

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 19, 2020

6 **Abstract**

7 Our moment-by-moment subject experience reflects external sensory information from each
8 moment, along with additional information from our past that we carry with us into that moment.
9 The blend of memories, knowledge, emotions, goals, and other internal perceptual and mental
10 states that color our subjective experience provides a *context* for interpreting new information
11 and conceptually linking what is happening now with our prior experiences. Because this con-
12 textual information is often person-specific, the subjective experience that each person encodes
13 into their memory is often idiosyncratic, even for shared experiences and sensory perspectives.
14 We sought to study which aspects of a shared naturalistic experience were preserved or distorted,
15 and how those distortions compared across individuals. To this end, we developed a geomet-

ric framework for mathematically characterizing the subjective conceptual content of dynamic naturalistic experiences. We model experiences as *trajectories* through word embedding spaces whose coordinates reflect the universe of thoughts under consideration. We also demonstrate how *memories* may also be modeled as trajectories through the same spaces. According to this view, encoding an experience into memory entails geometrically distorting or transforming the original experience’s trajectory. This translates qualitative neuropsychological questions about how we remember naturalistic experiences into quantitative geometric questions about the spatial configurations of trajectory shapes. We applied our framework to data collected as participants watched and verbally recounted a television episode while undergoing functional neuroimaging. We found that the trajectories of participants’ recounts of the episode nearly all captured the coarse spatial properties of the original episode’s trajectory (i.e., the essential plot points), but participants differed in their memory for fine 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 distort and transform our ongoing experiences when we encode them into episodic memories.

33 **Introduction**

34 What does it mean to *remember* something? In traditional episodic memory experiments (e.g.,
35 list-learning or trial-based experiments; Murdock, 1962; Kahana, 1996), remembering is often cast
36 as a discrete and binary operation: each studied item may be separated from the rest of one’s
37 experience and singularly labeled as having been recalled or forgotten. More nuanced studies
38 might incorporate self-reported confidence measures as a proxy for memory strength, or ask
39 participants to discriminate between “recollecting” the (contextual) details of an experience or
40 having a general feeling of “familiarity” (Yonelinas, 2002). Using well controlled, trial-based
41 experimental designs, the field has amassed a wealth of valuable information regarding human
42 episodic memory. However, there are fundamental properties of the external world and our

43 memories that trial-based experiments are not well suited to capture (for review, also see Koriat
44 and Goldsmith, 1994; Huk et al., 2018). First, our experiences and memories are continuous, rather
45 than discrete—isolating a (naturalistic) event from the context in which it occurs can substantially
46 change its meaning. Second, the specific language used to describe an experience has little bearing
47 on whether the experience should be considered to have been “remembered.” Asking whether
48 the rememberer has precisely reproduced a specific set of words to describe a given experience
49 is nearly orthogonal to whether or not they were actually able to remember it. In classic (e.g.,
50 list-learning) memory studies, by contrast, the number or proportion of exact recalls is often a
51 primary metric for assessing the quality of participants’ memories. Third, one might remember
52 the *essence* (or a general summary) of an experience but forget (or neglect to recount) particular
53 details. Capturing the essence of what happened is typically the main “point” of recounting a
54 memory to a listener, while the addition of highly specific details may add comparatively little to
55 successful conveyance of an experience.

56 How might one go about formally characterizing the *essence* of an experience, and whether it
57 has been recovered by the rememberer? Any given moment of an experience derives meaning
58 from surrounding moments, as well as from longer-range temporal associations (Lerner et al.,
59 2011; Manning, 2019, 2020). Therefore, the timecourse describing how an event unfolds is fun-
60 damental to its overall meaning. Further, this hierarchy formed by our subjective experiences
61 at different timescales defines a *context* for each new moment (e.g., Howard and Kahana, 2002;
62 Howard et al., 2014), and plays an important role in how we interpret that moment and remember
63 it later (for review see Manning et al., 2015; Manning, 2020). Our memory systems can leverage
64 these associations to form predictions that help guide our behaviors (Ranganath and Ritchey,
65 2012). For example, as we navigate the world, the features of our subjective experiences tend
66 to change gradually (e.g., the room or situation we are in at any given moment is strongly tem-
67 porally autocorrelated), allowing us to form stable estimates of our current situation and behave
68 accordingly (Zacks et al., 2007; Zwaan and Radvansky, 1998).

69 Occasionally, this gradual “drift” of our ongoing experience is punctuated by sudden changes,
70 or “shifts” (e.g., when we walk through a doorway; Radvansky and Zacks, 2017). Prior research

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

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

99 event (relative to other recalled events). We examine how these metrics relate to overall memory
100 performance, and discuss the ways in which they improve upon classic “proportion-recalled”
101 measures for analyzing naturalistic memory. Last, we utilize our framework to identify networks
102 of brain structures whose responses (as participants watched the episode) reflected the temporal
103 dynamics of either the episode or how participants would later recount it.

104 Results

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

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

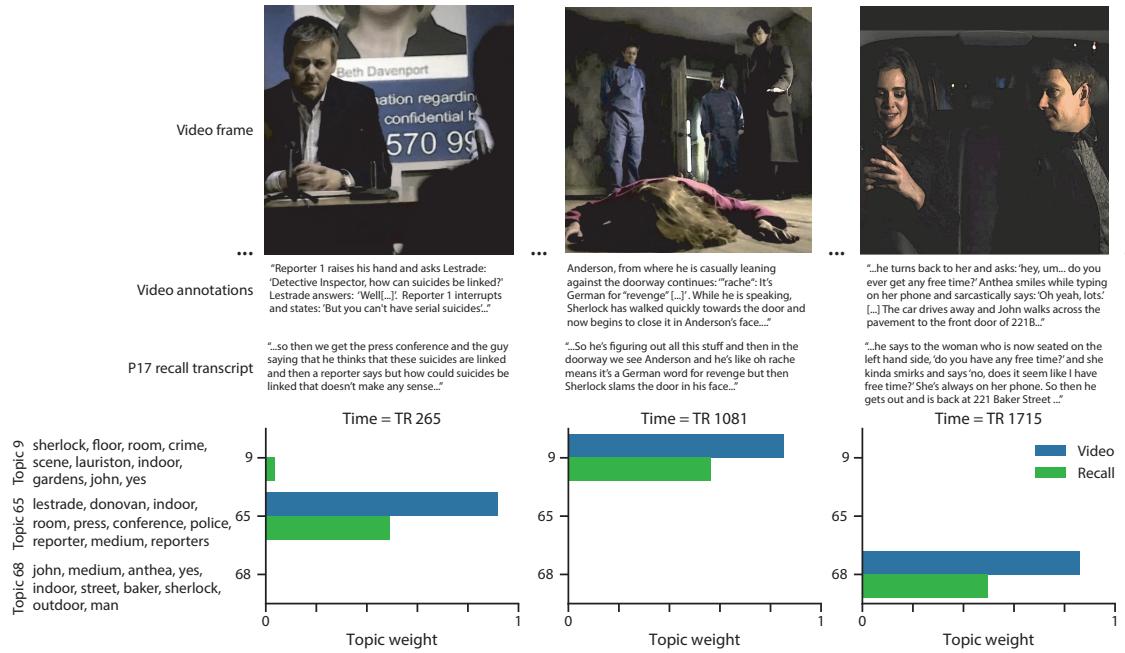


Figure 1: Methods overview. We used hand-annotated descriptions of each moment of video to fit a topic model. Three example video frames and their associated descriptions are displayed (top two rows). Participants later recalled the video (in the third row, we show example recalls of the same three scenes 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.

is similar in some respects to Dynamic Topic Models (Blei and Lafferty, 2006) in that we sought to characterize how the thematic content of the episode evolved over time. However, whereas Dynamic Topic Models are designed to characterize how the properties of *collections* of documents change over time, our sliding window approach allows us to examine the topic dynamics within a single document (or video). Specifically, our approach yielded (via the topic proportions matrix) a single *topic vector* for each sliding window of annotations transformed by the topic model. We then stretched (interpolated) the resulting windows-by-topics matrix to match the time series of the 1976 fMRI volumes collected as participants viewed the episode.

The 32 topics we found were heavily character-focused (i.e., the top-weighted word in each

topic was nearly always a character) and could be roughly divided into themes centered around Sherlock Holmes (the titular character), John Watson (Sherlock’s close confidant and assistant), supporting characters (e.g., Inspector Lestrade, Sergeant Donovan, or Sherlock’s brother Mycroft), or the interactions between various groupings of these characters (see Fig. S2). Several of the identified topics were highly similar, which we hypothesized might allow us to distinguish between subtle narrative differences if the distinctions between those overlapping topics were meaningful. The topic vectors for each timepoint were also *sparse*, in that only a small number (usually one or two) of topics tended to be “active” in any given timepoint (Fig. 2A). Further, the dynamics of the topic activations appeared to exhibit *persistence* (i.e., given that a topic was active in one timepoint, it was likely to be active in the following timepoint) along with *occasional rapid changes* (i.e., occasionally topics would appear to spring into or out of existence). These two properties of the topic dynamics may be seen in the block diagonal structure of the timepoint-by-timepoint correlation matrix (Fig. 2B) and reflect the gradual drift and sudden shifts fundamental to the temporal dynamics of real-world experiences. Given this observation, we adapted an approach devised by Baldassano et al. (2017), and used a hidden Markov model (HMM) to identify the *event boundaries* where the topic activations changed rapidly (i.e., the boundaries of the blocks in the temporal correlation matrix; event boundaries identified by the HMM are outlined in yellow in Fig. 2B). Part of our model fitting procedure required selecting an appropriate number of “events” into which the topic trajectory should be segmented. To accomplish this, we used an optimization procedure that maximized the difference between the topic weights for timepoints within an event versus timepoints across multiple events (see *Methods* for additional details). We then created a stable “summary” of the content within each video event by averaging the topic vectors across the timepoints spanned by each event (Fig. 2C).

Given that the time-varying content of the video could be segmented cleanly into discrete events, we wondered whether participants’ recalls of the video also displayed a similar structure. We applied the same topic model (already trained on the video annotations) to each participant’s recalls. Analogously to how we parsed the time-varying content of the video, to obtain similar estimates for each participant’s recall, we treated each overlapping window of (up to 10) sentences



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 from their transcript as a “document,” and computed the most probable mix of topics reflected in
163 each timepoint’s sentences. This yielded, for each participant, a number-of-windows by number-
164 of-topics topic proportions matrix that characterized how the topics identified in the original video
165 were reflected in the participant’s recalls. Note that an important feature of our approach is that
166 it allows us to compare participants’ recalls to events from the original video, despite different
167 participants using widely varying language to describe the events, and that those descriptions
168 often diverged in content and quality from the video annotations. This is a substantial benefit of
169 projecting the video and recalls into a shared “topic” space. An example topic proportions matrix
170 from one participant’s recalls is shown in Figure 2D.

