

Carryover effects in free recall reveal how prior experiences influence memories of new experiences

Jeremy R. Manning^{1,*}, Andrew C. Heusser^{1,2}, Kirsten Ziman^{1,3},
Emily Whitaker¹, and Paxton C. Fitzpatrick¹

¹Dartmouth College

²Akili Interactive

³Princeton University

*Corresponding author: jeremy.r.manning@dartmouth.edu

Abstract

We perceive, interpret, and remember ongoing experiences through the lens of our prior experiences. Inferring that we are in one type of situation versus another can lead us to interpret the same physical experience differently. In turn, this can affect how we focus our attention, form expectations of what will happen next, remember what is happening now, draw on our prior related experiences, and so on. To study these phenomena, we asked participants to perform simple word list learning tasks. Across different experimental conditions, we held the set of to-be-learned words constant, but we manipulated the orders in which the words were studied. We found that these order manipulations affected not only how the participants recalled the ordered lists, but also how they recalled later randomly ordered lists. Our work shows how structure in our ongoing experiences can exert influence on how we remember unrelated subsequent experiences.

17 Introduction

18 Experience is subjective: different people who encounter identical physical experiences
19 can take away very different meanings and memories. One reason is that our subjective ex-
20 periences in the moment are shaped in part the idiosyncratic prior experiences, memories,
21 goals, thoughts, expectations, and emotions that we bring with us into the present moment.
22 These factors collectively define a *context* for our experiences¹⁴. situation models: forming
23 expectations, predicting ambiguous future experiences The contexts we encounter help us
24 to construct *situation models*^{16,25} or *schemas*^{2,19} that describe how experiences are likely to
25 unfold based on our prior experiences with similar contextual cues. For example, when
26 we enter a sit-down restaurant, we might expect to be seated at a table, given a menu,
27 and served food. Priming someone to expect a particular situation or context can also
28 influence how they resolve potential ambiguities in their ongoing experiences, including
29 ambiguous movies and narratives³³.

30 Our understanding of how we form situation models and schemas, and how they in-
31 teract with our subjective experiences and memories, is constrained in part by substantial
32 differences in how we study these processes. Situation models and schemas are most often
33 studied using “naturalistic” stimuli such as narratives and movies^{21,35,36}. In contrast, our
34 understanding of how we organize our memories has been most widely studied using
35 more traditional paradigms like free recall of random word lists¹². In free recall, partici-
36 pants study lists of items and are instructed to recall the items in any order they choose.
37 The orders in which words come to mind can provide insights into how participants have
38 organized their memories of the studied words. Because random word lists are unstruc-
39 tured by design, it is not clear if or how non-trivial situation models might apply to these
40 stimuli. Nevertheless, there are *some* commonalities between memory for word lists and
41 memory for real-world experiences.

42 Like remembering real-world experiences, remembering words on a studied list re-
43 quires distinguishing the current list from the rest of one's experience. To model this
44 fundamental memory capability, cognitive scientists have posited the existence of a spe-
45 cial representation, called *context*, that is associated with each list. According to early
46 theories e.g.^{1,6} context representations are composed of many features which fluctuate
47 from moment to moment, slowly drifting through a multidimensional feature space. Dur-
48 ing recall, this representation forms part of the retrieval cue, enabling us to distinguish
49 list items from non-list items. Understanding the role of context in memory processes is
50 particularly important in self-cued memory tasks, such as *free recall*, where the retrieval
51 cue is "context" itself.

52 Over the past half-century, context-based models have enjoyed impressive success at
53 explaining many stereotyped behaviors observed during free recall and other list-learning
54 tasks^{6,7,9,13,22-24,28? -30}. These phenomena include the well-known recency and primacy
55 effects (superior recall of items from the end and, to a lesser extent, from the beginning of
56 the study list), as well as semantic and temporal clustering effects⁷. The contiguity effect
57 is an example of temporal clustering, which is perhaps the dominant form of organization
58 in free recall. This effect can be seen in the tendency for people to successively recall items
59 that occupied neighboring positions in the study list. For example, if a list contained the
60 sub-sequence "ABSENCE HOLLOW PUPIL" and the participant recalls the word "HOLLOW", it is
61 far more likely that the next response will be either "PUPIL" or "ABSENCE" than some other
62 list item¹¹. In addition, there is a strong forward bias in the contiguity effect: subjects
63 make forward transitions (i.e., "HOLLOW" followed by "PUPIL") about twice as often as
64 they make backward transitions, despite an overall tendency to begin recall at the end of
65 the list. There are also striking effects of semantic clustering^{3,4,10,15,26}, whereby the recall
66 of a given item is more likely to be followed by recall of a similar or related item than

