

<sup>1</sup> Memory for television episodes preserves event content  
<sup>2</sup> while introducing new across-event similarities

<sup>3</sup> Andrew C. Heusser<sup>1, 2, †</sup>, Paxton C. Fitzpatrick<sup>1, †</sup>, and Jeremy R. Manning<sup>1, \*</sup>

<sup>1</sup>Department of Psychological and Brain Sciences

Dartmouth College, Hanover, NH 03755, USA

<sup>2</sup>Akili Interactive

Boston, MA 02110

<sup>†</sup>Denotes equal contribution

<sup>\*</sup>Corresponding author: jeremy.r.manning@dartmouth.edu

<sup>4</sup> March 13, 2020

<sup>5</sup> **Abstract**

<sup>6</sup> Our experiences unfold over time defining unique *trajectories* through the relevant represen-  
<sup>7</sup> tational spaces. By casting our life events as temporally evolving trajectories, one can compare  
<sup>8</sup> the shape of the trajectory formed by an experience to that defined by our later remembering  
<sup>9</sup> of that experience. We propose a framework for mapping naturalistic experiences onto geomet-  
<sup>10</sup> ric spaces that characterize how experiences are segmented into discrete events, and how the  
<sup>11</sup> contents of event sequences evolve over time. We apply this approach to a naturalistic memory  
<sup>12</sup> experiment which had participants view and recount a television episode. The content of par-  
<sup>13</sup> ticipants' recounts of events from the original episode closely matched the original episode's  
<sup>14</sup> content. Further, we introduce two novel metrics for assessing memory quality (precision and  
<sup>15</sup> distinctiveness), both of which relate to participants' ability to recapitulate the experience. Lastly,  
<sup>16</sup> we identified a network of brain structures that are sensitive to the "shapes" of ongoing expe-

17       riences, and an overlapping network that is sensitive (at the time of encoding) to how people  
18       later remembered those experiences in relation to other experiences. In this way, modeling the  
19       content of richly structured experiences can reveal how (geometrically and conceptually) those  
20       experiences are segmented into events and integrated into our memories of other experiences.

21 **Introduction**

22       What does it mean to *remember* something? In traditional episodic memory experiments (e.g.,  
23       list-learning or trial-based experiments; Murdock, 1962; Kahana, 1996), remembering is often cast  
24       as a discrete and binary operation: each studied item may be separated from all others, and la-  
25       beled as having been recalled or forgotten. More nuanced studies might incorporate self-reported  
26       confidence measures as a proxy for memory strength, or ask participants to discriminate between  
27       “recollecting” the (contextual) details of an experience or having a general feeling of “familiarity”  
28       (Yonelinas, 2002). Using well-controlled, trial-based experimental designs, the field has amassed  
29       a wealth of valuable information regarding human episodic memory. However, there are funda-  
30       mental properties of the external world and our memories that trial-based experiments are not well  
31       suited to capture (for review also see Koriat and Goldsmith, 1994; Huk et al., 2018). First, our expe-  
32       riences and memories are continuous, rather than discrete—removing a (naturalistic) event from  
33       the context in which it occurs can substantially change its meaning. Second, the specific language  
34       used to describe an experience has little bearing on whether the experience should be considered to  
35       have been “remembered.” Asking whether the rememberer has precisely reproduced a specific set  
36       of words to describe a given experience is nearly orthogonal to whether they were actually able to  
37       remember it. In classic (e.g., list-learning) memory studies, by contrast, the number or proportion  
38       of precise recalls is often a primary metric for assessing the quality of participants’ memories.  
39       Third, one might remember the *essence* (or a general summary) of an experience but forget (or  
40       neglect to recount) particular details. Capturing the essence of what happened is typically the  
41       main “point” of recounting a memory to a listener, while the addition of highly specific details  
42       may add comparatively little to successful conveyance of an experience.

43 How might one go about formally characterizing the “essence” of an experience, or whether  
44 it has been recovered by the rememberer? Any given moment of an experience derives meaning  
45 from surrounding moments, as well as from longer-range temporal associations (Lerner et al.,  
46 2011; Manning, 2019). Therefore, the timecourse describing how an event unfolds is fundamental  
47 to its overall meaning. Further, this hierarchy formed by our subjective experiences at different  
48 timescales defines a *context* for each new moment (e.g., Howard and Kahana, 2002; Howard et al.,  
49 2014), and plays an important role in how we interpret that moment and remember it later (for  
50 review see Manning et al., 2015). Our memory systems can leverage these associations to form  
51 predictions that help guide our behaviors (Ranganath and Ritchey, 2012). For example, as we  
52 navigate the world, the features of our subjective experiences tend to change gradually (e.g., the  
53 room or situation we are in at any given moment is strongly temporally autocorrelated), allowing  
54 us to form stable estimates of our current situation and behave accordingly (Zacks et al., 2007;  
55 Zwaan and Radvansky, 1998).

56 Occasionally, this gradual “drift” of our ongoing experience is punctuated by sudden changes,  
57 or “shifts” (e.g., when we walk through a doorway; Radvansky and Zacks, 2017). Prior research  
58 suggests that these sharp transitions (termed *event boundaries*) help to discretize our experiences  
59 (and their mental representations) into *events* (Radvansky and Zacks, 2017; Brunec et al., 2018;  
60 Heusser et al., 2018a; Clewett and Davachi, 2017; Ezzyat and Davachi, 2011; DuBrow and Davachi,  
61 2013). The interplay between the stable (within-event) and transient (across-event) temporal  
62 dynamics of an experience also provides a potential framework for transforming experiences  
63 into memories that distill those experiences down to their essence. For example, prior work  
64 has shown that event boundaries can influence how we learn sequences of items (Heusser et al.,  
65 2018a; DuBrow and Davachi, 2013), navigate (Brunec et al., 2018), and remember and understand  
66 narratives (Zwaan and Radvansky, 1998; Ezzyat and Davachi, 2011). Prior research has implicated  
67 a network of brain regions (including the hippocampus and the medial prefrontal cortex) as playing  
68 a critical role in transforming experiences into structured and consolidated memories (Tomparry  
69 and Davachi, 2017).

70 Here we sought to examine how the temporal dynamics of a “naturalistic” experience were

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

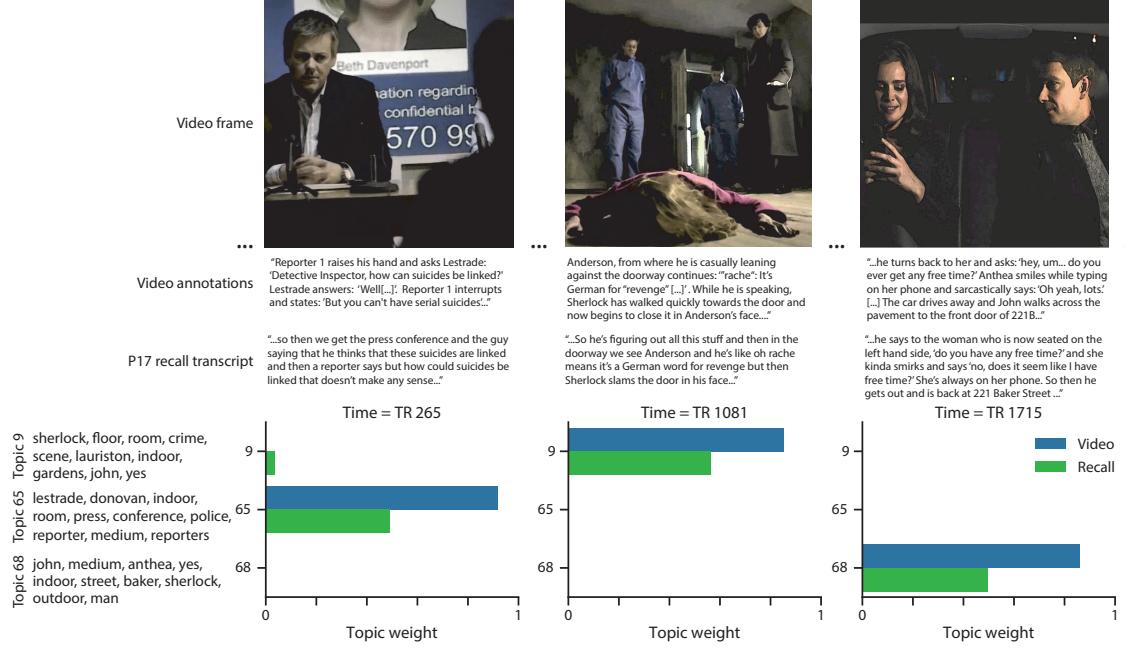
## 91 **Results**

92 To characterize the temporally dynamic contents of the *Sherlock* episode and participants' subse-  
93 quent recounts, we used a topic model (Blei et al., 2003) to discover the latent themes. Topic  
94 models take as inputs a vocabulary of words to consider and a collection of text documents, and  
95 return two output matrices. The first of these is a *topics matrix* whose rows are topics (latent  
96 themes) and whose columns correspond to words in the vocabulary. The entries of the topics ma-

97 trix define how each word in the vocabulary is weighted by each discovered topic. For example,  
98 a detective-themed topic might weight heavily on words like “crime,” and “search.” The second  
99 output is a *topic proportions matrix*, with one row per document and one column per topic. The topic  
100 proportions matrix describes what mixture of discovered topics is reflected in each document.