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

185 Two clear patterns emerged from this set of analyses. First, although every individual partic-
186 ipant’s recalls could be segmented into discrete events (i.e., every individual participant’s recall
187 correlation matrix exhibited clear block diagonal structure; Fig. S4), each participant appeared to
188 have a unique *recall resolution*, reflected in the sizes of those blocks. While some participants’ recall
189 topic proportions segmented into just a few events (e.g., Participants P4, P5, and P7), others’ seg-

190 mented into many shorter duration events (e.g., Participants P12, P13, and P17). This suggests that
191 different participants may be recalling the video with different levels of detail—i.e., some might
192 touch on just the major plot points, whereas others might attempt to recall every minor scene or
193 action. The second clear pattern present in every individual participant’s recall correlation matrix
194 was that, unlike in the video correlation matrix, there were substantial off-diagonal correlations.
195 Whereas each event in the original video was (largely) separable from the others (Fig. 2B), in
196 transforming those separable events into memory, participants appeared to be integrating across
197 multiple events, blending elements of previously recalled and not-yet-recalled content into each
198 newly recalled event (Figs. 2E, S4; also see Manning et al., 2011; Howard et al., 2012; Manning,
199 2019).

200 The above results indicate that both the structure of the original video and participants’ recalls
201 of the video exhibit event boundaries that can be identified automatically by characterizing the
202 dynamic content using a shared topic model and segmenting the content into events via HMMs.
203 Next, we asked whether some correspondence might be made between the specific content of the
204 events the participants experienced in the video, and the events they later recalled. One approach
205 to linking the experienced (video) and recalled events is to label each recalled event as matching
206 the video event with the most similar (i.e., most highly correlated) topic vector (Figs. 2G, S5). This
207 yields a sequence of “presented” events from the original video, and a (potentially differently
208 ordered) sequence of “recalled” events for each participant. Analogous to classic list-learning
209 studies, we can then examine participants’ recall sequences by asking which events they tended
210 to recall first (probability of first recall; Fig. 3A; Atkinson and Shiffrin, 1968; Postman and Phillips,
211 1965; Welch and Burnett, 1924); how participants most often transition between recalls of the
212 events as a function of the temporal distance between them (lag-conditional response probability;
213 Fig. 3B; Kahana, 1996); and which events they were likely to remember overall (serial position
214 recall analyses; Fig. 3C; Murdock, 1962). Interestingly, for two of these analyses (probability of
215 first recall and lag-conditional response probability curves) we observed patterns comparable to
216 classic effects from list-learning literature: namely, a higher probability of initiating recall with the
217 first event in the sequence (Fig. 3A) and a higher probability of transitioning to neighboring events

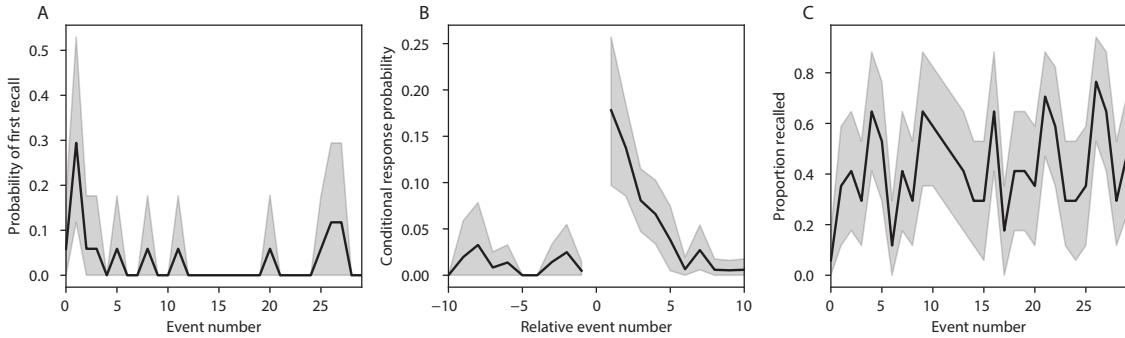


Figure 3: Naturalistic extensions of classic list-learning memory analyses. **A.** The probability of first recall as a function of the serial position of the event in the video. **B.** The probability of recalling each event, conditioned on having most recently recalled the event *lag* events away in 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 with an asymmetric forward bias (Fig. 3B). In contrast, we did not observe a pattern comparable
 219 to the serial position effect (Fig. 3C), but rather greater memory for specific events distributed
 220 approximately evenly throughout the video.

221 We can also apply two list-learning-native analyses that describe how participants group items
 222 in their recall sequences: temporal clustering and semantic clustering (Polyn et al., 2009, see
 223 *Methods* for details). Temporal clustering refers to the extent to which participants group their
 224 recall responses according to encoding position. Overall, we found that sequentially viewed video
 225 events were clustered heavily in participants' recall event sequences (mean clustering score: 0.767,
 226 SEM: 0.029), and that participants with higher temporal clustering scores tended to perform better
 227 according to both Chen et al. (2017)'s hand-annotated memory scores (Pearson's $r(15) = 0.62$, $p =$
 228 0.008) and our model's estimate (Pearson's $r(15) = 0.54$, $p = 0.024$). Semantic clustering measures
 229 the extent to which participants cluster their recall responses according to semantic similarity.
 230 We found that participants tended to recall semantically similar video events together (mean
 231 clustering score: 0.787, SEM: 0.018), and that semantic clustering score was also related to both
 232 hand-annotated (Pearson's $r(15) = 0.65$, $p = 0.004$) and model-derived (Pearson's $r(15) = 0.63$, $p =$
 233 0.007) memory performance.

234 Statistical models of memory studies often treat recall success as binary (in other words, an

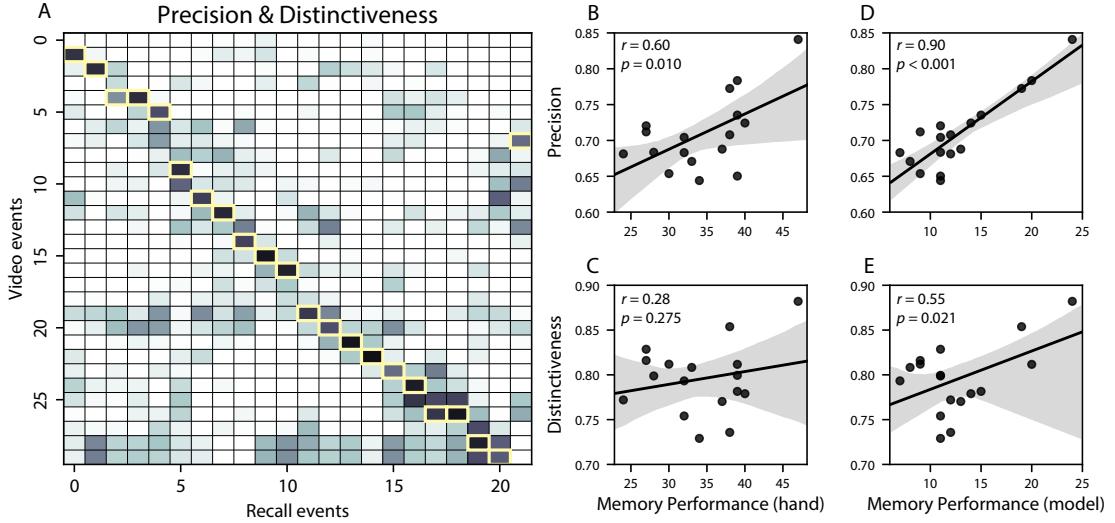


Figure 4: Novel content-based metrics of naturalistic memory: precision and distinctiveness. **A.** The video-recall correlation matrix for a representative participant (17). The yellow boxes highlight the maximum correlation in each column. The example participant's overall precision score was computed 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.

item either was or was not recalled), or occasionally categorical (e.g., to distinguish familiarity from recollection; Yonelinas et al., 2002). Such approaches are tenable in classical list-learning or recognition memory paradigms, as the presented stimuli tend to be very simple (e.g., a sequence 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 positively 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),

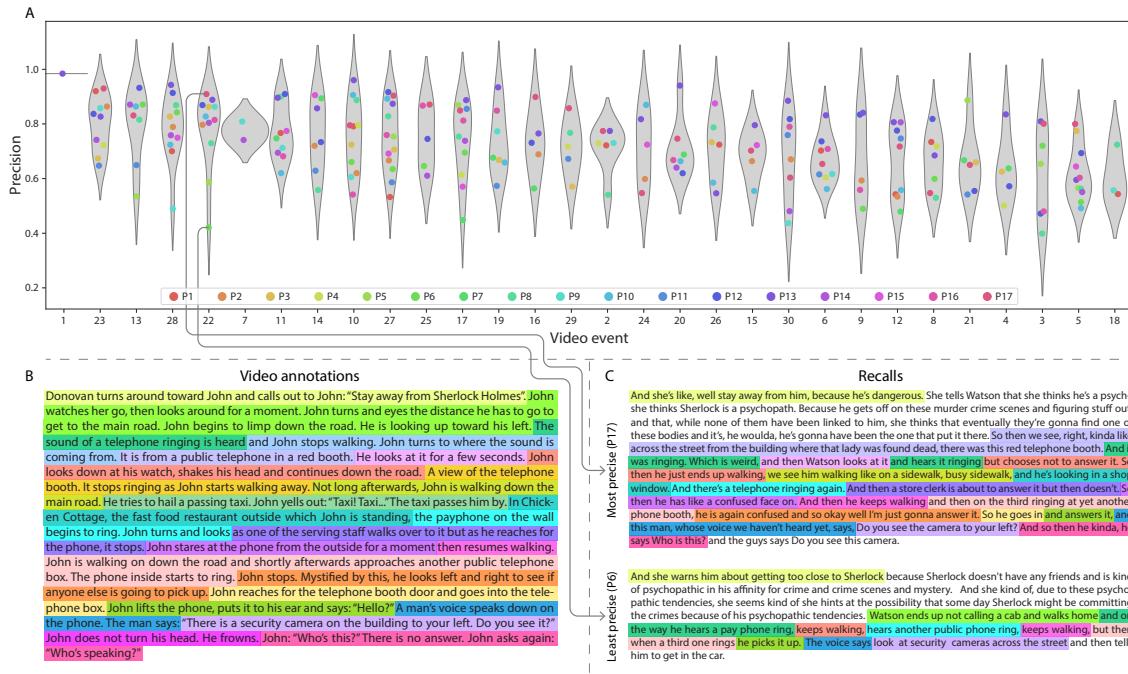


Figure 5: Precision metric reflects completeness of recall. **A.** Recall precision by video event. Grey violin plots display kernel density estimates for the distribution of recall precision scores for a single video event. Colored dots within each violin plot represent individual participants' recall 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.