67 a dissimilar or unrelated one. In general, people organize memories for words along a
68 wide variety of stimulus dimensions. As captured by models like the *Context Maintenance*
69 *and Retrieval Model*²³, the stimulus features associated with each word (e.g. the word's
70 meaning, font size, font color, location on the screen, size of the object the word represents,
71 etc.) are incorporated into the participant's mental context representation^{14,16–18,31}. During
72 a memory test, any of these features may serve as a memory cue, which in turn leads the
73 participant to recall in succession words that share stimulus features.

74 A key mystery is whether the sorts of situation models and schemas that people use to
75 organize their memories of real-world experiences might map onto the clustering effects
76 that reflect how people organize their memories for word lists. On one hand, situation
77 models and clustering effects both reflect statistical regularities in ongoing experience.
78 Our memory systems exploit these regularities when generating inferences about the
79 unobserved past and yet-to-be-experienced future^{5,20,25,27,32}. On the other hand, the rich
80 structure of real-world experiences and other naturalistic stimuli that enable people to
81 form deep and meaningful situation models and schemas have no obvious analog in
82 simple word lists. Often lists in free recall studies are explicitly *designed* to be devoid of
83 exploitable temporal structure, for example by sorting the words in a random order¹².

84 We designed an experimental paradigm to explore how people organize their mem-
85 ories for simple stimuli (word lists) whose temporal properties change across different
86 “situations,” analogous to how the content of real-world experiences change across dif-
87 ferent real-world situations. We asked participants to study and freely recall a series
88 of word lists (Fig. 1). Across the different conditions in the experiment, we varied the
89 lists' presentation orders in different ways across lists. The studied items (words) were
90 designed to vary along three general dimensions: semantic (word *category*, and physical
91 *size* of the referent), lexicographic (word *length* and *first letter*), and visual (font *color* and

the onscreen *location* of each word). In our main manipulation conditions, we asked participants to study and recall eight lists whose items were sorted by a target feature (e.g., word category). Next, we asked them to study and recall an additional eight lists whose items had the same features, but that were sorted in a random temporal order. We were interested in how these order manipulations affected participants' recall behaviors on early (sorted) lists, as well as how order manipulations on early lists affected recall behaviors on later (unsorted) lists. We used a series of control conditions as a baseline; in these control conditions all of the lists were sorted randomly, but we manipulated the presence or absence of the visual features. Finally, in an *adaptive* experimental condition we used participants' recall behaviors on early lists to manipulate, in real-time, the presentation orders of subsequent lists. In this adaptive condition, we sought to identify potential commonalities within and across participants in how people organized their memories and how those organizational tendencies affect overall performance.

Materials and methods

Participants

We enrolled a total of 491 Dartmouth undergraduate students across 11 experimental conditions. The conditions included two primary controls (feature rich, reduced), two secondary controls (reduced (early), reduced (late)), six order manipulation conditions (category, size, length, first letter, color, and location), and a final adaptive condition. Each of these conditions are described in the *Experimental design* subsection below.

Participants received course credit for enrolling in our study. We asked each participant to fill out a demographic survey that included information about their self-reported age, gender, ethnicity, race, education, vision, reading impairments, medications or recent

injuries, coffee consumption on the day of testing, and level of alertness at the time of testing. All components of the demographics survey were optional. One participant elected not to fill out any part of the demographic survey, and all other participants report some or all of their requested demographic information.

We aimed to run (to completion) at least 60 participants in each of the two primary control conditions and in the adaptive condition. In all other conditions we set a target enrollment of at least 30 participants. Because our data collection efforts were coordinated 12 researchers and multiple testing rooms and computers, it was not feasible for individual experimenters to know how many participants had been run in each experimental condition until the relevant databases were synchronized at the end of each working day. We also over-enrolled participants for each condition to help ensure that we met our minimum enrollment targets even if some participants dropped out of the study prematurely or did not show up for their testing session. This led us to exceed our target enrollments for several conditions.

