compared the difference in hippocampal activity for the 10 time points after an event boundary compared to 10 time points before an event boundary, both on the true boundaries and on shuffled event boundaries (using the duration-preserving procedure described above). The z value for this difference was computed to a p value, and then transformed to a false discovery rate q.

Reinstatement of event patterns during free recall

As in the movie-to-narration analysis, we compared the log-likelihood of the movie-recall fit to a null model in which the order of the event signatures was shuffled before fitting to the recall data, which yielded a z value that was converted to a p value. When measuring the match to human annotations, we compared to the same shuffled-event null models.

To assess the robustness of the encoding activity versus reinstatement correlations, we performed a bootstrap test, in which we resampled subjects (with replacement, yielding 17 subjects as in the original dataset) before taking the average and computing the correlation. The p value was defined as the fraction of 1000 resamples that yielded correlations with a different sign from the true correlation.

Anticipatory reinstatement for a familiar narrative

To determine if the correspondence with the movie was significantly different between the memory and no-memory conditions, we created null groups by averaging together a random half of the no-memory subjects with a random half of memory subjects, and then averaging together the remaining subjects from each group, yielding two group-averaged time courses whose correspondences should differ only by chance. We calculated a z value based on the correspondence difference for real versus null groups, which was converted to a p value and then corrected to a false discovery rate q. The analysis using only the narration conditions was performed similarly, computing a z value based on the expected deviation from the diagonal in the real versus null groups.

DATA AND SOFTWARE AVAILABILITY

All of the primary data used in this study are drawn from other papers (Chen et al., 2017; Zadbood et al., 2016), and the "Interleaved Stories" posterior cingulate cortex data are available on request. Implementations of our event segmentation model, along with simulated data examples, are available on GitHub at https://github.com/intelpni/brainiak (python) and at https://github.com/cbaldassano/Event-Segmentation (MATLAB).

Supplemental Information

Discovering Event Structure

in Continuous Narrative Perception and Memory

Christopher Baldassano, Janice Chen, Asieh Zadbood, Jonathan W. Pillow, Uri Hasson, and Kenneth A. Norman
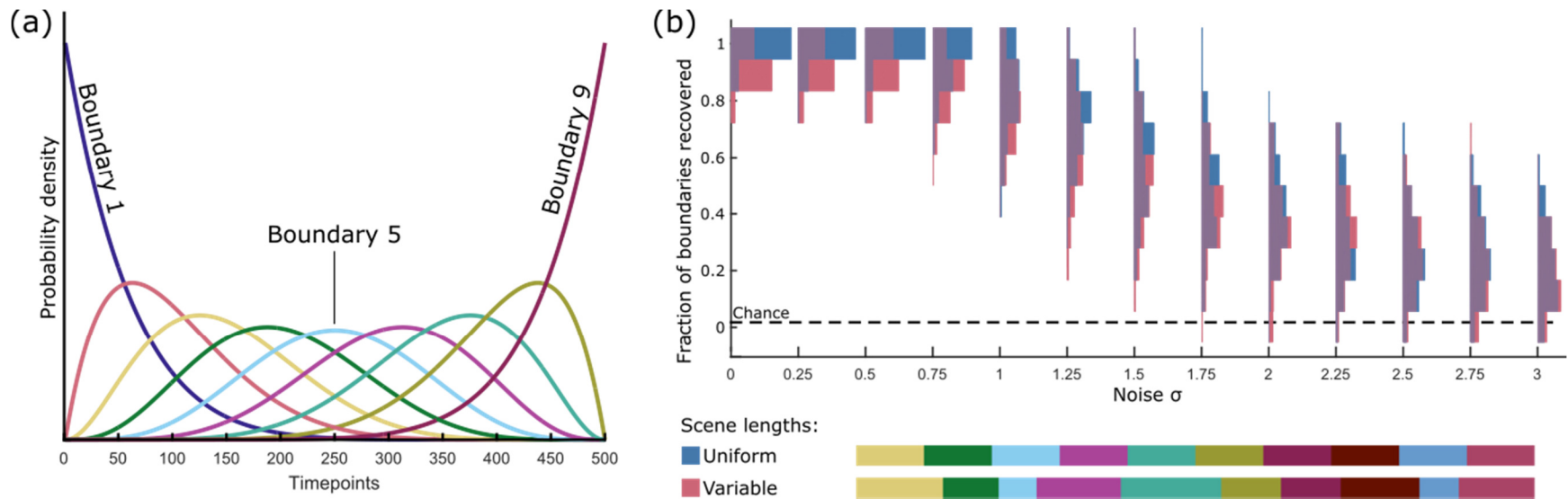
Figure S1: **Properties of the event segmentation model (related to Fig. 2).** (a) Our event segmentation model defines a uniform prior over all possible event segmentations in which every event occurs for at least one timepoint and all events occur in order. This induces a prior distribution over event boundaries, which depends on the number of timepoints T and the number of events K (T=500, K=10 in this figure). During the annealing process, the distribution of boundaries starts at this prior, which allows for a (highly uncertain) first estimate of the signature neural pattern for each event. Based on these patterns, the latent events for all timepoints are refit, and then the patterns are recalculated. The process continues, with the pattern variance slowly decreasing, until the log likelihood reaches a peak. (b) Simulated data with a discrete event structure obscured by varying levels of noise was input to the segmentation model, with T=500, K=10, and V=10. The model successfully recovers a majority of the underlying event boundaries at low noise levels, and can still identify an above-chance fraction of boundaries even at high noise levels that are as large as the differences between the event patterns. Having variable event lengths leads to only a small loss in performance, and does not change the overall performance curve.
