

1 Geometric models reveal behavioral and neural
2 signatures of how naturalistic experiences are
3 transformed into episodic memories

4 Andrew C. Heusser^{1, 2, †}, Paxton C. Fitzpatrick^{1, †}, and Jeremy R. Manning^{1, *}

¹Department of Psychological and Brain Sciences

Dartmouth College, Hanover, NH 03755, USA

²Akili Interactive

Boston, MA 02110

[†]Denotes equal contribution

^{*}Corresponding author: Jeremy.R.Manning@dartmouth.edu

5 August 18, 2020

6 **Abstract**

7 The ways in which our experiences unfold over time are, by nature, incalculably complex.
8 One popular approach to making sense of this complexity entails characterizing experiences as
9 unique *trajectories* through a particular geometric representational space whose dimensions are
10 features of our experiences. In word embedding spaces, for example, each dimension reflects a
11 concept. And by extension, any given coordinate reflects a weighted mixture of those concepts.
12 We propose a framework for projecting naturalistic experiences into word embedding spaces,
13 such that the conceptual content of each moment of an experience, and how different moments
14 of the experience relate, are captured by the *shape* of the experience's trajectory. By projecting
15 memories of those experiences into the same spaces, one may then geometrically compare the

shape of the original experience's trajectory to the shape of how it is remembered later. According to this view, encoding an experience into memory entails geometrically distorting or transforming the original experience's trajectory. This translates qualitative questions about how we remember naturalistic experiences into quantitative geometric comparisons. We applied our framework to data collected as participants watched and verbally recounted a television episode while undergoing functional neuroimaging. We found that the trajectory of each participant's recall reflected the high-level *essence* (i.e., large-scale narrative summaries) of the episode's trajectory, but participants differed markedly in their memories for low-level features (i.e., small-scale details). We also identified a network of brain structures that were sensitive to the shape of the episode's trajectory through word embedding space, and an overlapping network that predicted, at the time of encoding, how people would distort (transform) the episode's trajectory when they recounted the episode later. Our work provides insights into how our brains transform ongoing experiences when we encode them into episodic memories, and provides a formal geometric framework for characterizing the complex dynamic content of naturalistic experiences.

Introduction

What does it mean to *remember* something? In traditional episodic memory experiments (e.g., list-learning or trial-based experiments; Murdock, 1962; Kahana, 1996), remembering is often cast as a discrete and binary operation: each studied item may be separated from the rest of one's experience and singularly labeled as having been recalled or forgotten. More nuanced studies might incorporate self-reported confidence measures as a proxy for memory strength, or ask participants to discriminate between "recollecting" the (contextual) details of an experience or having a general feeling of "familiarity" (Yonelinas, 2002). Using well controlled, trial-based experimental designs, the field has amassed a wealth of valuable information regarding human episodic memory. However, there are fundamental properties of the external world and our memories that trial-based experiments are not well suited to capture (for review, also see Koriat and Goldsmith, 1994; Huk et al., 2018). First, our experiences and memories are continuous, rather than discrete—isolating a (naturalistic) event from the context in which it occurs can substantially

43 change its meaning. Second, the specific language used to describe an experience has little bearing
44 on whether the experience should be considered to have been “remembered.” Asking whether
45 the rememberer has precisely reproduced a specific set of words to describe a given experience
46 is nearly orthogonal to whether or not they were actually able to remember it. In classic (e.g.,
47 list-learning) memory studies, by contrast, the number or proportion of exact recalls is often a
48 primary metric for assessing the quality of participants’ memories. Third, one might remember
49 the *essence* (or a general summary) of an experience but forget (or neglect to recount) particular
50 details. Capturing the essence of what happened is typically the main “point” of recounting a
51 memory to a listener, while the addition of highly specific details may add comparatively little to
52 successful conveyance of an experience.

53 How might one go about formally characterizing the *essence* of an experience, and whether it
54 has been recovered by the rememberer? Any given moment of an experience derives meaning
55 from surrounding moments, as well as from longer-range temporal associations (Lerner et al., 2011;
56 Manning, 2019; ?). Therefore, the timecourse describing how an event unfolds is fundamental to
57 its overall meaning. Further, this hierarchy formed by our subjective experiences at different
58 timescales defines a *context* for each new moment (e.g., Howard and Kahana, 2002; Howard et al.,
59 2014), and plays an important role in how we interpret that moment and remember it later (for
60 review see Manning et al., 2015; ?). Our memory systems can leverage these associations to form
61 predictions that help guide our behaviors (Ranganath and Ritchey, 2012). For example, as we
62 navigate the world, the features of our subjective experiences tend to change gradually (e.g., the
63 room or situation we are in at any given moment is strongly temporally autocorrelated), allowing
64 us to form stable estimates of our current situation and behave accordingly (Zacks et al., 2007;
65 Zwaan and Radvansky, 1998).

66 Occasionally, this gradual “drift” of our ongoing experience is punctuated by sudden changes,
67 or “shifts” (e.g., when we walk through a doorway; Radvansky and Zacks, 2017). Prior research
68 suggests that these sharp transitions (termed *event boundaries*) help to discretize our experiences
69 (and their mental representations) into *events* (Radvansky and Zacks, 2017; Brunec et al., 2018;
70 Heusser et al., 2018a; Clewett and Davachi, 2017; Ezzyat and Davachi, 2011; DuBrow and Davachi,

71 2013). The interplay between the stable (within-event) and transient (across-event) temporal
72 dynamics of an experience also provides a potential framework for transforming experiences
73 into memories that distills those experiences down to their essence. For example, prior work
74 has shown that event boundaries can influence how we learn sequences of items (Heusser et al.,
75 2018a; DuBrow and Davachi, 2013), navigate (Brunec et al., 2018), and remember and understand
76 narratives (Zwaan and Radvansky, 1998; Ezzyat and Davachi, 2011). Prior research has implicated
77 a network of brain regions (including the hippocampus and the medial prefrontal cortex) as playing
78 a critical role in transforming experiences into structured and consolidated memories (Tomparay
79 and Davachi, 2017).

80 Here, we sought to examine how the temporal dynamics of a “naturalistic” experience were
81 later reflected in participants’ memories. We analyzed an open dataset that comprised behavioral
82 and functional Magnetic Resonance Imaging (fMRI) data collected as participants viewed and then
83 verbally recounted an episode of the BBC television series *Sherlock* (Chen et al., 2017). We developed
84 a computational framework for characterizing the temporal dynamics of the moment-by-moment
85 content of the episode, and of participants’ verbal recalls. Specifically, we use topic modeling (Blei
86 et al., 2003) to characterize the thematic conceptual (semantic) content present in each moment
87 of the episode and recalls, and hidden Markov models (Rabiner, 1989; Baldassano et al., 2017) to
88 discretize this evolving semantic content into events. In this way, we cast naturalistic experiences
89 (and memories of those experiences) as geometric *trajectories* that describe how the experiences
90 evolve over time. Under this framework, successful remembering entails verbally “traversing”
91 the content trajectory of the episode, thereby reproducing the shape (or essence) of the original
92 experience. Comparing the shapes of the topic trajectories of the episode and of participants’
93 retellings of the episode then reveals which aspects of the episode were preserved (or discarded) in
94 the translation into memory. We further introduce two novel metrics for assessing memory quality:
95 1) the *precision* with which a participant recounts each event, and 2) the *distinctiveness* of each recall
96 event (relative to other recalled events). We examine how these metrics relate to overall memory
97 performance, and discuss the ways in which they improve upon classic “proportion-recalled”
98 measures for analyzing naturalistic memory. Last, we utilize our framework to identify networks

99 of brain structures whose responses (as participants watched the episode) reflected the temporal
100 dynamics of either the episode or how participants would later recount it.

101 Results

102 To characterize the dynamic content of the *Sherlock* episode and participants' subsequent recounts,
103 we used a topic model (Blei et al., 2003) to discover the episode's latent themes. Topic models
104 take as inputs a vocabulary of words to consider and a collection of text documents, and return two
105 output matrices. The first of these is a *topics matrix* whose rows are *topics* (latent themes) and whose
106 columns correspond to words in the vocabulary. The entries of the topics matrix reflect how each
107 word in the vocabulary is weighted by each discovered topic. For example, a detective-themed
108 topic might weight heavily on words like "crime," and "search." The second output is a *topic*
109 *proportions matrix*, with one row per document and one column per topic. The topic proportions
110 matrix describes what mixture of discovered topics is reflected in each document.

111 Chen et al. (2017) collected hand-annotated information about each of 1000 (manually identified)
112 scenes spanning the roughly 50 minute video used in their experiment. This information included:
113 a brief narrative description of what was happening, the location where the scene took place, the
114 names of any characters on the screen, and other similar details (for a full list of annotated features,
115 see *Methods*). We took from these annotations the union of all unique words (excluding stop
116 words, such as "and," "or," "but," etc.) across all features and scenes as the "vocabulary" for the
117 topic model. We then concatenated the sets of words across all features contained in overlapping,
118 sliding windows of (up to) 50 scenes, and treated each window as a single "document" for the
119 purpose of fitting the topic model. Next, we fit a topic model with (up to) $K = 100$ topics to this
120 collection of documents. We found that 32 unique topics (with non-zero weights) were sufficient
121 to describe the time-varying content of the video (see *Methods*; Figs. 1, S2). Note that our approach
122 is similar in some respects to Dynamic Topic Models (Blei and Lafferty, 2006) in that we sought
123 to characterize how the thematic content of the episode evolved over time. However, whereas
124 Dynamic Topic Models are designed to characterize how the properties of *collections* of documents

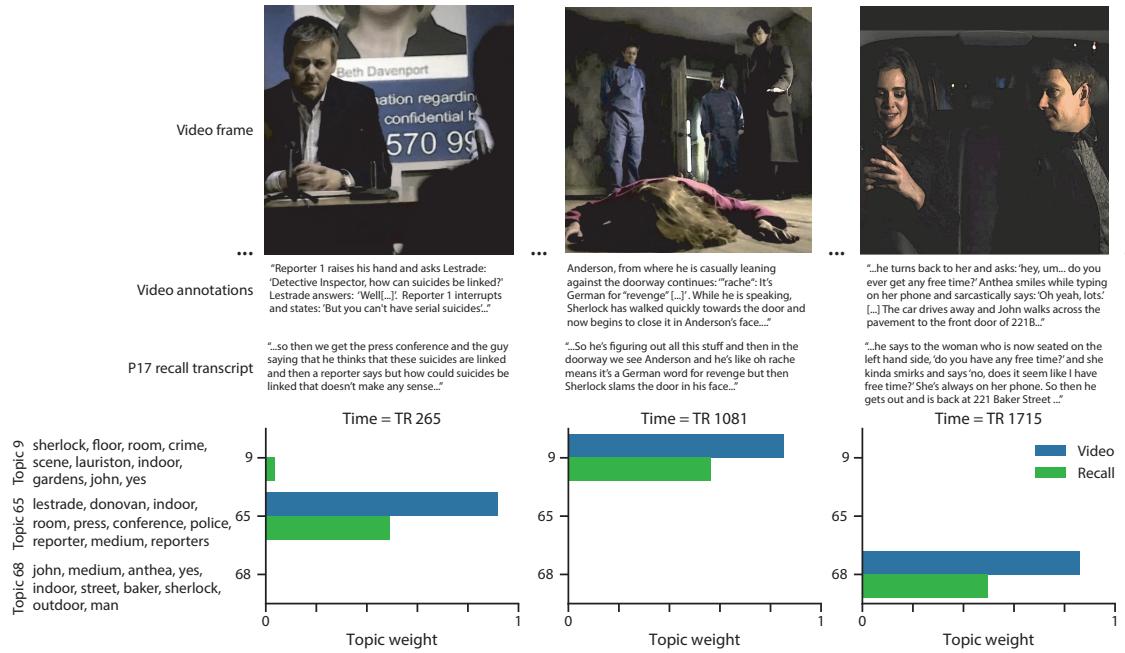


Figure 1: Methods overview. We used hand-annotated descriptions of each moment of video to fit a topic model. Three example video frames and their associated descriptions are displayed (top two rows). Participants later recalled the video (in the third row, we show example recalls of the same three scenes from participant 17). We used the topic model (fit to the annotations) to estimate topic vectors for each moment of video and each sentence the participants recalled. Example topic vectors are displayed in the bottom row (blue: video annotations; green: example participant's recalls). Three topic dimensions are shown (the highest-weighted topics for each of the three example scenes, respectively). We also show the 10 highest-weighted words for each topic. Figure S2 provides a full list of the top 10 words from each of the discovered topics.

125 change over time, our sliding window approach allows us to examine the topic dynamics within
 126 a single document (or video). Specifically, our approach yielded (via the topic proportions matrix)
 127 a single *topic vector* for each sliding window of annotations transformed by the topic model. We
 128 then stretched (interpolated) the resulting windows-by-topics matrix to match the time series of
 129 the 1976 fMRI volumes collected as participants viewed the episode.