Participants were assigned to experimental conditions based loosely on their date of participation. (This aspect of our procedure helped us to more easily synchronize the experiment databases across multiple testing computers.) Of the 490 participants who opted to fill out the demographics survey, reported ages ranged from 17 to 31 years (mean: 19.1; standard deviation: 1.356). A total of 318 participants reported their gender as female, 170 as male, and 2 participants declined to report their gender. A total of 442 participants reported their ethnicity as “not Hispanic or Latino,” 39 as “Hispanic or Latino,” and 9 declined to report their ethnicity. Participants reported their races as White (345 participants), Asian (120 participants), Black or African American (31 participants), American Indian or Alaska Native (11 participants), Native Hawaiian or Other Pacific Islander (4 participants), Mixed race (3 participants), Middle Eastern (1 participant), and Arab (1

140 participant). A total of 5 participants declined to report their race. We note that several
141 participants reported more than one of racial category. Participants reported their high-
142 est degrees achieved as “Some college” (359 participants), “High school graduate” (117
143 participants), “College graduate” (7 participants), “Some high school” (5 participants),
144 “Doctorate” (1 participant), and “Master’s degree” (1 participant). A total of 482 partici-
145 pants reported no reading impairments, and 8 reported mild reading impairments such
146 as mild dyslexia. A total of 489 participants reported having normal color vision and 1
147 participant reported that they were color blind. A total of 482 participants reported taking
148 no prescription medications and having no recent injuries; 4 participants reported having
149 ADHD, 1 reported having dyslexia, 1 reported having allergies, 1 reported a recently
150 torn ACL/MCL, and 1 reported a concussion from several months prior. The participants
151 reported consuming 0 – 3 cups of coffee prior to the testing session (mean: 0.32 cups;
152 standard deviation: 0.58 cups). Participants reported their current level of alertness, and
153 we converted their responses to numerical scores as follows: “very sluggish” (-2), “a little
154 sluggish” (-1), “neutral” (0), “a little alert” (1), and “very alert” (2). Across all partici-
155 pants, the full range of alertness levels were reported (range: -2 – 2; mean: 0.35; standard
156 deviation: 0.89) .

157 We dropped from our dataset the 1 participant who reported abnormal color vision, as
158 well as 39 participants whose data were corrupted due to technical failures while running
159 the experiment or during the daily database merges. In total, this left usable data from
160 452 participants, broken down by experimental condition as follows: feature rich (67
161 participants), reduced (61 participants), reduced (late) (41 participants), reduced (early),
162 (42 participants), category (30 participants), size (30 participants), length (30 participants),
163 first letter (30 participants), color (31 participants), location (30 participants), and adaptive
164 (60 participants). The participant who declined to fill out their demographic survey



Figure 1: Feature-rich free recall. After studying lists comprised of words that vary along several feature dimensions, participants verbally recall words in any order (microphone icon).

165 participated in the location condition, and we verified verbally that they had normal color
166 vision.

167 Experimental design

168 Our experiment is a variant of the classic free recall paradigm that we term *feature-rich free*
169 *recall*. In feature-rich free recall, participants study 16 lists, each comprised of 16 words
170 that vary along a number of stimulus dimensions: semantic category, object size, text
171 color, text location, word length, and starting letter (Fig. 1). Each list contains four words
172 from each of four different semantic categories and two object sizes; all other stimulus
173 features are randomized. After studying each list, the participant attempts to recall as
174 many words as they can from that list, in any order they choose. Because each individual
175 word is associated with several well-defined (and quantifiable) features, and because each
176 list incorporates a diverse mix of feature values along each dimension, this allows us to
177 evaluate participants' memory fingerprints in rich detail.

178 Stimuli

179 Stimuli in our paradigm were 256 English words selected in a previous study³⁴. The words
180 all referred to concrete nouns, and were chosen from 15 unique semantic categories: body
181 parts, building-related, cities, clothing, countries, flowers, fruits, insects, instruments,
182 kitchen-related, mammals, (US) states, tools, trees, and vegetables. We also tagged each

word according to the approximate size of the object the word referred to. Words were labeled as “small” if the corresponding object was likely able to “fit in a standard shoebox” or “large” if the object was larger than a shoebox. Semantic categories varied in how many object sizes they reflected (mean number of different sizes per category: 1.33; standard deviation: 0.49). The numbers of words in each semantic category also varied from 12 – 28 (mean number of words per category: 17.07; standard deviation number of words: 4.65). We also identified lexicographic features for each word, including the words’ first letters and lengths (i.e., number of letters). Across all categories, all possible first letters were represented except for ‘Q’ (average number of unique first letters per category: 11; standard deviation: 2 letters). Word lengths ranged from 3 – 12 letters (average: 6.17 letters; standard deviation: 2.06 letters).