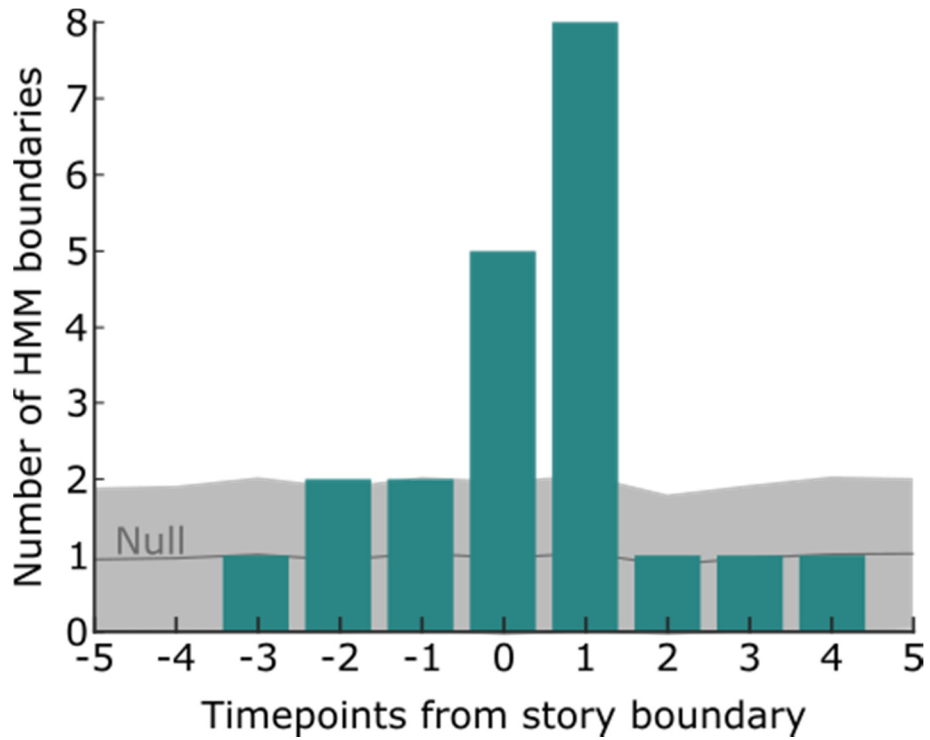
Figure S2: **The event segmentation model successfully identifies switches between stories (related to Fig 2).** Subjects listened to two stories, which were interleaved such that they alternated back and forth about every minute. Using data from PCC, an event segmentation model with 34 event transitions showed the best fit to held-out subjects (very close to the actual number of 32). Of these 34 transitions, the majority (20) were within 3 timepoints of a story switch. A null distribution was created by permuting the order of the events (preserving event lengths); under this null distribution the chance of having this many event boundaries close to true story switches was p<0.001.