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 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 these example transcripts reveals that the more precise recall captures more of the video event's

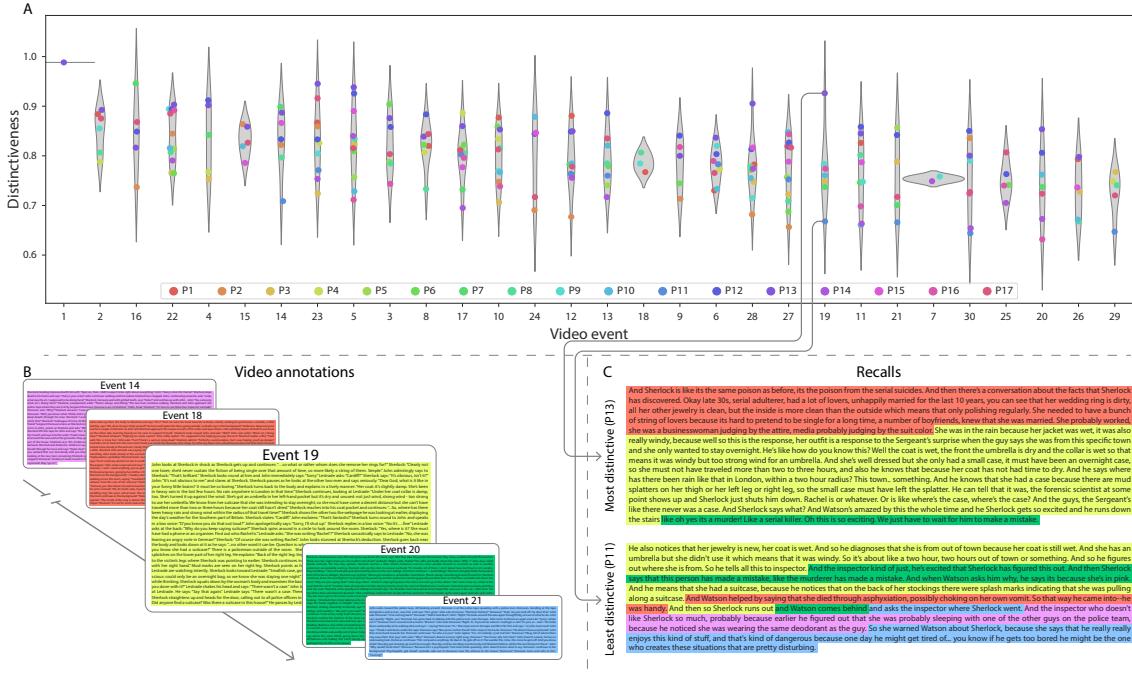


Figure 6: Distinctiveness metric reflects specificity of recall. A. Recall distinctiveness by video event. Kernel density estimates for each video event's distribution of recall distinctiveness scores, analogous to Fig. 5A. **B.** The sets of "Narrative Details" video annotations (generated by Chen et 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 each video event in panel B are highlighted with the corresponding color.

282 content, and with more detail.

283 Figure 6 similarly contrasts two example participants' recalls for a common video event (event
 284 19) to illustrate the tangible differences between high and low distinctiveness scores. Here, we
 285 have extracted the full set of sentences comprising the most distinctive recall event (P13) and least
 286 distinctive recall event (P11) matched to the example video event (Fig. 6C). We also extracted the
 287 annotations for the example video event, as well as those from each other video event whose content
 288 the example participants' single recall events described (Fig. 6B). We then shaded the annotation
 289 text for each video event with a different color, and shaded each word of the example participants'
 290 recall text by the color of the video event it describes. The majority of the most distinctive recall
 291 event text describes video event 19's content, with the first five and last one sentence describing

292 the video events immediately preceding and succeeding the current one, respectively. In contrast,
293 the least distinctive recall of video event 19 blends the content from five separate video events,
294 does not transition between them in order, and often combines descriptions of two video events'
295 content in the same sentence.

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

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

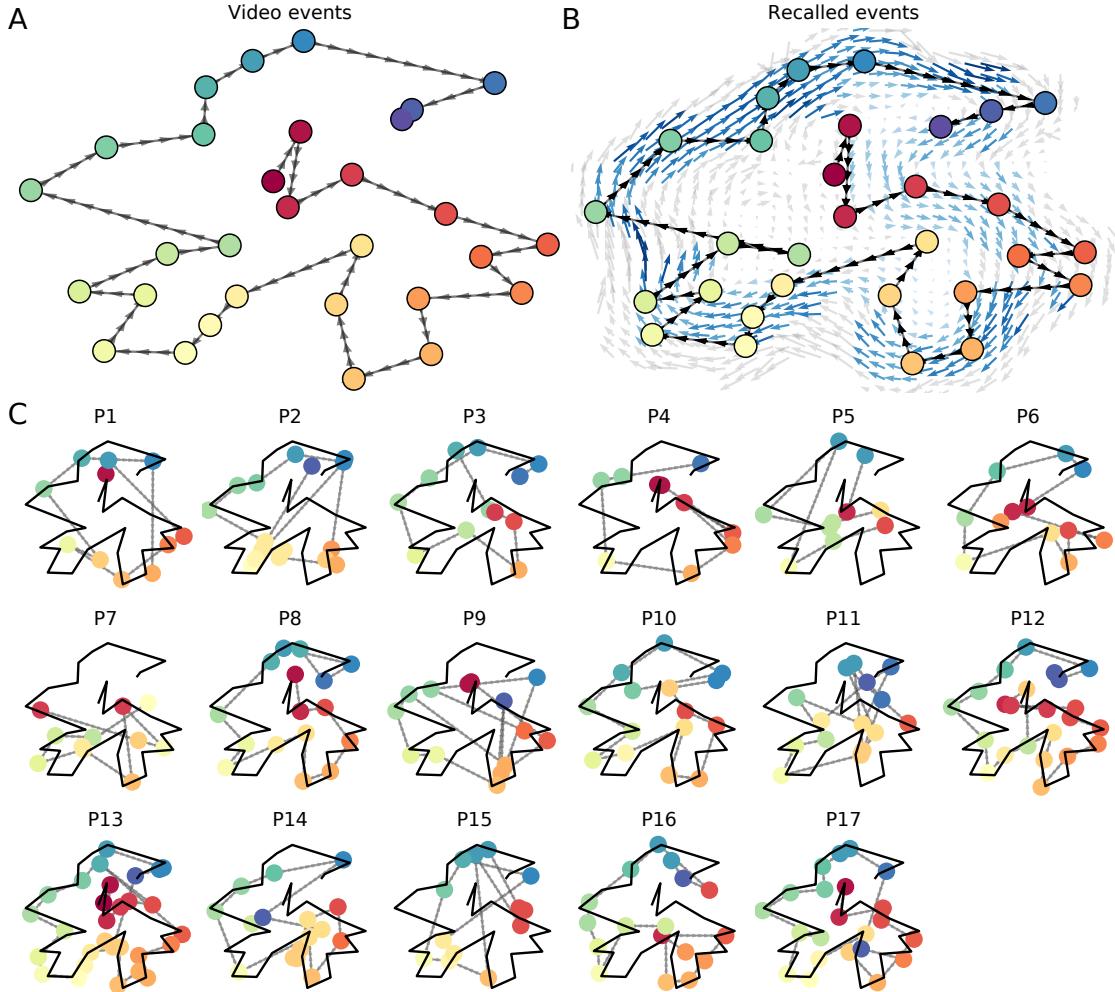


Figure 7: Trajectories through topic space capture the dynamic content of the video and recalls. All panels: the topic proportion matrices have been projected onto a shared two-dimensional space using UMAP. **A.** The two-dimensional topic trajectory taken by the episode of *Sherlock*. Each dot indicates 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 corrected). We observed that the locations traversed by nearly the entire video trajectory exhibited
321 such peaks. In other words, participants exhibited similar trajectories that also matched the
322 trajectory of the original video (Fig. 7C). This is especially notable when considering the fact that
323 the number of events participants recalled (dots in Fig. 7C) varied considerably across people, and
324 that every participant used different words to describe what they had remembered happening in
325 the video. Differences in the numbers of remembered events appear in participants' trajectories
326 as differences in the sampling resolution along the trajectory. We note that this framework also
327 provides a means of disentangling classic "proportion recalled" measures (i.e., the proportion
328 of video events described in participants' recalls) from participants' abilities to recapitulate the
329 overall unfolding of the original video's content (i.e., the similarity between the shapes of the
330 original video trajectory and that defined by each participant's recounting of the video).

331 In addition to the more "holistic" measure of memory described in the previous section, our
332 framework also affords the ability to drill down to individual words and quantify how each word
333 relates to the memorability of each event. The results displayed in Figures 3C and 5A suggest that
334 certain events were remembered better than others. Given this, we next asked whether the
335 events were generally remembered well or poorly tended to reflect particular content. Because
336 our analysis framework projects the dynamic video content and participants' recalls into a shared
337 space, and because the dimensions of that space represent topics (which are, in turn, sets of
338 weights over known words in the vocabulary), we are able to recover the weighted combination
339 of words that make up any point (i.e., topic vector) in this space. We first computed the average
340 precision with which participants recalled each of the 30 video events (Fig. 8A; note that this result
341 is analogous to a serial position curve created from our continuous recall quality metric). We
342 then computed a weighted average of the topic vectors for each video event, where the weights
343 reflected how reliably each event was recalled. To visualize the result, we created a "wordle"
344 image (Mueller et al., 2018) where words weighted more heavily by better-remembered topics
345 appear in a larger font (Fig. 8B, green box). Across the full video, content that reflected topics
346 necessary to convey the central focus of the video (e.g., the names of the two main characters,
347 "Sherlock" and "John," and the address of a major recurring location, "221B Baker Street") were

348 best remembered. An analogous analysis revealed which themes were poorly remembered. Here
349 in computing the weighted average over events' topic vectors, we weighted each event in *inverse*
350 proportion to how well it was remembered (Fig. 8B, red box). The least well-remembered video
351 content reflected information not necessary to later convey a general summary of the video, such
352 as the proper names of relatively minor characters (e.g., "Mike," "Molly," and "Lestrade") and
353 locations (e.g., "St. Bartholomew's Hospital").

