

1 **Abrupt remapping in human CA3/dentate gyrus**  
2 **signals resolution of memory interference**

3 Wanjia Guo<sup>1</sup>, Serra E. Favila<sup>2</sup>, Ghootae Kim<sup>3</sup>, Robert J. Molitor<sup>1</sup>, Brice A. Kuhl<sup>1</sup>

4 **Word Counts**

5 **Abstract:** 150

6 **Introduction, Results, Discussion:** 4703

7 **Methods:** 3009

8 **# of Figures:** 4

9 **1 Supplementary Table**

10 **Keywords:** hippocampus, episodic memory, pattern separation, repulsion, competition

11 **Acknowledgments:** This work was supported by NIH-NINDS R01 NS089729 awarded to B.A.K.

12 **Author Contributions:** W.G., G.K., and B.A.K. designed the experiment. W.G. and B.A.K. analyzed the  
13 data. S.E.F. consulted on data analyses. All authors wrote and edited the manuscript.

14 **ABSTRACT:**

15

16 Remapping refers to a decorrelation of hippocampal representations of similar spatial environments. While  
17 it has been speculated that remapping may contribute to the resolution of episodic memory interference in  
18 humans, direct evidence is surprisingly limited. Here, we tested this idea using high-resolution, pattern-  
19 based fMRI analyses. We show that activity patterns in human CA3/dentate gyrus exhibit an abrupt,  
20 temporally-specific decorrelation of highly similar memory representations that is precisely coupled with  
21 behavioral expressions of successful learning. Strikingly, the magnitude of this learning-related  
22 decorrelation was predicted by the amount of pattern overlap during initial stages of learning, with greater  
23 initial overlap leading to stronger decorrelation. Finally, we show that remapped activity patterns carry  
24 relatively more information about learned episodic associations compared to competing associations,  
25 further validating the learning-related significance of remapping. Collectively, these findings establish a  
26 critical link between hippocampal remapping and episodic memory interference and provide novel insight  
27 into why remapping occurs.

## 28 INTRODUCTION:

29

30 The hippocampus is critical for forming long-term, episodic memories<sup>1-3</sup>. However, one of the fundamental  
31 challenges that the hippocampus faces is that many experiences are similar, creating the potential for  
32 memory interference<sup>4,5</sup>. In rodents, it is well established that minor alterations to the environment can trigger  
33 sudden changes in hippocampal activity patterns—a phenomenon termed remapping<sup>6,7</sup>. An appealing  
34 possibility is that hippocampal remapping also occurs in human episodic memory, allowing for similar  
35 memories to be encoded in distinct activity patterns that prevent interference<sup>8</sup>. At present, however, there  
36 remains an important gap between evidence of place cell remapping in the rodent hippocampus and  
37 episodic memory interference in humans. To bridge this gap, it is informative to consider how properties of  
38 place cell remapping, as demonstrated in the rodent hippocampus, might translate to episodic memory  
39 interference in humans.

40

41 One of the most important properties of remapping in the rodent hippocampus is that it is characterized by  
42 abrupt transitions between representations<sup>9-12</sup>. These abrupt transitions, evidenced by decorrelations in  
43 patterns of neural activity, have most typically been observed as a function of the degree of environmental  
44 change<sup>9,11</sup>. However, abrupt remapping can also occur as a function of experience with a new  
45 environment<sup>10,12</sup>. Evidence of experience-dependent remapping<sup>6,13</sup> suggests an important point: that  
46 remapping fundamentally reflects changes in internal representations, as opposed to changes in  
47 environmental states<sup>14,15</sup>. An emphasis on internal representations lends itself well to human episodic  
48 memory in that it suggests that hippocampal remapping should occur as *memories change*. More  
49 specifically, this perspective makes the critical prediction that when two events are highly similar,  
50 hippocampal remapping will occur if, and when, corresponding memories become distinct. Testing this  
51 prediction requires repeatedly probing internal representations (memories) as well as hippocampal  
52 representations. However, standard approaches of averaging neuroimaging data across memories and  
53 participants can easily obscure or wash out abrupt changes in hippocampal representations if the timing of  
54 those changes varies across memories or participants.

55

56 Evidence of place cell remapping in rodents also motivates specific predictions regarding the relative  
57 contributions of hippocampal subfields, with a major distinction being between CA3/dentate gyrus and  
58 CA1<sup>8,16,17</sup>. In general, CA3 and dentate gyrus are thought to be more important than CA1 for discriminating  
59 between similar stimuli<sup>15,17-20</sup> and remapping has been shown to occur more abruptly in CA3 than in  
60 CA1<sup>10,12,21</sup>. Human fMRI studies also support this general distinction, with several studies specifically  
61 implicating CA3 and dentate gyrus in discriminating similar memories<sup>22-27</sup>. However, these studies have not  
62 directly established a link between temporally abrupt changes in CA3/dentate gyrus activity and changes  
63 in episodic memory states.

64

65 Here, we tested whether the resolution of interference between highly similar episodic memories is  
66 associated with an abrupt remapping of activity patterns in human CA3/dentate gyrus. We used an  
67 associative memory paradigm in which participants learned and were repeatedly tested on associations  
68 between scene images and object images<sup>28</sup>. The critical design feature was that the set of scene images  
69 included pairs of extremely similar scenes (**Fig. 1a**). These scene *pairmates* were intended to elicit  
70 associative memory interference. Across six rounds of learning, we tracked improvement in associative  
71 memory for each set of pairmates while also continuously tracking representational changes indexed by  
72 fMRI. Specifically, after each associative memory test round, participants were shown each scene image  
73 one at a time (exposure phase) which allowed us to measure the activity pattern evoked by each scene  
74 and, critically, the representational distance between scene pairmates. To preview, we find that behavioral

75 expressions of memory interference resolution are temporally-coupled to abrupt, stimulus-specific  
76 remapping of human CA3/dentate gyrus activity patterns. This remapping specifically exaggerated the  
77 representational distance between similar memories. In additional analyses, we show that the magnitude  
78 of remapping that individual memories experienced was predicted by the degree of initial pattern overlap  
79 among CA3/dentate gyrus representations and that remapped CA3/dentate gyrus representations carried  
80 increased and highly specific information about learned episodic associations.

81

## 82 **RESULTS:**

83

84 Participants completed six rounds of the experimental paradigm while inside an fMRI scanner. Each round  
85 included a study phase, an associative memory test phase, and a scene exposure phase (**Fig. 1b**). fMRI  
86 scanning was only conducted during the exposure phases. During the study phases, participants viewed  
87 scene-object associations one at a time. During the associative memory test phases, participants were  
88 shown scenes, one at a time, along with two very similar object choices (e.g., two guitars); one object was  
89 the target (i.e., the object that had been paired with the current scene) and the other object was the  
90 competitor (i.e., the object that had been paired with the scene pairmate). After selecting an object,  
91 participants indicated their confidence (high or low). During exposure phases, participants were shown each  
92 scene, along with novel scenes, and made a simple old/new judgment (mean  $\pm$  95% CI:  $d' = 5.40 \pm 0.88$ ;  
93 one-sample  $t$ -test vs. 0:  $t_{30} = 12.58$ ,  $p < 0.001$ , Cohen's  $d = 2.26$ ).

94

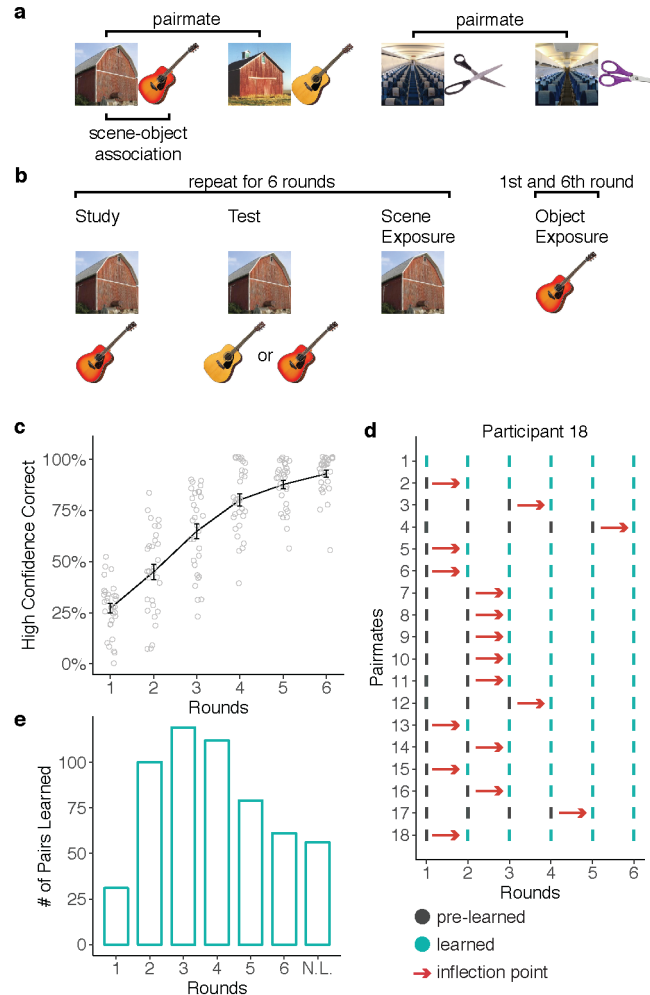
### 95 **Behavior.**

96

97 During the associative memory test phases, participants chose the correct object with above-chance  
98 accuracy in each of the 6 rounds ( $t_{30}$ 's  $\geq 2.65$ ,  $p$ 's  $\leq 0.013$ ,  $d$ 's  $\geq 0.48$ ; chance accuracy = 50%). Accuracy  
99 markedly improved across rounds, from a mean of 56.45%  $\pm$  4.98% in round 1 to a mean of 94.71%  $\pm$  2.21%  
100 in round 6 (main effect of round:  $F_{1,30} = 318.86$ ,  $p < 0.001$ ,  $\eta^2 = 0.91$ ). The rate of choosing the correct  
101 object with high-confidence also robustly increased across rounds, from a mean of 27.15%  $\pm$  4.71% in  
102 round 1 to 92.83%  $\pm$  3.58% in round 6 (main effect of round:  $F_{1,30} = 574.44$ ,  $p < 0.001$ ,  $\eta^2 = 0.95$ ; **Fig. 1c**).