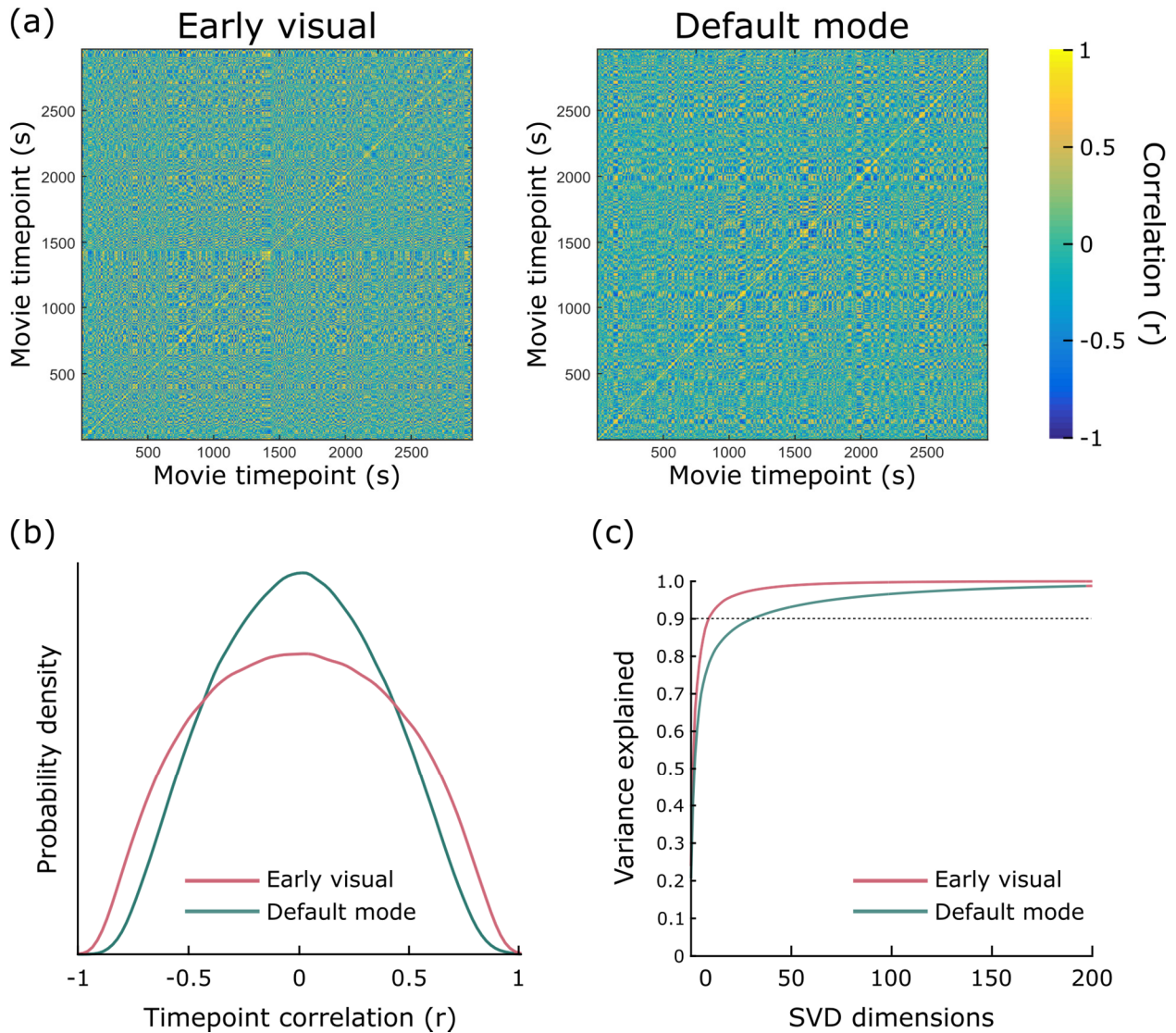
Figure S3. **Timepoint similarity matrices (related to Figs. 2, 3).** (a) For each pair of timepoints in the group average response to the 50-minute movie, similarity between activity patterns was assessed using Pearson correlation (in early visual cortex and default mode regions, angular gyrus and posterior cingulate cortex). These matrices exhibit blocks along the diagonal (characteristic of stable event patterns) as well as substantial off-diagonal structure. (b) These matrices both exhibit a range of correlation values, but the correlations are more concentrated near +/-1 in the early visual cortex. (c) Using SVD to estimate the dimensionality of the group timecourses, we find that capturing 90% of the variance requires 10 dimensions in visual cortex and 32 in default mode regions.
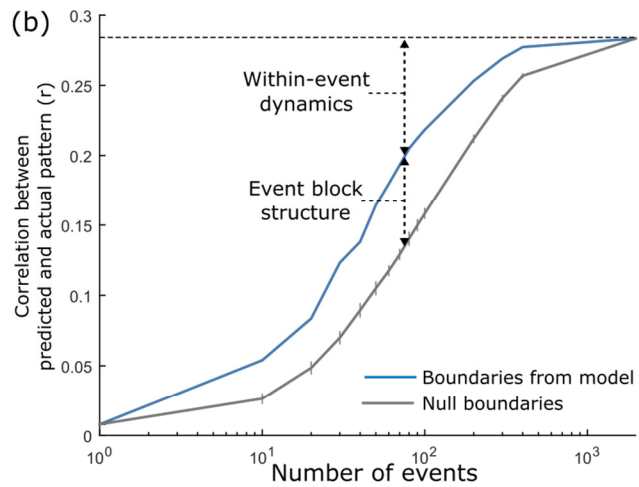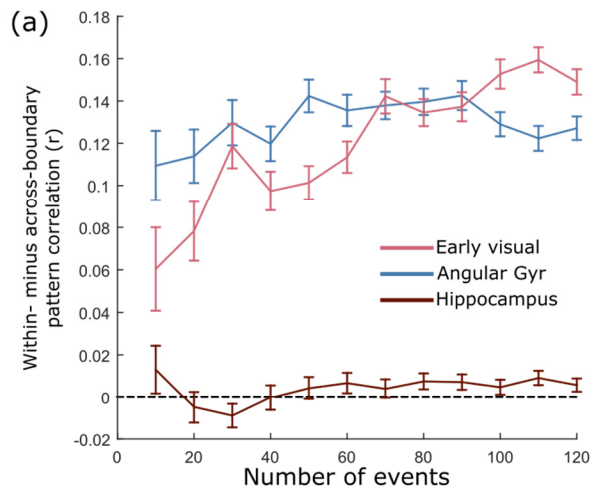
(a) (b) (c) (d)
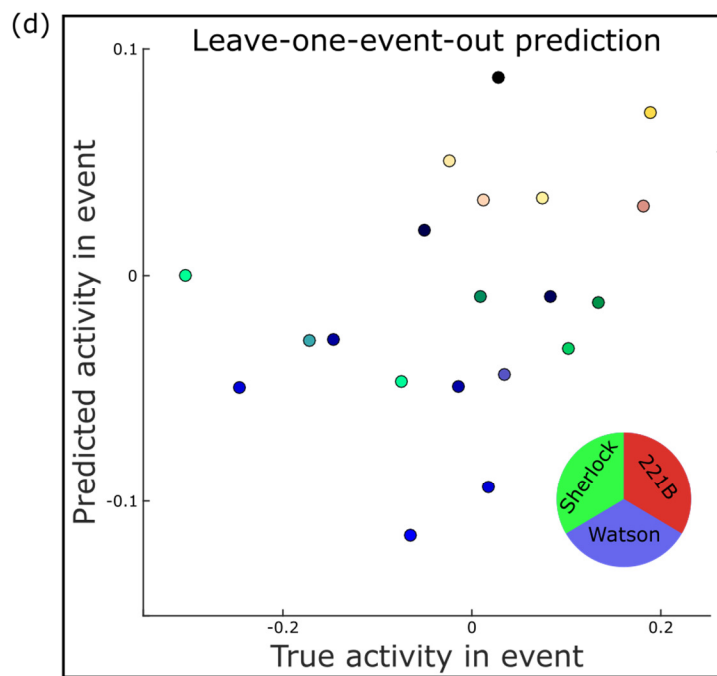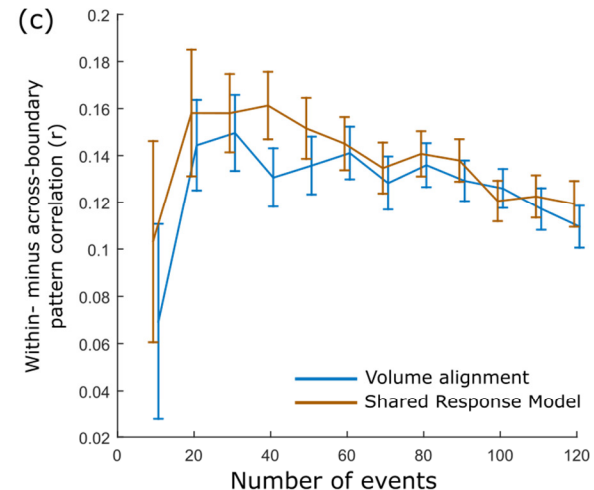
Figure S4. **Additional analyses of event structure evoked by the *Sherlock* movie (related to Fig 3).** (a) Early visual cortex and angular gyrus show similar goodness-of-fit (measured based on boundary prediction in a held-out subject) but have a different optimal number of events, with angular gyrus peaking around 50-90 events and visual cortex peaking around 110 events. The hippocampus did not show any significant event-structured activity patterns. Error bars indicate standard deviation of the null (scrambled-boundary) distribution. (b) To quantify the variance captured by the event segmentation model, we segmented the movie responses (in angular gyrus) from all but one of the subjects and used the average event activities to predict the responses in the held-out subject. At the optimal timescale (~75 events), this prediction is substantially better than a null model (in which the boundaries are permuted with the same event length distribution), indicating the existence of event structure at this timescale (error bars indicate 95% CI). There is also a gap between the event-structured