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

365 The results thus far inform us about which aspects of the dynamic content in the episode partic-
366 ipants watched were preserved or altered in participants' memories. We next carried out a series
367 of analyses aimed at understanding which brain structures might facilitate these preservations
368 and transformations between the external world and memory. In the first analysis, we sought
369 to identify brain structures that were sensitive to the dynamic unfolding of the video's content,
370 as characterized by its topic trajectory. We used a searchlight procedure to identify clusters of
371 voxels whose activity patterns displayed a proximal temporal correlation structure (as participants
372 watched the video) matching that of the original video's topic proportions (Fig. 9A; see *Methods* for
373 additional details). In a second analysis, we sought to identify brain structures whose responses
374 (during video viewing) reflected how each participant would later structure their recounting of the
375 video. We used an analogous searchlight procedure to identify clusters of voxels whose proximal

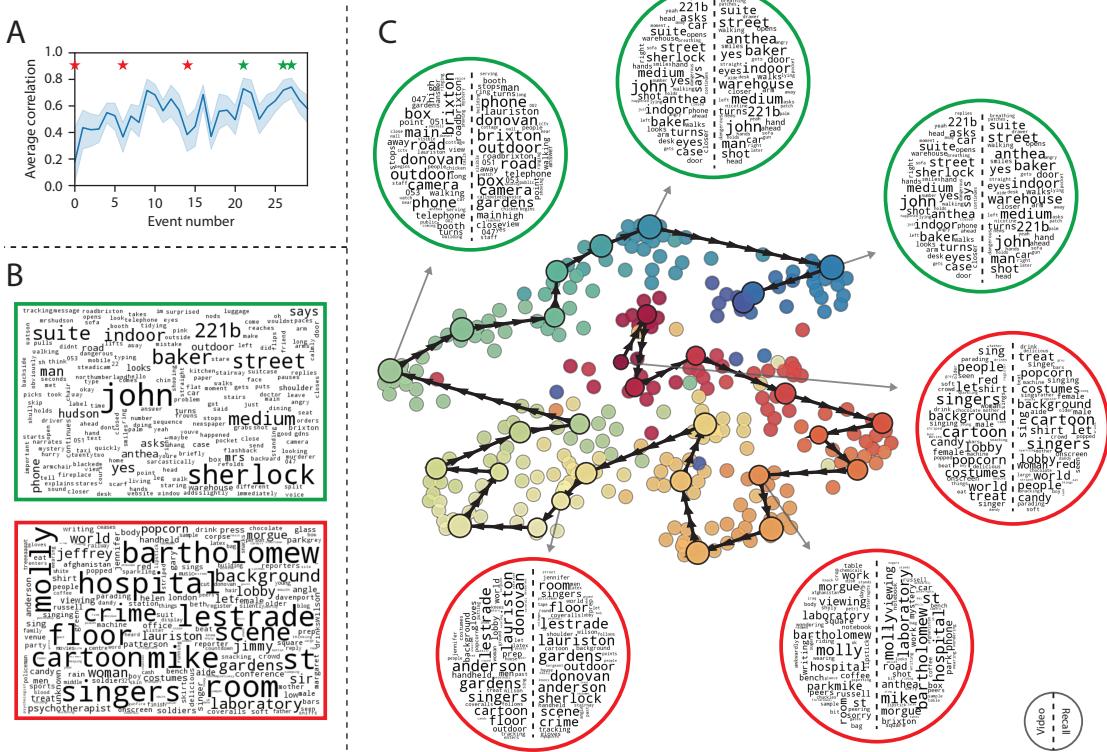


Figure 8: Language used in the most and least memorable events. **A.** Average precision (video event-recall event topic vector correlation) across participants for each video event. Error bars denote bootstrap-derived across-participant 95% confidence intervals. The stars denote the three best-remembered 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.

376 temporal correlation matrices matched that of the topic proportions for each individual's recall
377 (Figs. 9B; see *Methods* for additional details). To ensure our searchlight procedure identified re-
378 gions *specifically* sensitive to the temporal structure of the video or recalls (i.e., rather than those
379 with a temporal autocorrelation length similar to that of the video/recalls), we performed a phase
380 shift-based permutation correction (see *Methods* for additional details). As shown in Figure 9C, the
381 video-driven searchlight analysis revealed a distributed network of regions that may play a role in
382 processing information relevant to the narrative structure of the video. Similarly, the recall-driven
383 searchlight analysis revealed a second network of regions (Fig. 9D) that may facilitate a person-
384 specific transformation of one's experience into memory. In identifying regions whose responses
385 to ongoing experiences reflect how those experiences will be remembered later, this latter analysis
386 extends classic *subsequent memory analyses* (e.g., Paller and Wagner, 2002) to domain of naturalistic
387 stimuli.

388 The searchlight analyses described above yielded two distributed networks of brain regions,
389 whose activity timecourses mirrored to the temporal structure of the video (Fig. 9C) or participants'
390 eventual recalls (Fig. 9D). We next sought to gain greater insight into the structures and functional
391 networks our results reflected. To accomplish this, we performed an additional, exploratory
392 analysis using Neurosynth (Yarkoni et al., 2011). Given an arbitrary statistical map as input,
393 Neurosynth performs a massive automated meta-analysis, returning a ranked list of terms reported
394 in papers with similar significance maps. We ran Neurosynth on the significance maps for the video-
395 and recall-driven searchlight analyses. These maps, along with the 10 terms with maximally similar
396 meta-analysis images identified by Neurosynth are shown in Figure 10.

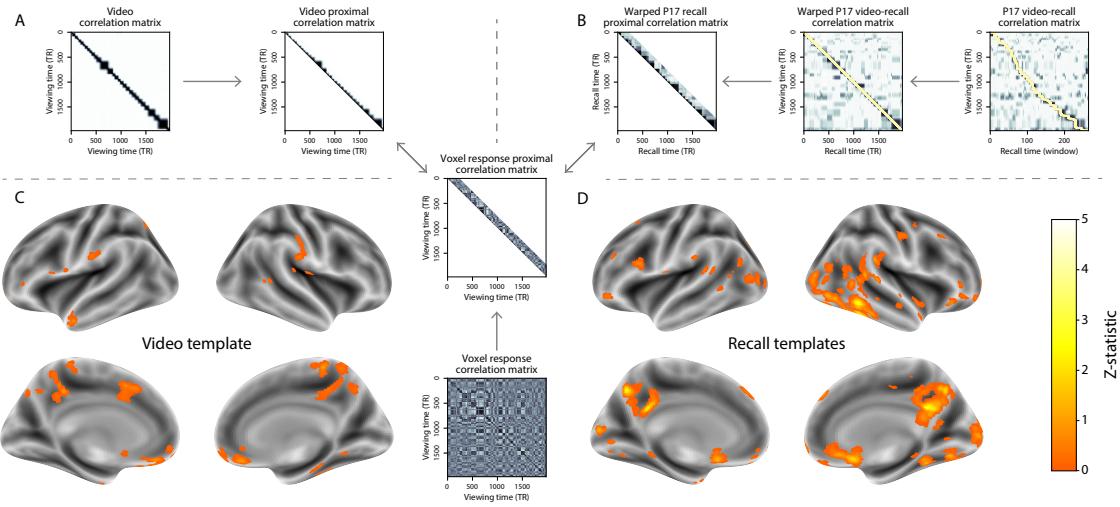


Figure 9: Brain structures that underlie the transformation of experience into memory. **A.** We isolated the proximal diagonals from the upper triangle of the video correlation matrix, and applied this same diagonal mask to the voxel response correlation matrix for each cube of voxels in the brain. 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 of 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.