101 Chen et al. (2017) collected hand-annotated information about each of 1000 (manually identified)  
102 time segments spanning the roughly 50 minute video used in their experiment. This information  
103 included: a brief narrative description of what was happening, the location where the scene  
104 took place, the names of any characters on the screen, and other similar details (for a full list of  
105 annotated features, see *Methods*). We took from these annotations the union of all unique words  
106 (excluding stop words, such as “and,” “or,” “but,” etc.) across all features and scenes as the  
107 “vocabulary” for the topic model. We then concatenated the sets of words across all features  
108 contained in overlapping, sliding windows of (up to) 50 scenes, and treated each window as a  
109 single “document” for the purpose of fitting the topic model. Next, we fit a topic model with (up  
110 to)  $K = 100$  topics to this collection of documents. We found that 32 unique topics (with non-zero  
111 weights) were sufficient to describe the time-varying content of the video (see *Methods*; Figs. 1, S2).  
112 Note that our approach is similar in some respects to Dynamic Topic Models (Blei and Lafferty,  
113 2006) in that we sought to characterize how the thematic content of the episode evolved over  
114 time. However, whereas Dynamic Topic Models are designed to characterize how the properties  
115 of *collections* of documents change over time, our sliding window approach allows us to examine  
116 the topic dynamics within a single document (or video). Specifically, our approach yielded (via the  
117 topic proportions matrix) a single *topic vector* for each sliding window of annotations transformed  
118 by the topic model. We then stretched (interpolated) the resulting windows-by-topics matrix to  
119 match the time series of the 1976 fMRI volumes collected as participants viewed the episode.

120 The 32 topics we found were heavily character-focused (i.e., the top-weighted word in each  
121 topic was nearly always a character) and could be roughly divided into themes centered around  
122 Sherlock Holmes (the titular character), John Watson (Sherlock’s close confidant and assistant),  
123 supporting characters (e.g., Inspector Lestrade, Sergeant Donovan, or Sherlock’s brother Mycroft),  
124 or the interactions between various pairs of these characters (see Fig. S2). Several of the identified



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

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

143 Given that the time-varying content of the video could be segmented cleanly into discrete  
144 events, we wondered whether participants’ recalls of the video also displayed a similar structure.  
145 We applied the same topic model (already trained on the video annotations) to each participant’s  
146 recalls. Analogous to how we parsed the time-varying content of the video, to obtain similar esti-  
147 mates for each participant’s recall, we treated each overlapping “window” of (up to 10) sentences  
148 from their transcript as a “document,” and computed the most probable mix of topics reflected in  
149 each timepoint’s sentences. This yielded, for each participant, a number-of-windows by number-  
150 of-topics topic proportions matrix that characterized how the topics identified in the original video  
151 were reflected in the participant’s recalls. Note that an important feature of our approach is that it  
152 allows us to compare participants’ recalls to events from the original video, despite different par-



**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).

<sup>153</sup> ticipants using widely varying language to describe the same event, and that those descriptions  
<sup>154</sup> may not match the original annotations. This is a substantial benefit of projecting the video and  
<sup>155</sup> recalls into a shared “topic” space. An example topic proportions matrix from one participant’s  
<sup>156</sup> recalls is shown in Figure 2D.

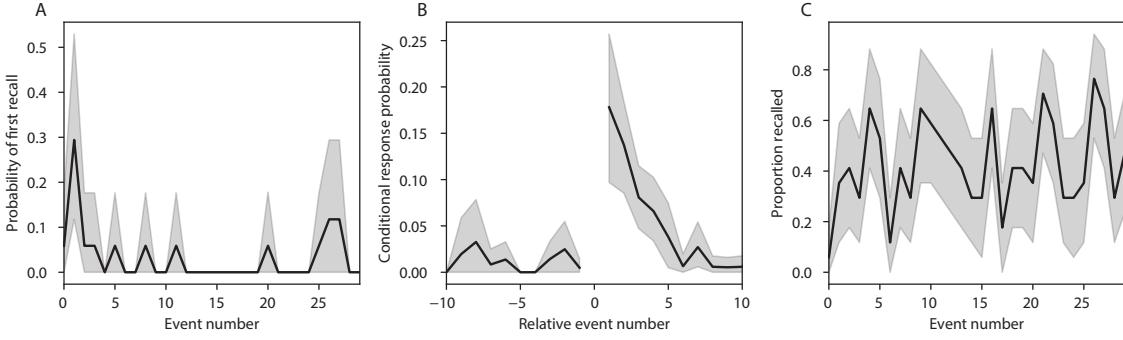
<sup>157</sup> Although the example participant’s recall topic proportions matrix has some visual similarity to  
<sup>158</sup> the video topic proportions matrix, the time-varying topic proportions for the example participant’s  
<sup>159</sup> recalls are not as sparse as those for the video (compare Figs. 2A and D). Similarly, although there do  
<sup>160</sup> appear to be periods of stability in the recall topic dynamics (i.e., most topics are active or inactive  
<sup>161</sup> over contiguous blocks of time), the individual topics’ overall timecourses are not as cleanly  
<sup>162</sup> delineated as the video topics’. To examine these patterns in detail, we computed the timepoint-  
<sup>163</sup> by-timepoint correlation matrix for the example participant’s recall topic trajectory (Fig. 2E). As in  
<sup>164</sup> the video correlation matrix (Fig. 2B), the example participant’s recall correlation matrix has a strong  
<sup>165</sup> block diagonal structure, indicating that their recalls are discretized into separated events. As for  
<sup>166</sup> the video correlation matrix, we leveraged an HMM-based optimization procedure (see *Methods*)  
<sup>167</sup> to determine how many events are reflected in the participant’s recalls and where specifically the  
<sup>168</sup> event boundaries fall (outlined in yellow). We carried out a similar analysis on all 17 participants’  
<sup>169</sup> recall topic proportions matrices (Fig. S4).

<sup>170</sup> Two clear patterns emerged from this set of analyses. First, although every individual partic-  
<sup>171</sup> ipant’s recalls could be segmented into discrete events (i.e., every individual participant’s recall  
<sup>172</sup> correlation matrix exhibited clear block diagonal structure; Fig. S4), each participant appeared to  
<sup>173</sup> have a unique *recall resolution*, reflected in the sizes of those blocks. While, some participants’ recall  
<sup>174</sup> topic proportions segmented into just a few events (e.g., Participants P4, P5, and P7), others’ seg-  
<sup>175</sup> mented into many shorter duration events (e.g., Participants P12, P13, and P17). This suggests that  
<sup>176</sup> different participants may be recalling the video with different levels of detail— e.g., some might  
<sup>177</sup> touch on just the major plot points, whereas others might attempt to recall every minor scene or ac-  
<sup>178</sup> tion. The second clear pattern present in every individual participant’s recall correlation matrix is  
<sup>179</sup> that, unlike in the video correlation matrix, there are substantial off-diagonal correlations. Whereas  
<sup>180</sup> each event in the original video was (largely) separable from the others (Fig. 2B), in transforming

<sup>181</sup> those separable events into memory, participants appear to be integrating across multiple events,  
<sup>182</sup> blending elements of previously recalled and not-yet-recalled content into each newly recalled  
<sup>183</sup> event (Figs. 2E, S4; also see Manning et al., 2011; Howard et al., 2012).