130 The 32 topics we found were heavily character-focused (i.e., the top-weighted word in each
 131 topic was nearly always a character) and could be roughly divided into themes centered around
 132 Sherlock Holmes (the titular character), John Watson (Sherlock's close confidant and assistant),
 133 supporting characters (e.g., Inspector Lestrade, Sergeant Donovan, or Sherlock's brother Mycroft),

134 or the interactions between various groupings of these characters (see Fig. S2). Several of the
135 identified topics were highly similar, which we hypothesized might allow us to distinguish between
136 subtle narrative differences if the distinctions between those overlapping topics were meaningful.
137 The topic vectors for each timepoint were also *sparse*, in that only a small number (usually one
138 or two) of topics tended to be “active” in any given timepoint (Fig. 2A). Further, the dynamics
139 of the topic activations appeared to exhibit *persistence* (i.e., given that a topic was active in one
140 timepoint, it was likely to be active in the following timepoint) along with *occasional rapid changes*
141 (i.e., occasionally topics would appear to spring into or out of existence). These two properties
142 of the topic dynamics may be seen in the block diagonal structure of the timepoint-by-timepoint
143 correlation matrix (Fig. 2B) and reflect the gradual drift and sudden shifts fundamental to the
144 temporal dynamics of real-world experiences. Given this observation, we adapted an approach
145 devised by Baldassano et al. (2017), and used a hidden Markov model (HMM) to identify the *event*
146 *boundaries* where the topic activations changed rapidly (i.e., the boundaries of the blocks in the
147 temporal correlation matrix; event boundaries identified by the HMM are outlined in yellow in
148 Fig. 2B). Part of our model fitting procedure required selecting an appropriate number of “events”
149 into which the topic trajectory should be segmented. To accomplish this, we used an optimization
150 procedure that maximized the difference between the topic weights for timepoints within an event
151 versus timepoints across multiple events (see *Methods* for additional details). We then created a
152 stable “summary” of the content within each video event by averaging the topic vectors across the
153 timepoints spanned by each event (Fig. 2C).

154 Given that the time-varying content of the video could be segmented cleanly into discrete
155 events, we wondered whether participants’ recalls of the video also displayed a similar structure.
156 We applied the same topic model (already trained on the video annotations) to each participant’s
157 recalls. Analogously to how we parsed the time-varying content of the video, to obtain similar
158 estimates for each participant’s recall, we treated each overlapping window of (up to 10) sentences
159 from their transcript as a “document,” and computed the most probable mix of topics reflected in
160 each timepoint’s sentences. This yielded, for each participant, a number-of-windows by number-
161 of-topics topic proportions matrix that characterized how the topics identified in the original video



Figure 2: Modelling naturalistic stimuli and recalls. All panels: darker colors indicate greater values; range: [0, 1]. **A.** Topic vectors ($K = 100$) for each of the 1976 video timepoints. **B.** Timepoint-by-timepoint correlation matrix of the topic vectors displayed in Panel A. Event boundaries discovered by the HMM are denoted in yellow (30 events detected). **C.** Average topic vectors for each of the 30 video events. **D.** Topic vectors for each of 265 sliding windows of sentences spoken by an example participant while recalling the video. **E.** Timepoint-by-timepoint correlation matrix of the topic vectors displayed in Panel D. Event boundaries detected by the HMM are denoted in yellow (22 events detected). For similar plots for all participants, see Figure S4. **F.** Average topic vectors for each of the 22 recalled events from the example participant. **G.** Correlations between the topic vectors for every pair of video events (Panel C) and recalled events (from the example participant; Panel F). For similar plots for all participants, see Figure S5. **H.** Average correlations between each pair of video events and recalled events (across all 17 participants). To create the figure, each recalled event was assigned to the video event with the most correlated topic vector (yellow boxes in panels G and H).

162 were reflected in the participant's recalls. Note that an important feature of our approach is that
163 it allows us to compare participants' recalls to events from the original video, despite different
164 participants using widely varying language to describe the events, and that those descriptions
165 often diverged in content and quality from the video annotations. This is a substantial benefit of
166 projecting the video and recalls into a shared "topic" space. An example topic proportions matrix
167 from one participant's recalls is shown in Figure 2D.

168 Although the example participant's recall topic proportions matrix has some visual similarity to
169 the video topic proportions matrix, the time-varying topic proportions for the example participant's
170 recalls are not as sparse as those for the video (compare Figs. 2A and D). Similarly, although
171 there do appear to be periods of stability in the recall topic dynamics (i.e., most topics are active
172 or inactive over contiguous blocks of time), the changes in topic activations that define event
173 boundaries appear less clearly delineated in participants' recalls than in the episode's annotations.
174 To examine these patterns in detail, we computed the timepoint-by-timepoint correlation matrix
175 for the example participant's recall trajectory (Fig. 2E). As in the video correlation matrix (Fig. 2B),
176 the example participant's recall correlation matrix has a strong block diagonal structure, indicating
177 that their recalls are discretized into separated events. As for the video correlation matrix, we
178 leveraged an HMM-based optimization procedure (see *Methods*) to determine how many events
179 are reflected in the participant's recalls and where specifically the event boundaries fall (outlined
180 in yellow). We carried out a similar analysis on all 17 participants' recall topic proportions matrices
181 (Fig. S4).

182 Two clear patterns emerged from this set of analyses. First, although every individual partic-
183 ipant's recalls could be segmented into discrete events (i.e., every individual participant's recall
184 correlation matrix exhibited clear block diagonal structure; Fig. S4), each participant appeared to
185 have a unique *recall resolution*, reflected in the sizes of those blocks. While some participants' recall
186 topic proportions segmented into just a few events (e.g., Participants P4, P5, and P7), others' seg-
187 mented into many shorter duration events (e.g., Participants P12, P13, and P17). This suggests that
188 different participants may be recalling the video with different levels of detail—i.e., some might
189 touch on just the major plot points, whereas others might attempt to recall every minor scene or

190 action. The second clear pattern present in every individual participant's recall correlation matrix
191 was that, unlike in the video correlation matrix, there were substantial off-diagonal correlations.
192 Whereas each event in the original video was (largely) separable from the others (Fig. 2B), in
193 transforming those separable events into memory, participants appeared to be integrating across
194 multiple events, blending elements of previously recalled and not-yet-recalled content into each
195 newly recalled event (Figs. 2E, S4; also see Manning et al., 2011; Howard et al., 2012; Manning,
196 2019).

197 The above results indicate that both the structure of the original video and participants' recalls
198 of the video exhibit event boundaries that can be identified automatically by characterizing the
199 dynamic content using a shared topic model and segmenting the content into events via HMMs.
200 Next, we asked whether some correspondence might be made between the specific content of the
201 events the participants experienced in the video, and the events they later recalled. One approach
202 to linking the experienced (video) and recalled events is to label each recalled event as matching
203 the video event with the most similar (i.e., most highly correlated) topic vector (Figs. 2G, S5). This
204 yields a sequence of "presented" events from the original video, and a (potentially differently
205 ordered) sequence of "recalled" events for each participant. Analogous to classic list-learning
206 studies, we can then examine participants' recall sequences by asking which events they tended
207 to recall first (probability of first recall; Fig. 3A; Atkinson and Shiffrin, 1968; Postman and Phillips,
208 1965; Welch and Burnett, 1924); how participants most often transition between recalls of the
209 events as a function of the temporal distance between them (lag-conditional response probability;
210 Fig. 3B; Kahana, 1996); and which events they were likely to remember overall (serial position
211 recall analyses; Fig. 3C; Murdock, 1962). Interestingly, for two of these analyses (probability of
212 first recall and lag-conditional response probability curves) we observed patterns comparable to
213 classic effects from list-learning literature: namely, a higher probability of initiating recall with the
214 first event in the sequence (Fig. 3A) and a higher probability of transitioning to neighboring events
215 with an asymmetric forward bias (Fig. 3B). In contrast, we did not observe a pattern comparable
216 to the serial position effect (Fig. 3C), but rather greater memory for specific events distributed
217 approximately evenly throughout the video.

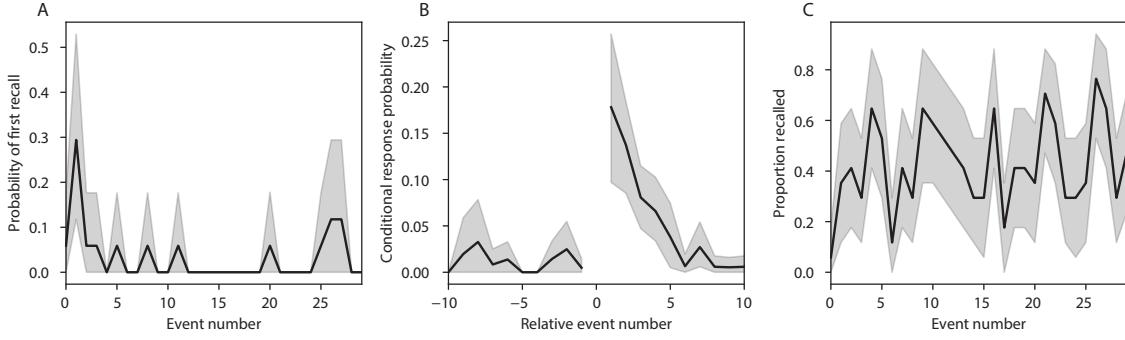


Figure 3: Naturalistic extensions of classic list-learning memory analyses. **A.** The probability of first recall as a function of the serial position of the event in the video. **B.** The probability of recalling each event, conditioned on having most recently recalled the event *lag* events away in the video. **C.** The proportion of participants who recalled each event, as a function of the serial position of the events in the video. All panels: error ribbons denote bootstrap-estimated standard error of the mean.

218 We can also apply two list-learning-native analyses that describe how participants group items
 219 in their recall sequences: temporal clustering and semantic clustering (Polyn et al., 2009, see
 220 *Methods* for details). Temporal clustering refers to the extent to which participants group their
 221 recall responses according to encoding position. Overall, we found that sequentially viewed video
 222 events were clustered heavily in participants' recall event sequences (mean clustering score: 0.767,
 223 SEM: 0.029), and that participants with higher temporal clustering scores tended to perform better
 224 according to both Chen et al. (2017)'s hand-annotated memory scores (Pearson's $r(15) = 0.62$, $p =$
 225 0.008) and our model's estimate (Pearson's $r(15) = 0.54$, $p = 0.024$). Semantic clustering measures
 226 the extent to which participants cluster their recall responses according to semantic similarity.
 227 We found that participants tended to recall semantically similar video events together (mean
 228 clustering score: 0.787, SEM: 0.018), and that semantic clustering score was also related to both
 229 hand-annotated (Pearson's $r(15) = 0.65$, $p = 0.004$) and model-derived (Pearson's $r(15) = 0.63$, $p =$
 230 0.007) memory performance.

231 Statistical models of memory studies often treat recall success as binary (in other words, an
 232 item either was or was not recalled), or occasionally categorical (e.g., to distinguish familiarity
 233 from recollection; Yonelinas et al., 2002). Such approaches are tenable in classical list-learning or
 234 recognition memory paradigms, as the presented stimuli tend to be very simple (e.g., a sequence

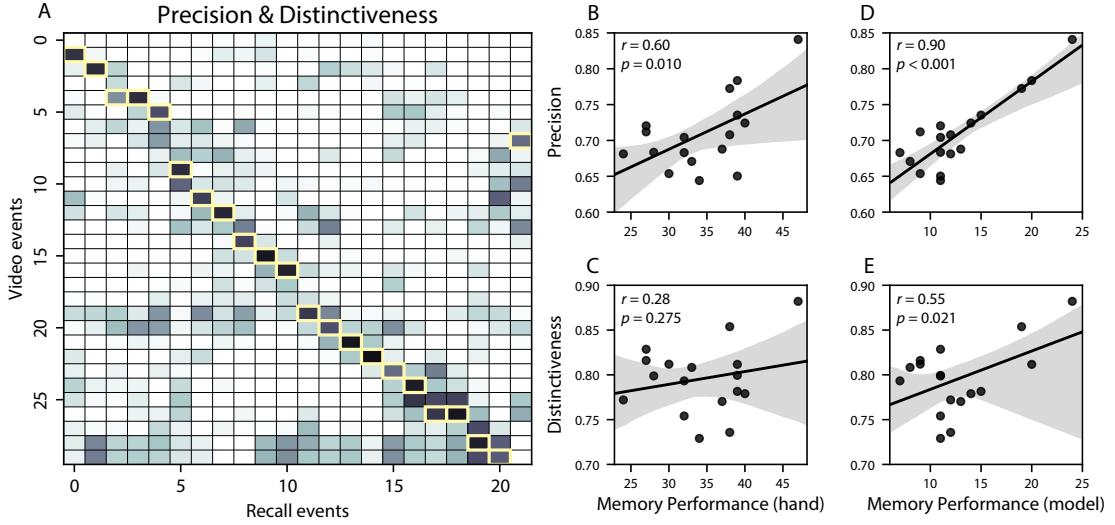


Figure 4: Novel content-based metrics of naturalistic memory: precision and distinctiveness. **A.** The video-recall correlation matrix for a representative participant (17). The yellow boxes highlight the maximum correlation in each column. The example participant's overall precision score was computed as the average across correlation values in the yellow boxes. Their distinctiveness score was computed as the average (over recall events) of 1 minus the average correlation between each recall event and all other recall events that do not display a box in the same row. **B.** The (Pearson's) correlation between precision and hand-annotated memory performance. **C.** The correlation between distinctiveness and hand-annotated memory performance. **D.** The correlation between precision and the number of video events successfully recalled, as determined by our model. **E.** The correlation between distinctiveness and the number of video events successfully recalled, as determined by our model.

of individual words or items). However, memory for naturalistic experiences is much more nuanced. For example, certain aspects of an experience might be correctly remembered at varying levels of detail, or distorted, or forgotten entirely. Further, each remembering is itself a richly structured phenomenon. Our framework produces a content-based model of individual video and recall events by projecting the dynamic content of the video and participants' recalls into a shared topic space. This allows for direct, quantitative comparisons between all stimulus and recall events, as well as between the recall events themselves. Leveraging these content-based models of the stimulus/recall events, we developed two novel, *continuous* metrics for analyzing naturalistic memory: *precision* and *distinctiveness*. Precision is intended to capture the "completeness" of recall, or how fully the presented content was recapitulated in memory. We define a recall event's

precision as the maximum correlation between the topic proportions of that recall event and any video event (Fig. 4). A second novel metric we introduce here is *distinctiveness*, which is intended to capture the “specificity” of recall. In other words, distinctiveness quantifies the extent to which a given recalled event reflects the most similar presented event moreso than it does other presented events. To compute a recall event’s distinctiveness, we first identify the video event to which its topic vector is most strongly correlated. We then define distinctiveness as one minus the average correlation between the given recall event and all *other* video events. In addition to individual events, one may also use these metrics to describe each participant’s overall performance by averaging across a participant’s event-wise precision or distinctiveness scores. Participants whose recall events are more veridical descriptions of what happened in the video event will presumably have higher precision scores. We find that, across participants, higher precision scores are positively correlated with both hand-annotated memory performance (as collected by Chen et al., 2017; Pearson’s $r(15) = 0.60, p = 0.010$) and the number of video events successfully remembered, as determined by our model (Pearson’s $r(15) = 0.90, p < 0.001$). We also hypothesized that participants who recounted events in a more distinctive way would display better overall memory. We find that participants’ distinctiveness scores were correlated with our model’s estimated number of recall events (Pearson’s $r(15) = 0.55, p = 0.021$). However, we found no evidence that distinctiveness scores were correlated with hand-annotated memory performance (Pearson’s $r(15) = 0.28, p = 0.275$). We elaborate on this potential discrepancy in the *Discussion* section.