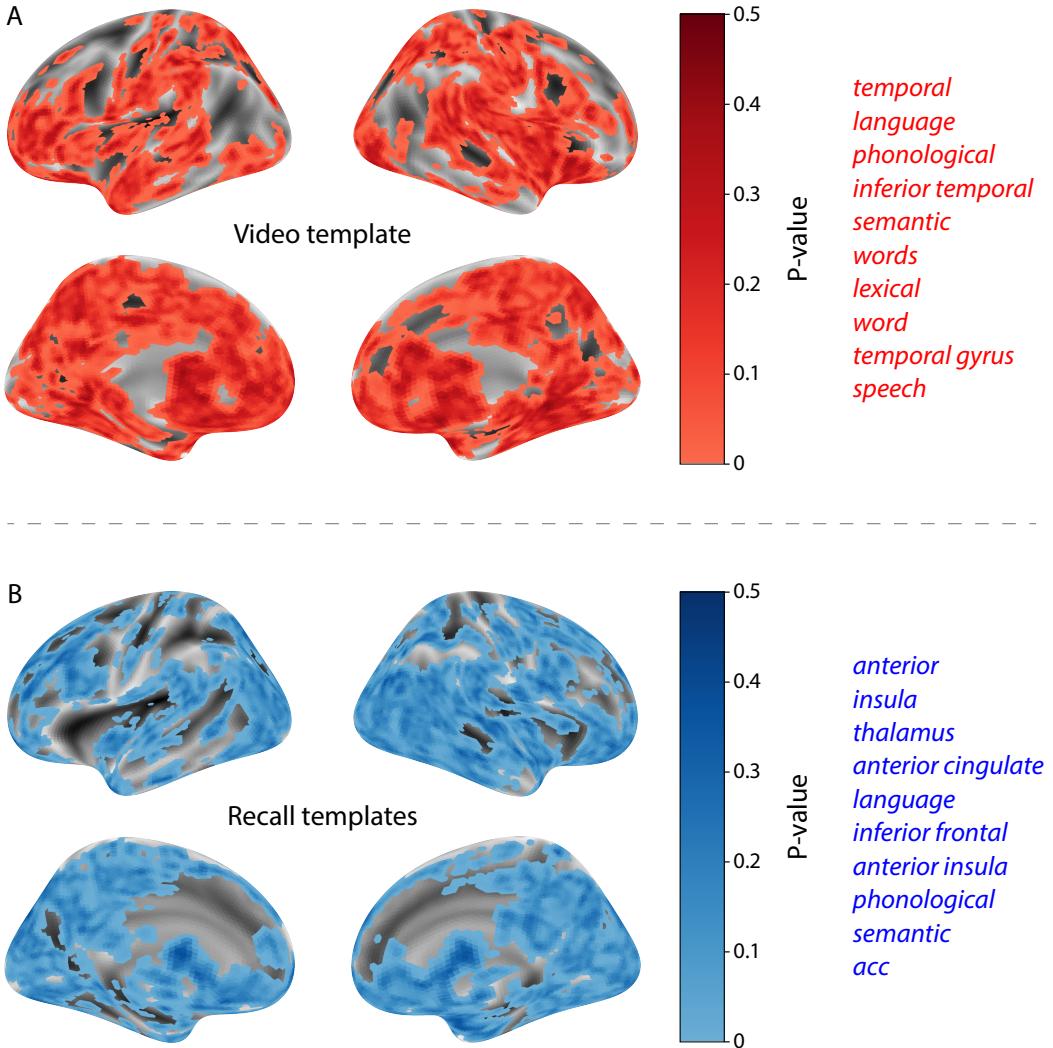


Figure 10: Decoding distributed statistical maps via Neurosynth meta-analyses. **A.** Video-searchlight significance and top 10 decoded terms. We constructed a map of the permutation-derived p -values for the video-driven searchlight analysis (Fig. 9A, C) at each voxel with a positive permutation-derived 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.

397 **Discussion**

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

419 Word embedding models are a rapidly growing area of machine learning research. Early ap-
420 proaches including latent semantic analysis (Landauer and Dumais, 1997) use word co-occurrence
421 statistics (i.e., how often pairs of words occur in the same documents contained in the corpus) to
422 derive a unique feature vector for each word. The feature vectors are constructed so that words
423 that co-occur more frequently have feature vectors that are closer (in Euclidean distance). Related

424 approaches, such as latent dirichlet allocation (Blei et al., 2003) attempt to explicitly model the
425 underlying causes of word co-occurrences by automatically identifying the set of themes or topics
426 reflected across the documents in the corpus. More recent work on these types of semantic mod-
427 els, including word2vec (Mikolov et al., 2013), the Universal Sentence Encoder (Cer et al., 2018),
428 GPT-2 (Radford et al., 2019), and GTP-3 (Brown et al., 2020) use deep neural networks to attempt
429 to identify the deeper conceptual representations underlying each word. Despite the growing
430 popularity of more sophisticated deep learning-based embedding models, here we leverage latent
431 dirichlet allocation (i.e., topic modelling) to embed video and recall text. This decision was mo-
432 tivated by several factors. First, topic models capture the *essence* of a text passage devoid of the
433 specific set and order of words used. This was an important feature of our model since different
434 people may accurately recall a scene using very different language. Second, words can mean dif-
435 ferent things in different contexts (e.g. “bat” may be the act of hitting a baseball, the object used for
436 that action, or as a flying mammal). Topic models are robust to this, allowing words to exist as part
437 of multiple topics. Last, topic models provide a straightforward means of recovering the weights
438 for the particular words comprising a topic, enabling straightforward interpretation of an event’s
439 contents (e.g. Fig. 8). Other models such as the Universal Sentence Encoder, GPT-2, and GPT-3
440 offer context-sensitive encoding of text passages, but the encoding space is complex and non-linear,
441 and thus recovering the original words used to fit the model is not straightforward. However, it is
442 worth pointing out that our general framework is divorced from the particular choice of language
443 model. Moreover, many of the aspects of our framework could be swapped out for other choices.
444 For example, the language model, the timeseries segmentation model and the video-recall match-
445 ing function could all be customized to suit a particular question space or application. Indeed for
446 some questions, recovery of the particular words used to describe a memory may not be necessary,
447 and thus other text-modeling approaches (including the deep learning-based embedding models
448 described above) may be preferable. Future work will explore the influence of particular model
449 choices on the framework’s efficacy.

450 In extending classical free recall analyses to our naturalistic memory framework, we recovered
451 two patterns of recall dynamics central to list-learning studies: a heightened probability of initiating

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

480 reflected in our findings' second departure from typical free recall dynamics: a lack of increased
481 probability of first recall for end-of-sequence events (Fig. 3A).

482 Because they disregard presentation order-dependent variability in the stimulus content, analyses
483 such as those in Figure 3 enable a more sensitive analysis of presentation order-dependent
484 temporal dynamics in free recall. Yet by the same token, they paint a wholly incomplete picture of
485 memory for naturalistic episodes. In an attempt to address this shortcoming, we have developed a
486 framework in the present study that characterizes the explicit semantic content of the stimulus and
487 subsequent recalls. However, sensitivity to stimulus and recall content introduces a new challenge:
488 distinguishing between levels of recall quality for a stimulus (e.g., an event) that is considered to
489 have been "remembered." When modeling memory in an experimental setting, recall quality for
490 individual events is often cast as binary (e.g., a given list item was simply either remembered or
491 not remembered). Various models of memory (e.g., Yonelinas, 2002) attempt to improve upon this
492 by including confidence ratings, rendering this binary judgement instead categorical. To better
493 evaluate naturalistic memory quality, we introduce a continuous metric (*precision*), which reflects
494 the level of completeness of a participant's recall for a feature-rich experience. Additionally, recall
495 quality for a single event is typically assessed independently from that for all other events (e.g., it
496 is difficult to "compare" a participant's binary recall success for list item 1 to that of list item 10).
497 The second novel metric we introduce (*distinctiveness*) is based on analyzing of the correlational
498 structure of an individual's full set of recall events, and reflects the specificity of their memory for
499 a single experienced event. We find that both of these metrics relate to the overall number of video
500 events participants successfully recalled, and that our precision metric additionally relates to Chen
501 et al. (2017)'s hand-annotated memory memory scores.

502 We did not find evidence that participants' average recall distinctiveness was related to their
503 hand-annotated memory scores computed by Chen et al. (2017). One possible explanation is that,
504 in hand-scoring each participant's verbal recall for each of 50 (manually-delimited) scenes, "[a]
505 scene was counted as recalled if the participant described any part of the scene" (Chen et al.,
506 2017). In other words, both an extensive description of a scene's content and a brief mention of
507 some subset of its content were (binarily) considered equally successful recalls. By contrast, we

508 identify the event structure in participants' recalls in an unsupervised manner, independent of
509 the video event-timeseries, prior to mapping between video and recall content. Our HMM-based
510 event-segmentation produces boundaries between timepoints where the topic proportions shift in
511 a substantial way, and because a small handful of words is unlikely to contribute significantly to
512 the topic proportions for any sliding window, such brief scene descriptions will most often not
513 result in a sufficiently large shift in the resulting topic proportions for the HMM to identify an
514 event boundary. Instead, they will be grouped with a neighboring event, consequently lowering
515 that event's distinctiveness score and by extension, the participant's overall distinctiveness score.
516 This is in essence the qualitative difference between distinctive and indistinctive recall, and reflects
517 the comparison shown in Figure 6C. Intriguingly, prior studies show that pattern separation, or the
518 ability to cleanly discriminate between similar experiences, is impaired in many cognitive disorders
519 as well as natural aging (Stark et al., 2010; Yassa et al., 2011; Yassa and Stark, 2011). Future work
520 might explore whether and how these metrics compare between cognitively impoverished groups
521 and healthy controls.

522 In the analyses outlined in Figure 9, we identified two networks of brain regions whose re-
523 sponses during video viewing were consistent with the temporal structure of the video and recall
524 topic trajectories, respectively. The network identified by the video trajectory analysis included the
525 ventromedial prefrontal cortex, left anterior temporal lobe, superior parietal and dorsal anterior
526 cingulate cortex. The network from the video-recall trajectory analysis also included the ventro-
527 medial prefrontal and superior parietal cortices, in addition to the posterior medial cortex (PMC)
528 and the inferior temporal regions. Notably, Chen et al. (2017) also observed the PMC in a number
529 of analyses including one that searched for regions whose activity patterns during encoding were
530 reinstated during free recall. The PMC has been consistently identified in studies involving stimuli
531 with meaningfully structured events (Cohn-Sheely and Ranganath, 2017). Further, the PMC is
532 part of the "posterior medial" system, a network of brain regions thought to represent situation
533 models (Zacks et al., 2007) in support of memory, spatial navigation and social cognition (Ran-
534 ganath and Ritchey, 2012). Given that we constructed our video-recall searchlight model to capture
535 temporal structure in the episode's semantic content (and how one's later recall aligns with that

536 structure), we speculate that the PMC may play a role in constructing mnemonic events from
537 meaningfully structured experiences.

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

548 Our second searchlight analysis identified a partially overlapping network of regions (Fig. 9D)
549 whose patterns of activity as participants viewed the video reflected the idiosyncratic structure of
550 each individual's later recalls. The associated significance map yielded a set of Neurosynth terms
551 that primarily reflected names of specific structural regions (such as "thalamus," "anterior insula,"
552 "anterior cingulate" and "inferior frontal"; Fig. 10B). Interestingly, these regions share membership
553 in a common, large-scale functional network (termed the "salience network") involved in detect-
554 ing and processing affective cues. In particular, the latter three regions have been implicated in
555 functions relevant to assigning personal meaning to an experience, including: ascribing subjective
556 value to raw, sensory input (Medford and Critchley, 2010); modulating semantic and phonological
557 processing in response to personally salient stimuli (Kelly et al., 2007); and directing and reallo-
558 cating attention and working memory resources towards the most relevant stimuli (Menon and
559 Uddin, 2010). This suggests that the network of structures displayed in Figure 9D may be play a roll
560 in transforming and restructuring ongoing experiences through the lens of one's prior experience
561 and subjective emotions as they are encoded in memory.

562 Our work has broad implications for how we characterize and assess memory in real-world
563 settings, such as the classroom or physician's office. For example, the most commonly used

564 classroom evaluation tools involve simply computing the proportion of correctly answered exam
565 questions. Our work indicates that this approach is only loosely related to what educators might
566 really want to measure: how well did the students understand the key ideas presented in the
567 course? Under this typical framework of assessment, the same exam score of 50% could be
568 ascribed to two very different students: one who attended the full course but struggled to learn
569 more than a broad overview of the material, and one who attended only half of the course but
570 understood the material perfectly. Instead, one could apply our computational framework to build
571 explicit content models of the course material and exam questions. This approach would provide
572 a more nuanced and specific view into which aspects of the material students had learned well
573 (or poorly). In clinical settings, memory measures that incorporate such explicit content models
574 might also provide more direct evaluations of patients' memories.