103

104 To test whether hippocampal remapping was temporally coupled with the resolution of memory interference,  
105 we identified, for each participant and for each set of pairmates, the learning round in which scene-object  
106 associations were recalled with high confidence (for both scenes in a pairmate). We refer to this timepoint  
107 as the 'learned round' (LR; see Methods). Of critical interest for our remapping analyses was the correlation  
108 of activity patterns evoked by scene images during the learned round (LR) with activity patterns evoked  
109 immediately prior to the learned round (LR-1). We refer to this transition (from pre-learned to learned) as  
110 the 'inflection point' (IP) in learning (**Fig. 1d**). For example, if the LR for a particular set of pairmates was  
111 round 4, then the IP was the transition from round 3 to 4. Our rationale for correlating activity patterns from  
112 LR-1 with LR was that this correlation would capture the critical *change* in hippocampal representations  
113 (remapping) that putatively supports learning.



**Figure 1. Experimental Design and Behavior.** **a.** Participants learned 36 scene-object associations. The 36 scenes comprised 18 scene pairmates which consisted of highly similar image pairs (e.g., ‘barn 1’ and ‘barn 2’). Scene pairmates were also associated with similar objects (e.g., ‘guitar 1’ and ‘guitar 2’). **b.** Participants completed 6 rounds of study, test, and exposure phases. During study, participants viewed scenes and associated objects. During test, participants were presented with scenes and had to select the associated object from a set of two choices, followed by a confidence rating (high or low confidence; not shown). During exposure, scenes (rounds 1-6) or objects (round 1 and 6) were presented and participants made an old/new judgment. fMRI data were only collected during the scene and object exposure phases. **c.** Mean percentage of high confidence correct responses for each test round. **d.** Data from a representative participant showing the ‘inflection point’ in learning, for each pairmate. The inflection point was defined as the point at which participants transitioned to high-confidence correct retrieval for both scenes within a pairmate—a transition from ‘pre-learned’ to ‘learned.’ **e.** The number of pairs that transitioned to a learned state at each round, aggregated across all participants and pairmates. N.L. indicates pairmates that were never learned. Notes: error bars reflect S.E.M.

114 **Remapping in CA3/dentate gyrus is time-locked to the inflection point in learning.**

115

116 For our fMRI analyses, our primary focus was on pattern similarity between scene pairmates. Pattern  
 117 similarity was measured by correlating patterns of fMRI activity evoked by each scene during the scene  
 118 exposure phases. Pairmate similarity was defined as the correlation between activity patterns evoked by  
 119 scene pairmates (e.g., ‘barn 1’ and ‘barn 2’; **Fig. 2b**). Correlations between scenes that were not pairmates

120 (e.g., ‘barn 1’ and ‘airplane 2’; **Fig. 2b**) provided an important baseline measure of non-pairmate similarity.  
121 We refer to the difference between these two measures (pairmate – non-pairmate similarity) as the *pairmate*  
122 *similarity score*<sup>28</sup>. A positive pairmate similarity score would indicate that visually similar scenes (e.g., two  
123 barns) are associated with more similar neural representations than two unrelated scenes. Critically,  
124 because pairmate similarity scores are a relative measure, they can be directly compared across different  
125 brain regions<sup>29</sup>—something that would be inadvisable with raw correlation values. For all pattern similarity  
126 analyses, correlations were always performed across learning rounds (e.g., correlating ‘barn 1’ at LR-1 with  
127 ‘barn 2’ at LR). This ensured independence of fMRI data<sup>30</sup>, but was also intended to capture *transitions* in  
128 hippocampal representations (remapping).

129  
130 Following a prior study that used similar stimuli and analyses<sup>28</sup>, fMRI analyses targeted the following regions  
131 of interest (ROIs): hippocampus, parahippocampal place area (PPA), and early visual cortex (EVC). PPA  
132 and EVC served as important control regions indexing high-level (PPA) and low-level (EVC) visual  
133 representations. We did not anticipate that these regions would demonstrate learning-related remapping.  
134 Within the hippocampus, we leveraged our high-resolution fMRI protocol to segment the hippocampus body  
135 into subfields comprising CA1 and CA2/CA3/dentate gyrus (CA23DG). Motivated by past empirical  
136 findings<sup>23,31</sup> and theoretical models<sup>8</sup>, we predicted that remapping would occur in CA23DG. More  
137 specifically, we predicted that CA23DG remapping would occur at the inflection point (IP) in learning. To  
138 test this prediction, we compared pairmate similarity scores at the IP to pairmate similarity scores at a  
139 timepoint *just prior* to the IP (pre-IP). Whereas pairmate similarity scores at the IP were based on  
140 correlations between activity patterns from the Learned Round (LR) and the preceding round (LR-1),  
141 pairmate similarity scores at the pre-IP were based on correlations shifted back one step in time: i.e.,  
142 between LR-1 and LR-2. Thus, whereas the IP captured the transition from pre-learned to learned, the pre-  
143 IP was an important reference point that corresponded to a ‘non-transition’ (pre-learned to pre-learned).

144  
145 An ANOVA with factors of behavioral state (pre-IP, IP) and ROI (CA1, CA23DG, PPA, EVC) revealed a  
146 significant main effect of ROI ( $F_{3,90} = 4.08$ ,  $p = 0.009$ ,  $\eta^2 = 0.04$ ), reflecting overall differences in pairmate  
147 similarity scores across ROIs. Scores were numerically lowest in CA23DG and numerically highest in EVC.  
148 There was no main effect of behavioral state ( $F_{1,30} = 2.71$ ,  $p = 0.110$ ,  $\eta^2 = 0.01$ ), indicating that learning did  
149 not have a global effect on representational structure across ROIs. Critically, however, the interaction  
150 between behavioral state and ROI was significant ( $F_{3,90} = 2.95$ ,  $p = 0.037$ ,  $\eta^2 = 0.04$ ), indicating that learning  
151 differentially influenced pairmate similarity scores across ROIs.

152  
153 Within CA23DG, pairmate similarity scores were significantly lower at the IP than the pre-IP ( $t_{30} = -2.24$ ,  $p$   
154  $= 0.033$ ,  $d = 0.40$ ,  $CI = [-0.012 \pm 0.011]$ ), consistent with our prediction that remapping would specifically  
155 occur at the behavioral inflection point. Importantly, we also confirmed via permutation test (see Methods)  
156 that CA23DG pairmate similarity scores at the IP were lower than would be expected if the mapping  
157 between pairmates and IP’s was shuffled within participants ( $p = 0.013$ , one-tailed; **Fig. 2d**).

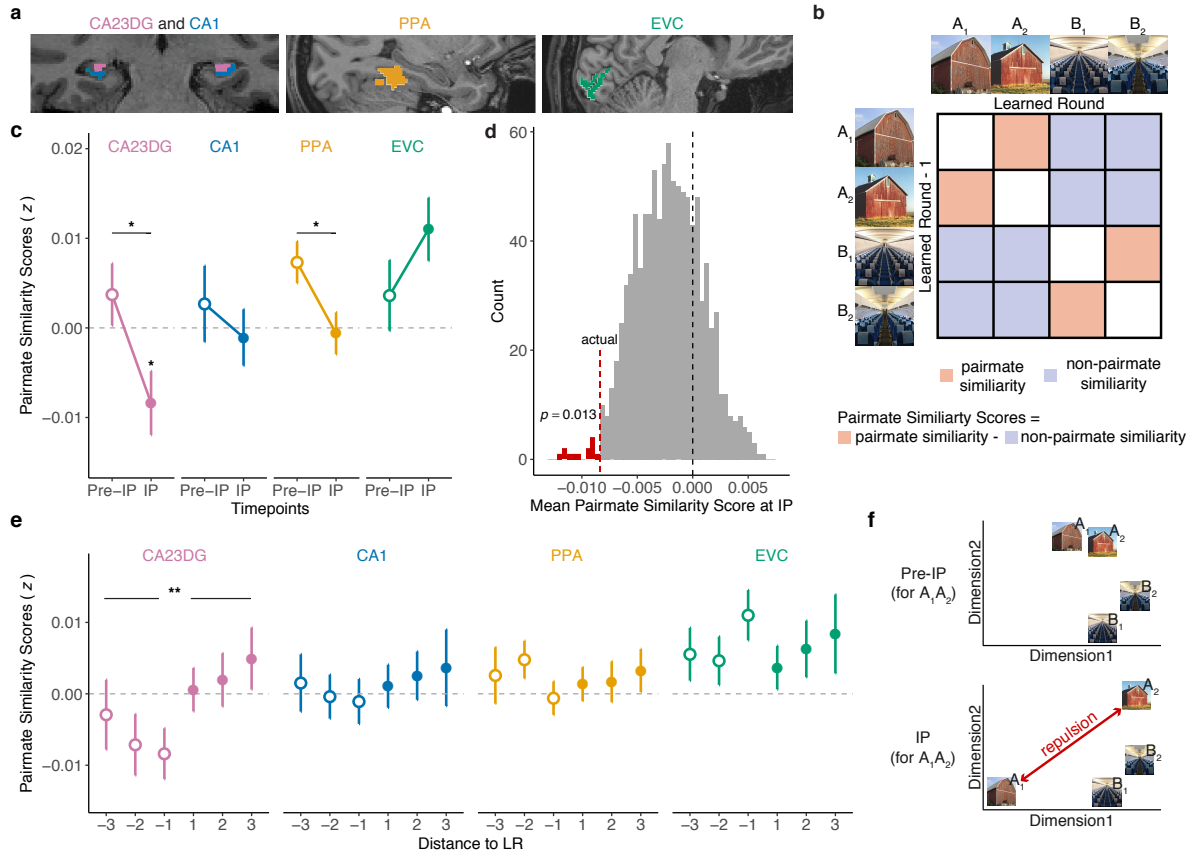
158  
159 Strikingly, CA23DG pairmate similarity scores not only decreased at the IP, but they were significantly *below*  
160 *0* at the IP ( $t_{30} = -2.36$ ,  $p = 0.025$ ,  $d = 0.19$ ,  $CI = [-0.008 \pm 0.007]$ ). In other words, pairs of scenes with  
161 extremely high visual similarity were represented as *less similar* than completely unrelated scenes in  
162 CA23DG. While seemingly counterintuitive, several recent fMRI studies have also found that, in certain  
163 situations, hippocampal pattern similarity is lower for similar than dissimilar events<sup>23,28,32</sup>. This has led to  
164 the proposal that similarity triggers a *repulsion* of hippocampal representations. That is, just as physical  
165 proximity triggers repulsion of like magnetic poles, representational proximity triggers repulsion of similar

166 memories (**Fig. 2f**). The present results, however, provide critical new evidence that this repulsion is time-  
167 locked to—and may, in fact, underlie—the resolution of interference between competing memories.