Further intuition for the behaviors captured by these two metrics may be gained by directly examining the content of the video and recalls our framework models. In Figure 5, we contrast recalls for the same video event (event 22) from two participants: one with a high precision score (P17), the other with a low precision score (P6). From the HMM-identified event boundaries, we recovered the set of annotations describing the content of an example video event (Fig. 5B), and divided them into different color-coded sections for each action or feature described. We then similarly recovered the set of sentences comprising the corresponding recall event for each of the two example participants. Because the recall sliding windows overlap heavily, and each

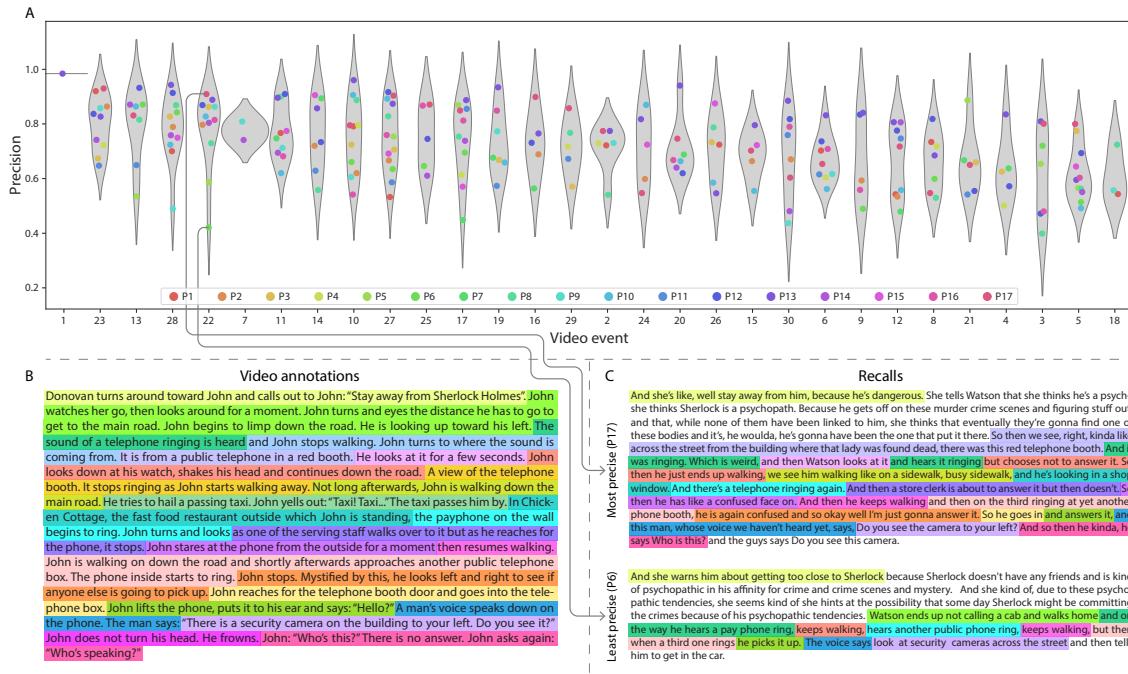


Figure 5: Precision metric reflects completeness of recall. **A.** Recall precision by video event. Grey violin plots display kernel density estimates for the distribution of recall precision scores for a single video event. Colored dots within each violin plot represent individual participants' recall precision for the given event. Video events are ordered along the *x*-axis by the average precision with which they were remembered. **B.** The set of "Narrative Details" video annotations (generated by Chen et al., 2017) for scenes comprising an example video event (22) identified by the HMM. Each action or feature is highlighted in a different color. **C.** A subset of the sentences comprising the most precise (P17) and least precise (P6) participants' recalls of video event 22. Descriptions of specific actions or features reflecting those highlighted in panel B are highlighted in the corresponding color.

recall event spans multiple recall timepoints (i.e., windows), we have stripped any sentences from the beginning and end that describe earlier or later video events for the sake of readability. In other words, Fig. 5C shows a subset of the full recall event text, comprising sentences between the first and last descriptions of content from the example video event. We then colored all words describing actions and features coded in panel B by their corresponding color. Visual comparison of the transcripts reveals that the most precise participant's recall both captures more of the video event's content, and does so with far more detail.

Figure 6 similarly contrasts two example participants' recalls for a common video event (event 19) to illustrate the tangible differences between high and low distinctiveness scores. Here, we

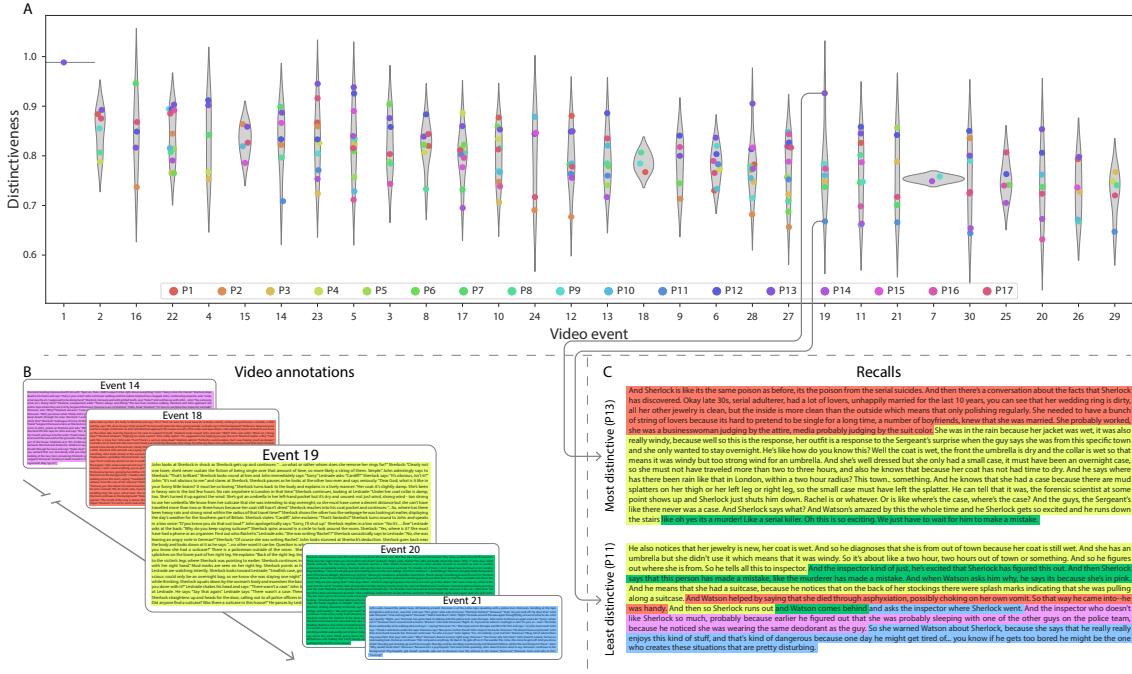


Figure 6: Distinctiveness metric reflects specificity of recall. A. Recall distinctiveness by video event. Kernel density estimates for each video event’s distribution of recall distinctiveness scores, analogous to Fig. 5A. B. The sets of “Narrative Details” video annotations (generated by Chen et al., 2017) for scenes comprising video events described by the example participants in panel C. Each event’s text is highlighted in a different color. C. The sentences comprising the most distinctive (P13) and least distinctive (P11) participants’ recalls of video event 19. Sections of recall describing each video event in panel B are highlighted with the corresponding color.

have extracted the full set of sentences comprising the most distinctive recall event (P13) and least distinctive recall event (P11) matched to the example video event (Fig. 6C). We also extracted the annotations for the example video event, as well as those from each other video event whose content the example participants' single recall events described (Fig. 6B). We then shaded the annotation text for each video event with a different color, and shaded each word of the example participants' recall text by the color of the video event it describes. The majority of the most distinctive recall event text describes video event 19's content, with the first five and last one sentence describing the video events immediately preceding and succeeding the current one, respectively. In contrast, the least precise participant's recall for video event 19 blends the content from five separate video events, does not transition between them in order, and often combines descriptions of two video

292 events' content in the same sentence.

293 The prior analyses leverage the correspondence between the 100-dimensional topic proportion
294 matrices for the video and participants' recalls to characterize recall. However, it is difficult to
295 gain deep insights into the content of (or relationships between) experiences and memories solely
296 by examining these topic proportions (e.g., Figs. 2A, D) or the corresponding correlation matrices
297 (Figs. 2B, E, S4). And while we can directly examine the original text underlying these topic
298 vectors (e.g., Figs. 5, 6) to show how relationships between them reflect real-world behavior, this
299 comparison becomes prohibitively cumbersome at larger timescales. To visualize the time-varying
300 high-dimensional content in a more intuitive way (Heusser et al., 2018b), we projected the topic
301 proportions matrices onto a two-dimensional space using Uniform Manifold Approximation and
302 Projection (UMAP; McInnes et al., 2018). In this lower-dimensional space, each point represents a
303 single video or recall event, and the distances between the points reflect the distances between the
304 events' associated topic vectors (Fig. 7). In other words, events that are nearer to each other in this
305 space are more semantically similar, and those that are farther apart are less so.

306 Visual inspection of the video and recall topic trajectories reveals a striking pattern. First, the
307 topic trajectory of the video (which reflects its dynamic content; Fig. 7A) is captured nearly perfectly
308 by the averaged topic trajectories of participants' recalls (Fig. 7B). To assess the consistency of these
309 recall trajectories across participants, we asked: given that a participant's recall trajectory had
310 entered a particular location in the reduced topic space, could the position of their *next* recalled
311 event be predicted reliably? For each location in the the reduced topic space, we computed the set of
312 line segments connecting successively recalled events (across all participants) that intersected that
313 location (see *Methods* for additional details). We then computed (for each location) the distribution
314 of angles formed by the lines defined by those line segments and a fixed reference line (the *x*-
315 axis). Rayleigh tests revealed the set of locations in topic space at which these across-participant
316 distributions exhibited reliable peaks (blue arrows in Fig. 7B reflect significant peaks at $p < 0.05$,
317 corrected). We observed that the locations traversed by nearly the entire video trajectory exhibited
318 such peaks. In other words, participants exhibited similar trajectories that also matched the
319 trajectory of the original video (Fig. 7C). This is especially notable when considering the fact that