<sup>184</sup> The above results indicate that both the structure of the original video and participants' recalls  
<sup>185</sup> of the video exhibit event boundaries that can be identified automatically by characterizing the  
<sup>186</sup> dynamic content using a shared topic model and segmenting the content into events via HMMs.  
<sup>187</sup> Next, we asked whether some correspondence might be made between the specific content of the  
<sup>188</sup> events the participants experienced in the video, and the events they later recalled. One approach  
<sup>189</sup> to linking the experienced (video) and recalled events is to label each recalled event as matching  
<sup>190</sup> the video event with the most similar (i.e., most highly correlated) topic vector (Figs. 2G, S5). This  
<sup>191</sup> yields a sequence of "presented" events from the original video, and a (potentially differently  
<sup>192</sup> ordered) sequence of "recalled" events for each participant. Analogous to classic list-learning  
<sup>193</sup> studies, we can then examine participants' recall sequences by asking which events they tended  
<sup>194</sup> to recall first (probability of first recall; Fig. 3A; Atkinson and Shiffrin, 1968; Postman and Phillips,  
<sup>195</sup> 1965; Welch and Burnett, 1924); how participants most often transition between recalls of the  
<sup>196</sup> events as a function of the temporal distance between them (lag-conditional response probability;  
<sup>197</sup> Fig. 3B; Kahana, 1996); and which events they were likely to remember overall (serial position  
<sup>198</sup> recall analyses; Fig. 3C; Murdock, 1962). Interestingly, for two of these analyses (probability of first  
<sup>199</sup> recall and lag-conditional response probability curves) we observe patterns comparable to classic  
<sup>200</sup> effects from the list-learning literature: namely, a higher probability of initiating recall with the  
<sup>201</sup> first event in the sequence (Fig. 3A) and a higher probability of transitioning to neighboring events  
<sup>202</sup> with an asymmetric forward bias (Fig. 3B). In contrast, we do not observe a pattern comparable to  
<sup>203</sup> the serial position effect (Fig. 3C), but rather we see higher memory for specific events distributed  
<sup>204</sup> somewhat evenly throughout the video.

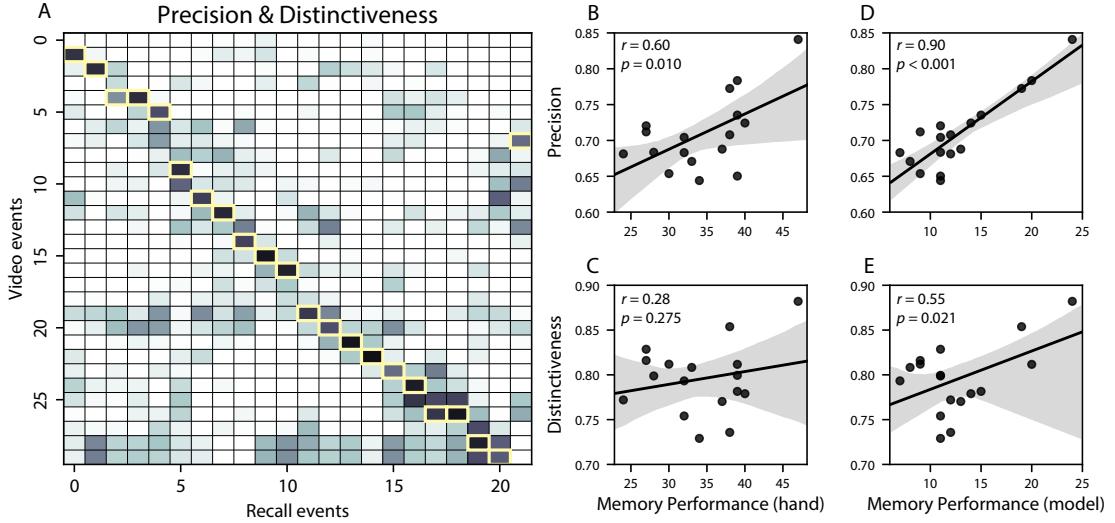
<sup>205</sup> We can also apply two list-learning-native analyses that describe how participants group items  
<sup>206</sup> in their recall sequences: temporal clustering and semantic clustering (Polyn et al., 2009, see  
<sup>207</sup> *Methods* for details). Temporal clustering refers to the extent to which participants group their  
<sup>208</sup> recall responses according to encoding position. Overall, we found that sequentially viewed video



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

events were clustered heavily in participants' recall event sequences (mean clustering score: 0.767, SEM: 0.029), and that participants with higher temporal clustering scores tended to perform better according to both Chen et al. (2017)'s hand-annotated memory scores (Pearson's  $r(15) = 0.62$ ,  $p = 0.008$ ) and our model's estimate (Pearson's  $r(15) = 0.54$ ,  $p = 0.024$ ). Semantic clustering measures the extent to which participants cluster their recall responses according to semantic similarity. We found that participants tended to recall semantically similar video events together (mean clustering score: 0.787, SEM: 0.018), and that semantic clustering score was also related to both hand-annotated (Pearson's  $r(15) = 0.65$ ,  $p = 0.004$ ) and model-derived (Pearson's  $r(15) = 0.63$ ,  $p = 0.007$ ) memory performance.

Statistical models of memory studies often treat recall success as binary (i.e., an 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, the feature-rich content of a naturalistic experiences may later be described with many, highly variable levels of success. Our framework produces a content-based model of individual stimulus and recall events by projecting the dynamic content of the video and participants' recalls into a shared topic space. This allows for direct, quantitative comparison



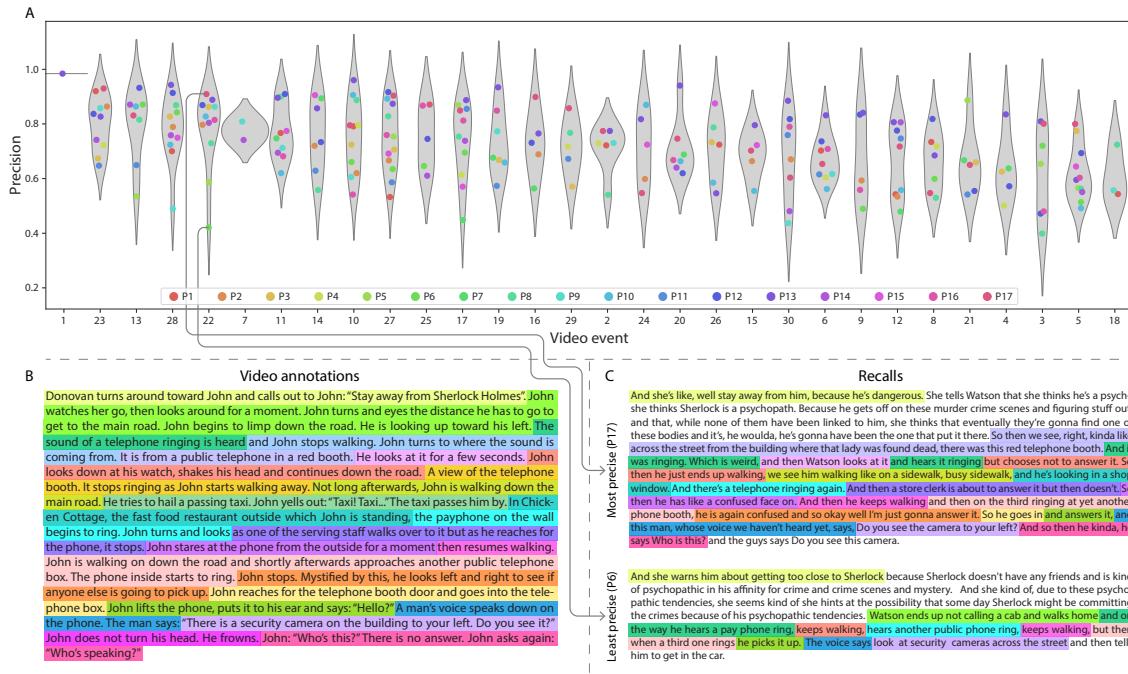
**Figure 4: Novel content-based metrics of naturalistic memory: precision and distinctiveness.** A. The video-recall correlation matrix for a representative participant (17). The yellow boxes highlight the maximum correlation in each column. The example participant's overall precision score was computed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

226 between all stimulus and recall events, as well as between the recall events themselves. Leveraging  
 227 these content-based models of the stimulus/recall events, we developed two novel, *continuous*  
 228 metrics for analyzing naturalistic memory: *precision* and *distinctiveness*. We define precision as  
 229 the “completeness” of recall, or how fully the presented content was recapitulated in memory.  
 230 Under our framework, we quantify this for a given recall event as the correlation between the  
 231 topic proportions of the recall event and the maximally correlated video event (Fig. 4). A second  
 232 novel metric we introduce here is *distinctiveness*, which we define as the “specificity” of recall,  
 233 or how unique the description of a given section of content was, compared to descriptions for  
 234 other sections of content. We quantify this for each recall event as 1 minus the average correlation  
 235 between the given recall event and all other recall events not matched to the same video event.