168  
169 In CA1, pairmate similarity scores did not significantly differ by learning state ( $t_{30} = -0.72$ ,  $p = 0.474$ ,  $d =$   
170  $0.13$ ,  $CI = [0.004 \pm 0.01]$ ) or differ from 0 either at the pre-IP ( $t_{30} = -0.63$ ,  $p = 0.531$ ,  $d = 0.11$ ,  $CI = [0.003 \pm$   
171  $0.009]$ ) or IP ( $t_{30} = -0.34$ ,  $p = 0.735$ ,  $d = 0.06$ ,  $CI = [-0.001 \pm 0.006]$ ). In PPA, pairmate similarity scores  
172 decreased from pre-IP to IP ( $t_{30} = -2.28$ ,  $p = 0.030$ ,  $d = 0.41$ ,  $CI = [0.008 \pm 0.007]$ ), with scores significantly  
173 greater than 0 in the pre-IP ( $t_{30} = 3.14$ ,  $p = 0.004$ ,  $d = 0.56$ ,  $CI = [0.007 \pm 0.005]$ ) but not different from 0 at  
174 the IP ( $t_{30} = -0.26$ ,  $p = 0.798$ ,  $d = 0.05$ ,  $CI = [-0.0006 \pm 0.005]$ ). In EVC, pairmate similarity scores did not  
175 significantly vary by learning state ( $t_{30} = -1.39$ ,  $p = 0.175$ ,  $d = 0.25$ ,  $CI = [-0.007 \pm 0.01]$ ); but there was a  
176 numerical increase from pre-IP to IP, with scores significantly above 0 at IP ( $t_{30} = 3.13$ ,  $p = 0.004$ ,  $d = 0.56$ ,  
177  $CI = [0.01 \pm 0.007]$ ) but not at pre-IP ( $t_{30} = 0.92$ ,  $p = 0.366$ ,  $d = 0.16$ ,  $CI = [0.004 \pm 0.008]$ ).

178  
179 The qualitative difference between CA23DG and EVC is striking in that, at the inflection point, these regions  
180 exhibited fully opposite representational structures: scene pairmates were *more similar* than non-pairmates  
181 in EVC, but *less similar* than non-pairmates in CA23DG. This finding parallels prior evidence of opposite  
182 representational structures in hippocampus and EVC<sup>28,32</sup> and argues against the possibility that CA23DG  
183 ‘inherited’ representational structure from early visual regions. More generally, it is striking that differences  
184 in pairmate similarity scores markedly varied across the four ROIs at the IP ( $F_{3,90} = 8.73$ ,  $p < 0.001$ ,  $\eta^2 =$   
185  $0.14$ ), but not at the pre-IP ( $F_{3,90} = 0.33$ ,  $p = 0.804$ ,  $\eta^2 = 0.008$ ), underscoring the influence of learning on  
186 representational structure.

187  
188 For the preceding fMRI analyses, the IP was defined as the correlation between the learned round (LR)  
189 and the immediately preceding round (LR-1). To more fully characterize how the representational state at  
190 the LR compared to other rounds, we additionally correlated representations at LR to representations at  
191 LR-2 and LR-3 (i.e., other rounds that preceded the LR) and also correlated LR with LR+1, LR+2, and LR+3  
192 (rounds that followed the LR) (**Fig. 2e**). Within CA23DG, pairmate similarity scores were significantly lower  
193 when correlating the LR with rounds that preceded learning compared to rounds that followed learning ( $t_{30}$   
194  $= -2.98$ ,  $p = 0.006$ ,  $d = 0.54$ ,  $CI = [-0.009 \pm 0.006]$ ). This striking asymmetry indicates that CA23DG  
195 representations expressed at the LR were systematically biased away from the initial representational  
196 position of competing memories. More generally, these data support the idea of an abrupt representational  
197 change (remapping) in CA23DG that was time-locked to the specific round at which learning occurred for  
198 individual pairmates. For CA1, PPA, and EVC, there were no significant differences in pairmate similarity  
199 scores when correlating the LR to rounds that preceded learning vs. followed learning ( $|t_{30}| \leq 0.79$ ,  $p$ 's  $\geq$   
200  $0.435$ ,  $d \leq 0.14$ ; **Fig. 2e**).



**Figure 2. Pairmate similarity scores change at the behavioral inflection point.** **a.** Regions of interest included CA23DG and CA1 in the hippocampus, the parahippocampal place area (PPA), and early visual cortex (EVC). **b.** Correlation matrix illustrating how pairmate similarity scores were computed for the behavioral inflection point. **c.** Pairmate similarity scores at the behavioral inflection point (IP) and just prior to the inflection point (pre-IP) across different regions of interest (ROIs). Pairmate similarity scores significantly varied by ROI ( $p = 0.009$ ) and there was a significant interaction between ROIs and behavioral state ( $p = 0.011$ ). **d.** A permutation test (1,000 iterations) was performed by shuffling, within participants, the mapping between the behavioral inflection point and scene pairmates. In CA23DG the actual mean group-level pairmate similarity score at the IP was lower than 98.70% of the permuted mean similarity scores. **e.** Pairmate similarity scores calculated by correlating the learned round (LR) with each of the three preceding rounds ( $-$  distance to LR) and each of the three succeeding rounds ( $+$  distance to LR). In CA23DG, pairmate similarity scores were significantly lower when the LR was correlated with preceding round compared to succeeding rounds ( $p = 0.006$ ). The difference was not significant for any other ROIs ( $p$ 's  $\geq 0.435$ ). **f.** Conceptual illustration of a decrease in pairmate similarity scores from pre-IP to IP. In the pre-IP state (top panel), A<sub>1</sub> and A<sub>2</sub> are nearby in representational space. In the IP state (bottom panel), the representational distance between A<sub>1</sub> and A<sub>2</sub> has been exaggerated. When pairmates (e.g., A<sub>1</sub> and A<sub>2</sub>) are farther apart in representational space than non-pairmates (e.g., A<sub>1</sub> and B<sub>2</sub>) the pairmate similarity score will be *negative* (i.e., pairmate similarity < non-pairmate similarity), consistent with a repulsion of competing representations. Notes: \*  $p < .05$ , \*\*  $p < .01$ , error bars reflect S.E.M.

## 201 **Overlap of CA23DG representations triggers remapping.**

202

203 The fact that pairmate similarity scores in CA23DG were negative at the IP (**Fig. 2c**) suggests that learning-  
 204 related remapping involved an active repulsion of competing hippocampal representations (**Fig. 2f**).  
 205 Conceptually, the key feature of a repulsion account is that separation of hippocampal representations is a  
 206 *reaction* to initial overlap among memories<sup>33</sup>. Here, because we measured representational states  
 207 throughout the course of learning, we were able to test this hypothesis directly. Specifically, we tested the

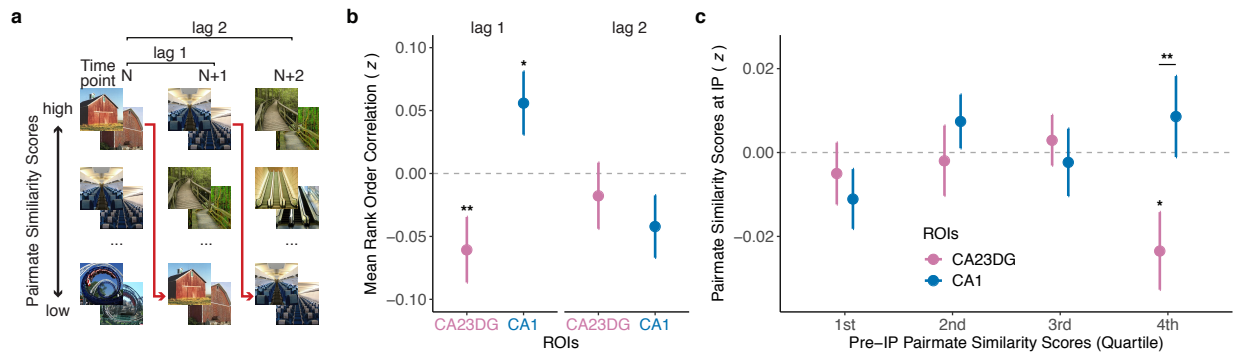


208 prediction that relatively greater pairmate similarity scores (i.e., higher overlap between memories) at a  
209 given timepoint is associated with relatively *lower* pairmate similarity scores (i.e., lower overlap between  
210 memories) at a successive timepoint.