We assigned the categorized words into a total of 16 lists with several constraints. First, we required that each list contained words from exactly 4 unique categories, each with exactly 4 exemplars from each category. Second, we required that (across all words on the list) at least one instance of both object sizes were represented. On average, each category was represented in 4.27 lists (standard deviation: 1.16 lists). Aside from these two constraints, we assigned each word to a unique list. After random assignment, each list contained words with an average of 11.13 unique starting letters (standard deviation: 1.15 letters) and an average word length of 6.17 letters (standard deviation: 0.34 letters).

The above assignments of words to lists was performed once across all participants, such that every participant studied the same set of 16 lists. We randomized the study order of these lists across participants. For participants in some conditions, on some lists, we also randomly varied two additional visual features to each word: the presentation font color, and the word’s onscreen location. These attributes were assigned independently for word (and for every participant) at the times the words were displayed onscreen. These

208 visual features were varied for words in all lists and conditions except for the “reduced”
209 condition (all lists), the first eight lists of the “reduced (early)” condition, and the last eight
210 lists of the “reduced (late)” condition. In these latter cases, words were all presented in
211 black at the center of the experimental computer’s display.

212 To assign a random font color to each word, we selected three integers uniformly
213 and at random between 0 and 255, corresponding to the red (r), green (g), and blue (b)
214 color channels for that word. To assign random presentation locations to each word, we
215 selected two floating point numbers uniformly at random (one for the word’s horizontal
216 x coordinate and the other for its vertical y coordinate). The bounds of these coordinates
217 were selected to cover the entire visible area of the display without cutting off any part of
218 the words. The words were shown on 27 in (diagonal) Retina 5K iMac displays (resolution:
219 5120×2880 pixels).

220 **Real-time speech-to-text processing**

221 Our experimental paradigm incorporates the Google Cloud Speech API speech-to-text
222 engine⁸ to automatically transcribe participants’ verbal recalls into text. This allows
223 recalls to be transcribed in real time— a distinguishing feature of the experiment; in typical
224 verbal recall experiments the audio data must be parsed manually. In prior work, we
225 used a similar experimental setup (equivalent to the “reduced” condition in the present
226 study) to verify that the automatically transcribed recalls were sufficiently close to human-
227 transcribed recalls to yield reliable data³⁴. This real-time speech processing component of
228 the paradigm plays an important role in the “adaptive” condition of the experiment, as
229 described below.