prediction and a model-free prediction based on the continuous average of all other subjects (dashed line, equivalent to setting the number of events equal to the number of timepoints), indicating continuous dynamics on the scale of ~5 TRs that are predictable across subjects but not tied to specific boundary timepoints (since they are also captured by the null boundaries). (c) To determine whether our results were sensitive to the way alignment was performed across subjects, we used the first half of the movie (from the angular gyrus) to fit a Shared Response Model (SRM) with 50 features (Chen et al., 2015). The SRM seeks to maximize functional correspondence across subjects by learning a projection matrix for each subject from their native voxel space to a shared, low-dimensional space. We then ran the event segmentation procedure on the second half of the movie, both in the original volume space and in the shared response space (using the learned projection matrices). In both cases significant event structure was found, peaking at 20-40 events in both cases (for the last 25 minutes of the movie), with even stronger effects using SRM. Error bars indicate standard deviation of the null (scrambled-boundary) distribution. (d) To investigate whether the model-identified event patterns were related to event content, we fit a linear model to predict voxel activation for each event based on human annotations. *Sherlock* was divided into 1000 fine-grained segments (approximately 3 seconds each), and each was annotated with its location and the characters that were visible in the shot. The three annotations with the highest variance were Sherlock Holmes, John Watson, and 221B Baker Street (Sherlock and John's apartment). Each of 20 model-identified events in posterior cingulate cortex was assigned a value between 0 and 1 for each of these three annotations, calculated as the fraction of segments in which each of these three annotations was present. Using linear ridge regression, we learned to predict the activity of each voxel from these annotations using all but one of the