575 Methods

576 Experimental design and data collection

577 Data were collected by Chen et al. (2017). In brief, participants ($n = 22$) viewed the first 48 minutes
578 of "A Study in Pink", the first episode of the BBC television series *Sherlock*, while fMRI volumes
579 were collected (TR = 1500 ms). Participants were pre-screened to ensure they had never seen any
580 episode of the show before. The stimulus was divided into a 23 min (946 TR) and a 25 min (1030 TR)
581 segment to mitigate technical issues related to the scanner. After finishing the clip, participants
582 were instructed to (quoting from Chen et al., 2017) "describe what they recalled of the [episode]
583 in as much detail as they could, to try to recount events in the original order they were viewed
584 in, and to speak for at least 10 minutes if possible but that longer was better. They were told that
585 completeness and detail were more important than temporal order, and that if at any point they
586 realized they had missed something, to return to it. Participants were then allowed to speak for
587 as long as they wished, and verbally indicated when they were finished (e.g., 'I'm done')." Five
588 participants were dropped from the original dataset due to excessive head motion (2 participants),

589 insufficient recall length (2 participants), or falling asleep during stimulus viewing (1 participant),
590 resulting in a final sample size of $n = 17$. For additional details about the experimental procedure
591 and scanning parameters, see Chen et al. (2017). The experimental protocol was approved by
592 Princeton University's Institutional Review Board.

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

599 The video stimulus was divided into 1,000 fine-grained “scenes” and annotated by an inde-
600 pendent coder. For each of these 1,000 scenes, the following information was recorded: a brief
601 narrative description of what was happening, the location where the scene took place, whether
602 that location was indoors or outdoors, the names of all characters on-screen, the name(s) of the
603 character(s) in focus in the shot, the name(s) of the character(s) currently speaking, the camera
604 angle of the shot, a transcription of any text appearing on-screen, and whether or not there was
605 music present in the background. Each scene was also tagged with its onset and offset time, in
606 both seconds and TRs.

607 **Data and code availability**

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

610 **Statistics**

611 All statistical tests performed in the behavioral analyses were two-sided. All statistical tests per-
612 formed in the neural data analyses were two-sided, except for the permutation-based thresholding,
613 which was one-sided. In this case, we were specifically interested in identifying voxels whose ac-

614 tivation time series reflected the temporal structure of the video and recall trajectories to a *greater*
615 extent than that of the phase-shifted trajectories.

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

617 **Topic modeling**

618 The input to the topic model we trained to characterize the dynamic content of the video comprised
619 998 hand-generated annotations of short (mean: 2.96s) scenes spanning the video clip (Chen
620 et al., 2017 generated 1000 annotations total; we removed two annotations referring to a break
621 between the first and second scan sessions, during which no fMRI data was collected). We
622 concatenated the text for all of the annotated features within each segment, creating a “bag of
623 words” describing each scene and performed some minor preprocessing (e.g., stemming possessive
624 nouns and removing punctuation). We then re-organized the text descriptions into overlapping
625 sliding windows spanning (up to) 50 scenes each. In other words, we estimated the “context”
626 for each scene using the text descriptions of the preceding 25 scenes, the present scene, and the
627 following 24 scenes. To model the context for scenes near the beginning of the video (i.e., within
628 25 scenes of the beginning or end), we created overlapping sliding windows that grew in size
629 from one scene to the full length. We also tapered the sliding window lengths at the end of the
630 video, whereby scenes within fewer than 24 scenes of the end of the video were assigned sliding
631 windows that extended to the end of the video. This procedure ensured that each scene’s content
632 was represented in the text corpus an equal number of times.

633 We trained our model using these overlapping text samples with `scikit-learn` (version 0.19.1;
634 Pedregosa et al., 2011), called from our high-dimensional visualization and text analysis software,
635 `HyperTools` (Heusser et al., 2018b). Specifically, we used the `CountVectorizer` class to transform
636 the text from each window into a vector of word counts (using the union of all words across all
637 scenes as the “vocabulary,” excluding English stop words); this yielded a number-of-windows
638 by number-of-words *word count* matrix. We then used the `LatentDirichletAllocation` class
639 (`topics=100`, `method='batch'`) to fit a topic model (Blei et al., 2003) to the word count matrix,

640 yielding a number-of-windows (1047) by number-of-topics (100) *topic proportions* matrix. The
641 topic proportions matrix describes the gradually evolving mix of topics (latent themes) present in
642 each scene. Next, we transformed the topic proportions matrix to match the 1976 fMRI volume
643 acquisition times. We assigned each topic vector to the timepoint (in seconds) midway between the
644 beginning of the first scene and the end of the last scene in its corresponding sliding text window.
645 By doing so, we warped the linear temporal distance between consecutive topic vectors to align
646 with the inconsistent temporal distance between consecutive annotations (whose durations varied
647 greatly). We then rescaled these timepoints to 1.5s TR units, and used linear interpolation to
648 estimate a topic vector for each TR. This resulted in a number-of-TRs (1976) by number-of-topics
649 (100) matrix.

650 We created similar topic proportions matrices using hand-annotated transcripts of each partici-
651 pant's verbal recall of the video (annotated by Chen et al., 2017). We tokenized the transcript into a
652 list of sentences, and then re-organized the list into overlapping sliding windows spanning (up to)
653 10 sentences each, analogously to how we parsed the video annotations. In turn, we transformed
654 each window's sentences into a word count vector (using the same vocabulary as for the video
655 model), and then we used the topic model already trained on the video scenes to compute the most
656 probable topic proportions for each sliding window. This yielded a number-of-windows (range:
657 83–312) by number-of-topics (100) topic proportions matrix for each participant. These reflected
658 the dynamic content of each participant's recalls. Note: for details on how we selected the video
659 and recall window lengths and number of topics, see *Supporting Information* and Figure S1.

660 **Parsing topic trajectories into events using Hidden Markov Models**

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

667 to implement this segmentation.

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

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

677 where a was the distribution of within-state topic vector correlations, and b was the distribution of
678 across-state topic vector correlations . We computed the first Wasserstein distance (W_1 ; also known
679 as *Earth mover's distance*; Dobrushin, 1970; Ramdas et al., 2017) between these distributions for a
680 large range of possible K -values (range [2, 50]), and selected the K that yielded the maximum value.
681 Figure 2B displays the event boundaries returned for the video, and Figure S4 displays the event
682 boundaries returned for each participant's recalls. See Figure S6 for the optimization functions
683 for the video and recalls. After obtaining these event boundaries, we created stable estimates of
684 the content represented in each event by averaging the topic vectors across timepoints between
685 each pair of event boundaries. This yielded a number-of-events by number-of-topics matrix for
686 the video and recalls from each participant.

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

688 In traditional list-learning experiments, participants view a list of items (e.g., words) and then recall
689 the items later. Our video-recall event matching approach affords us the ability to analyze memory
690 in a similar way. The video and recall events can be treated analogously to studied and recalled

691 “items” in a list-learning study. We can then extend classic analyses of memory performance and
692 dynamics (originally designed for list-learning experiments) to the more naturalistic video recall
693 task used in this study.

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

703 **Probability of first recall (PFR).** PFR curves (Welch and Burnett, 1924; Postman and Phillips,
704 1965; Atkinson and Shiffrin, 1968) reflect the probability that an item will be recalled first as a
705 function of its serial position during encoding. To carry out this analysis, we initialized a number-
706 of-participants (17) by number-of-video-events (30) matrix of zeros. Then for each participant, we
707 found the index of the video event that was recalled first (i.e., the video event whose topic vector
708 was most strongly correlated with that of the first recall event) and filled in that index in the matrix
709 with a 1. Finally, we averaged over the rows of the matrix, resulting in a 1 by 30 array representing
710 the proportion of participants that recalled an event first, as a function of the order of the event’s
711 appearance in the video (Fig. 3A).

712 **Lag conditional probability curve (lag-CRP).** The lag-CRP curve (Kahana, 1996) reflects the
713 probability of recalling a given item after the just-recalled item, as a function of their relative
714 encoding positions (or *lag*). In other words, a lag of 1 indicates that a recalled item was presented
715 immediately after the previously recalled item, and a lag of -3 indicates that a recalled item came 3
716 items before the previously recalled item. For each recall transition (following the first recall), we

717 computed the lag between the current recall event and the next recall event, normalizing by the
718 total number of possible transitions. This yielded a number-of-participants (17) by number-of-lags
719 (-29 to +29; 58 lags total excluding lags of 0) matrix. We averaged over the rows of this matrix to
720 obtain a group-averaged lag-CRP curve (Fig. 3B).

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

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

740 **Semantic clustering scores.** Semantic clustering describes a participant's tendency to recall se-
741 mantically similar presented items together in their recall sequences. Here, we used the topic
742 vectors for each event as a proxy for its semantic content. Thus, the similarity between the seman-

tic content for two events can be computed by correlating their respective topic vectors. For each recall event transition, we sorted all not-yet-recalled events according to how correlated the topic vector of the closest-matching video event was to the topic vector of the closest-matching video event to the just-recalled event. We then computed the percentile rank of the observed next recall. We averaged these percentile ranks across all of the participant's recalls to obtain a single semantic clustering score for the participant.

749 Novel naturalistic memory metrics

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

756 **Distinctiveness.** We also considered the *distinctiveness* of each recalled event. That is, how unique
757 a participant's description of a video event was, versus their descriptions of other video events.
758 We hypothesized that participants with high memory performance might describe each event in
759 a more distinctive way (relative to those with lower memory performance who might describe
760 events in a more general way). To test this hypothesis we define a distinctiveness score for each
761 recall event i as

$$d(i) = 1 - \frac{1}{N-1} \sum_{j=i} \text{corr}(\text{event}_i, \text{event}_j)$$

762 where the average is taken over the correlation between the recall event i 's topic vector and the
763 topic vectors from all other recall events from that participant. We averaged these distinctiveness
764 scores across all of the events recalled by the given participant to get the participant's distinctiveness
765 score. We correlated these distinctiveness scores with hand-annotated and model-derived memory

766 performance scores across-subjects, as above.