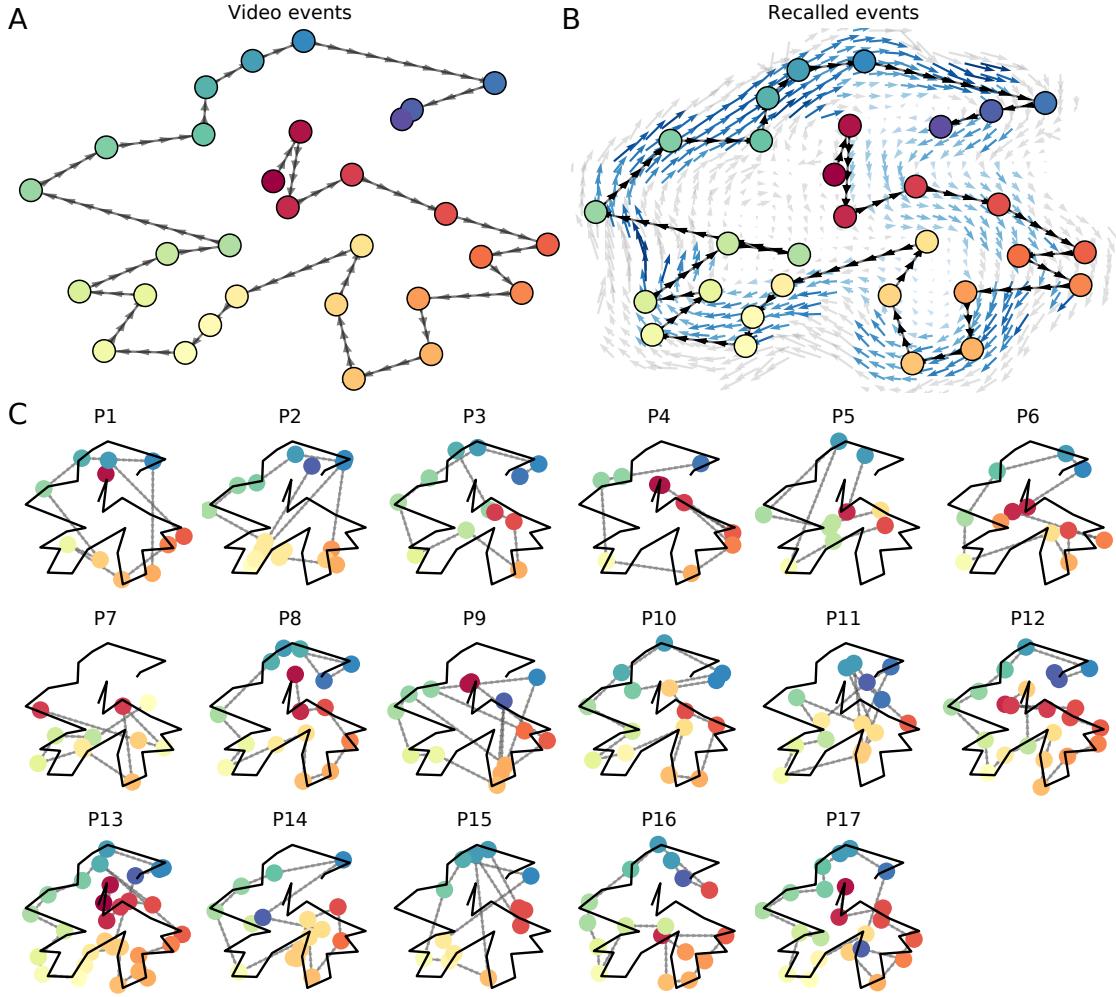


Figure 7: Trajectories through topic space capture the dynamic content of the video and recalls. All panels: the topic proportion matrices have been projected onto a shared two-dimensional space using UMAP. **A.** The two-dimensional topic trajectory taken by the episode of *Sherlock*. Each dot indicates an event identified using the HMM (see *Methods*); the dot colors denote the order of the events (early events are in red; later events are in blue), and the connecting lines indicate the transitions between successive events. **B.** The average two-dimensional trajectory captured by participants' recall sequences, with the same format and coloring as the trajectory in Panel A. To compute the event positions, we matched each recalled event with an event from the original video (see *Results*), and then we averaged the positions of all events with the same label. The arrows reflect the average transition direction through topic space taken by any participants whose trajectories crossed that part of topic space; blue denotes reliable agreement across participants via a Rayleigh test ($p < 0.05$, corrected). **C.** The recall topic trajectories (gray) taken by each individual participant (P1–P17). The video's trajectory is shown in black for reference. Here, events (dots) are colored by their matched video event (Panel A).

320 the number of events participants recalled (dots in Fig. 7C) varied considerably across people, and
321 that every participant used different words to describe what they had remembered happening in
322 the video. Differences in the numbers of remembered events appear in participants' trajectories
323 as differences in the sampling resolution along the trajectory. We note that this framework also
324 provides a means of disentangling classic "proportion recalled" measures (i.e., the proportion
325 of video events described in participants' recalls) from participants' abilities to recapitulate the
326 overall unfolding of the original video's content (i.e., the similarity between the shapes of the
327 original video trajectory and that defined by each participant's recounting of the video).

328 In addition to the more "holistic" measure of memory described in the previous section, our
329 framework also affords the ability to drill down to individual words and quantify how each word
330 relates to the memorability of each event. The results displayed in Figures 3C and 5A suggest that
331 certain events were remembered better than others. Given this, we next asked whether the
332 events were generally remembered well or poorly tended to reflect particular content. Because
333 our analysis framework projects the dynamic video content and participants' recalls into a shared
334 space, and because the dimensions of that space represent topics (which are, in turn, sets of weights
335 over words in the vocabulary), we are able to recover the weighted combination of words that make
336 up any point (i.e., topic vector) in this space. We first computed the average precision with which
337 participants recalled each of the 30 video events (Fig. 8A; note that this result is analogous to a serial
338 position curve created from our continuous recall quality metric). We then computed a weighted
339 average of the topic vectors for each video event, where the weights reflected how reliably each
340 event was recalled. To visualize the result, we created a "wordle" image (Mueller et al., 2018) where
341 words weighted more heavily by better-remembered topics appear in a larger font (Fig. 8B, green
342 box). Across the full video, content that reflected topics necessary to convey the central focus of the
343 video (e.g., the names of the two main characters, "Sherlock" and "John", and the address of a major
344 recurring location, "221B Baker Street") were best remembered. An analogous analysis revealed
345 which themes were poorly remembered. Here in computing the weighted average over events'
346 topic vectors, we weighted each event in *inverse* proportion to how well it was remembered (Fig. 8B,
347 red box). The least well-remembered video content reflected information not necessary to later

348 convey a general summary of the video, such as the proper names of relatively minor characters
349 (e.g., "Mike," "Molly," and "Lestrade") and locations (e.g., "St. Bartholomew's Hospital").

350 A similar result emerged from assessing the topic vectors for individual video and recall events
351 (Fig. 8C). Here, for each of the three best- and worst-remembered video events, we have constructed
352 two wordles: one from the original video event's topic vector (left) and a second from the average
353 recall topic vector for that event (right). The three best-remembered events (circled in green)
354 correspond to scenes important to the central plot-line: a mysterious figure spying on John in a
355 phone booth; John meeting Sherlock at Baker St. to discuss the murders; and Sherlock laying
356 a trap to catch the killer. Meanwhile, the three worst-remembered events (circled in red) reflect
357 scenes that are non-essential to summarizing the narrative's structure: the video of singing cartoon
358 characters participants viewed prior to the main episode; John asking Molly about Sherlock's habit
359 of over-analyzing people; and Sherlock noticing evidence of Anderson's and Donovan's affair.

360 The results thus far inform us about which aspects of the dynamic content in the episode partic-
361 ipants watched were preserved or altered in participants' memories. We next carried out a series
362 of analyses aimed at understanding which brain structures might facilitate these preservations
363 and transformations between the external world and memory. In the first analysis, we sought
364 to identify brain structures that were sensitive to the dynamic unfolding of the video's content,
365 as characterized by its topic trajectory. We used a searchlight procedure to identify clusters of
366 voxels whose activity patterns displayed a proximal temporal correlation structure (as participants
367 watched the video) matching that of the original video's topic proportions (Fig. 9A; see *Methods* for
368 additional details). In a second analysis, we sought to identify brain structures whose responses
369 (during video viewing) reflected how each participant would later structure their recounting of the
370 video. We used an analogous searchlight procedure to identify clusters of voxels whose proximal
371 temporal correlation matrices matched that of the topic proportions for each individual's recall
372 (Figs. 9B; see *Methods* for additional details). To ensure our searchlight procedure identified re-
373 gions *specifically* sensitive to the temporal structure of the video or recalls (i.e., rather than those
374 with a temporal autocorrelation length similar to that of the video/recalls), we performed a phase
375 shift-based permutation correction (see *Methods* for additional details). As shown in Figure 9C, the

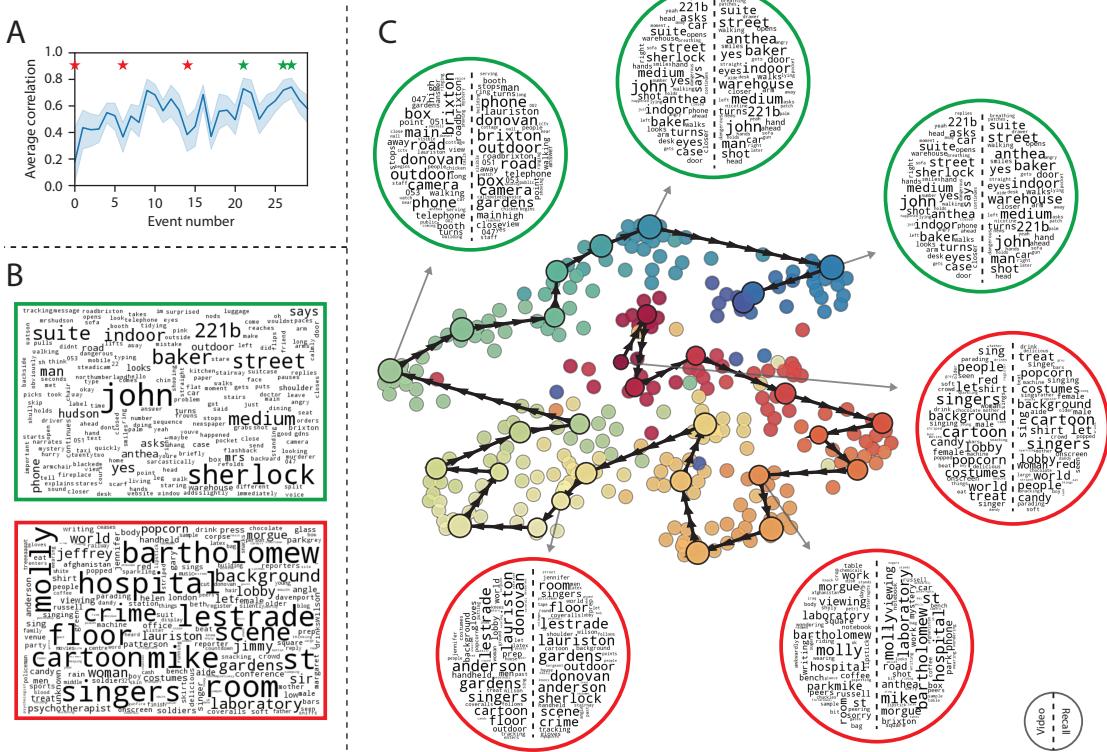


Figure 8: Language used in the most and least memorable events. **A.** Average precision (video event-recall event topic vector correlation) across participants for each video event. Error bars denote bootstrap-derived across-participant 95% confidence intervals. The stars denote the three best-remembered events (green) and worst-remembered events (red). **B.** Wordles comprising the top 200 highest-weighted words reflected in the weighted-average topic vector across video events. Green: video events were weighted by how well the topic vectors derived from recalls of those events matched the video events' topic vectors (Panel A). Red: video events were weighted by the inverse of how well their topic vectors matched the recalled topic vectors. **C.** The set of all video and recall events is projected onto the two-dimensional space derived in Figure 7. The dots outlined in black denote video events (dot size reflects the average correlation between the video event's topic vector and the topic vectors from the closest matching recalled events from each participant; bigger dots denote stronger correlations). The dots without black outlines denote recalled events. All dots are colored using the same scheme as Figure 7A. Wordles for several example events are displayed (green: three best-remembered events; red: three worst-remembered events). Within each circular wordle, the left side displays words associated with the topic vector for the video event, and the right side displays words associated with the (average) recall event topic vector, across all recall events matched to the given video event.

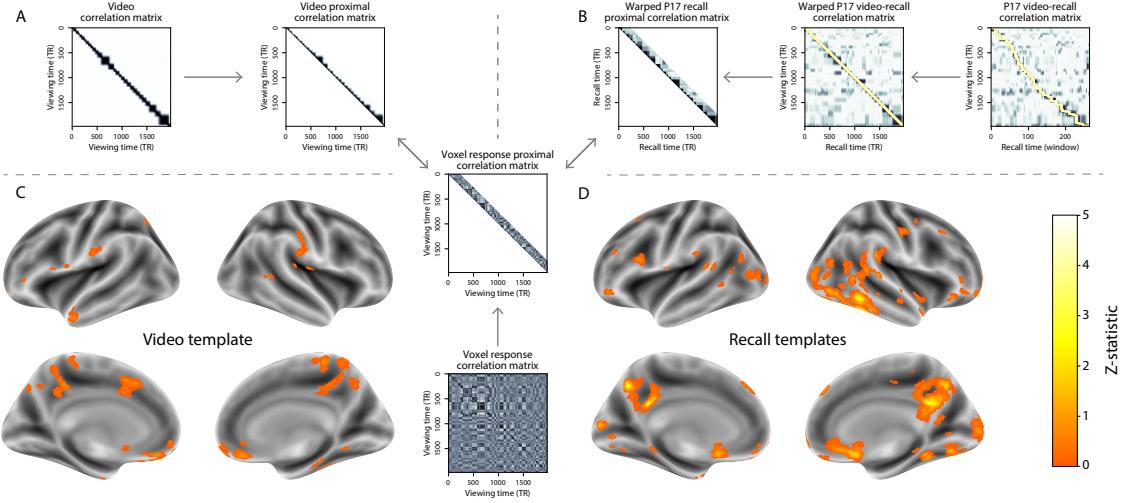


Figure 9: Brain structures that underlie the transformation of experience into memory. **A.** We isolated the proximal diagonals from the upper triangle of the video correlation matrix, and applied this same diagonal mask to the voxel response correlation matrix for each cube of voxels in the brain. We then searched for brain regions whose activation timeseries consistently exhibited a similar proximal correlational structure to the video model, across participants. **B.** We used dynamic time warping (Berndt and Clifford, 1994) to align each participant's recall timeseries to the TR timeseries of the video. We then applied the same diagonal mask used in Panel A to isolate the proximal temporal correlations and searched for brain regions whose activation timeseries for an individual consistently exhibited a similar proximal correlational structure to each individual's recall. **C.** We identified a network of regions sensitive to the narrative structure of participants' ongoing experience. The map shown is thresholded at $p < 0.05$, corrected. **D.** We also identified a network or regions sensitive to how individuals would later structure the video's content in their recalls. The map shown is thresholded at $p < 0.05$, corrected.