events, and then tested our regression model on a held-out event. A prediction for an example voxel is shown in the box on the left, where each point indicates one event and has an RGB color corresponding to its three annotations (note that this voxel was selected post-hoc for illustration). Prediction performance showed a gradient throughout the region; splitting along the long axis of PCC, we found above-chance prediction performance (coefficient of determination greater than zero for held-out events) for 92.9% of voxels in the posterior/inferior half (significantly better than the permuted model, which showed 26.0% predictive voxels on average, p=0.023) and 46.2% of voxels in the anterior/superior half (not significantly different from the permuted model, which showed 25.1% predictive voxels on average, p=0.23). For a more detailed encoding-model analysis of information represented in this region, see J. Chen et al. (2017, Figure S6).
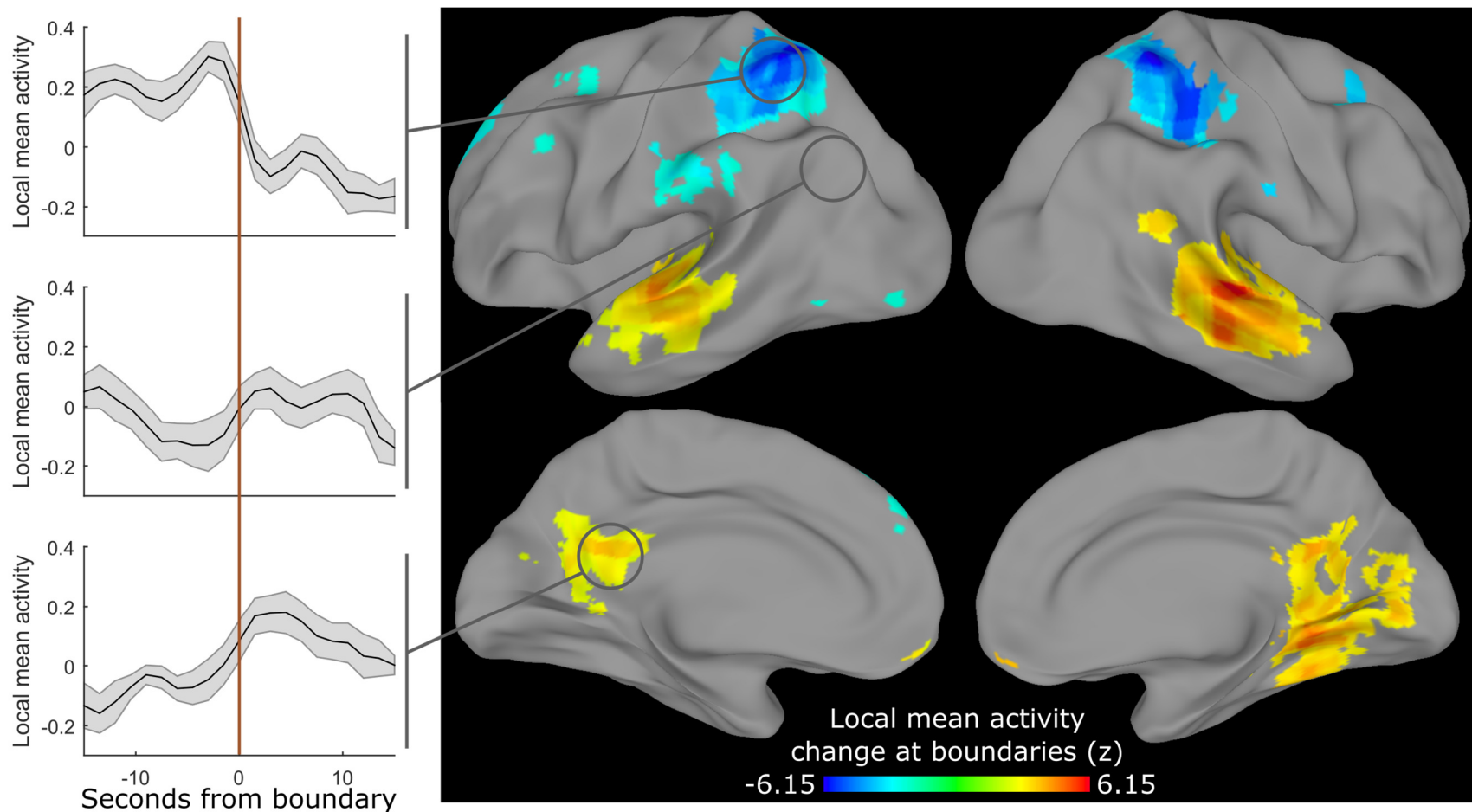
Figure S5. **Relationship between changes in activity patterns and local univariate activity (related to Fig. 6).** In Fig. 6, we identified regions whose event boundaries were related to increases in hippocampal activity. To determine if this effect could be driven by a relationship between multivariate event boundaries and *local* univariate activity, here we identified event boundaries for each searchlight using all but one subject (at that searchlight's optimal timescale) and then measured changes in the mean univariate activity of this searchlight around these boundaries in the held-out subject. Interestingly, there was no simple relationship between multivariate changes and local univariate activity, with some regions such as posterior cingulate cortex (bottom inset) showing *increases* in activity after boundaries, others such as the superior parietal lobule (top inset) showing *decreases* in activity, and others such as angular gyrus (middle inset) showing no significant change (shaded region indicates 95% confidence interval). Note that the searchlight insets were selected post-hoc for illustration.