211  
212 To test this hypothesis, we first translated the 6 learning rounds into 5 ‘timepoints’ (see Methods). Each  
213 timepoint corresponded to the set of scene pair similarity scores obtained by correlating activity patterns  
214 across consecutive learning rounds [e.g., timepoint 1 =  $r(\text{round } 1, \text{round } 2)$ ]. These scores reflected the  
215 representational structure at each timepoint (i.e., which pairmates were relatively similar, which pairmates  
216 were relatively dissimilar). We then rank correlated the pairmate similarity scores across successive  
217 timepoints [ $r(\text{timepoint } 1, \text{timepoint } 2)$ ]. Whereas a positive rank correlation would indicate that  
218 representational structure is preserved across time points, a negative rank correlation would indicate that  
219 representational structure is *inverted* across time points. Critically, an inversion of representational structure  
220 is precisely what would be predicted if initial overlap among activity patterns (i.e., high pairmate similarity  
221 scores) triggers a repulsion of activity patterns (i.e., low pairmate similarity scores).

222  
223 Strikingly, the rank correlation in CA23DG was significantly negative ( $t_{30} = -2.99, p = 0.006, d = 0.54, CI =$   
224  $[-0.06 \pm 0.04]$ ). In contrast, the rank correlation in CA1 was significantly positive ( $t_{30} = 2.11, p = 0.043, d =$   
225  $0.38, CI = [0.06 \pm 0.05]$ ). The difference between CA23DG and CA1 was also significant ( $t_{30} = 3.73, p <$   
226  $0.001, d = 0.67, CI = [0.12 \pm 0.06]$ ). Importantly, the negative correlation in CA23DG cannot be explained  
227 by regression to the mean (see Methods). Moreover, when we tested correlations at a lag of 2 [ $r(\text{timepoint}$   
228  $N, \text{timepoint } N+2)$ ], correlations did not significantly differ from 0 for either CA23DG ( $t_{30} = -0.71, p = 0.485,$   
229  $d = 0.13, CI = [-0.02 \pm 0.05]$ ) or CA1 ( $t_{30} = -1.60, p = 0.120, d = 0.29, CI = [-0.04 \pm 0.05]$ ). Further, the  
230 interaction between lag (1, 2) and ROI (CA23DG, CA1) was also significant ( $F_{1,30} = 7.09, p = 0.012, \eta^2 =$   
231  $0.06$ ), indicating that the dissociation between CA23DG and CA1 was relatively stronger at lag 1  
232 (consecutive timepoints) than lag 2 (non-consecutive timepoints). Thus, representational structure at a  
233 given time point specifically predicted representational structure at a successive timepoint. Rank  
234 correlations did not differ from 0 in either PPA or EVC, either for lag 1 or lag 2 ( $|t_{30}|$ 's  $\leq 1.12, p$ 's  $\geq 0.272,$   
235  $d$ 's  $\leq 0.20$ ).

236  
237 While the negative correlation in CA23DG was fully consistent with our prediction—and with the idea that  
238 high pattern overlap triggers repulsion—the negative correlation could alternatively be explained by  
239 pairmates with relatively low pairmate similarity at timepoint N tending to have relatively high similarity at  
240 timepoint N+1. Additionally, because our analysis was entirely agnostic to behavioral data, it does not  
241 specifically establish that the negative pairmate similarity scores that we observed at the behavioral IP (**Fig.**  
242 **2c** and **2e**) were triggered by pattern overlap at IP-1. Thus, as a complementary analysis, we binned all  
243 pairmates, by quartiles, according to pairmate similarity scores at IP-1, with the 4<sup>th</sup> quartile representing  
244 pairmates with the highest pairmate similarity scores. We then computed the mean pairmate similarity  
245 scores for those bins at the IP. Again, this analysis was separately performed for CA23DG and CA1. An  
246 ANOVA with factors of ROI (CA23DG, CA1) and pairmate similarity scores at IP-1 (4 quartiles) revealed a  
247 significant interaction ( $F_{3,90} = 3.19, p = 0.027, \eta^2 = 0.03$ ). Critically, this interaction was driven by a marked  
248 difference between CA23DG and CA1 when considering the bin with the *highest overlap* at IP-1 (i.e., 4th  
249 quartile:  $t_{30} = -2.87, p = 0.008, d = 0.51, CI = [-0.03 \pm 0.02]$ , **Fig. 3c**). For CA23DG, pairmate similarity  
250 scores at the IP were significantly below 0 and numerically lowest for pairmates whose similarity scores at  
251 IP-1 were in the 4<sup>th</sup> quartile (comparison to 0:  $t_{30} = -2.54, p = 0.017, d = 0.46, CI = [-0.023 \pm 0.019]$ ); the  
252 pattern in CA1 was qualitatively opposite. Collectively, these results provide novel, theory-consistent  
253 evidence that remapping of competing representations is actively triggered by initial representational  
254 overlap.



### 255 CA23DG scene representations differentiate between competing object associations.

256

257 Thus far, we have focused on similarity among neural representations evoked while viewing the scene  
 258 images (scene exposure phase). However, our paradigm also included two fMRI runs during which  
 259 participants viewed each of the objects associated with the scene images (object exposure phase; see  
 260 Methods). This allowed us to test whether hippocampal activity patterns evoked while viewing the scenes  
 261 resembled—or came to resemble—activity patterns evoked while viewing corresponding object images.

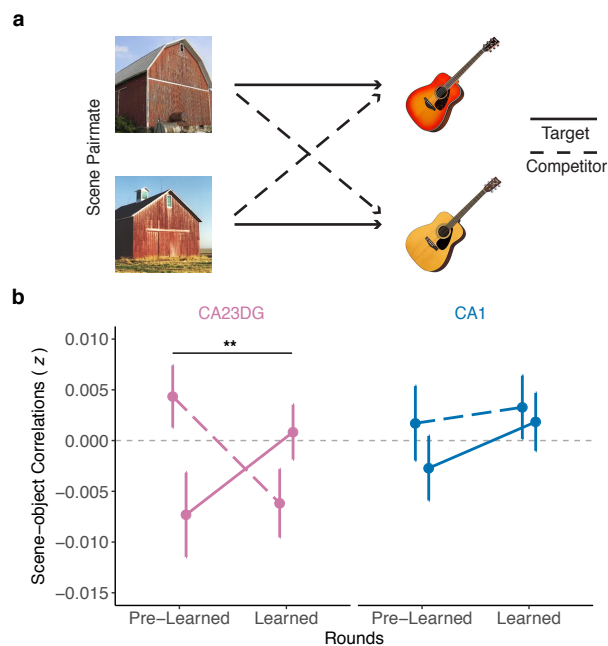
262

263 Whereas, pairmate similarity scores were computed by correlating activity patterns across rounds of the  
 264 scene exposure phase (e.g., LR-1 and LR), here we computed correlations between a single round of the  
 265 scene exposure phase (e.g., LR) and the average of the two object rounds (see Methods). For this analysis,  
 266 there were three important factors that we considered. First, we considered whether scene representations  
 267 were in a ‘pre-learned’ state (LR-1) or ‘learned’ state (LR). Second, we separately tested correlations  
 268 between each scene and (a) the target object (e.g., ‘guitar 1’) vs. (b) the competing object (e.g., ‘guitar 2’)  
 269 (Fig. 4a). Third, we again compared results in CA23DG vs. CA1.

270

271 A repeated measures ANOVA with factors of ROI (CA23DG, CA1), behavioral state (pre-learned, learned),  
 272 and object relevance (target, competitor) revealed a significant interaction between behavioral state and  
 273 object relevance ( $F_{1,30} = 12.42$ ,  $p = 0.001$ ,  $\eta^2 = 0.02$ ). Qualitatively, this interaction reflected a learning-  
 274 related change wherein hippocampal representations of scene images became relatively more similar to

275 target objects and less similar to competitor objects. However, this 2-way interaction between behavioral  
276 state and object relevance was qualified by a trend toward a 3-way interaction between behavioral state,  
277 object relevance, and ROI ( $F_{1,30} = 4.07$ ,  $p = 0.053$ ,  $\eta^2 = 0.01$ ). Specifically, the interaction between  
278 behavioral state (pre-learned, learned) and object relevance (target, competitor) was significant in CA23DG  
279 ( $F_{1,30} = 11.98$ ,  $p = 0.002$ ,  $\eta^2 = 0.06$ ) but not in CA1 ( $F_{1,30} = 0.44$ ,  $p = 0.510$ ,  $\eta^2 = 0.002$ ) (**Fig. 4b**). For  
280 CA23DG, there was a qualitative increase, from the pre-learned to learned state, in similarity between  
281 scenes and target objects and a qualitative decrease, from the pre-learned to learned state, in similarity  
282 between scenes and competing objects. In other words, the remapping of CA23DG scene representations  
283 that occurred at the learned round yielded a relative strengthening of information related to target object  
284 associations and a relative weakening of information related to competing object associations. This  
285 dissociation in CA23DG is striking when considering that target and competitor objects were extremely  
286 similar (see **Fig. 1a**, **Fig. 4a**) and even more so when considering that during the scene and object exposure  
287 phases participants were not instructed or required in any way to recall the corresponding images. The 2-  
288 way interaction between behavioral state and object relevance was not significant for PPA or EVC [ $F_{1,30}$ 's  
289  $\leq 3.23$ ,  $p$ 's  $\geq 0.082$ ,  $\eta^2$ 's  $\leq 0.02$ ].