236 In addition to individual events, one may also use these metrics to describe each participant's  
237 overall performance (i.e., by averaging across a participant's event-wise precision or distinctiveness  
238 scores). Participants whose recall events are more veridical descriptions of what happened in the  
239 video event will presumably have higher precision scores. We find that, across participants,  
240 a higher precision score is correlated to both hand-annotated memory performance (Pearson's  
241  $r(15) = 0.60, p = 0.010$ ) and the number of video events successfully remembered, as determined  
242 by our model (Pearson's  $r(15) = 0.90, p < 0.001$ ). We also hypothesized that participants who  
243 recounted events in a more distinctive way would display better overall memory. We find that  
244 this distinctiveness score is related to our model's estimated number of recalled events (Pearson's  
245  $r(15) = 0.55, p = 0.021$ ), and while we do not find distinctiveness to be related to hand-annotated  
246 memory performance (Pearson's  $r(15) = 0.28, p = 0.275$ ), this is not entirely surprising given how  
247 the hand-annotated memory scores were computed (see *Methods* and *Discussion* for details).

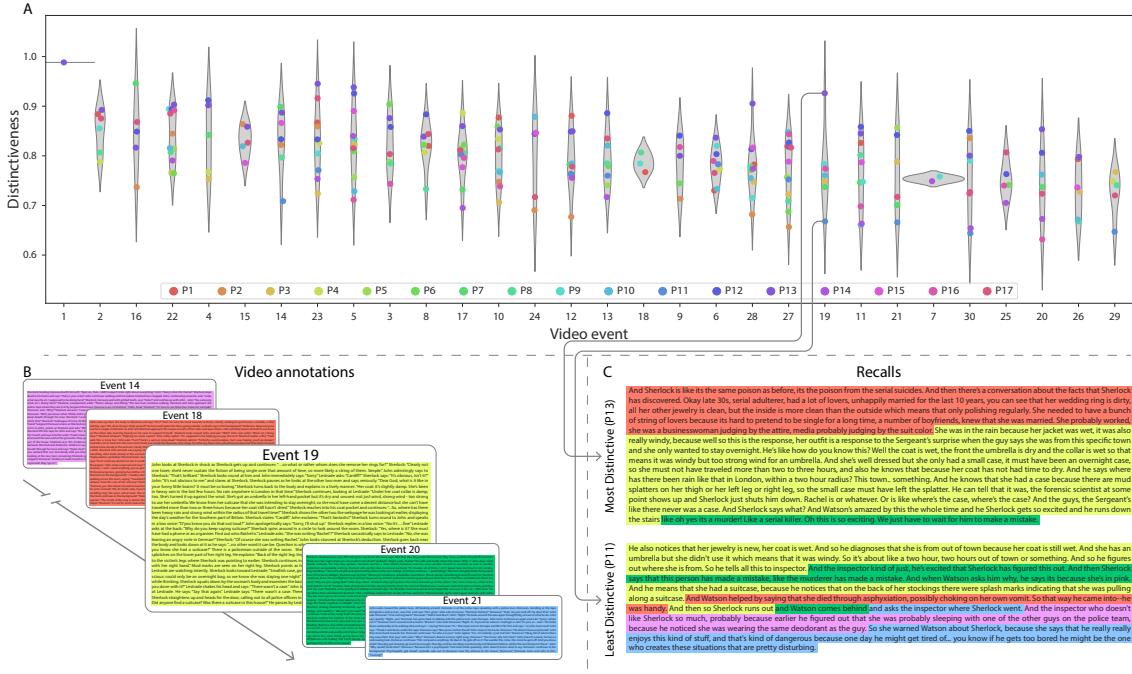
248 Further intuition for the behaviors captured by these two metrics may be gained by directly  
249 examining the content of the video and recalls our framework models. In Figure 5, we contrast  
250 recalls for the same video event (event 22) from two participants: one with a high precision score  
251 (P17), the other with a low precision score (P6). From the HMM-identified event boundaries,  
252 we recovered the set of annotations describing the content of an example video event (Fig. 5B),  
253 and divided them into different color-coded sections for each action or feature described. We  
254 then similarly recovered the set of sentences comprising the corresponding recall event for each  
255 of the two example participants. Because the recall sliding windows overlap heavily, and each  
256 recall event spans multiple recall timepoints (i.e., windows), we have stripped any sentences from  
257 the beginning and end that describe earlier or later video events for the sake of readability. In  
258 other words, Fig. 5C shows a subset of the full recall event text, comprising sentences between  
259 the first and last descriptions of content from the example video event. We then colored all words  
260 describing actions and features coded in panel B by their corresponding color. Visual comparison  
261 of the transcripts reveals that the most precise participant's recall both captures more of the video  
262 event's content, and does so with far more detail.

263 Figure 6 similarly contrasts two example participants' recalls for a common video event (event



**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.

19) to illustrate the tangible differences between high and low distinctiveness scores. Here, we  
 20) have extracted the full set of sentences comprising the most distinctive recall event (P13) and least  
 21) distinctive recall event (P11) matched to the example video event (Fig. 6C). We also extracted the  
 22) annotations for the example video event, as well as those from each other video event whose content  
 23) the example participants' single recall events described (Fig. 6B). We then shaded the annotation  
 24) text for each video event with a different color, and shaded each word of the example participants'  
 25) recall text by the color of the video event it describes. The majority of the most distinctive recall  
 26) event text describes video event 19's content, with the first five and last one sentence describing  
 27) the video events immediately preceding and succeeding the current one, respectively. In contrast,



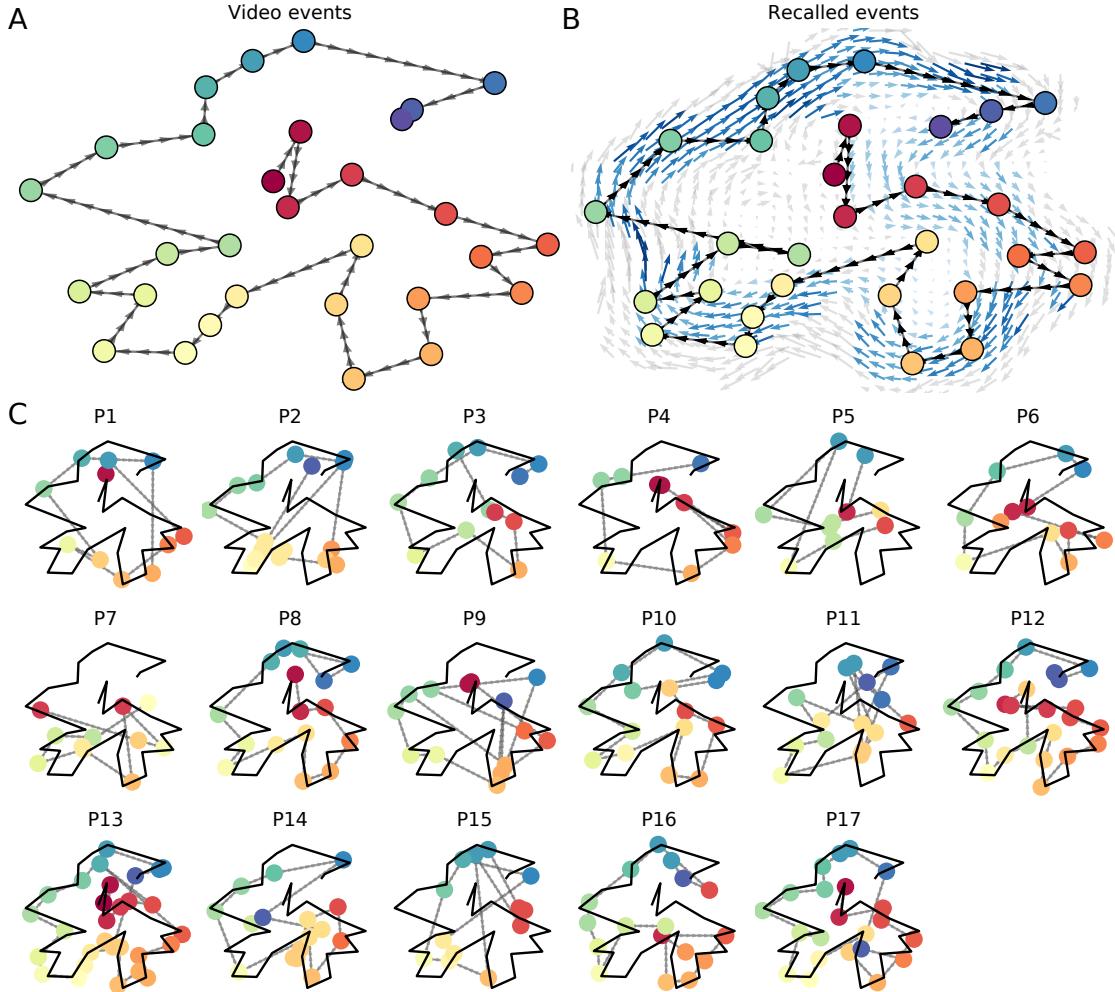
**Figure 6: Distinctiveness metric reflects specificity of recall.** A. Recall distinctiveness by video event. Kernel density estimates for each video event’s distribution of recall distinctiveness scores, analogous to Fig. 5A. B. The sets of “Narrative Details” video annotations (generated by Chen et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

273 the least precise participant's recall for video event 19 blends the content from five separate video  
274 events, does not transition between them in order, and often combines descriptions of two video  
275 events' content in the same sentence.