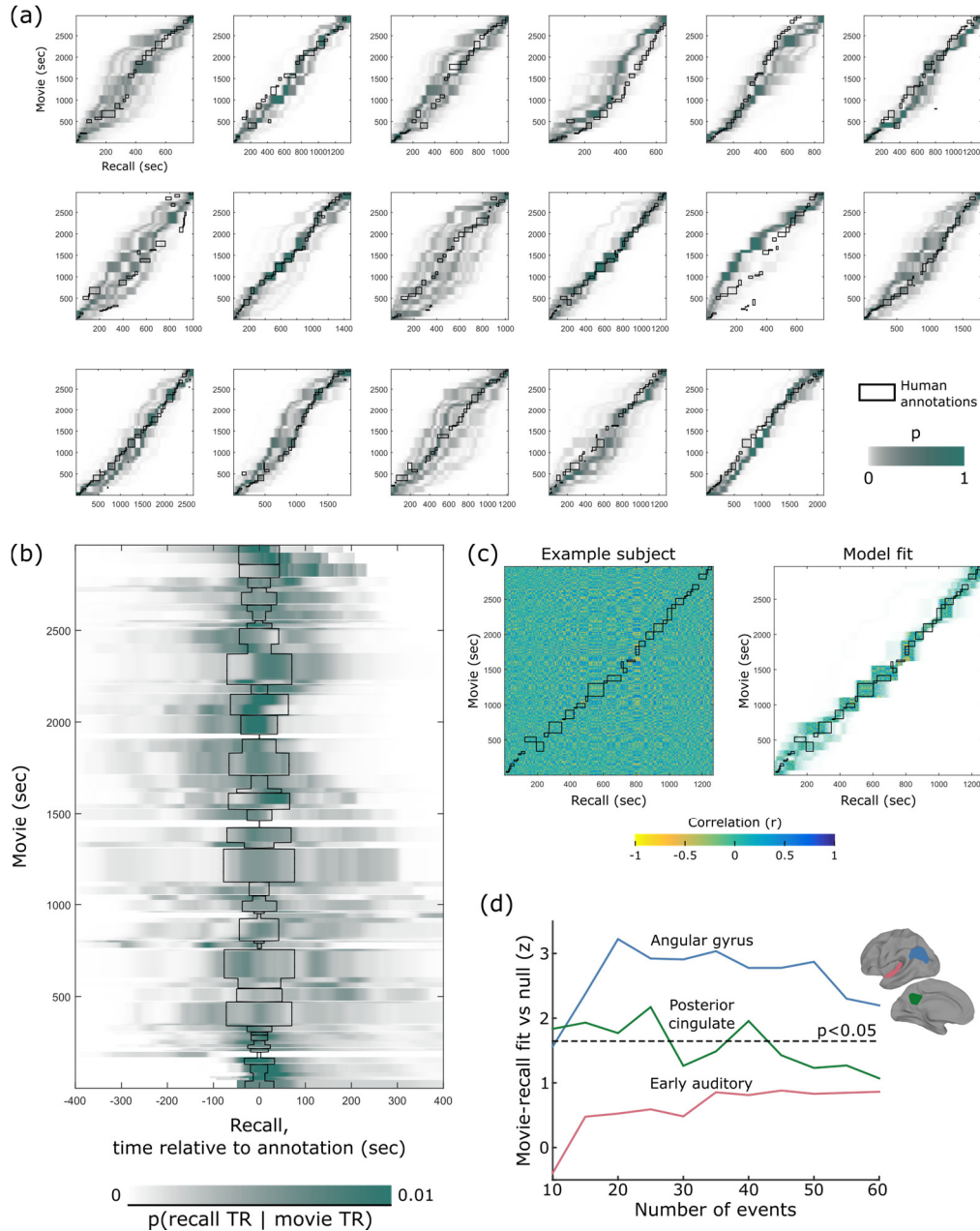
6

Figure S6: **Additional analyses of movie-recall correspondence (related to Fig. 7).** (a) The correspondence between movie-watching and free recall for each subject was estimated using the event segmentation model (using a combined ROI of angular gyrus and posterior cingulate cortex). Black boxes indicate the correspondence annotated by human observers. (b) To combine results across subjects, we re-centered each row of these correspondences relative to the human annotations and then averaged across subjects (for each movie timepoint, only subjects that recalled that timepoint of the movie were included in the average). Black boxes indicate the maximum duration of the recall for each movie scene, across subjects. This group correspondence shows that the model places the bulk of its probability mass in or near the human annotations, despite not using annotations during the fitting process. (c) When the timepoint-timepoint correlation matrix (left) is masked by the model-identified correspondence (right), we observe that the model selects a path with positive correlation values. (d) The results shown in main Fig. 6b hold for most choices of the number of latent events between 10 and 40, with decreasing goodness-of-fit for larger numbers of events. Note that the best fits were achieved with models having approximately 20-25 events, similar to the minimum number of human-labeled events recalled by the subjects (24, table S1 in J. Chen et al., 2017).