376 video-driven searchlight analysis revealed a distributed network of regions that may play a role in
 377 processing information relevant to the narrative structure of the video. Similarly, the recall-driven
 378 searchlight analysis revealed a second network of regions (Fig. 9D) that may facilitate a person-
 379 specific transformation of one's experience into memory. In identifying regions whose responses
 380 to ongoing experiences reflect how those experiences will be remembered later, this latter analysis
 381 extends classic *subsequent memory analyses* (e.g., Paller and Wagner, 2002) to domain of naturalistic
 382 stimuli.

383 The searchlight analyses described above yielded two distributed networks of brain regions,
 384 whose activity timecourses mirrored to the temporal structure of the video (Fig. 9C) or participants'
 385 eventual recalls (Fig. 9D). We next sought to gain greater insight into the structures and functional

386 networks our results reflected. To accomplish this, we performed an additional, exploratory
387 analysis using Neurosynth (Yarkoni et al., 2011). Given an arbitrary statistical map as input,
388 Neurosynth performs a massive automated meta-analysis, returning a ranked list of terms reported
389 in papers with similar significance maps. We ran Neurosynth on the significance maps for the video-
390 and recall-driven searchlight analyses. These maps, along with the 10 terms with maximally similar
391 meta-analysis images identified by Neurosynth are shown in Figure 10.

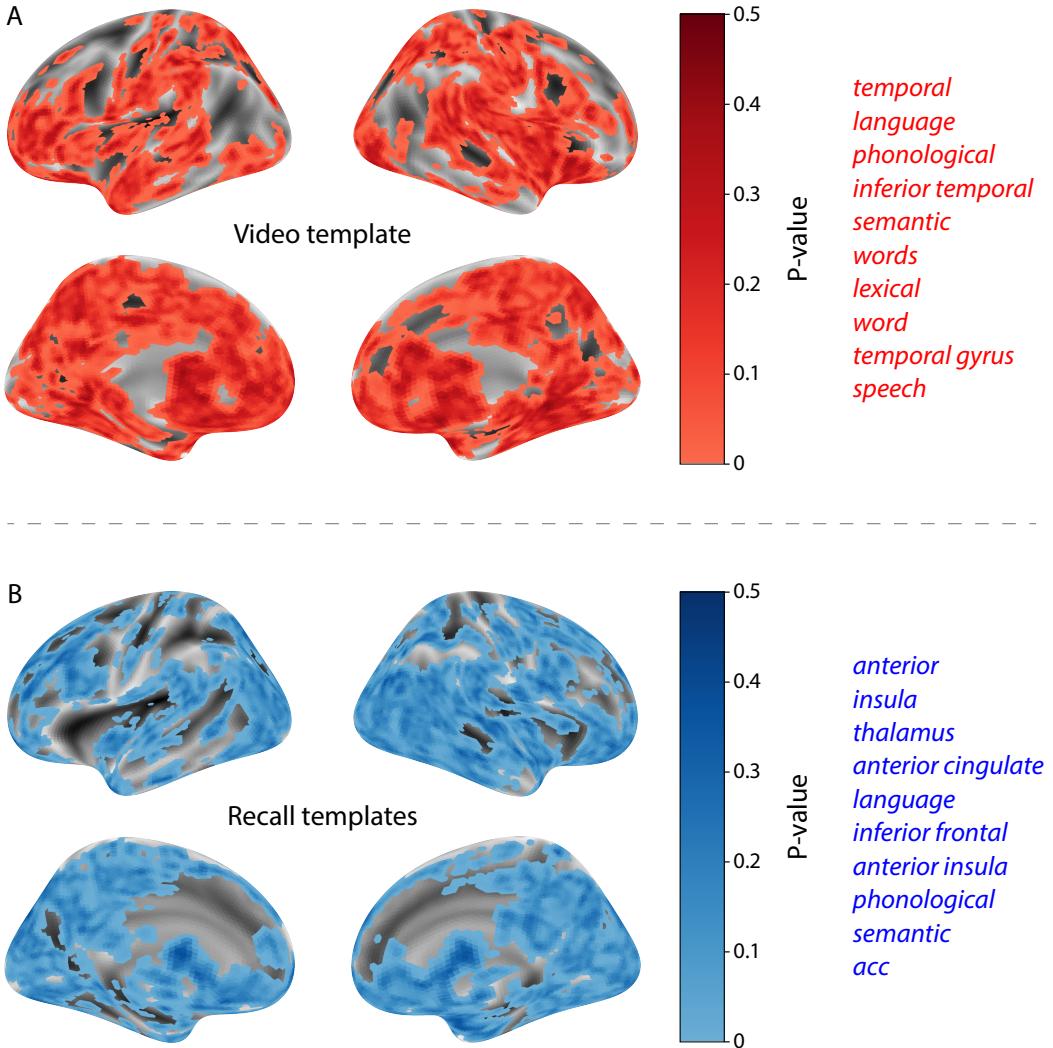


Figure 10: Decoding distributed statistical maps via Neurosynth meta-analyses. **A.** Video-searchlight significance and top 10 decoded terms. We constructed a map of the permutation-derived p -values for the video-driven searchlight analysis (Fig. 9A, C) at each voxel with a positive permutation-derived z -score. The top 10 terms decoded from this significance map are shown in red. **B.** Recall-searchlight significance and top 10 decoded terms. We constructed a map of the permutation-derived p -values for the recall-driven searchlight analysis (Fig. 9A, C) at each voxel with a positive permutation-derived z -score. The top 10 terms decoded from this significance map are shown in blue.

392 **Discussion**

393 Our work casts remembering as reproducing (behaviorally and neurally) the topic trajectory, or
394 shape, of an experience. This view draws inspiration from prior work aimed at elucidating
395 the neural and behavioral underpinnings of how we process dynamic naturalistic experiences
396 and remember them later. One approach to identifying neural responses to naturalistic stimuli
397 (including experiences) entails building a model of the stimulus and searching for brain regions
398 whose responses are consistent with the model. In prior work, a series of studies from Uri Hasson's
399 group (Lerner et al., 2011; Simony et al., 2016; Chen et al., 2017; Baldassano et al., 2017; Zadbood
400 et al., 2017) have extended this approach with a clever twist: rather than building an explicit
401 stimulus model, these studies instead search for brain responses (while experiencing the stimulus)
402 that are reliably similar across individuals. So called *inter-subject correlation* (ISC) and *inter-subject*
403 *functional connectivity* (ISFC) analyses effectively treat other people's brain responses to the stimulus
404 as a "model" of how its features change over time. By contrast, in our present work, we use topic
405 models to construct an explicit content model directly from the stimulus (i.e., the topic trajectory
406 of the video). Projecting each participant's recall into a space shared by both the stimulus and
407 other participants then allows us to compare recalls both directly to the stimulus and to each other.
408 Similarly, prior work introducing the use of HMMs to discover latent event structure in naturalistic
409 stimuli and recall (Baldassano et al., 2017) used between-subjects cross-validation to identify event
410 boundaries shared across participants, and between stimulus and recall. Our framework allows
411 us to break from the restriction of a common, shared event-timeseries and identify the unique
412 *resolution* of each participant's recall event structure, and how that may differ from the video and
413 that of other participants.

414 While a large number of language models exist (e.g., WAS, LSA, word2vec, universal sentence
415 encoder; Steyvers et al., 2004; Landauer et al., 1998; Mikolov et al., 2013; Cer et al., 2018), here
416 we use latent dirichlet allocation (LDA)-based topic models for a few reasons. First, topic models
417 capture the *essence* of a text passage devoid of the specific set and order of words used. This was
418 an important feature of our model since different people may accurately recall a scene using very

419 different language. Second, words can mean different things in different contexts (e.g. “bat” may
420 be the act of hitting a baseball, the object used for that action, or as a flying mammal). Topic
421 models are robust to this, allowing words to exist as part of multiple topics. Last, topic models
422 provide a straightforward means to recover the weights for the particular words comprising a topic,
423 enabling easy interpretation of an event’s contents (e.g. Fig. 8). Other models such as Google’s
424 Universal Sentence Encoder offer a context-sensitive encoding of text passages, but the encoding
425 space is complex and non-linear, and thus recovering the original words used to fit the model is
426 not straightforward. However, it’s worth pointing out that our framework is divorced from the
427 particular choice of language model. Moreover, many of the aspects of our framework could be
428 swapped out for other choices. For example, the language model, the timeseries segmentation
429 model and the video-recall matching function could all be customized for the particular problem.
430 Indeed for some problems, recovery of the particular recall words may not be necessary, and thus
431 other text-modeling approaches (such as universal sentence encoder) may be preferable. Future
432 work will explore the influence of particular model choices on the framework’s efficacy.

433 In extending classical free recall analyses to our naturalistic memory framework, we recovered
434 two patterns of recall dynamics central to list-learning studies: a heightened probability of initiating
435 recall with the first presented “item” (in our case, video events; Fig. 3A) and a strong bias toward
436 transitioning from recalling a given event to recalling the one immediately following it (Fig. 3B).
437 However, equally noteworthy are the typical free recall results *not* recovered in these analyses,
438 as each highlights a fundamental difference between the list-learning paradigm and naturalistic
439 memory paradigms like the one employed in the present study. The most noticeable departure
440 from hallmark free recall dynamics in these findings is the apparent lack of a serial position effect in
441 Figure 3C, which instead shows greater and lesser recall probabilities for events distributed across
442 the video. Stimuli in free recall experiments most often comprise lists of simple, common words,
443 presented to participants in a random order. (In fact, numerous word pools have been developed
444 based on these criteria; e.g., Friendly et al., 1982). These stimulus qualities enable two assumptions
445 that are central to word list analyses, but frequently do not hold for real-world experiences. First,
446 researchers conducting list-learning studies may assume that the content at each presentation index

is essentially equal, and does not possess attributes that would render it, on average, more or less memorable than others. Such is rarely the case with real-world experiences or experiments meant to approximate them, and the effects of both intrinsic and observer-dependent factors on stimulus memorability are well established (for review see Chun and Turk-Browne, 2007; Bylinskii et al., 2015; Tyng et al., 2017). Second, the random ordering of list items ensures that (across participants, on average) there is no relationship between the thematic similarity of individual stimuli and their presentation positions—in other words, two successively presented items are no more likely to be highly semantically similar than they are to be highly dissimilar. In most cases, the exact opposite is true of real-world episodes. Our internal thoughts, our actions, and the physical state of the world around us all tend to follow a direct, causal progression. As a result, each moment of our experience tends to be inherently more similar to surrounding moments than to those in the distant past or future. Memory literature has termed this strong temporal autocorrelation “context,” and in various media that depict real-world events (e.g., movies or written stories), we recognize it as a *narrative structure*. While a random word list (by definition) has no such structure, the logical progression between ideas and actions in a naturalistic stimulus prompts the rememberer to recount presented events in order, starting with the beginning. This tendency is reflected in our findings’ second departure from typical free recall dynamics: a lack of increased probability of first recall for end-of-sequence events (Fig. 3A).

Because they disregard presentation order-dependent variability in the stimulus content, analyses such as those in Figure 3 enable a more sensitive analysis of presentation order-dependent temporal dynamics in free recall. Yet by the same token, they paint a wholly incomplete picture of memory for naturalistic episodes. In an attempt to address this shortcoming, we have developed a framework in the present study that characterizes the explicit semantic content of the stimulus and subsequent recalls. However, sensitivity to stimulus and recall content introduces a new challenge: distinguishing between levels of recall quality for a stimulus (e.g., an event) that is considered to have been “remembered.” When modeling memory in an experimental setting, recall quality for individual events is often cast as binary (e.g., a given list item was simply either remembered or not remembered). Various models of memory (e.g., Yonelinas, 2002) attempt to improve upon this

475 by including confidence ratings, rendering this binary judgement instead categorical. To better
476 evaluate naturalistic memory quality, we introduce a continuous metric (*precision*), which reflects
477 the level of completeness of a participant’s recall for a feature-rich experience. Additionally, recall
478 quality for a single event is typically assessed independently from that for all other events (e.g., it
479 is difficult to “compare” a participant’s binary recall success for list item 1 to that of list item 10).
480 The second novel metric we introduce (*distinctiveness*) is based on analyzing of the correlational
481 structure of an individual’s full set of recall events, and reflects the specificity of their memory
482 for a single experienced event. We find that both of these metrics relate to the overall number of
483 video events participants successfully recalled, and that our precision metric additionally relates to
484 Chen et al. (2017)’s hand-annotated memory memory scores. Though we do not find participants’
485 average recall distinctiveness related to the hand-annotated memory scores, this is not entirely
486 surprising given the divergence of behavior they capture. In hand-scoring each participant’s ver-
487 bal recall for each of 50 (manually-delimited) scenes, “[a] scene was counted as recalled if the
488 participant described any part of the scene” (Chen et al., 2017). In other words, both an extensive
489 description of a scene’s content and a brief mention of some subset of its content were (binarily)
490 considered equally successful recalls. By contrast, we identify the event structure in participants’
491 recalls in an unsupervised manner, independent of the video event-timeseries, prior to mapping
492 between video and recall content. Our HMM-based event-segmentation produces boundaries
493 between timepoints where the topic proportions shift in a substantial way, and because a small
494 handful of words is unlikely to contribute significantly to the topic proportions for any sliding win-
495 dow, such brief scene descriptions will most often not begat a sufficiently large shift in the resulting
496 topic proportions for the HMM to identify an event boundary. Instead, they will be grouped with
497 a neighboring event, consequently lowering that event’s distinctiveness score and by extension,
498 the participant’s overall distinctiveness score. This is in essence the qualitative difference between
499 distinctive and indistinctive recall, and reflects the comparison shown in Figure 6C. Intriguingly,
500 prior studies show that pattern separation, or the ability to cleanly discriminate between similar
501 experiences, is impaired in many cognitive disorders as well as natural aging (Stark et al., 2010;
502 Yassa et al., 2011; Yassa and Stark, 2011). Future work might explore whether and how these

503 metrics compare between cognitively impoverished groups and healthy controls.