The prior analyses leverage the correspondence between the 100-dimensional topic proportion matrices for the video and participants' recalls to characterize recall. However, it is difficult to gain deep insights into the content of (or relationships between) experiences and memories solely by examining these topic proportions (e.g., Figs. 2A, D) or the corresponding correlation matrices (Figs. 2B, E, S4). And while we can directly examine the original text underlying these topic vectors (e.g., Figs. 5, 6) to show how relationships between them reflect real-world behavior, this comparison becomes prohibitively cumbersome at larger timescales. To visualize the time-varying

283 high-dimensional content in a more intuitive way (Heusser et al., 2018b), we projected the topic  
284 proportions matrices onto a two-dimensional space using Uniform Manifold Approximation and  
285 Projection (UMAP; McInnes et al., 2018). In this lower-dimensional space, each point represents a  
286 single video or recall event, and the distances between the points reflect the distances between the  
287 events' associated topic vectors (Fig. 7). In other words, events that are nearer to each other in this  
288 space are more semantically similar, and those that are farther apart are less so.

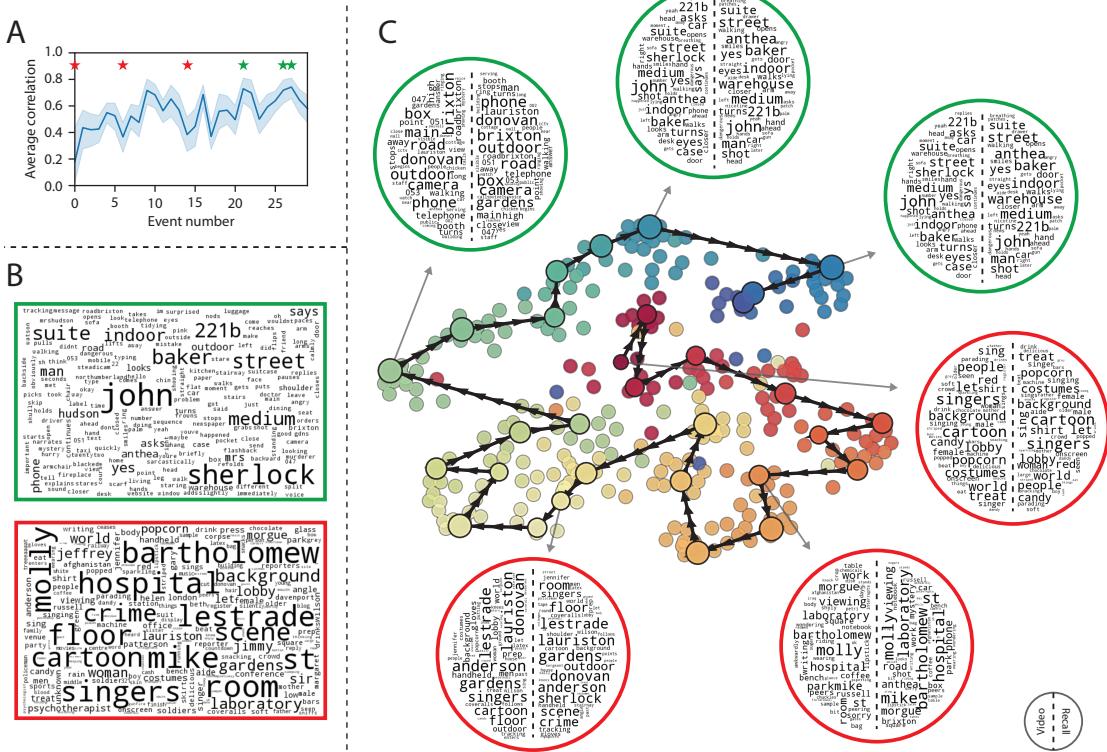
289 Visual inspection of the video and recall topic trajectories reveals a striking pattern. First, the  
290 topic trajectory of the video (which reflects its dynamic content; Fig. 7A) is captured nearly perfectly  
291 by the averaged topic trajectories of participants' recalls (Fig. 7B). To assess the consistency of these  
292 recall trajectories across participants, we asked: given that a participant's recall trajectory had  
293 entered a particular location in the reduced topic space, could the position of their *next* recalled  
294 event be predicted reliably? For each location in the the reduced topic space, we computed the set of  
295 line segments connecting successively recalled events (across all participants) that intersected that  
296 location (see *Methods* for additional details). We then computed (for each location) the distribution  
297 of angles formed by the lines defined by those line segments and a fixed reference line (the *x*-  
298 axis). Rayleigh tests revealed the set of locations in topic space at which these across-participant  
299 distributions exhibited reliable peaks (blue arrows in Fig. 7B reflect significant peaks at  $p < 0.05$ ,  
300 corrected). We observed that the locations traversed by nearly the entire video trajectory exhibited  
301 such peaks. In other words, participants exhibited similar trajectories that also matched the  
302 trajectory of the original video (Fig. 7C). This is especially notable when considering the fact that  
303 the number of events participants recalled (dots in Fig. 7C) varied considerably across people, and  
304 that every participant used different words to describe what they had remembered happening in  
305 the video. Differences in the numbers of remembered events appear in participants' trajectories  
306 as differences in the sampling resolution along the trajectory. We note that this framework also  
307 provides a means of disentangling classic "proportion recalled" measures (i.e., the proportion  
308 of video events described in participants' recalls) from participants' abilities to recapitulate the  
309 overall unfolding of the original video's content (i.e., the similarity between the shapes of the  
310 original video trajectory and that defined by each participant's recounting of the video).



**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).

311 In addition to the more “holistic” measure of memory described in the previous section, our  
312 framework also affords the ability to drill down to individual words and quantify how each word  
313 relates to the memorability of each event. The results displayed in Figures 3C and 5A suggest that  
314 certain events were remembered better than others. Given this, we next asked whether the  
315 events were generally remembered well or poorly tended to reflect particular content. Because  
316 our analysis framework projects the dynamic video content and participants’ recalls into a shared  
317 space, and because the dimensions of that space represent topics (which are, in turn, sets of weights  
318 over words in the vocabulary), we are able to recover the weighted combination of words that make  
319 up any point (i.e., topic vector) in this space. We first computed the average precision with which  
320 participants recalled each of the 30 video events (Fig. 8A; note that this result is analogous to a serial  
321 position curve created from our continuous recall quality metric). We then computed a weighted  
322 average of the topic vectors for each video event, where the weights reflected how reliably each  
323 event was recalled. To visualize the result, we created a “wordle” image (Mueller et al., 2018) where  
324 words weighted more heavily by better-remembered topics appear in a larger font (Fig. 8B, green  
325 box). Across the full video, content that reflected topics necessary to convey the central focus of the  
326 video (e.g., the names of the two main characters, “Sherlock” and “John”, and the address of a major  
327 recurring location, “221B Baker Street”) were best remembered. An analogous analysis revealed  
328 which themes were poorly remembered. Here in computing the weighted average over events’  
329 topic vectors, we weighted each event in *inverse* proportion to how well it was remembered (Fig. 8B,  
330 red box). The least well-remembered video content reflected information not necessary to later  
331 convey a general summary of the video, such as the proper names of relatively minor characters  
332 (e.g., “Mike,” “Molly,” and “Lestrade”) and locations (e.g., “St. Bartholomew’s Hospital”).