**Figure 4. Scene-object similarity as a function of behavioral state.** **a.** Example associations between scene pairmates and objects. Scene-object similarity was calculated by correlating activity patterns evoked during the scene exposure phases (at different behavioral states) and the object exposure phases. Target similarity refers to correlations between a given scene and the object with which it was studied. Competitor similarity refers to correlations between a given scene and the object with which its pairmate was studied. **b.** Scene-object similarity as a function of object relevance (target, competitor), ROI (CA23DG, CA1), and behavioral state (pre-learned, learned). Correlations between unrelated scenes and objects (across pairmate similarity; not shown) was subtracted from target and competitor similarity values. For CA23DG, there was a significant interaction between behavioral state and object relevance ( $p = 0.002$ ). Notes: \*\*  $p < .01$ , error bars reflect S.E.M.

290 **DISCUSSION:**

291

292 Here, we show that learning to discriminate competing episodic memories is associated with an abrupt  
293 remapping of activity patterns in CA3/dentate gyrus. Specifically, fMRI pattern similarity in CA3/dentate  
294 gyrus decreased precisely when behavioral expressions of learning emerged. Additionally, the degree to  
295 which remapping occurred in CA3/dentate gyrus was predicted by the degree of initial pattern overlap  
296 among competing memories. Finally, remapped CA3/dentate gyrus representations contained relatively  
297 stronger information about relevant episodic associations and relatively weaker information about  
298 competing episodic associations, confirming the learning-related significance of the remapping effect.

299

300 Our findings complement recent demonstrations of remapping-like phenomena in the human  
301 hippocampus<sup>34,35</sup> as well as evidence of abrupt remapping in the rodent hippocampus<sup>9–12</sup>. However, our  
302 findings provide unique and direct support for the proposal that hippocampal remapping is associated with  
303 the resolution of human episodic memory interference<sup>8</sup>. Specifically, we demonstrate an abrupt transition  
304 in hippocampal representations that occurred at an important inflection point in learning—the point at which  
305 participants were able to correctly discriminate similar memories and retrieve associations with high  
306 confidence. Notably, this finding was only possible because (a) we repeatedly probed episodic memory and  
307 hippocampal representations over the course of learning and (b) we identified inflection points in a  
308 participant- and pairmate-specific manner. Indeed, inflection points varied considerably across and within  
309 participants (**Fig. 1d** and **Sup. table 1**) and the observed hippocampal remapping effect was significantly  
310 weaker when the specific mapping between behavior and fMRI data was shuffled within participants (**Fig.**  
311 **2d**).

312

313 The fact that CA23DG remapping occurred precisely at the inflection point in learning strongly suggests  
314 that remapping was related to learning. This argument is also reinforced by our independent finding that  
315 remapped CA23DG activity patterns, evoked while participants viewed individual scene images, carried  
316 more information (compared to the pre-learning state) about target versus competing object associations.  
317 In other words, the inflection point defined from behavioral expressions of associative memory also  
318 captured a critical change in associative representations encoded in CA23DG activity patterns. The fact  
319 that CA23DG exaggerated the representational distance between competing scenes (remapping) while  
320 simultaneously reflecting learned associations (scene-object similarity) is consistent with the idea that CA3  
321 balances both pattern separation and pattern completion mechanisms<sup>4,17,36,37</sup>. The fact that remapped  
322 activity patterns contained information about learned associations is also consistent with the argument that  
323 hippocampal remapping does not simply reflect changes in the external environment—which did not change  
324 over the course of the experiment—but instead fundamentally reflects changes in internal models of the  
325 environment<sup>14,15</sup>.

326

327 One aspect of our findings which does not, to our knowledge, have a direct analog in rodent studies of  
328 remapping is the negative pairmate similarity score we observed at the inflection point in CA23DG. The  
329 negative score indicates that scene pairmates—which were extremely similar images—were associated  
330 with *less overlapping* CA23DG representations than completely unrelated scenes. In rodents, the most  
331 extreme version of remapping occurs when two similar environments are associated with fully independent  
332 place codes<sup>8</sup>. In our study, however, if each scene was associated with an independent representation,  
333 then the similarity between pairmates would be equal to, but not lower than, the similarity between non-  
334 pairmates. Instead, the negative pairmate similarity score requires a *dependence* between competing  
335 hippocampal representations wherein a given memory representation systematically moves away from the  
336 representational position of a competing memory (**Fig. 2f**). We refer to this dependence as ‘repulsion’ in

337 order to emphasize the oppositional influence that competing memories exerted. Several recent human  
338 fMRI studies have reported conceptually similar effects in the hippocampus<sup>28,32,38</sup>—and in CA3/dentate  
339 gyrus, specifically<sup>22–26</sup>. However, the current findings are the first to directly establish that the repulsion of  
340 competing hippocampal representations is temporally coupled to the resolution of memory interference.

341  
342 Based on computational models<sup>33,39,40</sup>, our prediction was that the repulsion effect in CA23DG was a direct  
343 consequence of initial overlap among activity patterns. Indeed, a recent study found that hippocampal  
344 repulsion was more likely to occur for behaviorally-confusable memories<sup>32</sup>, potentially because confusable  
345 memories are associated with greater pattern overlap during initial learning. In the current study, we  
346 tested—and confirmed—this account directly. Specifically, we found that the representational structure  
347 (relative pairmate similarity) in CA23DG at a given timepoint was *negatively correlated* with representational  
348 structure at an immediately following timepoint. This negative relationship is highly consistent with the idea  
349 that overlap, itself, triggers plasticity that ‘punishes’ those features which are shared across  
350 memories<sup>24,33,39,40</sup>. While our study does not afford inferences about the causal relationship between  
351 repulsion and learning, the idea that repulsion (or remapping more generally) is triggered by  
352 representational overlap, combined with the fact that remapping was associated with learning, is consistent  
353 with the possibility that repulsion of CA3/dentate gyrus representations is a causal factor in learning.

354  
355 Across multiple analyses, we observed dissociations between CA3/dentate gyrus and CA1. The fact that  
356 the remapping effects were selective to CA3/dentate gyrus is consistent with evidence from rodent studies  
357 of remapping and pattern separation<sup>8,16,36</sup> and with several human fMRI studies<sup>22–25,36</sup>. Perhaps the most  
358 striking dissociation between CA23DG and CA1 comes from our analysis of representational structure  
359 across time points. Whereas CA23DG exhibited a negative rank correlation across successive timepoints,  
360 CA1 exhibited a positive rank correlation (**Fig. 3b**). Thus, in contrast to CA23DG, CA1 was characterized  
361 by stability (though only modest stability) of representational structure across timepoints<sup>4</sup>. This dissociation  
362 between CA23DG and CA1 is consistent with the idea that CA3, in particular, supports rapid plasticity that  
363 allows for changes in memory representations on short time scales<sup>41</sup> and is also consistent with evidence  
364 of faster remapping in CA3/dentate gyrus than in CA1<sup>10,12,21</sup>. It is also notable that the remapping effect we  
365 observed in CA23DG at the inflection point in learning strongly contrasted with the pattern of data in early  
366 visual cortex. Whereas CA23DG exhibited a negative pairmate similarity score at the inflection point, EVC  
367 exhibited a significant, positive pairmate similarity score at the inflection point. This finding makes the  
368 important point that CA23DG was not inheriting representational structure from early sensory regions (e.g.,  
369 due to visual attention)—rather, CA23DG fully inverted the representational structure that was expressed  
370 in early visual cortex<sup>28</sup>.

371  
372 Taken together, our findings constitute novel evidence for a remapping of human CA3/dentate gyrus  
373 representations that is temporally-coupled to the resolution of episodic memory interference. These findings  
374 were motivated by—and complement—existing evidence of remapping in the rodent hippocampus. Yet,  
375 our findings also go beyond existing rodent or human studies by establishing a direct link between  
376 remapping and changes in internal memory states<sup>14,15</sup>. Additionally, our conclusion that overlap among  
377 CA3/dentate gyrus representations actively triggers a repulsion of memory representations has important  
378 implications for theoretical accounts of how the hippocampus resolves memory interference<sup>5,8,36,39</sup> and will  
379 hopefully inspire targeted new analyses that test for similar mechanisms in rodent models.

380 **REFERENCES:**

- 381 1. Eichenbaum, H. A cortical–hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41–50  
382 (2000).
- 383 2. Squire, L. & Zola-Morgan, S. The medial temporal lobe memory system. *Science* **253**, 1380–1386  
384 (1991).
- 385 3. O’Keefe, J. & Nadel, L. *The hippocampus as a cognitive map*. (Clarendon Press ; Oxford University  
386 Press, 1978).
- 387 4. Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M. & Norman, K. A. Complementary learning  
388 systems within the hippocampus: a neural network modelling approach to reconciling episodic memory  
389 with statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160049 (2017).
- 390 5. O’Reilly, R. C. & Norman, K. A. Hippocampal and neocortical contributions to memory: advances in the  
391 complementary learning systems framework. *Trends Cogn. Sci.* **6**, 505–510 (2002).
- 392 6. Bostock, E., Muller, R. U. & Kubie, J. L. Experience-dependent modifications of hippocampal place cell  
393 firing. *Hippocampus* **1**, 193–205 (1991).
- 394 7. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of  
395 hippocampal complex-spike cells. *J. Neurosci. Off. J. Soc. Neurosci.* **7**, 1951–1968 (1987).
- 396 8. Colgin, L. L., Moser, E. I. & Moser, M.-B. Understanding memory through hippocampal remapping.  
397 *Trends Neurosci.* **31**, 469–477 (2008).
- 398 9. Colgin, L. L. *et al.* Attractor-Map Versus Autoassociation Based Attractor Dynamics in the  
399 Hippocampal Network. *J. Neurophysiol.* **104**, 35–50 (2010).
- 400 10. Leutgeb, S., Leutgeb, J. K., Moser, E. I. & Moser, M.-B. Fast rate coding in hippocampal CA3 cell  
401 ensembles. *Hippocampus* **16**, 765–774 (2006).
- 402 11. Wills, T. J. Attractor Dynamics in the Hippocampal Representation of the Local Environment.  
403 *Science* **308**, 873–876 (2005).
- 404 12. Lee, I., Rao, G. & Knierim, J. J. A Double Dissociation between Hippocampal Subfields:  
405 Differential Time Course of CA3 and CA1 Place Cells for Processing Changed Environments. *Neuron*  
406 **42**, 803–815 (2004).