504 In the analyses outlined in Figure 9, we identified two networks of brain regions whose re-
505 sponses during video viewing were consistent with the temporal structure of the video and recall
506 topic trajectories, respectively. The network identified by the video trajectory analysis included the
507 ventromedial prefrontal cortex, left anterior temporal lobe, superior parietal and dorsal anterior
508 cingulate cortex. The network from the video-recall trajectory analysis also included the ventro-
509 medial prefrontal and superior parietal cortices, in addition to the posterior medial cortex (PMC)
510 and the inferior temporal regions. Notably, Chen et al. (2017) also observed the PMC in a number
511 of analyses including one that searched for regions whose activity patterns during encoding were
512 reinstated during free recall. The PMC has been consistently identified in studies involving stimuli
513 with meaningfully structured events Cohn-Sheely and Ranganath (2017). Further, the PMC is part
514 of the “posterior medial” system, a network of brain regions thought to represent situation models
515 Zacks et al. (2007) in support of memory, spatial navigation and social cognition (Ranganath and
516 Ritchey, 2012). Given that we constructed our video-recall searchlight model to capture temporal
517 structure in the episode’s semantic content (and how one’s later recall aligns with that structure),
518 we speculate that the PMC may play a role in constructing mnemonic events from meaningfully
519 structured experiences.

520 Decoding the associated significance maps with Neurosynth revealed two intriguing results.
521 First, the top 10 terms returned for the video-driven searchlight significance map were centered
522 around themes of language and semantic meaning (Fig. 10A). In other words, the voxels identified
523 as more reflective of the video’s temporal structure (i.e., voxels with lower permutation correction-
524 derived p -values), as defined by our model, were most likely to be reported as active in studies
525 focused on the the neural underpinnings of semantic processing. This finding is interesting, as our
526 model specifically captures the temporal structure of the video’s *semantic* content (e.g., as opposed
527 to that of the visual, auditory, or affective content). This suggests that the network of structures
528 displayed in Figure 9C may play a roll in processing the evolving semantic structure of ongoing
529 experiences.

530 Our second searchlight analysis identified a largely separate network of regions (Fig. 9D)

whose patterns of activity as participants viewed the video reflected the idiosyncratic structure of each individual's later recall. Decoding the associated significance map yielded a set of terms that primarily reflected names of specific structural regions (such as "thalamus," "anterior insula," "anterior cingulate" and "inferior frontal"; Fig. 10B). Interestingly, these regions share membership in a common, large-scale functional network (termed the "salience network") involved in detecting and processing affective cues. In particular, the latter three regions have been implicated in functions relevant to assigning personal meaning to an experience, including: ascribing subjective value to raw, sensory input (Medford and Critchley, 2010); modulating semantic and phonological processing in response to personally salient stimuli (Kelly et al., 2007); and directing and reallocating attention and working memory resources towards the most relevant stimuli (Menon and Uddin, 2010). This suggests that the network of structures displayed in Figure 9D may play a role in transforming and restructuring ongoing experiences through the lens of one's own personal values as they are encoded in memory.

Our work has broad implications for how we characterize and assess memory in real-world settings, such as the classroom or physician's office. For example, the most commonly used classroom evaluation tools involve simply computing the proportion of correctly answered exam questions. Our work indicates that this approach is only loosely related to what educators might really want to measure: how well did the students understand the key ideas presented in the course? Under this typical framework of assessment, the same exam score of 50% could be ascribed to two very different students: one who attended the full course but struggled to learn more than a broad overview of the material, and one who attended only half of the course but understood the material perfectly. Instead, one could apply our computational framework to build explicit content models of the course material and exam questions. This approach would provide a more nuanced and specific view into which aspects of the material students had learned well (or poorly). In clinical settings, memory measures that incorporate such explicit content models might also provide more direct evaluations of patients' memories.

557 **Methods**

558 **Experimental design and data collection**

559 Data were collected by Chen et al. (2017). In brief, participants ($n = 22$) viewed the first 48 minutes
560 of “A Study in Pink”, the first episode of the BBC television series *Sherlock*, while fMRI volumes
561 were collected (TR = 1500 ms). Participants were pre-screened to ensure they had never seen any
562 episode of the show before. The stimulus was divided into a 23 min (946 TR) and a 25 min (1030 TR)
563 segment to mitigate technical issues related to the scanner. After finishing the clip, participants
564 were instructed to (quoting from Chen et al., 2017) “describe what they recalled of the [episode]
565 in as much detail as they could, to try to recount events in the original order they were viewed
566 in, and to speak for at least 10 minutes if possible but that longer was better. They were told that
567 completeness and detail were more important than temporal order, and that if at any point they
568 realized they had missed something, to return to it. Participants were then allowed to speak for
569 as long as they wished, and verbally indicated when they were finished (e.g., ‘I’m done’).” Five
570 participants were dropped from the original dataset due to excessive head motion (2 participants),
571 insufficient recall length (2 participants), or falling asleep during stimulus viewing (1 participant),
572 resulting in a final sample size of $n = 17$. For additional details about the experimental procedure
573 and scanning parameters, see Chen et al. (2017). The experimental protocol was approved by
574 Princeton University’s Institutional Review Board.

575 After preprocessing the fMRI data and warping the images into a standard (3 mm³ MNI) space,
576 the voxel activations were z-scored (within voxel) and spatially smoothed using a 6 mm (full width
577 at half maximum) Gaussian kernel. The fMRI data were also cropped so that all video-viewing
578 data were aligned across participants. This included a constant 3 TR (4.5 s) shift to account for the
579 lag in the hemodynamic response. (All of these preprocessing steps followed Chen et al., 2017,
580 where additional details may be found.)

581 The video stimulus was divided into 1,000 fine-grained “scenes” and annotated by an inde-
582 pendent coder. For each of these 1,000 scenes, the following information was recorded: a brief
583 narrative description of what was happening, the location where the scene took place, whether

584 that location was indoors or outdoors, the names of all characters on-screen, the name(s) of the
585 character(s) in focus in the shot, the name(s) of the character(s) currently speaking, the camera
586 angle of the shot, a transcription of any text appearing on-screen, and whether or not there was
587 music present in the background. Each scene was also tagged with its onset and offset time, in
588 both seconds and TRs.

589 **Data and code availability**

590 The fMRI data we analyzed are available online [here](#). The behavioral data and all of our analysis
591 code may be downloaded [here](#).

592 **Statistics**

593 All statistical tests performed in the behavioral analyses were two-sided. All statistical tests per-
594 formed in the neural data analyses were two-sided, except for the permutation-based thresholding,
595 which was one-sided. In this case, we were specifically interested in identifying voxels whose ac-
596 tivation time series reflected the temporal structure of the video and recall trajectories to a *greater*
597 extent than that of the phase-shifted trajectories.

598 **Modeling the dynamic content of the video and recall transcripts**

599 **Topic modeling**

600 The input to the topic model we trained to characterize the dynamic content of the video comprised
601 998 hand-generated annotations of short (mean: 2.96s) scenes spanning the video clip (Chen et al.,
602 2017 generated 1000 annotations total; we removed two referring to the break between the first
603 and second scan sessions, during which no fMRI data was collected). We concatenated the text
604 for all of the annotated features within each segment, creating a “bag of words” describing each
605 scene and performed some minor preprocessing (e.g., stemming possessive nouns and removing
606 punctuation). We then re-organized the text descriptions into overlapping sliding windows span-
607 ning (up to) 50 scenes each. In other words, we created a “context” for each scene comprising the

608 text descriptions of the preceding 25 scenes, the present scene, and the following 24 scenes. To
609 model the “context” for scenes near the beginning and end of the video (i.e., within 25 scenes of
610 the beginning or end), we created overlapping sliding windows that grew in size from one scene
611 to the full length, then similarly tapered their length at the end. This additionally ensured that
612 each scene’s content was represented in the text corpus an equal number of times.

613 We trained our model using these overlapping text samples with `scikit-learn` (version 0.19.1;
614 Pedregosa et al., 2011), called from our high-dimensional visualization and text analysis software,
615 `HyperTools` (Heusser et al., 2018b). Specifically, we used the `CountVectorizer` class to transform
616 the text from each window into a vector of word counts (using the union of all words across all
617 scenes as the “vocabulary,” excluding English stop words); this yielded a number-of-windows
618 by number-of-words *word count* matrix. We then used the `LatentDirichletAllocation` class
619 (`topics=100, method='batch'`) to fit a topic model (Blei et al., 2003) to the word count matrix,
620 yielding a number-of-windows (1047) by number-of-topics (100) *topic proportions* matrix. The
621 topic proportions matrix describes the gradually evolving mix of topics (latent themes) present in
622 each scene. Next, we transformed the topic proportions matrix to match the 1976 fMRI volume
623 acquisition times. We assigned each topic vector to the timepoint (in seconds) midway between the
624 beginning of the first scene and the end of the last scene in its corresponding sliding text window.
625 By doing so, we warped the linear temporal distance between consecutive topic vectors to align
626 with the inconsistent temporal distance between consecutive annotations (whose durations varied
627 greatly). We then rescaled these timepoints to 1.5s TR units, and used linear interpolation to
628 estimate a topic vector for each TR. This resulted in a number-of-TRs (1976) by number-of-topics
629 (100) matrix.

630 We created similar topic proportions matrices using hand-annotated transcripts of each par-
631 ticipant’s recall of the video (annotated by Chen et al., 2017). We tokenized the transcript into a
632 list of sentences, and then re-organized the list into overlapping sliding windows spanning (up
633 to) 10 sentences each, analogously to how we parsed the video annotations. In turn, we trans-
634 formed each window’s sentences into a word count vector (using the same vocabulary as for the
635 video model), then used the topic model already trained on the video scenes to compute the most

probable topic proportions for each sliding window. This yielded a number-of-windows (range: 83–312) by number-of-topics (100) topic proportions matrix for each participant. These reflected the dynamic content of each participant’s recalls. Note: for details on how we selected the video and recall window lengths and number of topics, see *Supporting Information* and Figure S1.

640 Parsing topic trajectories into events using Hidden Markov Models

We parsed the topic trajectories of the video and participants’ recalls into events using Hidden Markov Models (Rabiner, 1989). Given the topic proportions matrix (describing the mix of topics at each timepoint) and a number of states, K , an HMM recovers the set of state transitions that segments the timeseries into K discrete states. Following Baldassano et al. (2017), we imposed an additional set of constraints on the discovered state transitions that ensured that each state was encountered exactly once (i.e., never repeated). We used the BrainIAK toolbox (Capota et al., 2017) to implement this segmentation.

We used an optimization procedure to select the appropriate K for each topic proportions matrix. Prior studies on narrative structure and processing have shown that we both perceive and internally represent the world around us at multiple, hierarchical timescales (e.g., Hasson et al., 2008; Lerner et al., 2011; Hasson et al., 2015; Chen et al., 2017; Baldassano et al., 2017, 2018). However, for the purposes of our framework, we sought to identify the single timeseries of event-representations that is emphasized *most heavily* in the temporal structure of the video and of each participant’s recall. We quantified this as the set of K states that maximized the similarity between topic vectors for timepoints comprising each state, while minimizing the similarity between topic vectors for timepoints across different states. Specifically, we computed (for each matrix)

$$\operatorname{argmax}_K [W_1(a, b)],$$

where a was the distribution of within-state topic vector correlations, and b was the distribution of across-state topic vector correlations. We computed the first Wasserstein distance (W_1 ; also known as “earth mover’s distance”; Dobrushin, 1970; Ramdas et al., 2017) between these distributions for a

660 large range of possible K -values (range [2,50]), and selected the K that yielded the maximum value.
661 Figure 2B displays the event boundaries returned for the video, and Figure S4 displays the event
662 boundaries returned for each participant's recalls. See Figure S6 for the optimization functions
663 for the video and recalls. After obtaining these event boundaries, we created stable estimates of
664 the content represented in each event by averaging the topic vectors across timepoints between
665 each pair of event boundaries. This yielded a number-of-events by number-of-topics matrix for
666 the video and recalls from each participant.

667 **Naturalistic extensions of classic list-learning analyses**

668 In traditional list-learning experiments, participants view a list of items (e.g., words) and then recall
669 the items later. Our video-recall event matching approach affords us the ability to analyze memory
670 in a similar way. The video and recall events can be treated analogously to studied and recalled
671 "items" in a list-learning study. We can then extend classic analyses of memory performance and
672 dynamics (originally designed for list-learning experiments) to the more naturalistic video recall
673 task used in this study.