767 **Averaging correlations** In all instances where we performed statistical tests involving precision
768 or distinctiveness scores, we used the Fisher z -transformation (Fisher, 1925) to stabilize the variance
769 across the distribution of correlation values prior to performing the test. Similarly, when averaging
770 precision or distinctiveness scores, we z -transformed the scores prior to computing the mean, and
771 inverse z -transformed the result.

772 **Visualizing the video and recall topic trajectories**

773 We used the UMAP algorithm (McInnes et al., 2018) to project the 100-dimensional topic space onto
774 a two-dimensional space for visualization (Figs. 7, 8). To ensure that all of the trajectories were
775 projected onto the *same* lower dimensional space, we computed the low-dimensional embedding
776 on a “stacked” matrix created by vertically concatenating the events-by-topics topic proportions
777 matrices for the video, across-participants average recall and all 17 individual participants’ recalls.
778 We then separated the rows of the result (a total-number-of-events by two matrix) back into
779 individual matrices for the video topic trajectory, across-participant average recall trajectory and the
780 trajectories for each individual participant’s recalls (Fig. 7). This general approach for discovering
781 a shared low-dimensional embedding for a collections of high-dimensional observations follows
782 Heusser et al. (2018b).

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

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

793 In Figure 7B, we present an analysis aimed at characterizing locations in topic space that dif-
794 ferent participants move through in a consistent way (via their recall topic trajectories). The
795 two-dimensional topic space used in our visualizations (Fig. 7) comprised a 60×60 (arbitrary
796 units) square. We tiled this space with a 50×50 grid of evenly spaced vertices, and defined a
797 circular area centered on each vertex whose radius was two times the distance between adjacent
798 vertices (i.e., 2.4 units). For each vertex, we examined the set of line segments formed by connecting
799 each pair successively recalled events, across all participants, that passed through this circle. We
800 computed the distribution of angles formed by those segments and the x -axis, and used a Rayleigh
801 test to determine whether the distribution of angles was reliably “peaked” (i.e., consistent across
802 all transitions that passed through that local portion of topic space). To create Figure 7B we drew
803 an arrow originating from each grid vertex, pointing in the direction of the average angle formed
804 by the line segments that passed within 2.4 units. We set the arrow lengths to be inversely propor-
805 tional to the p -values of the Rayleigh tests at each vertex. Specifically, for each vertex we converted
806 all of the angles of segments that passed within 2.4 units to unit vectors, and we set the arrow
807 lengths at each vertex proportional to the length of the (circular) mean vector. We also indicated
808 any significant results ($p < 0.05$, corrected using the Benjamani-Hochberg procedure) by coloring
809 the arrows in blue (darker blue denotes a lower p -value, i.e., a longer mean vector); all tests with
810 $p \geq 0.05$ are displayed in gray and given a lower opacity value.

811 **Searchlight fMRI analyses**

812 In Figure 9, we present two analyses aimed at identifying brain regions whose responses (as par-
813 ticipants viewed the video) exhibited a particular temporal structure. We developed a searchlight
814 analysis wherein we constructed a $5 \times 5 \times 5$ cube of voxels (following Chen et al., 2017) centered on
815 each voxel in the brain, and for each of these cubes, computed the temporal correlation matrix of
816 the voxel responses during video viewing. Specifically, for each of the 1976 volumes collected dur-
817 ing video viewing, we correlated the activity patterns in the given cube with the activity patterns

818 (in the same cube) collected during every other timepoint. This yielded a 1976 by 1976 correlation
819 matrix for each cube. Note: participant 5's scan ended 75s early, and in Chen et al., 2017's publicly
820 released dataset, their scan data was padded to match the length of the other participants'. For
821 our searchlight analyses, we removed this padded data (i.e., the last 50 TRs), resulting in a 1925 by
822 1925 correlation matrix for each cube in participant 5's brain.

823 Next, we constructed a series of "template" matrices. The first template reflected the timecourse
824 of the video's topic trajectory, and the others reflected the timecourse of each participant's recall
825 trajectory. To construct the video template, we computed the correlations between the topic
826 proportions estimated for every pair of TRs (prior to segmenting the trajectory into discrete events;
827 i.e., the correlation matrix shown in Figs. 2B and 9A). We constructed similar temporal correlation
828 matrices for each participant's recall topic trajectory (Figs. 2D, S4). However, to correct for length
829 differences and potential non-linear transformations between viewing time and recall time, we
830 first used dynamic time warping (Berndt and Clifford, 1994) to temporally align participants'
831 recall topic trajectories with the video topic trajectory. An example correlation matrix before and
832 after warping is shown in Fig. 9B. This yielded a 1976 by 1976 correlation matrix for the video
833 template and for each participant's recall template.

834 The temporal structure of the video's content (as described by our model) is captured in the
835 block-diagonal structure of the video's temporal correlation matrix (e.g., Figs. 2B, 9A), with time
836 periods of thematic stability represented as dark blocks of varying sizes. Inspecting the video
837 correlation matrix suggests that the video's semantic content is highly temporally specific (i.e., the
838 correlations between topic vectors from distant timepoints are almost all near zero). By contrast,
839 the activity patterns of individual (cubes of) voxels can encode relatively limited information
840 on their own, and their activity frequently contributes to multiple separate functions (Freedman
841 et al., 2001; Sigman and Dehaene, 2008; Charron and Koechlin, 2010; Rishel et al., 2013). By
842 nature, these two attributes give rise to similarities in activity across large timescales that may not
843 necessarily reflect a single task. To enable a more sensitive analysis of brain regions whose shifts
844 in activity patterns mirrored shifts in the semantic content of the video or recalls, we restricted
845 the temporal correlations we considered to the timescale of semantic information captured by our

846 model. Specifically, we isolated the upper triangle of the video correlation matrix and created a
847 “proximal correlation mask” that included only diagonals from the upper triangle of the video
848 correlation matrix up to the first diagonal that contained no positive correlations. Applying this
849 mask to the full video correlation matrix was analogous to excluding diagonals beyond the corner
850 of the largest diagonal block. In other words, the timescale of temporal correlations we considered
851 corresponded to the longest period of thematic stability in the video, and by extension the longest
852 expected period of thematic stability in participants’ recalls and the longest period of stability we
853 might expect to see in voxel activity arising from processing or encoding video content. Figure 9
854 shows this proximal correlation mask applied to the temporal correlation matrices for the video,
855 an example participant’s (warped) recall, and an example cube of voxels from our searchlight
856 analyses.

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

863 We further sought to ensure that our analysis identified regions where the activations’ temporal
864 structure specifically reflected that of the video, rather than regions whose activity was simply
865 autocorrelated at a width similar to the video template’s diagonal. To achieve this, we used a phase
866 shift-based permutation procedure, whereby we circularly shifted the video’s topic trajectory by
867 a random number of timepoints, computed the resulting “null” video template, and re-ran the
868 searchlight analysis, in full. (For each of the 100 permutations, the same random shift was used for
869 all participants). We z-scored the observed (unshifted) result at each voxel against the distribution
870 of permutation-derived “null” results, and estimated a *p*-value by computing the proportion of
871 shifted results that yielded larger values. To create the map in Figure 9C, we thresholded out
872 any voxels whose similarity to the unshifted video’s structure fell below the 95th percentile of the
873 permutation-derived similarity results.

874 We used an analogous procedure to identify which voxels' responses reflected the recall tem-
875 plates. For each participant, we correlated the proximal diagonals from the upper triangle of the
876 correlation matrix for each cube of voxels with the proximal diagonals from the upper triangle
877 of their (time-warped) recall correlation matrix. As in the video template analysis, this yielded a
878 voxelwise map of correlation coefficients per participant. However, whereas the video analysis
879 compared every participant's responses to the same template, here the recall templates were unique
880 for each participant. As in the analysis described above, we *t*-scored the (Fisher *z*-transformed)
881 voxelwise correlations, and used the same permutation procedure we developed for the video
882 responses to ensure specificity to the recall timeseries and assign significance values. To create the
883 map in Figure 9D we again thresholded out any voxels whose scores were below the 95th percentile
884 of the permutation-derived null distribution.

885 **Neurosynth decoding analyses**

886 Neurosynth parses a massive online database of over 14,000 neuroimaging studies and constructs
887 meta-analysis images for over 13,000 psychology- and neuroscience-related terms, based on NIfTI
888 images accompanying studies where those terms appear at a high frequency. Given a novel image
889 (tagged with its value type; e.g., *t*-, *F*- or *p*-statistics), Neurosynth returns a list of terms whose
890 meta-analysis images are most similar. Our permutation procedure yielded, for each of the two
891 searchlight analyses, a voxelwise map of significance (*p*-statistic) values. These maps describe the
892 extent to which each voxel *specifically* reflected the temporal structure of the video or individuals'
893 recalls (i.e., for each voxel, the proportion of phase-shifted topic vector correlation matrices less
894 similar to the voxel activity correlation matrix than the unshifted video's correlation matrix). We
895 inputted the two statistical maps described above to Neurosynth to create a list of the 10 most
896 representative terms for each map.

897 **References**

- 898 Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control
899 processes. In Spence, K. W. and Spence, J. T., editors, *The psychology of learning and motivation*,
900 volume 2, pages 89–105. Academic Press, New York.
- 901 Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017).
902 Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3):709–
903 721.
- 904 Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of real-world event schemas
905 during narrative perception. *Journal of Neuroscience*, 38(45):9689–9699.
- 906 Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In
907 *KDD workshop*, volume 10, pages 359–370.
- 908 Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International
909 Conference on Machine Learning*, ICML '06, pages 113–120, New York, NY, US. ACM.
- 910 Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine
911 Learning Research*, 3:993 – 1022.
- 912 Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam,
913 P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R.,
914 Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S.,
915 Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020).
916 Language models are few-shot learners. *arXiv*, 2005.14165.
- 917 Brunec, I. K., Moscovitch, M. M., and Barene, M. D. (2018). Boundaries shape cognitive represen-
918 tations of spaces and events. *Trends in Cognitive Sciences*, 22(7):637–650.
- 919 Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., and Oliva, A. (2015). Intrinsic and extrinsic
920 effects on image memorability. *Vision Research*, 116:165–178.