- 407 13. Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. Long-term plasticity in hippocampal  
408 place-cell representation of environmental geometry. *Nature* **416**, 90–94 (2002).
- 409 14. Sanders, H., Wilson, M. A. & Gershman, S. J. Hippocampal remapping as hidden state inference.  
410 *eLife* **9**, e51140 (2020).
- 411 15. Keinath, A. T., Nieto-Posadas, A., Robinson, J. C. & Brandon, M. P. DG–CA3 circuitry mediates  
412 hippocampal representations of latent information. *Nat. Commun.* **11**, 3026 (2020).
- 413 16. Duncan, K. D. & Schlichting, M. L. Hippocampal representations as a function of time, subregion,  
414 and brain state. *Neurobiol. Learn. Mem.* **153**, 40–56 (2018).
- 415 17. Guzowski, J. F., Knierim, J. J. & Moser, E. I. Ensemble Dynamics of Hippocampal Regions CA3  
416 and CA1. *Neuron* **44**, 581–584 (2004).
- 417 18. McHugh, T. J. *et al.* Dentate Gyrus NMDA Receptors Mediate Rapid Pattern Separation in the  
418 Hippocampal Network. *Science* **317**, 94–99 (2007).
- 419 19. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B. & Moser, E. I. Distinct Ensemble Codes in  
420 Hippocampal Areas CA3 and CA1. *Science* **305**, 1295–1298 (2004).
- 421 20. Vazdarjanova, A. & Guzowski, J. F. Differences in Hippocampal Neuronal Population Responses  
422 to Modifications of an Environmental Context: Evidence for Distinct, Yet Complementary, Functions of  
423 CA3 and CA1 Ensembles. *J. Neurosci.* **24**, 6489–6496 (2004).
- 424 21. van Dijk, M. T. & Fenton, A. A. On How the Dentate Gyrus Contributes to Memory Discrimination.  
425 *Neuron* **98**, 832-845.e5 (2018).
- 426 22. Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A. & Preston, A. R. Memory reactivation  
427 during learning simultaneously promotes dentate gyrus/CA2,3 pattern differentiation and CA1 memory  
428 integration. *J. Neurosci.* (2020) doi:10.1523/JNEUROSCI.0394-20.2020.
- 429 23. Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P. & Ranganath, C. CA1 and  
430 CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal  
431 subfields. *Nat. Commun.* **9**, 294 (2018).
- 432 24. Kim, G., Norman, K. A. & Turk-Browne, N. B. Neural Differentiation of Incorrectly Predicted  
433 Memories. *J. Neurosci.* **37**, 2022–2031 (2017).

- 434 25. Copara, M. S. *et al.* Complementary Roles of Human Hippocampal Subregions during Retrieval  
435 of Spatiotemporal Context. *J. Neurosci.* **34**, 6834–6842 (2014).
- 436 26. Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. Shaping of Object Representations in the  
437 Human Medial Temporal Lobe Based on Temporal Regularities. *Curr. Biol.* **22**, 1622–1627 (2012).
- 438 27. Bakker, A., Kirwan, C. B., Miller, M. & Stark, C. E. L. Pattern separation in the human  
439 hippocampal CA3 and dentate gyrus. *Science* **319**, 1640–1642 (2008).
- 440 28. Favila, S. E., Chanales, A. J. H. & Kuhl, B. A. Experience-dependent hippocampal pattern  
441 differentiation prevents interference during subsequent learning. *Nat. Commun.* **7**, 11066 (2016).
- 442 29. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems  
443 neuroscience. *Front. Syst. Neurosci.* (2008) doi:10.3389/neuro.06.004.2008.
- 444 30. Mumford, J. A., Davis, T. & Poldrack, R. A. The impact of study design on pattern estimation for  
445 single-trial multivariate pattern analysis. *NeuroImage* **103**, 130–138 (2014).
- 446 31. Leutgeb, J. K., Leutgeb, S., Moser, M.-B. & Moser, E. I. Pattern Separation in the Dentate Gyrus  
447 and CA3 of the Hippocampus. *Science* **315**, 961–966 (2007).
- 448 32. Chanales, A. J. H., Oza, A., Favila, S. E. & Kuhl, B. A. Overlap among Spatial Memories Triggers  
449 Repulsion of Hippocampal Representations. *Curr. Biol.* **27**, 2307-2317.e5 (2017).
- 450 33. Hulbert, J. C. & Norman, K. A. Neural Differentiation Tracks Improved Recall of Competing  
451 Memories Following Interleaved Study and Retrieval Practice. *Cereb. Cortex* **25**, 3994–4008 (2015).
- 452 34. Kyle, C. T., Stokes, J. D., Lieberman, J. S., Hassan, A. S. & Ekstrom, A. D. Successful retrieval of  
453 competing spatial environments in humans involves hippocampal pattern separation mechanisms.  
454 *eLife* **4**, e10499 (2015).
- 455 35. Steemers, B. *et al.* Hippocampal Attractor Dynamics Predict Memory-Based Decision Making.  
456 *Curr. Biol.* **26**, 1750–1757 (2016).
- 457 36. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–  
458 525 (2011).
- 459 37. Hindy, N. C., Ng, F. Y. & Turk-Browne, N. B. Linking pattern completion in the hippocampus to  
460 predictive coding in visual cortex. *Nat. Neurosci.* **19**, 665–667 (2016).



- 461 38. Jiang, J., Wang, S.-F., Guo, W., Fernandez, C. & Wagner, A. D. Prefrontal reinstatement of  
462 contextual task demand is predicted by separable hippocampal patterns. *Nat. Commun.* **11**, 2053  
463 (2020).
- 464 39. Ritvo, V. J. H., Turk-Browne, N. B. & Norman, K. A. Nonmonotonic Plasticity: How Memory  
465 Retrieval Drives Learning. *Trends Cogn. Sci.* **23**, 726–742 (2019).
- 466 40. Norman, K. A., Newman, E. L. & Detre, G. A neural network model of retrieval-induced forgetting.  
467 *Psychol. Rev.* **114**, 887–953 (2007).
- 468 41. Rebola, N., Carta, M. & Mulle, C. Operation and plasticity of hippocampal CA3 circuits:  
469 implications for memory encoding. *Nat. Rev. Neurosci.* **18**, 208–220 (2017).
- 470

471 **METHODS:**

472

473 **Participants.**

474 Thirty-six participants (21 female; mean age = 23.69 yrs, range = 18 – 34 yrs) were enrolled in the  
475 experiment following procedures approved by the University of Oregon Institutional Review Board. All  
476 participants were right-handed native-English speakers with normal or corrected-to-normal vision, with no  
477 self-reported psychiatric or neurological disease. One participant was excluded due to excess motion in the  
478 scanner (max FD > 3.5 mm); another 4 participants were excluded due to low behavioral performance (see  
479 Results for more details). The final analysis included 31 participants. All participants received monetary  
480 compensation for participating.

481

482 **Stimuli.**

483 Thirty-six images of scenes and 36 images of everyday objects were used in the experiment. The set of 36  
484 scenes and the set of 36 objects were each comprised of 18 ‘pairmates’ of visually and semantically similar  
485 images (**Fig. 1a**). An additional 36 scenes and 12 objects were used as lures for the scene and object  
486 exposure phases of the study, respectively. Separately for each participant, scene pairmates were  
487 randomly assigned to object pairmates (**Fig. 1a**). For example, if ‘barn 1’ was assigned to ‘guitar 1’, then  
488 ‘barn 2’ would be assigned to ‘guitar 2.’

489

490 **Experimental procedure.**

491 After providing consent and reviewing the instructions, participants entered the MRI scanner. Inside the  
492 scanner, participants completed 6 rounds of the experimental paradigm (**Fig. 1b**). The first round and the  
493 last round included 4 phases: study, test, scene exposure (scanned), and object exposure (scanned).  
494 Rounds 2–5 were the same, except they did not include the object exposure phase. Across all phases,  
495 stimuli were displayed on a grey background, projected from the back of the scanner. After exiting the  
496 scanner, participants completed a separate memory task that involved learning new scene-object  
497 associations (not reported here). The experiment was implemented in PsychoPy<sup>1</sup> and lasted approximately  
498 3 hrs, with about 2 hrs 15 min inside the scanner.

499

500 *Study Phase.* During the study phases, participants learned 36 scene-object associations, one association  
501 at a time. Each trial began with the presentation of a scene image (1000 ms), followed by a white fixation  
502 cross (200 ms), the associated object image (1000 ms) and then another white fixation cross (1200 ms)  
503 until the start of the next trial. The order in which the 36 scene-object associations were studied was  
504 randomized for each round and for each participant.

505

506 *Test Phase.* During the test phases, participants attempted to retrieve the object associated with each of  
507 the 36 scenes. Each trial began with the presentation of a scene (1000 ms), followed by a white fixation  
508 cross (200 ms), and then the presentation of two object pairmates (e.g., ‘Guitar 1’ and ‘Guitar 2’). One of  
509 the object images was the ‘target’ (i.e., the object associated with the cued scene) and the other object  
510 image was the ‘competitor’ (i.e., the object associated with the cued scene’s pairmate). Participants had a  
511 maximum of 4000 ms to select the correct object image (target) via a button box in their right hand. If no  
512 response was made, the next trial began after a white fixation cross was displayed for 1200 ms. If a  
513 response was made, a confidence rating then appeared beneath the objects and participants had a  
514 maximum of 3000 ms to indicate whether their response was a “Guess” or “Sure.” After indicating their  
515 confidence (or after time ran out), a white fixation cross appeared (1200 ms) until the start of the next trial.  
516 The location of the correct object (left or right) and the order in which each of the 36 scene-object  
517 associations were tested were randomized for each round and for each participant.