230 **Order manipulation conditions**

231 **Feature dimensions.**

232 **Constructing feature-sorted lists.**

233 **Random conditions**

234 **Adaptive conditions**

235 **Online “fingerprint” analysis.**

236 **Ordering “stabilize” lists by an estimated fingerprint.**

237 **Ordering “destabilize” lists by an estimated fingerprint.**

238 **Analysis**

239 **Probability of first recall and probability of n^{th} recall**

240 **Lag conditional response probability**

241 **Computing clustering scores and memory fingerprints**

242 **Identifying event boundaries**

243 **Serial position curves and recall accuracy**

244 **Computing low-dimensional embeddings of memory fingerprints**

245 **Results**

246 **Figure S3.**

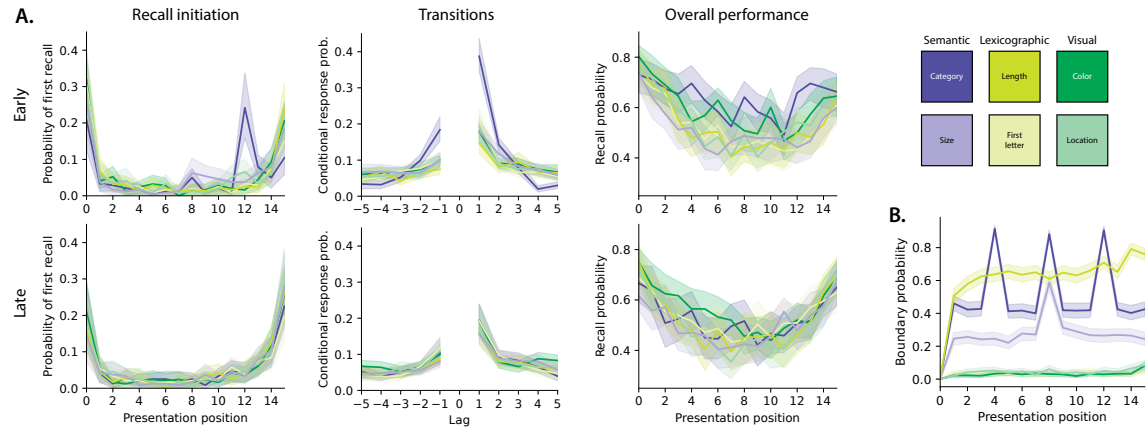


Figure 2: Recall dynamics in feature rich free recall (order manipulation conditions). **A.** Behavioral plots. **Left panels.** The probabilities of initiating recall with each word are plotted as a function of presentation position. **Middle panels.** The conditional probabilities of recalling each word are plotted as a function of the relative position (Lag) to the words recalled just-prior. **Right panels.** The overall probabilities of recalling each word are plotted as a function of presentation position. **All panels.** Error ribbons denote bootstrap-estimated 95% confidence intervals (calculated across participants). Top panels display the recall dynamics for early (order manipulation) lists in each condition (color). Bottom panels display the recall dynamics for late (randomly ordered) lists. See Figures S1 and S2 for analogous plots for the random (control) and adaptive conditions. **B.** Proportion of event boundaries (see *Methods*) for each condition's feature of focus, plotted as a function of presentation position.

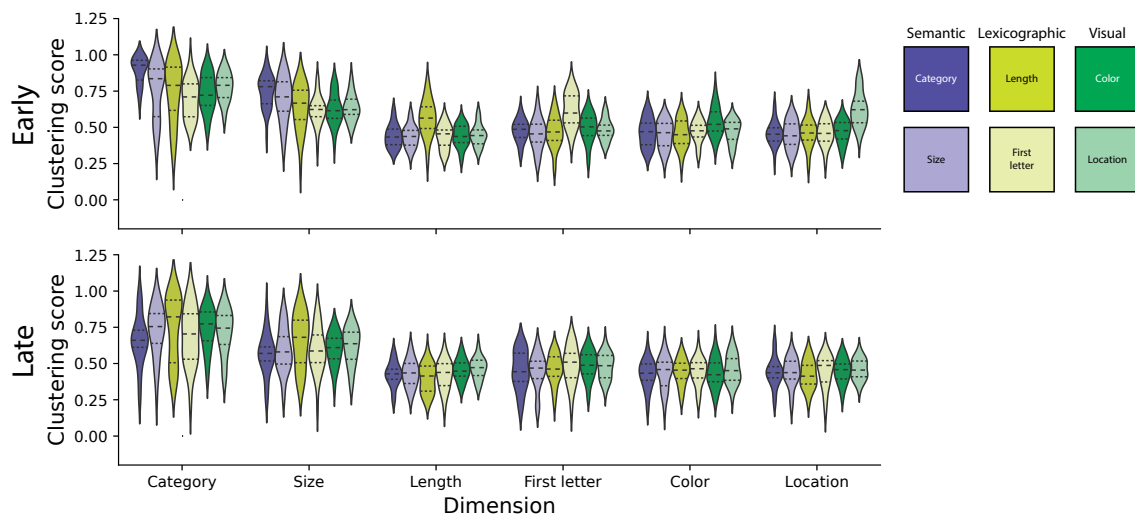


Figure 3: Memory “fingerprints” (order manipulation conditions). The across-participant distributions of clustering scores for each feature type (x -coordinate) are displayed for each experimental condition (color), separately for order manipulation (early, top) and randomly ordered (late, bottom) lists. See Figures S5 and S6 for analogous plots for the random (control) and adaptive conditions.

247 Figure S7.

248 Figure S4.

249 Discussion

250 References

- 251 [1] Anderson, J. R. and Bower, G. H. (1972). Recognition and retrieval processes in free
252 recall. *Psychological Review*, 79(2):97–123.
- 253 [2] Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of real-world
254 event schemas during narrative perception. *The Journal of Neuroscience*, 38(45):9689–
255 9699.