333 A similar result emerged from assessing the topic vectors for individual video and recall events  
334 (Fig. 8C). Here, for each of the three best- and worst-remembered video events, we have constructed  
335 two wordles: one from the original video event’s topic vector (left) and a second from the average  
336 recall topic vector for that event (right). The three best-remembered events (circled in green)  
337 correspond to scenes important to the central plot-line: a mysterious figure spying on John in a  
338 phone booth; John meeting Sherlock at Baker St. to discuss the murders; and Sherlock laying

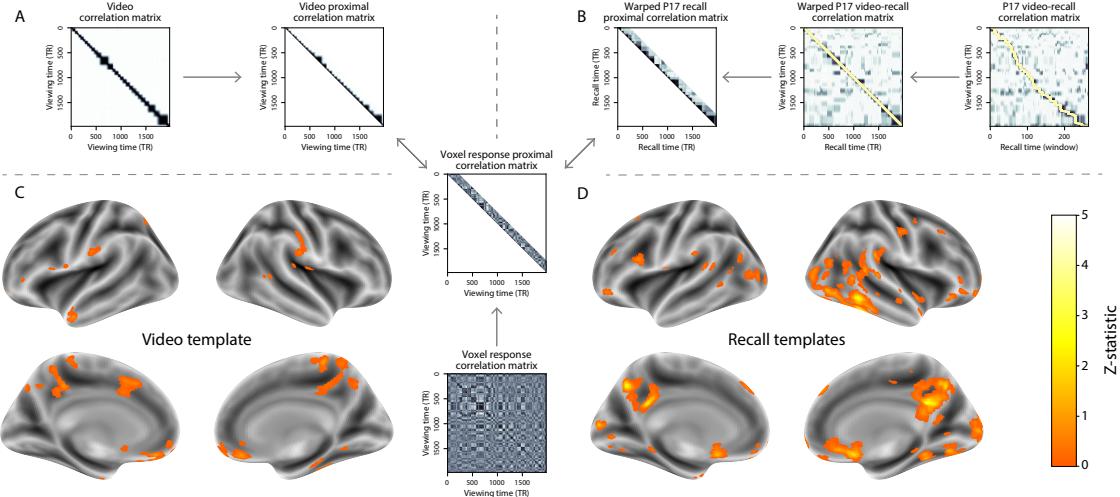


**Figure 8: Language used in the most and least memorable events.** **A.** Average precision (video event-recall event topic vector correlation) across participants for each video event. Error bars denote bootstrap-derived across-participant 95% confidence intervals. The stars denote the three best-remembered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

339 a trap to catch the killer. Meanwhile, the three worst-remembered events (circled in red) reflect  
340 scenes that are non-essential to summarizing the narrative's structure: the video of singing cartoon  
341 characters participants viewed prior to the main episode; John asking Molly about Sherlock's habit  
342 of over-analyzing people; and Sherlock noticing evidence of Anderson's and Donovan's affair.

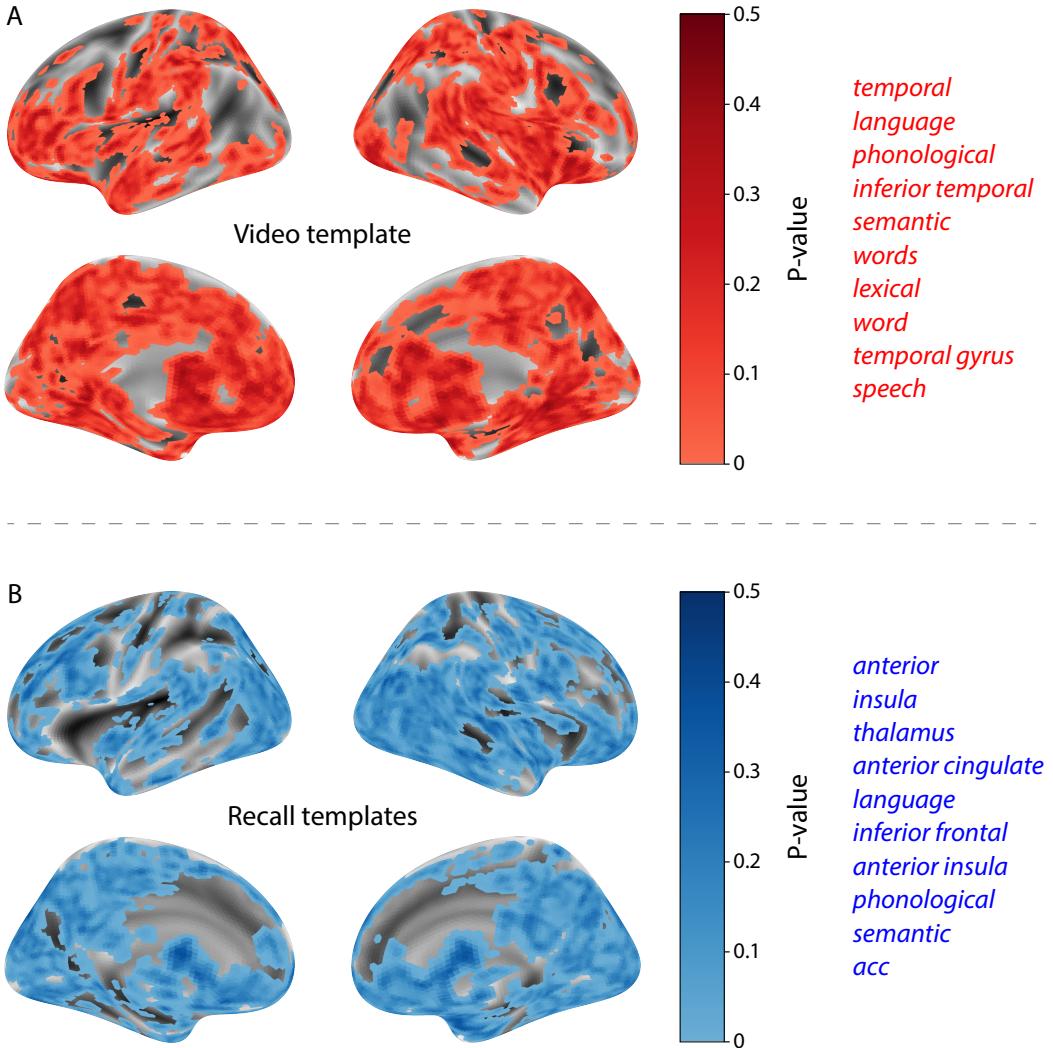
343 The results thus far inform us about which aspects of the dynamic content in the episode partic-  
344 ipants watched were preserved or altered in participants' memories. We next carried out a series  
345 of analyses aimed at understanding which brain structures might facilitate these preservations  
346 and transformations between the external world and memory. In the first analysis, we sought  
347 to identify brain structures that were sensitive to the dynamic unfolding of the video's content,  
348 as characterized by its topic trajectory. We used a searchlight procedure to identify clusters of  
349 voxels whose activity patterns displayed a proximal temporal correlation structure (as participants  
350 watched the video) matching that of the original video's topic proportions (Fig. 9A; see *Methods* for  
351 additional details). In a second analysis, we sought to identify brain structures whose responses  
352 (during video viewing) reflected how each participant would later structure their recounting of the  
353 video. We used an analogous searchlight procedure to identify clusters of voxels whose proximal  
354 temporal correlation matrices matched that of the topic proportions for each individual's recall  
355 (Figs. 9B; see *Methods* for additional details). To ensure our searchlight procedure identified re-  
356 gions *specifically* sensitive to the temporal structure of the video or recalls (i.e., rather than those  
357 with a temporal autocorrelation length similar to that of the video/recalls), we performed a phase  
358 shift-based permutation correction (see *Methods* for additional details). As shown in Figure 9C, the  
359 video-driven searchlight analysis revealed a distributed network of regions that may play a role in  
360 processing information relevant to the narrative structure of the video. Similarly, the recall-driven  
361 searchlight analysis revealed a second network of regions (Fig. 9D) that may facilitate a person-  
362 specific transformation of one's experience into memory. In identifying regions whose responses  
363 to ongoing experiences reflect how those experiences will be remembered later, this latter analysis  
364 extends classic *subsequent memory analyses* (e.g., Paller and Wagner, 2002) to domain of naturalistic  
365 stimuli.

366 The searchlight analyses described above yielded two distributed networks of brain regions,



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

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



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

375 **Discussion**

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

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

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

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

430 is essentially equal, and does not possess attributes that would render it, on average, more or less  
431 memorable than others. Such is rarely the case with real-world experiences or experiments meant  
432 to approximate them, and the effects of both intrinsic and observer-dependent factors on stimulus  
433 memorability are well established (for review see Chun and Turk-Browne, 2007; Bylinskii et al.,  
434 2015; Tyng et al., 2017). Second, the random ordering of list items ensures that (across participants,  
435 on average) there is no relationship between the thematic similarity of individual stimuli and their  
436 presentation positions—in other words, two successively presented items are no more likely to be  
437 highly semantically similar than they are to be highly dissimilar. In most cases, the exact opposite  
438 is true of real-world episodes. Our internal thoughts, our actions, and the physical state of the  
439 world around us all tend to follow a direct, causal progression. As a result, each moment of our  
440 experience tends to be inherently more similar to surrounding moments than to those in the distant  
441 past or future. Memory literature has termed this strong temporal autocorrelation “context,” and  
442 in various media that depict real-world events (e.g., movies or written stories), we recognize  
443 it as a *narrative structure*. While a random word list (by definition) has no such structure, the  
444 logical progression between ideas and actions in a naturalistic stimulus prompts the rememberer  
445 to recount presented events in order, starting with the beginning. This tendency is reflected in our  
446 findings’ second departure from typical free recall dynamics: a lack of increased probability of first  
447 recall for end-of-sequence events (Fig. 3A).