518

519 Scene Exposure Phase. During the scene exposure phases, which were conducted during fMRI scanning,  
520 participants saw 39 scene images in each of two blocks (78 scenes per round). Each block included the 36  
521 studied scenes and 3 novel lure scenes. Participants made an old/new judgment for each scene. Each trial  
522 began with the presentation of a scene image (500 ms), followed by a red fixation cross (1500 ms) which  
523 represented the response window. Participants again responded using the button box. After the red fixation  
524 cross, a white fixation cross (2000 ms) was presented until the start of the next trial. The order of the 39  
525 scene trials within each block was randomized for each block, round, and participant. Between the two  
526 blocks of 39 trials, participants performed a short odd/even judgment task (4 trials). Each odd/even trial  
527 consisted of a single-digit number displayed on the screen (500 ms), followed by a red fixation cross (1000  
528 ms) which represented the response window, and then a white fixation cross (1000 ms) until the start of the  
529 next trial.

530

531 Object Exposure Phase. The object exposure phase (conducted during fMRI scanning) was only included  
532 in the first and sixth rounds and followed an identical structure and procedure as the scene exposure phase.  
533 The only difference was that the 39 trials in each block corresponded to the 36 studied objects and 3 novel  
534 lure objects.

535

### 536 **MRI acquisition.**

537 All images were acquired on a Siemens 3T Skyra MRI system in the Lewis Center for Neuroimaging at the  
538 University of Oregon. Functional data were acquired with a T2\*-weighted echo-planar imaging sequence  
539 with partial-brain coverage that prioritized full coverage of the hippocampus and early visual cortex  
540 (repetition time = 2000 ms, echo time = 36 ms, flip angle = 90°, 72 slices, 1.7x1.7x1.7mm voxels). A total  
541 of 8 functional scans were acquired. Each functional scan comprised 177 volumes and included 10 s of  
542 lead-in time and 10 s of lead-out time at the beginning and end of each scan, respectively. The 8 functional  
543 scans corresponded to 6 rounds of the scene exposure phase (scans 1 and 3–7) and 2 rounds of the object  
544 exposure phase (scans 2 and 8). Anatomical scans included a whole-brain high-resolution T1-weighted  
545 magnetization prepared rapid acquisition gradient echo anatomical volume (1x1x1mm voxels) and a high-  
546 resolution (coronal direction) T2-weighted scan (0.43x0.43x2mm voxels) to facilitate segmentation of  
547 hippocampal subfields.

548

### 549 **Anatomical data preprocessing.**

550 Preprocessing was performed using *fMRIPrep* 1.5.0<sup>2,3</sup> (RRID:SCR\_016216), which is based  
551 on *Nipype* 1.2.2<sup>4,5</sup> (RRID:SCR\_002502). The T1-weighted (T1w) image was corrected for intensity non-  
552 uniformity (INU) with *N4BiasFieldCorrection*<sup>6</sup> (ANTs 2.2.0<sup>7</sup>, RRID:SCR\_004757), and used as the T1w-  
553 reference throughout the workflow. The T1w-reference was skull-stripped with the  
554 *antsBrainExtraction.sh* workflow (ANTs) in *Nipype*, using OASIS30ANTs as target template. Brain tissue  
555 segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the  
556 brain-extracted T1w using *fast*<sup>8</sup> (FSL 5.0.9, RRID:SCR\_002823). Volume-based spatial normalization to  
557 one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration  
558 with *antsRegistration* (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w  
559 template. ICBM 152 Nonlinear Asymmetrical template version 2009c<sup>9</sup> (RRID:SCR\_008796; TemplateFlow  
560 ID: MNI152NLin2009cAsym) was used for spatial normalization.

561

### 562 **Functional data preprocessing.**

563 For each of the 8 BOLD scans per participant, the following preprocessing was performed. First, a reference  
564 volume and its skull-stripped version were generated using *fMRIPrep*. A deformation field to correct for

565 susceptibility distortions was estimated based on two echo-planar imaging (EPI) references with opposing  
566 phase-encoding directions, using 3dQwarp, AFNI<sup>10</sup>. Based on the estimated susceptibility distortion, an  
567 unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference.  
568 The BOLD reference was then co-registered to the T1w reference using bbrregister (FreeSurfer) which  
569 implements boundary-based registration<sup>11</sup>. Co-registration was configured with six degrees of freedom.  
570 Head-motion parameters with respect to the BOLD reference (transformation matrices, and six  
571 corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering  
572 using mcflirt FSL 5.0.9<sup>12</sup>. BOLD scans were slice-time corrected using 3dTshift AFNI<sup>10</sup>(RRID:SCR\_005927).  
573 The BOLD time-series (including slice-timing correction when applied) were resampled onto their original,  
574 native space by applying a single, composite transform to correct for head-motion and susceptibility  
575 distortions. Framewise displacement (FD) confounding time-series were calculated based on  
576 the resampled BOLD time-series for each functional scan<sup>13</sup>.

577

### 578 **fMRI first-level general linear model (GLM) analyses.**

579 After *fMRIPrep* preprocessing, the first 5 volumes (10 s) of each functional scan were discarded. Then, the  
580 brain mask generated by *fMRIPrep* from the T1 anatomical image was used to perform brain extraction for  
581 each of the 8 functional scans. Each functional scan was then median centered. For the 6 scans of the  
582 scene exposure phase and 2 scans of the object exposure phase, all first level GLMs were performed in  
583 participants' native space with *FSL* using a Double-Gamma HRF with temporal derivatives, implemented  
584 with *Nipype*. GLMs were calculated using a variation of the Least Squares – Separate method<sup>14</sup>: a separate  
585 GLM was calculated for each of the 36 scenes (for scene exposure phases) or objects (for object exposure  
586 phases) across both repeats within a scan. For each GLM, there was one regressor of interest (representing  
587 a single scene or object image across its two repetitions per scan). All other trials (including lure images),  
588 framewise displacement, xyz translation and xyz rotation were represented with nuisance regressors.  
589 Additionally, a high pass filter (128 Hz) was applied for each GLM. This model resulted in 36 beta-maps  
590 per scan (one map per scene/object) which were converted to *t*-maps that represented the pattern of activity  
591 elicited by each scene/object for each scan.

592

### 593 **Regions of interest.**

594 A region of interest (ROI) for early visual cortex (EVC) was created from the probabilistic maps of Visual  
595 Topography<sup>15</sup> in the MNI space with a 0.5 threshold. This ROI was transformed into each participant's  
596 native space using inverse T1w-to-MNI non-linear transformation. For each participant, the top 300 EVC  
597 voxels were then selected by averaging the *t*-maps of all scenes and objects and then choosing the voxels  
598 with the highest *t*-statistics (i.e., the voxels most responsive to visual stimuli). An ROI for the  
599 parahippocampal place area (PPA) was created by first using an automated meta-analysis in Neurosynth  
600 with the key term "place". Then, clusters were created using voxels with a z-score > 2 based on the  
601 Neurosynth associative tests. Since these clusters were generated through an automated meta-analysis  
602 and were not anatomically exclusive to PPA, we visually inspected the results and manually selected the  
603 two largest clusters that were spatially consistent with PPA. One cluster was in the right hemisphere (voxel  
604 size = 247) and one cluster was in the left hemisphere (voxel size = 163). These clusters were combined  
605 into a single PPA mask. This mask was then transformed into each participant's native space using the  
606 inverse T1w-to-MNI transformation. For each participant, a final PPA ROI was generated by averaging the  
607 *t*-maps of all scene exposure phase scans and then selecting the 300 voxels with the highest average *t*-  
608 statistics (i.e., the most scene-responsive voxels). To create hippocampal ROIs, we used the Automatic  
609 Segmentation of Hippocampal Subfields (ASHS)<sup>16</sup> toolbox with the upenn2017 atlas to generate subfield  
610 ROIs in each participant's hippocampal body, including CA23DG—the combination of CA2, CA3 and  
611 dentate gyrus—and CA1. The most anterior and posterior slices of the hippocampal body were manually

612 determined for each participant based on the T2-weighted anatomical structure. Each participant's subfield  
613 segmentations were also manually inspected to ensure accuracy of the segmentation protocol. Then, each  
614 subfield ROI was transformed into each participant's native space using the T2-to-T1w transformation,  
615 calculated with FLIRT (fsl) with 6 degrees of freedom, implemented with *Nipype*. All ROIs were again  
616 visually inspected following the transformation to native space to ensure the ROIs were anatomically correct.

617

#### 618 **fMRI pattern similarity analyses.**

619 *Pairmate Similarity Scores.* Pattern similarity was calculated as the Fisher z-transformed Pearson  
620 correlation between *t*-maps within each ROI. All pattern similarity analyses were performed by correlating  
621 the *t*-maps for stimuli across scans (i.e., correlations were never performed within the same scan). For our  
622 primary analyses related to pattern similarity between scene images, of critical interest was mean similarity  
623 between pairmate scenes (*pairmate similarity*) relative to mean similarity between non-pairmate scenes  
624 (*non-pairmate similarity*). For example, the correlation between the *t*-maps for 'barn 1' from scan 3 and  
625 'barn 2' from scan 4 would reflect pairmate similarity, whereas the correlation between the *t*-maps for 'barn  
626 1' from scan 3 and 'airplane 2' from scan 4 would reflect non-pairmate similarity. We then calculated the  
627 mean difference between pairmate similarity and non-pairmate similarity, which we refer to as the *pairmate*  
628 *similarity score*.