674 Perhaps the simplest and most widely used measure of memory performance is *accuracy*—i.e.,
675 the proportion of studied (experienced) items (in this case, video events) that the participant later
676 remembered. Chen et al. (2017) used this method to rate each participant's memory quality by
677 computing the proportion of (50, manually identified) scenes mentioned in their recall. We found a
678 strong across-participants correlation between these independent ratings and the proportion of (30,
679 HMM-identified) video events matched to participants' recalls (Pearson's $r(15) = 0.71, p = 0.002$).
680 We further considered a number of more nuanced memory performance measures that are typically
681 associated with list-learning studies. We also provide a software package, Quail, for carrying out
682 these analyses (Heusser et al., 2017).

683 **Probability of first recall (PFR).** PFR curves (Welch and Burnett, 1924; Postman and Phillips,
684 1965; Atkinson and Shiffrin, 1968) reflect the probability that an item will be recalled first as a
685 function of its serial position during encoding. To carry out this analysis, we initialized a number-

686 of-participants (17) by number-of-video-events (30) matrix of zeros. Then for each participant, we
687 found the index of the video event that was recalled first (i.e., the video event whose topic vector
688 was most strongly correlated with that of the first recall event) and filled in that index in the matrix
689 with a 1. Finally, we averaged over the rows of the matrix, resulting in a 1 by 30 array representing
690 the proportion of participants that recalled an event first, as a function of the order of the event's
691 appearance in the video (Fig. 3A).

692 **Lag conditional probability curve (lag-CRP).** The lag-CRP curve (Kahana, 1996) reflects the
693 probability of recalling a given item after the just-recalled item, as a function of their relative
694 encoding positions (or *lag*). In other words, a lag of 1 indicates that a recalled item was presented
695 immediately after the previously recalled item, and a lag of -3 indicates that a recalled item came
696 3 items before the previously recalled item. For each recall transition (following the first recall),
697 we computed the lag between the current recall event and the next recall event, normalizing by
698 the total number of possible transitions. This yielded a number-of-participants (17) by number-
699 of-lags (-29 to +29; 59 lags total) matrix. We averaged over the rows of this matrix to obtain a
700 group-averaged lag-CRP curve (Fig. 3B).

701 **Serial position curve (SPC).** SPCs (Murdock, 1962) reflect the proportion of participants that
702 remember each item as a function of the items' serial positions during encoding. We initialized
703 a number-of-participants (17) by number-of-video-events (30) matrix of zeros. Then, for each
704 recalled event, for each participant, we found the index of the video event that the recalled event
705 most closely matched (via the correlation between the events' topic vectors) and entered a 1 into
706 that position in the matrix. This resulted in a matrix whose entries indicated whether or not each
707 event was recalled by each participant (depending on whether the corresponding entires were
708 set to one or zero). Finally, we averaged over the rows of the matrix to yield a 1 by 30 array
709 representing the proportion of participants that recalled each event as a function of the events'
710 order appearance in the video (Fig. 3C).

711 **Temporal clustering scores.** Temporal clustering describes a participant’s tendency to organize
712 their recall sequences by the learned items’ encoding positions. For instance, if a participant
713 recalled the video events in the exact order they occurred (or in exact reverse order), this would
714 yield a score of 1. If a participant recalled the events in random order, this would yield an expected
715 score of 0.5. For each recall event transition (and separately for each participant), we sorted
716 all not-yet-recalled events according to their absolute lag (i.e., distance away in the video). We
717 then computed the percentile rank of the next event the participant recalled. We averaged these
718 percentile ranks across all of the participant’s recalls to obtain a single temporal clustering score
719 for the participant.

720 **Semantic clustering scores.** Semantic clustering describes a participant’s tendency to recall se-
721 mantically similar presented items together in their recall sequences. Here, we used the topic
722 vectors for each event as a proxy for its semantic content. Thus, the similarity between the seman-
723 tic content for two events can be computed by correlating their respective topic vectors. For each
724 recall event transition, we sorted all not-yet-recalled events according to how correlated the topic
725 vector of the *closest-matching* video event was to the topic vector of the closest-matching video event
726 to the just-recalled event. We then computed the percentile rank of the observed next recall. We
727 averaged these percentile ranks across all of the participant’s recalls to obtain a single semantic
728 clustering score for the participant.

729 **Novel naturalistic memory metrics**

730 **Precision.** We tested whether participants who recalled more events were also more *precise* in
731 their recollections. For each participant, we computed the average correlation between the topic
732 vectors for each recall event and those of its closest-matching video event. This gave a single value
733 per participant representing the average precision across all recalled events. We then correlated
734 these values with both hand-annotated and model-derived (i.e., the number of unique video events
735 matched by a participant’s recall events) memory performance.

736 **Distinctiveness.** We also considered the *distinctiveness* of each recalled event. That is, how unique
737 a participant’s description of a video event was, versus their descriptions of other video events.
738 We hypothesized that participants with high memory performance might describe each event in
739 a more distinctive way (relative to those with lower memory performance who might describe
740 events in a more general way). To test this hypothesis we define a distinctiveness score for each
741 recall event as

$$d(\text{event}) = 1 - \bar{c}(\mathbb{P} \setminus \{\text{event}\}),$$

742 where $\bar{c}(\mathbb{P} \setminus \{\text{event}\})$ is the average correlation between the given recall event’s topic vector and
743 the topic vectors from all other recall events not matched to the same video event (for a single
744 participant). We then averaged these distinctiveness scores across all of the events recalled by the
745 given participant and correlated resulting values with hand-annotated and model derived memory
746 performance scores across-subjects, as above.

747 Note: in all instances where we performed statistical tests involving precision or distinctiveness
748 scores, we used Fisher’s *z*-transformation (Fisher, 1925) to stabilize the variance across the dis-
749 tribution of correlation values prior to performing the test. Similarly, when averaging precision
750 or distinctiveness scores, we *z*-transformed the scores prior to computing the mean, and inverse
751 *z*-transformed the result.

752 Visualizing the video and recall topic trajectories

753 We used the UMAP algorithm (McInnes et al., 2018) to project the 100-dimensional topic space onto
754 a two-dimensional space for visualization (Figs. 7, 8). Importantly, to ensure that all of the trajec-
755 tories were projected onto the *same* lower dimensional space, we computed the low-dimensional
756 embedding on a “stacked” matrix created by vertically concatenating the events-by-topics topic
757 proportions matrices for the video, across-participants average recall and all 17 individual partici-
758 pants’ recalls. We then divided the rows of the result (a total-number-of-events by two matrix) back
759 into separate matrices for the video topic trajectory, across-participant average recall trajectory and

760 the trajectories for each individual participant's recalls (Fig. 7). This general approach for dis-
761 covering a shared low-dimensional embedding for a collections of high-dimensional observations
762 follows Heusser et al. (2018b).

763 We optimized the manifold space for visualization based on two criteria: First, that the 2D
764 embedding of the video trajectory should reflect its original 100-dimensional structure as faithfully
765 as possible. Second, that the path traversed by the embedded video trajectory should intersect
766 itself a minimal number of times. The first criteria helps bolster the validity of visual intuitions
767 about relationships between sections of video content, based on their locations in the embedding
768 space. The second criteria was motivated by the observed low off-diagonal values in the video
769 trajectory's temporal correlation matrix (suggesting that the same topic-space coordinates should
770 not be revisited; see Figure 2A in the main text). For further details on how we created this
771 low-dimensional embedding space, see *Supporting Information*.

772 **Estimating the consistency of flow through topic space across participants**

773 In Figure 7B, we present an analysis aimed at characterizing locations in topic space that dif-
774 ferent participants move through in a consistent way (via their recall topic trajectories). The
775 two-dimensional topic space used in our visualizations (Fig. 7) comprised a 60 x 60 (arbitrary
776 units) square. We tiled this space with a 50 x 50 grid of evenly spaced vertices, and defined a
777 circular area centered on each vertex whose radius was two times the distance between adjacent
778 vertices (i.e., 2.4 units). For each vertex, we examined the set of line segments formed by connecting
779 each pair successively recalled events, across all participants, that passed through this circle. We
780 computed the distribution of angles formed by those segments and the x -axis, and used a Rayleigh
781 test to determine whether the distribution of angles was reliably "peaked" (i.e., consistent across
782 all transitions that passed through that local portion of topic space). To create Figure 7B we drew
783 an arrow originating from each grid vertex, pointing in the direction of the average angle formed
784 by the line segments that passed within its circular radius. We set the arrow lengths to be inversely
785 proportional to the p -values of the Rayleigh tests at each vertex. Specifically, for each vertex we
786 converted all of the angles of segments that passed within 2.4 units to unit vectors, and we set

787 the arrow lengths at each vertex proportional to the length of the (circular) mean vector. We also
788 indicated any significant results ($p < 0.05$, corrected using the Benjamani-Hochberg procedure) by
789 coloring the arrows in blue (darker blue denotes a lower p -value, i.e., a longer mean vector); all
790 tests with $p \geq 0.05$ are displayed in gray and given a lower opacity value.

791 **Searchlight fMRI analyses**

792 In Figure 9, we present two analyses aimed at identifying brain regions whose responses (as par-
793 ticipants viewed the video) exhibited a particular temporal structure. We developed a searchlight
794 analysis wherein we constructed a $5 \times 5 \times 5$ cube of voxels (following Chen et al., 2017) centered on
795 each voxel in the brain, and for each of these cubes, computed the temporal correlation matrix of
796 the voxel responses during video viewing. Specifically, for each of the 1976 volumes collected dur-
797 ing video viewing, we correlated the activity patterns in the given cube with the activity patterns
798 (in the same cube) collected during every other timepoint. This yielded a 1976 by 1976 correlation
799 matrix for each cube. Note: participant 5's scan ended 75s early, and in Chen et al., 2017's publicly
800 released dataset, their scan data was padded to match the length of the other participants'. For
801 our searchlight analyses, we removed this padded data (i.e., the last 50 TRs), resulting in a 1925 by
802 1925 correlation matrix for each cube in participant 5's brain.

803 Next, we constructed a series of "template" matrices: the first reflecting the timecourse of
804 video's topic trajectory, and the others reflecting that of each participant's recall topic trajectory.
805 To construct the video template, we computed the correlations between the topic proportions
806 estimated for every pair of TRs (prior to segmenting the trajectory into discrete events; i.e., the
807 correlation matrix shown in Figs. 2B and 9A). We constructed similar temporal correlation matrices
808 for each participant's recall topic trajectory (Figs. 2D, S4). However, to correct for length differences
809 and potential non-linear transformations between viewing time and recall time, we first used
810 dynamic time warping (Berndt and Clifford, 1994) to temporally align participants' recall topic
811 trajectories with the video topic trajectory. An example correlation matrix before and after warping
812 is shown in Fig. 9B. This yielded a 1976 by 1976 correlation matrix for the video template and for
813 each participant's recall template.

814 The temporal structure of the video’s content (as described by our model) is captured in the
815 block-diagonal structure of the video’s temporal correlation matrix (e.g., Figs. 2B, 9A), with time
816 periods of thematic stability represented as dark blocks of varying sizes. Inspecting the video
817 correlation matrix suggests that the video’s semantic content is highly temporally specific (i.e.,
818 the correlations between topic vectors from distant timepoints are almost entirely near-zero).
819 By contrast, the activity patterns of individual (cubes of) voxels can encode relatively limited
820 information on their own, and their activity frequently contributes to multiple separate functions
821 (Freedman et al., 2001; Sigman and Dehaene, 2008; Charron and Koechlin, 2010; Rishel et al., 2013).
822 By nature, these two attributes give rise to similarities in activity across large timescales that may
823 not necessarily reflect a single task. To enable a more sensitive analysis of brain regions whose shifts
824 in activity patterns mirrored shifts in the semantic content of the video or recalls, we restricted the
825 temporal correlations we considered to timescale of semantic information captured by our model.
826 Specifically, we isolated the upper triangle of the video correlation matrix and created a “proximal
827 correlation mask” that included only diagonals from the upper triangle of the video correlation
828 matrix up to the first that contained no positive correlations. Applying this mask to the full video
829 correlation matrix was analogous to excluding diagonals beyond the corner of the largest diagonal
830 block. In other words, the timescale of temporal correlations we considered corresponded to the
831 longest period of thematic stability in the video, and by extension the longest expected period
832 of thematic stability in participants’ recalls and the longest period of stability we might expect
833 to see in voxel activity arising from processing or encoding video content. Figure 9 shows this
834 proximal correlation mask applied to the temporal correlation matrices for the video, an example
835 participant’s (warped) recall, and an example cube of voxels from our searchlight analyses.

836 To determine which (cubes of) voxel responses matched the video template, we correlated the
837 proximal diagonals from the upper triangle of the voxel correlation matrix for each cube with the
838 proximal diagonals from video template matrix (Kriegeskorte et al., 2008). This yielded, for each
839 participant, a voxelwise map of correlation values. We then performed a one-sample t -test on the
840 distribution of (Fisher z -transformed) correlations at each voxel, across participants. This resulted
841 in a value for each voxel (cube), describing how reliably its timecourse mirrored that of the video.