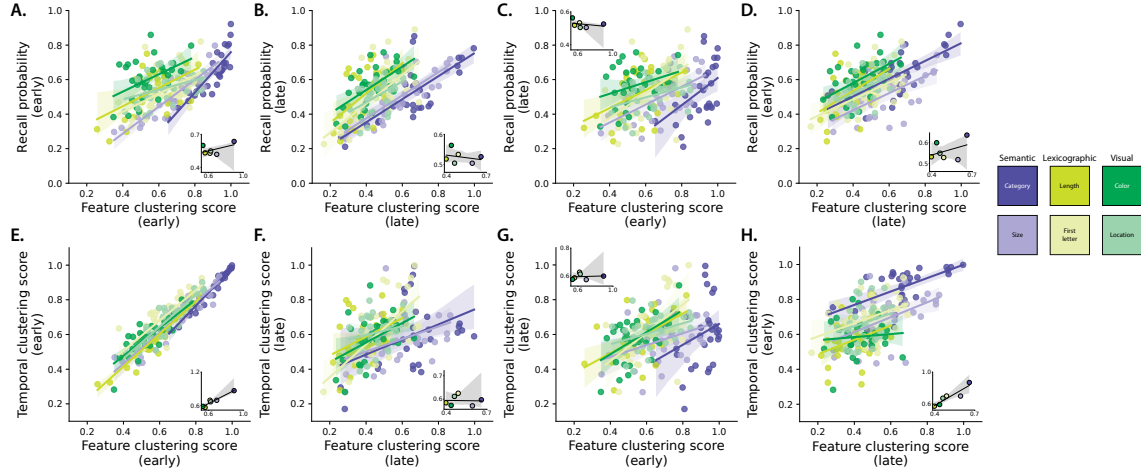


Figure 4: Interactions between feature clustering, recall probability, and contiguity. A. Recall probability versus feature clustering scores for order manipulation (early) lists. B. Recall probability versus feature clustering for randomly ordered (late) lists. C. Recall probability on late lists versus feature clustering on early lists. D. Recall probability on early lists versus feature clustering on late lists. E. Temporal clustering scores (contiguity) versus feature clustering scores on early lists. F. Temporal clustering scores versus feature clustering scores on late lists. G. Temporal clustering scores on late lists versus feature clustering scores on early lists. H. Temporal clustering scores on early lists versus feature clustering scores on late lists. **All panels.** Each dot in the main scatterplots denotes the average scores for one participant. The colored regression lines are computed across participants. The inset displays condition-averaged results, where each dot reflects a single condition and the regression line is computed across experimental conditions. All error ribbons denote bootstrap-estimated 95% confidence intervals.

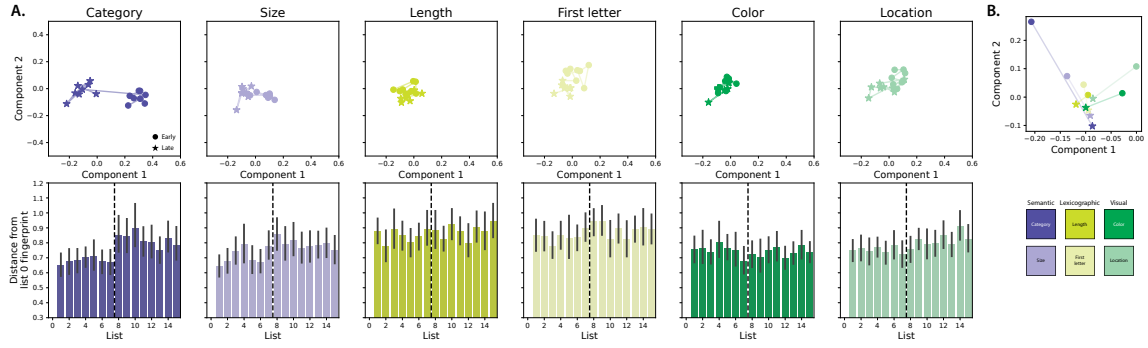


Figure 5: Memory fingerprint dynamics (order manipulation conditions). **A.** Each column (and color) reflects an experimental condition. In the top panels, each marker displays a 2D projection of the (across-participant) average memory fingerprint for one list. Order manipulation (early) lists are denoted by circles and randomly ordered (late) lists are denoted by stars. All of the fingerprints (across all conditions and lists) are projected into a common space. The bar plots in the bottom panels display the Euclidean distances of the per-list memory fingerprints to the list 0 fingerprint, for each condition. Error bars denote bootstrap-estimated 95% confidence intervals. The dotted vertical lines denote the boundaries between early and late lists. **B.** In this panel, the fingerprints for early (circle) and late (star) lists are averaged across lists and participants before projecting the fingerprints into a (new) 2D space. See Figure S8 for analogous plots for the random (control) conditions.