629

630 *Learned Round.* To relate pairmate similarity scores to behavioral measures of learning, we identified the  
631 *Learned Round* (LR) for each pairmate, separately for each participant. The LR was based on performance  
632 in the associative memory test. Specifically, the LR was defined as the first round in which the target object  
633 was selected with high confidence for both scenes in a pairmate, with the additional requirement that  
634 performance remained stable in all subsequent rounds. It was therefore possible that both scenes in a  
635 pairmate were associated with high confidence correct responses in round N, not in round N+1, and then  
636 (again) in round N+2 and thereafter; in this case, the LR would be round N+2.

637

638 *Inflection Point.* The *inflection point* (IP) was defined as the transition from LR – 1 to LR (i.e., the transition  
639 from 'pre-learned' to 'learned'). Thus, pattern similarity analyses of the IP refer to the correlation of *t*-maps  
640 from LR-1 to *t*-maps from LR. We hypothesized that the behavioral state change from LR-1 to LR would  
641 correspond to a reduction in pattern similarity between pairmates. Pattern similarity analyses at the IP were  
642 contrasted against the 'pre-IP' state, which was based on the correlation of *t*-maps from LR-2 and LR-1  
643 (i.e., a non-transition from 'not learned' to 'not learned') (**Fig. 2c**). Pairmates for which participants never  
644 reached and sustained high-confidence correct responses (mean  $\pm$  s.d.,  $1.81 \pm 2.27$  per participant) and  
645 pairmates that were learned in the 1<sup>st</sup> round (LR = 1; mean  $\pm$  s.d.,  $1.00 \pm 1.26$ ) were excluded from the IP  
646 analysis because neither the pre-IP nor IP states could be measured. For pairmates that were learned in  
647 the 2<sup>nd</sup> round (LR = 2; mean  $\pm$  s.d.,  $3.23 \pm 2.80$ ), pattern similarity at the IP was calculated and included in  
648 the analyses, but pattern similarity at the pre-IP state could not be calculated because an LR – 2 did not  
649 exist. For rest of the pairmates (LR = 3, 4, 5, or 6), we calculated pattern similarity for both pre-IP and IP  
650 (**Fig. 1e**). Similar restrictions applied to correlations between LR and LR-3, LR + 1, LR + 2, and LR + 3 (**Fig.**  
651 **2e**). The number of pairmates included in each comparison and for each participant are reported in  
652 **Supplementary Table 1**.

653

654 *Representational Structure Across Time Points.* To test whether representational overlap triggered  
655 remapping (related to **Fig. 3**), the 6 learning rounds were translated into 5 timepoints. Each timepoint  
656 corresponded to a pair of consecutive learning rounds ([1,2], [2,3], [3,4], [4,5], [5,6]). For each timepoint,  
657 pairmate similarity scores were calculated, as described above, by correlating activity patterns from  
658 consecutive learning rounds (e.g., pairmate similarity scores at timepoint 1 were based on correlations

659 between round 1 and round 2). This yielded a set of pairmate similarity scores at each of the 5 timepoints.  
660 These sets of similarity scores reflected the representational structure at each timepoint (i.e., which  
661 pairmates were relatively similar and which pairmates were relatively dissimilar). Pairmate similarity scores  
662 were then correlated across timepoints using Spearman's rank correlation (Fisher z transformed). Lag 1  
663 correlations refer to rank correlations between successive timepoints whereas lag 2 correlations refer to  
664 correlations between timepoints two steps apart. To facilitate a direct comparison between lag 1 vs. lag 2  
665 correlations, correlations were computed for the following timepoints: Lag 1 =  $r(\text{timepoint } 1, 2)$ ,  $r(\text{timepoint } 2, 3)$ ,  $r(\text{timepoint } 3, 4)$ ; Lag 2 =  $r(\text{timepoint } 1, 3)$ ,  $r(\text{timepoint } 2, 4)$ ,  $r(\text{timepoint } 3, 5)$ . It is important to  
667 emphasize that we did not correlate initial pairmate similarity scores with the *change* in pairmate similarity  
668 as this would produce an artifactual correlation (via regression to the mean). In contrast, a negative rank  
669 correlation (as we observed in CA23DG) cannot be explained by regression to the mean. Mathematically,  
670 if all values at timepoint  $N$  *partially* regressed toward the mean at timepoint  $N+1$ , this would yield a *positive*  
671 rank correlation (i.e., representational structure would be partially preserved). If all values *fully* regressed  
672 toward the mean (i.e., variance at timepoint  $N+1 = 0$ ), this would yield a null correlation ( $r = 0$ ;  
673 representational structure fully abolished).

674  
675 *Scene-Object Similarity.* To calculate pattern similarity between scenes and objects (related to **Fig. 4**),  
676 activation patterns for objects were first generated by averaging  $t$ -maps across the two object exposure  
677 phases, resulting in a single, mean activity pattern for each object. These object-specific activity patterns  
678 were then correlated with activity patterns from the scene exposure phases at LR - 1 (i.e., the pre-learned  
679 state) and LR (i.e., the learned state). Correlations were separated into three groups: (1) *target* correlations  
680 refer to the correlation between a scene and the object it was associated with during the study phase (e.g.,  
681 'barn 1' and 'guitar 1'), (2) *competitor* correlations refer to the correlation between a scene and the object  
682 that was associated with that scene's pairmate during the study phase (e.g., 'barn 1' and 'guitar 2'), and (3)  
683 *across pairmate* correlations refer to correlations between a scene and an object that was not associated  
684 with that scene or its pairmate during the study phase (e.g., 'barn 1' and 'scissors 1'). Target and competitor  
685 correlations were expressed *relative to* across pairmate correlations.

#### 686 687 **Statistics.**

688 To compare pairmate similarity scores and other measures across ROIs and learning states, repeated  
689 measures ANOVAs and paired-samples  $t$ -tests were used. To test whether pairmate similarity scores and  
690 other measures were significantly positive or negative (i.e., above/below 0), one-sample  $t$ -tests were used.  
691 To test whether the negative pairmate similarity score observed in CA23DG at the inflection point depended  
692 on the specific mapping between behavioral and fMRI measures, we randomly shuffled the mapping  
693 between the behavioral inflection point and scene pairmate, within each participant (see **Fig. 1d**), and then  
694 computed the group-level mean pairmate similarity score at the permuted inflection point. This was  
695 repeated 1,000 times, producing a distribution of 1,000 permuted means. The observed pairmate similarity  
696 score at the inflection point was then compared against this distribution of permuted means.

#### 697 698 **Data Availability.**

699 The data that support the findings of this study are available from the corresponding author upon reasonable  
700 request.

701 **METHODS REFERENCES:**

- 702 1. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203  
703 (2019).
- 704 2. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–  
705 116 (2019).
- 706 3. Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong  
707 Ma, Ayse Ilkay Isik, et al. 2018. “fMRIPrep.” *Software*.  
708 Zenodo. <https://doi.org/10.5281/zenodo.852659>.
- 709 4. Gorgolewski, K. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing  
710 Framework in Python. *Front. Neuroinformatics* **5**, (2011).
- 711 5. Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis,  
712 Michael Philipp Notter, Dorota Jarecka, et al. 2018. “Nipype.” *Software*.  
713 Zenodo. <https://doi.org/10.5281/zenodo.596855>.
- 714 6. Tustison, N. J. *et al.* N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320  
715 (2010).
- 716 7. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration  
717 with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med.*  
718 *Image Anal.* **12**, 26–41 (2008).
- 719 8. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random  
720 field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
- 721 9. Fonov, V., Evans, A., McKinstry, R., Alml, C. & Collins, D. Unbiased nonlinear average age-  
722 appropriate brain templates from birth to adulthood. *NeuroImage* **47**, S102 (2009).
- 723 10. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.*  
724 **10**, 171–178 (1997).
- 725 11. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based  
726 registration. *NeuroImage* **48**, 63–72 (2009).

- 727 12. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and  
728 Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825–841 (2002).
- 729 13. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state  
730 fMRI. *NeuroImage* **84**, 320–341 (2014).
- 731 14. Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving BOLD activation in  
732 event-related designs for multivoxel pattern classification analyses. *NeuroImage* **59**, 2636–2643  
733 (2012).
- 734 15. Wang, L., Mruczek, R. E. B., Arcaro, M. J. & Kastner, S. Probabilistic Maps of Visual Topography  
735 in Human Cortex. *Cereb. Cortex N. Y. N 1991* **25**, 3911–3931 (2015).
- 736 16. Yushkevich, P. A. *et al.* Automated volumetry and regional thickness analysis of hippocampal  
737 subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* **36**,  
738 258–287 (2015).
- 739



## Supplementary information

Participant # \ Round	1	2	3	4	5	6	Never Learned
1	1	7	6	4	0	0	0
2	1	1	4	4	6	2	0
3	1	7	5	5	0	0	0
4	0	3	0	5	4	3	3
5	0	2	3	6	4	2	1
6	3	6	2	6	0	1	0
7	0	6	4	3	3	1	1
8	0	2	5	4	5	1	1
9	0	1	1	2	2	2	10
10	0	0	8	2	5	2	1
11	3	3	4	3	2	2	1
12	0	1	2	5	2	5	3
13	1	1	2	4	7	2	1
14	0	0	3	4	4	5	2
15	1	6	7	2	1	1	0
16	1	2	6	1	2	4	2
17	2	3	3	5	3	2	0
18	5	3	2	3	4	0	1
19	0	0	2	7	6	2	1
20	0	1	6	2	1	4	4
21	0	1	3	3	4	7	0
22	1	3	4	2	3	1	4
23	0	6	5	4	1	2	0
24	3	4	7	1	2	1	0
25	1	10	4	3	0	0	0
26	0	0	2	9	2	1	4
27	3	0	4	2	2	1	6
28	1	8	4	3	0	0	2
29	0	6	2	1	1	2	6
30	2	6	6	1	0	2	1
31	1	1	3	6	3	3	1

**Table1.** Number of pairmates that transitioned to learned round ('LR') status, for each participant and each round. Note: pairmates that were learned in the first round or never learned were excluded from fMRI analyses.