842 We further sought to ensure that our analysis identified regions where the activations' temporal
843 structure specifically reflected that of the video, rather than regions whose activity was simply
844 autocorrelated at a width similar to the video template's diagonal. To achieve this, we used a phase
845 shift-based permutation procedure, wherein we circularly shifted the video's topic trajectory by
846 a random number of timepoints, computed the resulting "null" video template, and re-ran the
847 searchlight analysis, in full. (For each of the 100 permutations, the same random shift was used for
848 all participants). We z -scored the observed (unshifted) result at each voxel against the distribution
849 of permutation-derived "null" results, and estimated a p -value by computing the proportion of
850 shifted results that yielded larger values. To create the map in Figure 9C, we thresholded out
851 any voxels whose similarity to the unshifted video's structure fell below the 95th percentile of the
852 permutation-derived similarity results.

853 We used an analogous procedure to identify which voxels' responses reflected the recall tem-
854 plates. For each participant, we correlated the proximal diagonals from the upper triangle of the
855 correlation matrix for each cube of voxels with the proximal diagonals from the upper triangle
856 of their (time-warped) recall correlation matrix. As in the video template analysis, this yielded a
857 voxelwise map of correlation coefficients per participant. However, whereas the video analysis
858 compared every participant's responses to the same template, here the recall templates were unique
859 for each participant. As in the analysis described above, we t -scored the (Fisher z -transformed)
860 voxelwise correlations, and used the same permutation procedure we developed for the video
861 responses to ensure specificity to the recall timeseries and assign significance values. To create the
862 map in Figure 9D we again thresholded out any voxels whose correspondence values fell below
863 the 95th percentile of the permutation-derived null distribution.

864 Neurosynth decoding analyses

865 Neurosynth parses a massive online database of over 14,000 neuroimaging studies and constructs
866 meta-analysis images for over 13,000 psychology- and neuroscience-related terms, based on NIfTI
867 images accompanying studies where those terms appear at a high frequency. Then, given a novel
868 image (tagged with its value type; e.g., t -, F - or p -statistics), Neurosynth returns a list of terms whose

meta-analysis images are most similar to this new data. Our permutation procedure yielded, for each of the two searchlight analyses, a voxelwise map of significance (p -statistic) values. These maps describe the extent to which each voxel *specifically* reflected the temporal structure of the video or individuals' recalls (i.e., for each voxel, the proportion of phase-shifted topic vector correlation matrices less similar to the voxel activity correlation matrix than the unshifted video's correlation matrix). We input the two statistical maps described above to Neurosynth to create a list of the 10 most representative terms for each map.

References

- Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. and Spence, J. T., editors, *The psychology of learning and motivation*, volume 2, pages 89–105. Academic Press, New York.
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3):709–721.
- Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, 38(45):9689–9699.
- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370.
- Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 113–120, New York, NY, US. ACM.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993 – 1022.
- Brunec, I. K., Moscovitch, M. M., and Barense, M. D. (2018). Boundaries shape cognitive representations of spaces and events. *Trends in Cognitive Sciences*, 22(7):637–650.

- 893 Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., and Oliva, A. (2015). Intrinsic and extrinsic
894 effects on image memorability. *Vision Research*, 116:165–178.
- 895 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., Keller, B., Wang, Y., and
896 Shin, Y. S. (2017). Brain imaging analysis kit.
- 897 Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes,
898 M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal sentence encoder.
899 *arXiv*, 1803.11175.
- 900 Charron, S. and Koechlin, E. (2010). Divided representations of current goals in the human frontal
901 lobes. *Science*, 328(5976):360–363.
- 902 Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared
903 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,
904 20(1):115.
- 905 Chun, M. and Turk-Browne, N. (2007). Interactions between attention and memory. *Current opinion*
906 *in neurobiology*, 17(2):177–184.
- 907 Clewett, D. and Davachi, L. (2017). The ebb and flow of experience determines the temporal
908 structure of memory. *Curr Opin Behav Sci*, 17:186–193.
- 909 Cohn-Sheely, B. I. and Ranganath, C. (2017). Time regained: how the human brain constructs
910 memory for time. *Current Opinion in Behavioral Sciences*, 17:169–177.
- 911 Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions.
912 *Theory of Probability & Its Applications*, 15(3):458–486.
- 913 DuBrow, S. and Davachi, L. (2013). The influence of contextual boundaries on memory for the
914 sequential order of events. *Journal of Experimental Psychology: General*, 142(4):1277–1286.
- 915 Ezzyat, Y. and Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological*
916 *Science*, 22(2):243–252.

- 917 Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.
- 918 Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of
919 visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316.
- 920 Friendly, M., Franklin, P. E., Hoffman, D., and Rubin, D. C. (1982). The Toronto Word Pool:
921 Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080
922 words. *Behavior Research Methods and Instrumentation*, 14:375–399.
- 923 Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral
924 component of information processing. *Trends in Cognitive Science*, 19(6):304–315.
- 925 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal
926 receptive windows in human cortex. *Journal of Neuroscience*, 28(10):2539–2550.
- 927 Heusser, A. C., Ezzyat, Y., Shiff, I., and Davachi, L. (2018a). Perceptual boundaries cause mnemonic
928 trade-offs between local boundary processing and across-trial associative binding. *Journal of
929 Experimental Psychology Learning, Memory, and Cognition*, 44(7):1075–1090.
- 930 Heusser, A. C., Fitzpatrick, P. C., Field, C. E., Ziman, K., and Manning, J. R. (2017). Quail: a
931 Python toolbox for analyzing and plotting free recall data. *The Journal of Open Source Software*,
932 10.21105/joss.00424.
- 933 Heusser, A. C., Ziman, K., Owen, L. L. W., and Manning, J. R. (2018b). HyperTools: a Python
934 toolbox for gaining geometric insights into high-dimensional data. *Journal of Machine Learning
935 Research*, 18(152):1–6.
- 936 Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal
937 of Mathematical Psychology*, 46:269–299.
- 938 Howard, M. W., MacDonald, C. J., Tiganj, Z., Shankar, K. H., Du, Q., Hasselmo, M. E., and H., E.
939 (2014). A unified mathematical framework for coding time, space, and sequences in the medial
940 temporal lobe. *Journal of Neuroscience*, 34(13):4692–4707.

- 941 Howard, M. W., Viskontas, I. V., Shankar, K. H., and Fried, I. (2012). Ensembles of human MTL
942 neurons “jump back in time” in response to a repeated stimulus. *Hippocampus*, 22:1833–1847.
- 943 Huk, A., Bonnen, K., and He, B. J. (2018). Beyond trial-based paradigms: continuous behavior, on-
944 going neural activity, and naturalistic stimuli. *Journal of Neuroscience*, 10.1523/JNEUROSCI.1920-
945 17.2018.
- 946 Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24:103–109.
- 947 Kelly, S., Lloyd, D., Nurmikko, T., and Roberts, N. (2007). Retrieving autobiographical memories
948 of painful events activates the anterior cingulate cortex and inferior frontal gyrus. *THe Journal of
949 Pain*, 8(4):307–314.
- 950 Kriat, A. and Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distin-
951 guishing accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of
952 Experimental Psychology: General*, 123(3):297–315.
- 953 Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – con-
954 necting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:1 – 28.
- 955 Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to latent semantic analysis.
956 *Discourse Processes*, 25:259–284.
- 957 Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy
958 of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8):2906–2915.
- 959 Manning, J. R. (2019). Episodic memory: mental time travel or a quantum ‘memory wave’ function?
960 *PsyArXiv*, doi:10.31234/osf.io/6zjwb.
- 961 Manning, J. R., Norman, K. A., and Kahana, M. J. (2015). The role of context in episodic memory.
962 In Gazzaniga, M., editor, *The Cognitive Neurosciences, Fifth edition*, pages 557–566. MIT Press.
- 963 Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns
964 in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National
965 Academy of Sciences, USA*, 108(31):12893–12897.

- 966 McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and
967 projection for dimension reduction. *arXiv*, 1802(03426).
- 968 Medford, N. and Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate
969 cortex: awareness and response. *Brain Structure and Function*, 214(5-6):535–549.
- 970 Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of
971 insula function. *Brain Structure and Function*, 214(5-6):655–667.
- 972 Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations
973 in vector space. *arXiv*, 1301.3781.
- 974 Mueller, A., Fillion-Robin, J.-C., Boidol, R., Tian, F., Nechifor, P., yoonsubKim, Peter, Rampin, R.,
975 Corvellec, M., Medina, J., Dai, Y., Petrushev, B., Langner, K. M., Hong, Alessio, Ozsvald, I.,
976 vkolmakov, Jones, T., Bailey, E., Rho, V., IgorAPM, Roy, D., May, C., foobuzz, Piyush, Seong,
977 L. K., Goey, J. V., Smith, J. S., Gus, and Mai, F. (2018). WordCloud 1.5.0: a little word cloud
978 generator in Python. *Zenodo*, <https://zenodo.org/record/1322068#.W4tPKZNKh24>.
- 979 Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*,
980 64:482–488.
- 981 Paller, K. A. and Wagner, A. D. (2002). Observing the transformation of experience into memory.
982 *Trends in Cognitive Sciences*, 6(2):93–102.
- 983 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Pretten-
984 hofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,
985 M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine
986 Learning Research*, 12:2825–2830.
- 987 Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model
988 of organizational processes in free recall. *Psychological Review*, 116(1):129–156.
- 989 Postman, L. and Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal
990 of Experimental Psychology*, 17:132–138.

- 991 Rabiner, L. (1989). A tutorial on Hidden Markov Models and selected applications in speech
992 recognition. *Proceedings of the IEEE*, 77(2):257–286.
- 993 Radvansky, G. A. and Zacks, J. M. (2017). Event boundaries in memory and cognition. *Curr Opin*
994 *Behav Sci*, 17:133–140.
- 995 Ramdas, A., Trillos, N., and Cuturi, M. (2017). On wasserstein two-sample testing and related
996 families of nonparametric tests. *Entropy*, 19(2):47.
- 997 Ranganath, C. and Ritchey, M. (2012). Two cortical systems for memory-guided behavior. *Nature*
998 *Reviews Neuroscience*, 13:713 – 726.
- 999 Rishel, C. A., Huang, G., and Freedman, D. J. (2013). Independent category and spatial encoding
1000 in parietal cortex. *Neuron*, 77(5):969–979.
- 1001 Sigman, M. and Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during
1002 dual-task performance. *Journal of Neuroscience*, 28(30):7585–7589.
- 1003 Simony, E., Honey, C. J., Chen, J., and Hasson, U. (2016). Uncovering stimulus-locked network
1004 dynamics during narrative comprehension. *Nature Communications*, 7(12141):1–13.
- 1005 Stark, S. M., Yassa, M. A., and Stark, C. E. L. (2010). Individual differences in spatial pattern
1006 separation performance associated with healthy aging in humans. *Learning & Memory*, 17(6):284–
1007 288.
- 1008 Steyvers, M., Shiffrin, R. M., and Nelson, D. L. (2004). Word association spaces for predicting
1009 semantic similarity effects in episodic memory. In Healy, A. F., editor, *Cognitive Psychology and*
1010 *its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. American
1011 Psychological Association, Washington, DC.
- 1012 Tompry, A. and Davachi, L. (2017). Consolidation promotes the emergence of representational
1013 overlap in the hippocampus and medial prefrontal cortex. *Neuron*, 96(1):228–241.
- 1014 Tyng, C. M., Amin, H. U., Saad, M. N. M., and S, M. A. (2017). The influences of emotion on
1015 learning and memory. *Frontiers in psychology*, 8:1454.

- 1016 Welch, G. B. and Burnett, C. T. (1924). Is primacy a factor in association-formation. *American Journal*
1017 *of Psychology*, 35:396–401.
- 1018 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale
1019 automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665.
- 1020 Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., and Stark, C. E. L. (2011). Pattern
1021 separation deficits associated with increased hippocampal ca3 and dentate gyrus activity in
1022 nondemented older adults. *Hippocampus*, 21(9):968–979.
- 1023 Yassa, M. A. and Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends In Neuro-*
1024 *sciences*, 34(10):515–525.
- 1025 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research.
1026 *Journal of Memory and Language*, 46:441–517.
- 1027 Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., and
1028 Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection
1029 and familiarity. *Nature Neuroscience*, 5(11):1236–41.
- 1030 Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception:
1031 a mind-brain perspective. *Psychological Bulletin*, 133:273–293.
- 1032 Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit
1033 memories to other brains: Constructing shared neural representations via communication. *Cereb*
1034 *Cortex*, 27(10):4988–5000.
- 1035 Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and
1036 memory. *Psychological Bulletin*, 123(2):162 – 185.

1037 **Supporting information**

- 1038 Supporting information is available in the online version of the paper.

1039 **Acknowledgements**

1040 We thank Luke Chang, Janice Chen, Chris Honey, Lucy Owen, Emily Whitaker, and Kirsten Ziman
1041 for feedback and scientific discussions. We also thank Janice Chen, Yuan Chang Leong, Kenneth
1042 Norman, and Uri Hasson for sharing the data used in our study. Our work was supported in part
1043 by NSF EPSCoR Award Number 1632738. The content is solely the responsibility of the authors
1044 and does not necessarily represent the official views of our supporting organizations.

1045 **Author contributions**

1046 Conceptualization: A.C.H. and J.R.M.; Methodology: A.C.H., P.C.F. and J.R.M.; Software: A.C.H.,
1047 P.C.F. and J.R.M.; Analysis: A.C.H., P.C.F. and J.R.M.; Writing, Reviewing, and Editing: A.C.H.,
1048 P.C.F. and J.R.M.; Supervision: J.R.M.

1049 **Author information**

1050 The authors declare no competing financial interests. Correspondence and requests for materials
1051 should be addressed to J.R.M. (jeremy.r.manning@dartmouth.edu).