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

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

486 metrics compare between cognitively impoverished groups and healthy controls.

487 In the analyses outlined in Figure 9, we identified two networks of brain regions whose re-  
488 sponses during video viewing were consistent with the temporal structure of the video and recall  
489 topic trajectories, respectively. The network identified by the video trajectory analysis included the  
490 ventromedial prefrontal cortex, left anterior temporal lobe, superior parietal and dorsal anterior  
491 cingulate cortex. The network from the video-recall trajectory analysis also included the ventro-  
492 medial prefrontal and superior parietal cortices, in addition to the posterior medial cortex (PMC)  
493 and the inferior temporal regions. Notably, Chen et al. (2017) also observed the PMC in a number  
494 of analyses including one that searched for regions whose activity patterns during encoding were  
495 reinstated during free recall. The PMC has been consistently identified in studies involving stimuli  
496 with meaningfully structured events ?. Further, the PMC is part of the "posterior medial" system,  
497 a network of brain regions thought to represent situation models Zacks et al. (2007) in support  
498 of memory, spatial navigation and social cognition (Ranganath and Ritchey, 2012). Given that  
499 we constructed our video-recall searchlight model to capture temporal structure in the episode's  
500 semantic content (and how one's later recall aligns with that structure), we speculate that the PMC  
501 may play a role in constructing mnemonic events from meaningfully structured experiences.

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

512 Our second searchlight analysis identified a largely separate network of regions (Fig. 9D)  
513 whose patterns of activity as participants viewed the video reflected the idiosyncratic structure

514 of each individual’s later recall. Decoding the associated significance map yielded a set of terms  
515 that primarily reflected names of specific structural regions (such as “thalamus,” “anterior insula,”  
516 “anterior cingulate” and “inferior frontal”; Fig. 10B). Interestingly, these regions share membership  
517 in a common, large-scale functional network (termed the “salience network”) involved in detecting  
518 and processing affective cues. In particular, the latter three regions have been implicated in  
519 functions relevant to assigning personal meaning to an experience, including: ascribing subjective  
520 value to raw, sensory input (?); modulating semantic and phonological processing in response  
521 to personally salient stimuli (?); and directing and reallocating attention and working memory  
522 resources towards the most relevant stimuli (Menon and Uddin, 2010). This suggests that the  
523 network of structures displayed in Figure 9D may play a role in transforming and restructuring  
524 ongoing experiences through the lens of one’s own personal values as they are encoded in memory.

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

538 **Methods**

539 **Experimental design and data collection**

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

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

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

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

## 570 **Data and code availability**

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

## 573 **Statistics**

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

## 579 **Modeling the dynamic content of the video and recall transcripts**

### 580 **Topic modeling**

581 The input to the topic model we trained to characterize the dynamic content of the video comprised  
582 998 hand-generated annotations of short (mean: 2.96s) scenes spanning the video clip (Chen et al.,  
583 2017 generated 1000 annotations total; we removed two referring to the break between the first  
584 and second scan sessions, during which no fMRI data was collected). We concatenated the text  
585 for all of the annotated features within each segment, creating a “bag of words” describing each  
586 scene and performed some minor preprocessing (e.g., stemming possessive nouns and removing  
587 punctuation). We then re-organized the text descriptions into overlapping sliding windows span-  
588 ning (up to) 50 scenes each. In other words, we created a “context” for each scene comprising the

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

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

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

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

## 621 Parsing topic trajectories into events using Hidden Markov Models

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

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

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

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

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

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

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

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

664 **Probability of first recall (PFR).** PFR curves (Welch and Burnett, 1924; Postman and Phillips,  
665 1965; Atkinson and Shiffrin, 1968) reflect the probability that an item will be recalled first as a  
666 function of its serial position during encoding. To carry out this analysis, we initialized a number-

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

673 **Lag conditional probability curve (lag-CRP).** The lag-CRP curve (Kahana, 1996) reflects the  
674 probability of recalling a given item after the just-recalled item, as a function of their relative  
675 encoding positions (or *lag*). In other words, a lag of 1 indicates that a recalled item was presented  
676 immediately after the previously recalled item, and a lag of -3 indicates that a recalled item came  
677 3 items before the previously recalled item. For each recall transition (following the first recall),  
678 we computed the lag between the current recall event and the next recall event, normalizing by  
679 the total number of possible transitions. This yielded a number-of-participants (17) by number-  
680 of-lags (-29 to +29; 61 lags total) matrix. We averaged over the rows of this matrix to obtain a  
681 group-averaged lag-CRP curve (Fig. 3B).

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

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

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

710 **Novel naturalistic memory metrics**

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

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

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

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

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

### 733 Visualizing the video and recall topic trajectories

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

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

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

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

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

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

772 **Searchlight fMRI analyses**

773 In Figure 9, we present two analyses aimed at identifying brain regions whose responses (as par-  
774 ticipants viewed the video) exhibited a particular temporal structure. We developed a searchlight  
775 analysis wherein we constructed a  $5 \times 5 \times 5$  cube of voxels (following Chen et al., 2017) centered on  
776 each voxel in the brain, and for each of these cubes, computed the temporal correlation matrix of  
777 the voxel responses during video viewing. Specifically, for each of the 1976 volumes collected dur-  
778 ing video viewing, we correlated the activity patterns in the given cube with the activity patterns  
779 (in the same cube) collected during every other timepoint. This yielded a 1976 by 1976 correlation  
780 matrix for each cube. Note: participant 5's scan ended 75s early, and in Chen et al., 2017's publicly  
781 released dataset, their scan data was padded to match the length of the other participants'. For  
782 our searchlight analyses, we removed this padded data (i.e., the last 50 TRs), resulting in a 1925 by  
783 1925 correlation matrix for each cube in participant 5's brain.

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

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

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

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

834 We used an analogous procedure to identify which voxels' responses reflected the recall tem-  
835 plates. For each participant, we correlated the proximal diagonals from the upper triangle of the  
836 correlation matrix for each cube of voxels with the proximal diagonals from the upper triangle  
837 of their (time-warped) recall correlation matrix. As in the video template analysis, this yielded a  
838 voxelwise map of correlation coefficients per participant. However, whereas the video analysis  
839 compared every participant's responses to the same template, here the recall templates were unique  
840 for each participant. As in the analysis described above, we  $t$ -scored the (Fisher  $z$ -transformed)  
841 voxelwise correlations, and used the same permutation procedure we developed for the video  
842 responses to ensure specificity to the recall timeseries and assign significance values. To create the  
843 map in Figure 9D we again thresholded out any voxels whose correspondence values fell below  
844 the 95<sup>th</sup> percentile of the permutation-derived null distribution.

## 845 **Neurosynth decoding analyses**

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

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

## References

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

- 874 Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., and Oliva, A. (2015). Intrinsic and extrinsic  
875 effects on image memorability. *Vision Research*, 116:165–178.
- 876 Capota, M., Turek, J., Chen, P.-H., Zhu, X., Manning, J. R., Sundaram, N., Keller, B., Wang, Y., and  
877 Shin, Y. S. (2017). Brain imaging analysis kit.
- 878 Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes,  
879 M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal sentence encoder.  
880 *arXiv*, 1803.11175.
- 881 Charron, S. and Koechlin, E. (2010). Divided representations of current goals in the human frontal  
882 lobes. *Science*, 328(5976):360–363.
- 883 Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared  
884 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,  
885 20(1):115.
- 886 Chun, M. and Turk-Browne, N. (2007). Interactions between attention and memory. *Current opinion*  
887 in neurobiology, 17(2):177–184.
- 888 Clewett, D. and Davachi, L. (2017). The ebb and flow of experience determines the temporal  
889 structure of memory. *Curr Opin Behav Sci*, 17:186–193.
- 890 Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions.  
891 *Theory of Probability & Its Applications*, 15(3):458–486.
- 892 DuBrow, S. and Davachi, L. (2013). The influence of contextual boundaries on memory for the  
893 sequential order of events. *Journal of Experimental Psychology: General*, 142(4):1277–1286.
- 894 Ezzyat, Y. and Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological*  
895 *Science*, 22(2):243–252.
- 896 Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