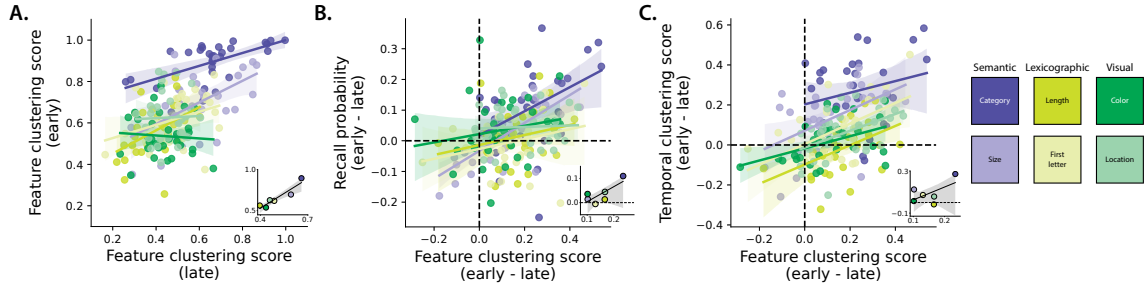


Figure 6: Feature clustering carryover effects. **A.** Feature clustering scores for order manipulation (early) versus randomly ordered (late) lists. **B.** Accuracy differences (on early versus late lists) versus feature clustering “carryover” (defined as the differences between the average clustering scores on early and late lists). **C.** Temporal clustering differences (on early versus late lists) versus feature clustering carryover. **All panels.** Each dot in the main scatterplots denotes the average scores for one participant. The colored regression lines are computed across participants. The inset displays condition-averaged results, where each dot reflects a single condition and the regression line is computed across experimental conditions. All error ribbons denote bootstrap-estimated 95% confidence intervals.

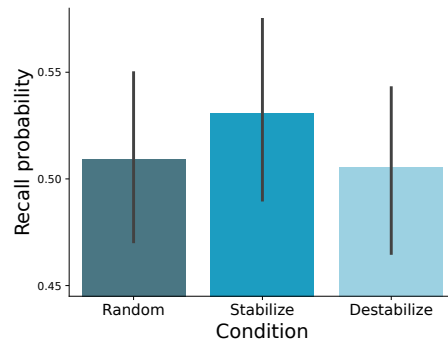


Figure 7: Recall performance (adaptive conditons). The bars display the average probability of recall (taken across words, lists, and participants) for lists from each adaptive condition. Error bars denote bootstrap-estimated 95% confidence intervals. For additional details about participants' behavior and performance during the adaptive conditions, see Figure S2.

- 256 [3] Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged
257 associates. *Journal of General Psychology*, 49:229–240.
- 258 [4] Bousfield, W. A., Sedgewick, C. H., and Cohen, B. H. (1954). Certain temporal charac-
259 teristics of the recall of verbal associates. *American Journal of Psychology*, 67:111–118.
- 260 [5] Bower, G. H., Black, J. B., and Turner, T. J. (1979). Scripts in memory for text. *Cognitive*
261 *Psychology*, 11(2):177–220.
- 262 [6] Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psycho-*
263 *logical Review*, 62:145–154.
- 264 [7] Glenberg, A. M., Bradley, M. M., Kraus, T. A., and Renzaglia, G. J. (1983). Studies of
265 the long-term recency effect: support for a contextually guided retrieval theory. *Journal*
266 *of Experimental Psychology: Learning, Memory, and Cognition*, 12:413–418.
- 267 [8] Halpern, Y., Hall, K. B., Schogol, V., Riley, M., Roark, B., Skobeltsyn, G., and Bäuml,
268 M. (2016). Contextual prediction models for speech recognition. In *Interspeech*, pages
269 2338–2342.