- 921 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., Keller, B., Wang, Y., and
922 Shin, Y. S. (2017). Brain imaging analysis kit.
- 923 Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes,
924 M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal sentence encoder.
925 *arXiv*, 1803.11175.
- 926 Charron, S. and Koechlin, E. (2010). Divided representations of current goals in the human frontal
927 lobes. *Science*, 328(5976):360–363.
- 928 Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared
929 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,
930 20(1):115.
- 931 Chun, M. and Turk-Browne, N. (2007). Interactions between attention and memory. *Current opinion
932 in neurobiology*, 17(2):177–184.
- 933 Clewett, D. and Davachi, L. (2017). The ebb and flow of experience determines the temporal
934 structure of memory. *Curr Opin Behav Sci*, 17:186–193.
- 935 Cohn-Sheely, B. I. and Ranganath, C. (2017). Time regained: how the human brain constructs
936 memory for time. *Current Opinion in Behavioral Sciences*, 17:169–177.
- 937 Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions.
938 *Theory of Probability & Its Applications*, 15(3):458–486.
- 939 DuBrow, S. and Davachi, L. (2013). The influence of contextual boundaries on memory for the
940 sequential order of events. *Journal of Experimental Psychology: General*, 142(4):1277–1286.
- 941 Ezzyat, Y. and Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological
942 Science*, 22(2):243–252.
- 943 Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

- 944 Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of
945 visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316.
- 946 Friendly, M., Franklin, P. E., Hoffman, D., and Rubin, D. C. (1982). The Toronto Word Pool:
947 Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080
948 words. *Behavior Research Methods and Instrumentation*, 14:375–399.
- 949 Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral
950 component of information processing. *Trends in Cognitive Science*, 19(6):304–315.
- 951 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal
952 receptive windows in human cortex. *Journal of Neuroscience*, 28(10):2539–2550.
- 953 Heusser, A. C., Ezzyat, Y., Shiff, I., and Davachi, L. (2018a). Perceptual boundaries cause mnemonic
954 trade-offs between local boundary processing and across-trial associative binding. *Journal of*
955 *Experimental Psychology Learning, Memory, and Cognition*, 44(7):1075–1090.
- 956 Heusser, A. C., Fitzpatrick, P. C., Field, C. E., Ziman, K., and Manning, J. R. (2017). Quail: a
957 Python toolbox for analyzing and plotting free recall data. *The Journal of Open Source Software*,
958 10.21105/joss.00424.
- 959 Heusser, A. C., Ziman, K., Owen, L. L. W., and Manning, J. R. (2018b). HyperTools: a Python
960 toolbox for gaining geometric insights into high-dimensional data. *Journal of Machine Learning*
961 *Research*, 18(152):1–6.
- 962 Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal*
963 *of Mathematical Psychology*, 46:269–299.
- 964 Howard, M. W., MacDonald, C. J., Tiganj, Z., Shankar, K. H., Du, Q., Hasselmo, M. E., and H., E.
965 (2014). A unified mathematical framework for coding time, space, and sequences in the medial
966 temporal lobe. *Journal of Neuroscience*, 34(13):4692–4707.
- 967 Howard, M. W., Viskontas, I. V., Shankar, K. H., and Fried, I. (2012). Ensembles of human MTL
968 neurons “jump back in time” in response to a repeated stimulus. *Hippocampus*, 22:1833–1847.

- 969 Huk, A., Bonnen, K., and He, B. J. (2018). Beyond trial-based paradigms: continuous behavior, on-
970 going neural activity, and naturalistic stimuli. *Journal of Neuroscience*, 10.1523/JNEUROSCI.1920-
971 17.2018.
- 972 Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24:103–109.
- 973 Kelly, S., Lloyd, D., Nurmikko, T., and Roberts, N. (2007). Retrieving autobiographical memories
974 of painful events activates the anterior cingulate cortex and inferior frontal gyrus. *The Journal of
975 Pain*, 8(4):307–314.
- 976 Kriat, A. and Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distin-
977 guishing accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of
978 Experimental Psychology: General*, 123(3):297–315.
- 979 Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – con-
980 necting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:1 – 28.
- 981 Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato’s problem: the latent semantic
982 analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*,
983 104:211–240.
- 984 Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy
985 of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8):2906–2915.
- 986 Manning, J. R. (2019). Episodic memory: mental time travel or a quantum ‘memory wave’ function?
987 *PsyArXiv*, doi:10.31234/osf.io/6zjwb.
- 988 Manning, J. R. (2020). Context reinstatement. In Kahana, M. J. and Wagner, A. D., editors, *Handbook
989 of Human Memory*. Oxford University Press.
- 990 Manning, J. R., Norman, K. A., and Kahana, M. J. (2015). The role of context in episodic memory.
991 In Gazzaniga, M., editor, *The Cognitive Neurosciences, Fifth edition*, pages 557–566. MIT Press.

- 992 Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns
993 in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National*
994 *Academy of Sciences, USA*, 108(31):12893–12897.
- 995 McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and
996 projection for dimension reduction. *arXiv*, 1802(03426).
- 997 Medford, N. and Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate
998 cortex: awareness and response. *Brain Structure and Function*, 214(5-6):535–549.
- 999 Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of
1000 insula function. *Brain Structure and Function*, 214(5-6):655–667.
- 1001 Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations
1002 in vector space. *arXiv*, 1301.3781.
- 1003 Mueller, A., Fillion-Robin, J.-C., Boidol, R., Tian, F., Nechifor, P., yoonsubKim, Peter, Rampin, R.,
1004 Corvellec, M., Medina, J., Dai, Y., Petrushev, B., Langner, K. M., Hong, Alessio, Ozsvald, I.,
1005 vkolmakov, Jones, T., Bailey, E., Rho, V., IgorAPM, Roy, D., May, C., foobuzz, Piyush, Seong,
1006 L. K., Goey, J. V., Smith, J. S., Gus, and Mai, F. (2018). WordCloud 1.5.0: a little word cloud
1007 generator in Python. *Zenodo*, <https://zenodo.org/record/1322068#.W4tPKZNKh24>.
- 1008 Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*,
1009 64:482–488.
- 1010 Paller, K. A. and Wagner, A. D. (2002). Observing the transformation of experience into memory.
1011 *Trends in Cognitive Sciences*, 6(2):93–102.
- 1012 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Pretten-
1013 hofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,
1014 M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine*
1015 *Learning Research*, 12:2825–2830.

- 1016 Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model
1017 of organizational processes in free recall. *Psychological Review*, 116(1):129–156.
- 1018 Postman, L. and Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal*
1019 of *Experimental Psychology*, 17:132–138.
- 1020 Rabiner, L. (1989). A tutorial on Hidden Markov Models and selected applications in speech
1021 recognition. *Proceedings of the IEEE*, 77(2):257–286.
- 1022 Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are
1023 unsupervised multitask learners. *OpenAI Blog*, 1(8).
- 1024 Radvansky, G. A. and Zacks, J. M. (2017). Event boundaries in memory and cognition. *Curr Opin*
1025 *Behav Sci*, 17:133–140.
- 1026 Ramdas, A., Trillos, N., and Cuturi, M. (2017). On wasserstein two-sample testing and related
1027 families of nonparametric tests. *Entropy*, 19(2):47.
- 1028 Ranganath, C. and Ritchey, M. (2012). Two cortical systems for memory-guided behavior. *Nature*
1029 *Reviews Neuroscience*, 13:713 – 726.
- 1030 Rishel, C. A., Huang, G., and Freedman, D. J. (2013). Independent category and spatial encoding
1031 in parietal cortex. *Neuron*, 77(5):969–979.
- 1032 Sigman, M. and Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during
1033 dual-task performance. *Journal of Neuroscience*, 28(30):7585–7589.
- 1034 Simony, E., Honey, C. J., Chen, J., and Hasson, U. (2016). Uncovering stimulus-locked network
1035 dynamics during narrative comprehension. *Nature Communications*, 7(12141):1–13.
- 1036 Stark, S. M., Yassa, M. A., and Stark, C. E. L. (2010). Individual differences in spatial pattern
1037 separation performance associated with healthy aging in humans. *Learning & Memory*, 17(6):284–
1038 288.

- 1039 Tomrary, A. and Davachi, L. (2017). Consolidation promotes the emergence of representational
1040 overlap in the hippocampus and medial prefrontal cortex. *Neuron*, 96(1):228–241.
- 1041 Tyng, C. M., Amin, H. U., Saad, M. N. M., and S, M. A. (2017). The influences of emotion on
1042 learning and memory. *Frontiers in psychology*, 8:1454.
- 1043 Welch, G. B. and Burnett, C. T. (1924). Is primacy a factor in association-formation. *American Journal
1044 of Psychology*, 35:396–401.
- 1045 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale
1046 automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665.
- 1047 Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., and Stark, C. E. L. (2011). Pattern
1048 separation deficits associated with increased hippocampal ca3 and dentate gyrus activity in
1049 nondemented older adults. *Hippocampus*, 21(9):968–979.
- 1050 Yassa, M. A. and Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends In Neuro-
1051 sciences*, 34(10):515–525.
- 1052 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research.
1053 *Journal of Memory and Language*, 46:441–517.
- 1054 Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., and
1055 Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection
1056 and familiarity. *Nature Neuroscience*, 5(11):1236–41.
- 1057 Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception:
1058 a mind-brain perspective. *Psychological Bulletin*, 133:273–293.
- 1059 Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit
1060 memories to other brains: Constructing shared neural representations via communication. *Cereb
1061 Cortex*, 27(10):4988–5000.
- 1062 Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and
1063 memory. *Psychological Bulletin*, 123(2):162 – 185.

1064 **Supporting information**

1065 Supporting information is available in the online version of the paper.

1066 **Acknowledgements**

1067 We thank Luke Chang, Janice Chen, Chris Honey, Lucy Owen, Emily Whitaker, and Kirsten Ziman
1068 for feedback and scientific discussions. We also thank Janice Chen, Yuan Chang Leong, Kenneth
1069 Norman, and Uri Hasson for sharing the data used in our study. Our work was supported in part
1070 by NSF EPSCoR Award Number 1632738. The content is solely the responsibility of the authors
1071 and does not necessarily represent the official views of our supporting organizations.

1072 **Author contributions**

1073 Conceptualization: A.C.H. and J.R.M.; Methodology: A.C.H., P.C.F. and J.R.M.; Software: A.C.H.,
1074 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.,
1075 P.C.F. and J.R.M.; Supervision: J.R.M.

1076 **Author information**

1077 The authors declare no competing financial interests. Correspondence and requests for materials
1078 should be addressed to J.R.M. (jeremy.r.manning@dartmouth.edu).