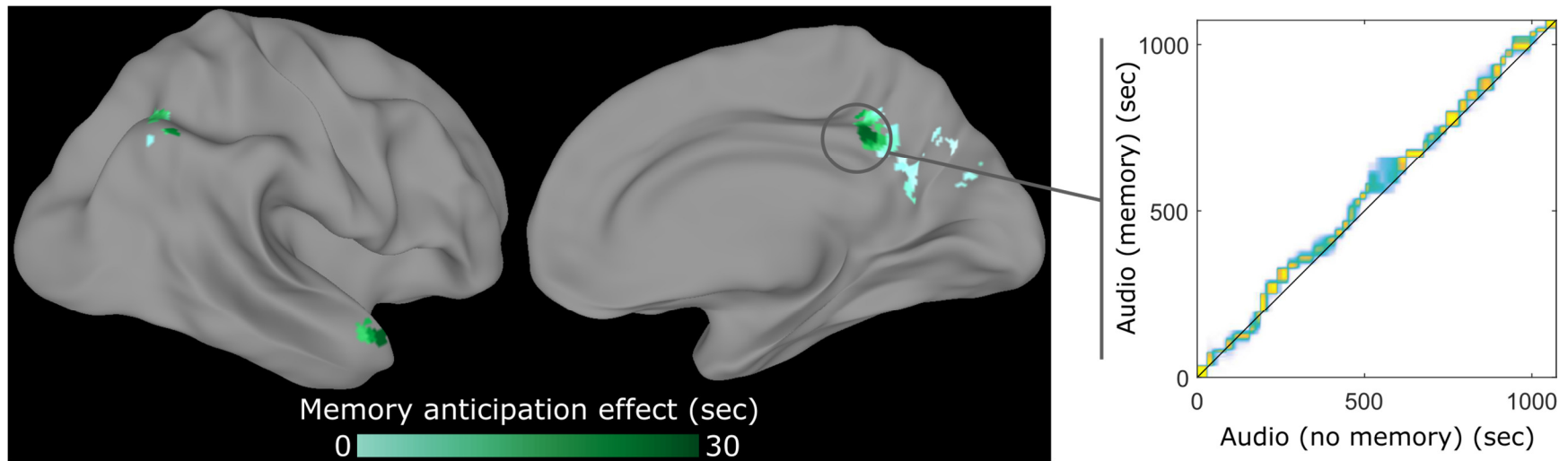
Figure S7. **Measuring anticipation effects without reference to the movie data (related to Fig. 8).** Rather than aligning the movie-watching, listening with memory, and listening without memory datasets as in Fig. 8, we can instead align just the two listening conditions and measure the deviation of the model correspondence from a straight diagonal line. For a given probabilistic correspondence, we computed the expected distance from the diagonal and compared to a null model in which subjects were randomly permuted between the memory and no-memory conditions. The probability that the memory group significantly led the no-memory group was converted to a false discovery rate and thresholded at $q < 0.05$. Since this analysis uses less data, the results are noisier, but still show the same basic result, with the strongest effects in posterior cingulate cortex, and smaller effects on the angular gyrus and anterior temporal lobe (the medial frontal lobe also showed an effect at a weaker threshold). Note that the searchlight inset was selected post-hoc for illustration.