- 897 Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of  
898 visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316.
- 899 Friendly, M., Franklin, P. E., Hoffman, D., and Rubin, D. C. (1982). The Toronto Word Pool:  
900 Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080  
901 words. *Behavior Research Methods and Instrumentation*, 14:375–399.
- 902 Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral  
903 component of information processing. *Trends in Cognitive Science*, 19(6):304–315.
- 904 Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal  
905 receptive windows in human cortex. *Journal of Neuroscience*, 28(10):2539–2550.
- 906 Heusser, A. C., Ezzyat, Y., Shiff, I., and Davachi, L. (2018a). Perceptual boundaries cause mnemonic  
907 trade-offs between local boundary processing and across-trial associative binding. *Journal of*  
908 *Experimental Psychology Learning, Memory, and Cognition*, 44(7):1075–1090.
- 909 Heusser, A. C., Fitzpatrick, P. C., Field, C. E., Ziman, K., and Manning, J. R. (2017). Quail: a  
910 Python toolbox for analyzing and plotting free recall data. *The Journal of Open Source Software*,  
911 10.21105/joss.00424.
- 912 Heusser, A. C., Ziman, K., Owen, L. L. W., and Manning, J. R. (2018b). HyperTools: a Python  
913 toolbox for gaining geometric insights into high-dimensional data. *Journal of Machine Learning*  
914 *Research*, 18(152):1–6.
- 915 Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal*  
916 *of Mathematical Psychology*, 46:269–299.
- 917 Howard, M. W., MacDonald, C. J., Tiganj, Z., Shankar, K. H., Du, Q., Hasselmo, M. E., and H., E.  
918 (2014). A unified mathematical framework for coding time, space, and sequences in the medial  
919 temporal lobe. *Journal of Neuroscience*, 34(13):4692–4707.
- 920 Howard, M. W., Viskontas, I. V., Shankar, K. H., and Fried, I. (2012). Ensembles of human MTL  
921 neurons “jump back in time” in response to a repeated stimulus. *Hippocampus*, 22:1833–1847.

- 922 Huk, A., Bonnen, K., and He, B. J. (2018). Beyond trial-based paradigms: continuous behavior, on-  
923 going neural activity, and naturalistic stimuli. *Journal of Neuroscience*, 10.1523/JNEUROSCI.1920-  
924 17.2018.
- 925 Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24:103–109.
- 926 Koriat, A. and Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distin-  
927 guishing accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of*  
928 *Experimental Psychology: General*, 123(3):297–315.
- 929 Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – con-  
930 nnecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:1 – 28.
- 931 Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to latent semantic analysis.  
932 *Discourse Processes*, 25:259–284.
- 933 Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy  
934 of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8):2906–2915.
- 935 Manning, J. R. (2019). Episodic memory: mental time travel or a quantum ‘memory wave’ function?  
936 *PsyArXiv*, doi:10.31234/osf.io/6zjwb.
- 937 Manning, J. R., Norman, K. A., and Kahana, M. J. (2015). The role of context in episodic memory.  
938 In Gazzaniga, M., editor, *The Cognitive Neurosciences, Fifth edition*, pages 557–566. MIT Press.
- 939 Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns  
940 in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National*  
941 *Academy of Sciences, USA*, 108(31):12893–12897.
- 942 McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and  
943 projection for dimension reduction. *arXiv*, 1802(03426).
- 944 Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of  
945 insula function. *Brain Structure and Function*, 214(5-6):655–667.

- 946 Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations  
947 in vector space. *arXiv*, 1301.3781.
- 948 Mueller, A., Fillion-Robin, J.-C., Boidol, R., Tian, F., Nechifor, P., yoonsubKim, Peter, Rampin, R.,  
949 Corvellec, M., Medina, J., Dai, Y., Petrushev, B., Langner, K. M., Hong, Alessio, Ozsvald, I.,  
950 vkolmakov, Jones, T., Bailey, E., Rho, V., IgorAPM, Roy, D., May, C., foobuzz, Piyush, Seong,  
951 L. K., Goey, J. V., Smith, J. S., Gus, and Mai, F. (2018). WordCloud 1.5.0: a little word cloud  
952 generator in Python. *Zenodo*, <https://zenodo.org/record/1322068#.W4tPKZNKh24>.
- 953 Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*,  
954 64:482–488.
- 955 Paller, K. A. and Wagner, A. D. (2002). Observing the transformation of experience into memory.  
956 *Trends in Cognitive Sciences*, 6(2):93–102.
- 957 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Pretten-  
958 hofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,  
959 M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine  
960 Learning Research*, 12:2825–2830.
- 961 Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model  
962 of organizational processes in free recall. *Psychological Review*, 116(1):129–156.
- 963 Postman, L. and Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal  
964 of Experimental Psychology*, 17:132–138.
- 965 Rabiner, L. (1989). A tutorial on Hidden Markov Models and selected applications in speech  
966 recognition. *Proceedings of the IEEE*, 77(2):257–286.
- 967 Radvansky, G. A. and Zacks, J. M. (2017). Event boundaries in memory and cognition. *Curr Opin  
968 Behav Sci*, 17:133–140.
- 969 Ramdas, A., Trillos, N., and Cuturi, M. (2017). On wasserstein two-sample testing and related  
970 families of nonparametric tests. *Entropy*, 19(2):47.

- 971 Ranganath, C. and Ritchey, M. (2012). Two cortical systems for memory-guided behavior. *Nature*  
972 *Reviews Neuroscience*, 13:713 – 726.
- 973 Rishel, C. A., Huang, G., and Freedman, D. J. (2013). Independent category and spatial encoding  
974 in parietal cortex. *Neuron*, 77(5):969–979.
- 975 Sigman, M. and Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during  
976 dual-task performance. *Journal of Neuroscience*, 28(30):7585–7589.
- 977 Simony, E., Honey, C. J., Chen, J., and Hasson, U. (2016). Uncovering stimulus-locked network  
978 dynamics during narrative comprehension. *Nature Communications*, 7(12141):1–13.
- 979 Stark, S. M., Yassa, M. A., and Stark, C. E. L. (2010). Individual differences in spatial pattern  
980 separation performance associated with healthy aging in humans. *Learning & Memory*, 17(6):284–  
981 288.
- 982 Steyvers, M., Shiffrin, R. M., and Nelson, D. L. (2004). Word association spaces for predicting  
983 semantic similarity effects in episodic memory. In Healy, A. F., editor, *Cognitive Psychology and*  
984 *its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. American  
985 Psychological Association, Washington, DC.
- 986 Tompry, A. and Davachi, L. (2017). Consolidation promotes the emergence of representational  
987 overlap in the hippocampus and medial prefrontal cortex. *Neuron*, 96(1):228–241.
- 988 Tyng, C. M., Amin, H. U., Saad, M. N. M., and S, M. A. (2017). The influences of emotion on  
989 learning and memory. *Frontiers in psychology*, 8:1454.
- 990 Welch, G. B. and Burnett, C. T. (1924). Is primacy a factor in association-formation. *American Journal*  
991 *of Psychology*, 35:396–401.
- 992 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale  
993 automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665.

- 994 Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., and Stark, C. E. L. (2011). Pattern  
995 separation deficits associated with increased hippocampal ca3 and dentate gyrus activity in  
996 nondemented older adults. *Hippocampus*, 21(9):968–979.
- 997 Yassa, M. A. and Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends In Neuro-*  
998 *sciences*, 34(10):515–525.
- 999 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research.  
1000 *Journal of Memory and Language*, 46:441–517.
- 1001 Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., and  
1002 Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection  
1003 and familiarity. *Nature Neuroscience*, 5(11):1236–41.
- 1004 Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception:  
1005 a mind-brain perspective. *Psychological Bulletin*, 133:273–293.
- 1006 Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit  
1007 memories to other brains: Constructing shared neural representations via communication. *Cereb*  
1008 *Cortex*, 27(10):4988–5000.
- 1009 Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and  
1010 memory. *Psychological Bulletin*, 123(2):162 – 185.

## 1011 Supporting information

1012 Supporting information is available in the online version of the paper.

## 1013 Acknowledgements

1014 We thank Luke Chang, Janice Chen, Chris Honey, Lucy Owen, Emily Whitaker, and Kirsten Ziman  
1015 for feedback and scientific discussions. We also thank Janice Chen, Yuan Chang Leong, Kenneth

1016 Norman, and Uri Hasson for sharing the data used in our study. Our work was supported in part  
1017 by NSF EPSCoR Award Number 1632738. The content is solely the responsibility of the authors  
1018 and does not necessarily represent the official views of our supporting organizations.

1019 **Author contributions**

1020 Conceptualization: A.C.H. and J.R.M.; Methodology: A.C.H., P.C.F. and J.R.M.; Software: A.C.H.,  
1021 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.,  
1022 P.C.F. and J.R.M.; Supervision: J.R.M.

1023 **Author information**

1024 The authors declare no competing financial interests. Correspondence and requests for materials  
1025 should be addressed to J.R.M. (jeremy.r.manning@dartmouth.edu).