- 270 [9] Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal
271 context. *Journal of Mathematical Psychology*, 46:269–299.
- 272 [10] Jenkins, J. J. and Russell, W. A. (1952). Associative clustering during recall. *Journal of*
273 *Abnormal and Social Psychology*, 47:818–821.
- 274 [11] Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and*
275 *Cognition*, 24:103–109.
- 276 [12] Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press, New
277 York, NY.
- 278 [13] Kimball, D. R., Smith, T. A., and Kahana, M. J. (2007). The fSAM model of false recall.
279 *Psychological Review*, 114(4):954–993.
- 280 [14] Manning, J. R. (2020). Context reinstatement. In Kahana, M. J. and Wagner, A. D.,
281 editors, *Handbook of Human Memory*. Oxford University Press.
- 282 [15] Manning, J. R. and Kahana, M. J. (2012). Interpreting semantic clustering effects in
283 free recall. *Memory*, 20(5):511–517.
- 284 [16] Manning, J. R., Norman, K. A., and Kahana, M. J. (2015). The role of context in
285 episodic memory. In Gazzaniga, M., editor, *The Cognitive Neurosciences*, pages 557–566.
286 MIT Press.
- 287 [17] Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscil-
288 latory patterns in temporal lobe reveal context reinstatement during memory search.
289 *Proceedings of the National Academy of Sciences, USA*, 108(31):12893–12897.
- 290 [18] Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., and Kahana, M. J.

- (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *The Journal of Neuroscience*, 32(26):8871–8878.
- [19] Masís-Obando, R., Norman, K. A., and Baldassano, C. (2022). Scheme representations in distinct brain networks support narrative memory during encoding and retrieval. *eLife*, 11:e70445.
- [20] Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., and Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behavior*, 1:680–692.
- [21] Nastase, S. A., Goldstein, A., and Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 15(222):117254–117261.
- [22] Polyn, S. M. and Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, 12:24–30.
- [23] Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). Task context and organization in free recall. *Neuropsychologia*, 47:2158–2163.
- [24] Raaijmakers, J. G. W. and Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In Bower, G. H., editor, *The Psychology of Learning and Motivation: Advances in Research and Theory*, volume 14, pages 207–262. Academic Press, New York, NY.
- [25] Ranganath, C. and Ritchey, M. (2012). Two cortical systems for memory-guided behavior. *Nature Reviews Neuroscience*, 13:713–726.
- [26] Romney, A. K., Brewer, D. D., and Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, 4:28–34.

- 314 [27] Schapiro, A. and Turk-Browne, N. (2015). Statistical learning. *Brain Mapping: An*
315 *Encyclopedic Reference*, 3:501–506.
- 316 [28] Sederberg, P. B., Howard, M. W., and Kahana, M. J. (2008). A context-based theory of
317 recency and contiguity in free recall. *Psychological Review*, 115(4):893–912.
- 318 [29] Shankar, K. H. and Howard, M. W. (2012). A scale-invariant internal representation
319 of time. *Neural Computation*, 24:134–193.
- 320 [30] Sirotin, Y. B., Kimball, D. R., and Kahana, M. J. (2005). Going beyond a single list:
321 modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin and*
322 *Review*, 12(5):787–805.
- 323 [31] Smith, S. M. and Vela, E. (2001). Environmental context-dependent memory: a review
324 and meta-analysis. *Psychonomic Bulletin and Review*, 8(2):203–220.
- 325 [32] Xu, X., Zhu, Z., and Manning, J. R. (2022). The psychological arrow of time drives
326 temporal asymmetries in retrodicting versus predicting narrative events. *PsyArXiv*,
327 page doi.org/10.31234/osf.io/yp2qu.
- 328 [33] Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., and
329 Hasson, U. (2017). Same story, different story: the neural representation of interpretive
330 frameworks. *Psychological Science*, 28(3):307–319.
- 331 [34] Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., and Manning, J. R. (2018).
332 Is automatic speech-to-text transcription ready for use in psychological experiments?
333 *Behavior Research Methods*, 50:2597–2605.
- 334 [35] Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). The construction of
335 situation models in narrative comprehension: an event-indexing model. *Psychological*
336 *Science*, 6(5):292–297.

- 337 [36] Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language compre-
338 hension and memory. *Psychological Bulletin*, 123(2):162–185.