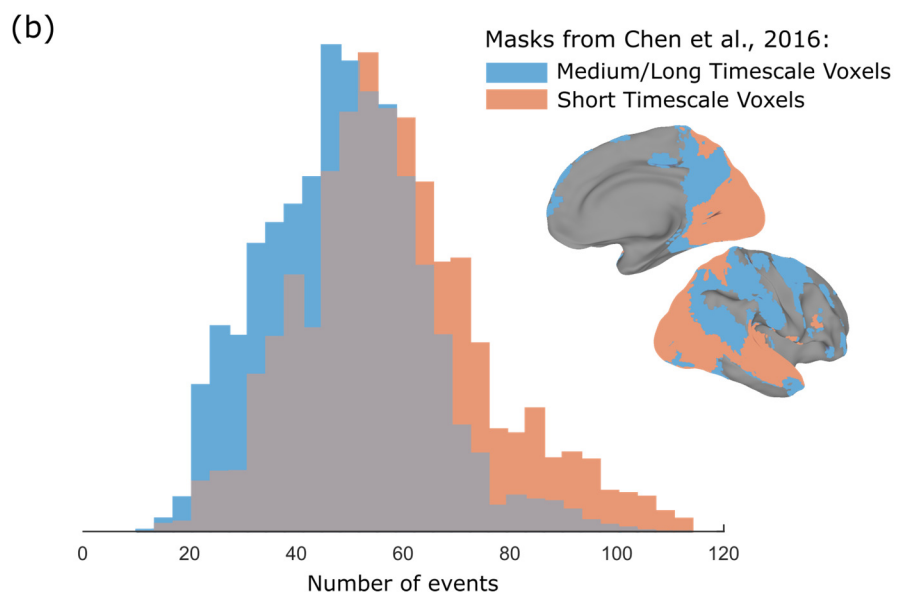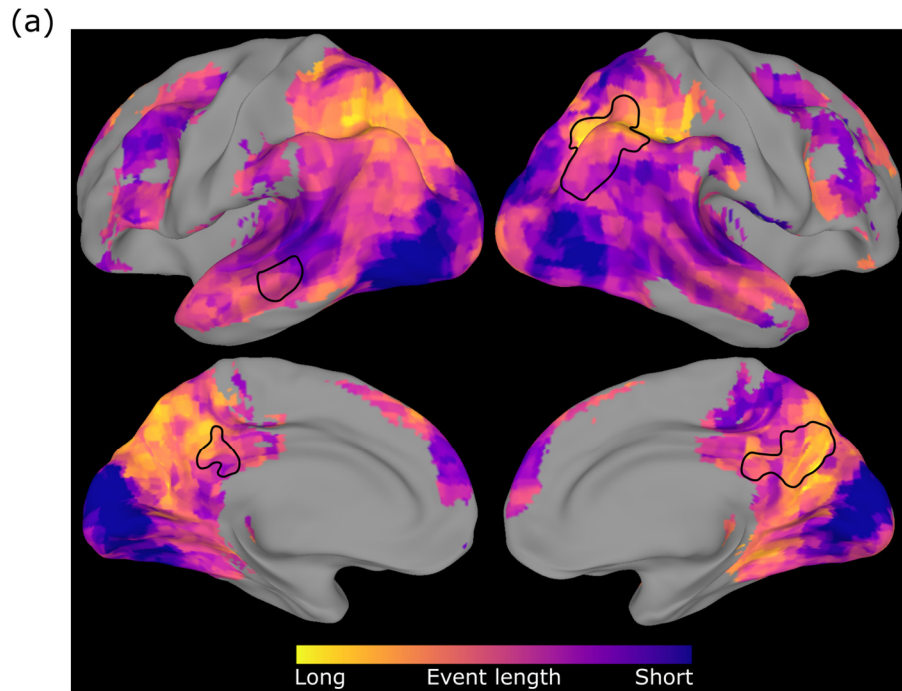
Figure S8: **Comparison of cortical topographies across analyses and with previous work (related to Figs. 3, 5, 6, 8).** (a) The optimal number of events during movie watching (from Fig. 3) is overlaid with the regions (black outlines) the show effects in all three maps in Fig. 5 (regions showing across-modality correspondence), Fig. 6 (regions with event boundaries that predict hippocampal activity increases), and Fig. 8 (regions showing anticipatory reinstatement). The primary regions of overlap are the angular gyrus and posterior medial cortex (and inferior temporal cortex), all of which exhibit long event segments. (b) The optimal number of events during movie watching was compared to a map of voxel timescales (J. Chen et al., 2016), which was defined based on sensitivity to temporal scrambling of a movie. Although derived from very different types of experimental data, these two approaches yield similar topographies. The majority of regions with a small number of (long) events have medium/long temporal receptive windows (blue), while the majority of regions with a large number of (short) events have short temporal receptive windows (orange).