

Research Articles: Behavioral/Cognitive

### Adaptive memory distortions are predicted by feature representations in parietal cortex

https://doi.org/10.1523/JNEUROSCI.2875-20.2021

Cite as: J. Neurosci 2021; 10.1523/JNEUROSCI.2875-20.2021

Received: 12 November 2020 Revised: 26 January 2021 Accepted: 27 January 2021

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2021 the authors

|                    | 1  | Adaptive memory distortions are predicted by   |
|--------------------|----|--|
|                    | 2  | feature representations in parietal cortex   |
| bt                 | 3  |  |
|                    | 4  | Yufei Zhao (赵雨菲) <sup>1</sup> , Avi J. H. Chanales <sup>2</sup> , & Brice A. Kuhl <sup>1</sup> |
| Ö                  | 5  | Department of Psychology, University of Oregon <sup>1</sup> , 97401                            |
| S                  | 6  | Department of Psychology, New York University <sup>2</sup> , 10016                             |
| $\square$          | 7  | 1  |
| an                 | 8  | Abbreviated title: Memory distortions in parietal cortex                                       |
| Ž                  | 9  | Corresponding author email address: Yufei Zhao (yzhao17@uoregon.edu) or Brice Kuhl             |
| σ                  | 10 | (bkuhl@uoregon.edu)  |
| te                 | 11 | Pages: 35  |
| Q                  | 12 | Figures: 4   |
| Ð                  | 13 | Tables: 1  |
| Q                  | 14 | Number of words: Abstract (152), Introduction (696), discussion (1448)                         |
| O                  | 15 |  |
|                    | 16 | Acknowledgements: This work was supported by NIH grant NINDS R01-NS107727 and NSF CAREER       |
|                    | 17 | Award BCS-1752921 to B.A.K.  |
|                    | 18 |  |
| $\hat{\mathbf{O}}$ | 19 |  |
| Z                  | 20 |  |
|                    | 21 |  |
| Ð                  | 22 |  |
| Ζ                  | 23 |  |
|                    | 24 |  |
|                    |    | <sup>1</sup> The authors declare no competing financial interests.                             |

### 25 ABSTRACT

26

27 Similarity between memories is a primary cause of interference and forgetting. Exaggerating subtle 28 differences between memories is therefore a potential mechanism for reducing interference. Here, we 29 report a human fMRI study (n = 29, 19 female) that tested whether behavioral and neural expressions of 30 memories are adaptively distorted to reduce interference. Participants learned and repeatedly retrieved 31 object images, some of which were identical except for subtle color differences. Behavioral measures of 32 color memory revealed exaggeration of differences between similar objects. Importantly, greater memory 33 exaggeration was associated with lower memory interference. fMRI pattern analyses revealed that color 34 information in parietal cortex was stronger during memory recall when color information was critical for 35 discriminating competing memories. Moreover, greater representational distance between competing 36 memories in parietal cortex predicted greater color memory exaggeration and lower memory interference. 37 Together, these findings reveal that competition between memories induces adaptive, feature-specific 38 distortions in parietal representations and corresponding behavioral expressions.

39

40 Keywords: episodic memory, interference, repulsion, fMRI, pattern similarity

41

### 42 Significance Statement (120 words)

43

Similarity between memories is a primary cause of interference and forgetting. Here, we show that when remembering highly similar objects, subtle differences in the features of these objects are exaggerated in memory in order to reduce interference. These memory distortions are reflected in, and predicted by, overlap of activity patterns in lateral parietal cortex. These findings provide unique insight into how memory interference is resolved and specifically implicate lateral parietal cortex in representing featurespecific memory distortions.

### 50 INTRODUCTION

51

52 Given the vast number of memories that humans store, overlap between memories is inevitable. For 53 example, one may have taken multiple vacations to the same town or parked in the same garage on 54 many occasions. There is a long history of behavioral studies in psychology documenting the many 55 contexts in which this type of overlap leads to memory interference and forgetting (Anderson & Spellman, 56 1995; Barnes & Underwood, 1959; Mensink & Raaijmakers, 1988; Osgood, 1949; Wixted, 2004). As a 57 result, a primary focus of theoretical models of memory has been to specify the computational 58 mechanisms by which interference is resolved (Colgin, Moser, & Moser, 2008; O'Reilly & McClelland, 59 1994; Treves & Rolls, 1994). These models have largely focused on how memories are encoded so that 60 the content of memories is protected against interference. An alternative perspective, however, is that 61 instead of protecting memories from interference, there is adaptive value in allowing the content of 62 memories to be shaped by interference (Hulbert & Norman, 2015; Kim, Norman, & Turk-Browne, 2017). 63 Specifically, to the extent that overlap across memories is the root cause of interference, then distorting 64 memories to reduce this overlap is a potentially effective remedy.

65 Evidence from recent neuroimaging studies hints at the idea that memory representations are 66 distorted as an adaptive response to interference. Namely, several studies have found that when similar 67 events are encoded into memory, this triggers a targeted exaggeration of differences in patterns of 68 activity in the hippocampus (Ballard, Wagner, & McClure, 2019; Chanales, Oza, Favila, & Kuhl, 2017; 69 Dimsdale-Zucker, Ritchey, Ekstrom, Yonelinas, & Ranganath, 2018; Favila, Chanales, & Kuhl, 2016; 70 Hulbert & Norman, 2015; Kim et al., 2017; Schapiro, Kustner, & Turk-Browne, 2012; Schlichting, 71 Mumford, & Preston, 2015). The key observation in these studies is that similar memories 'move apart' 72 from each other in representational space, suggesting a form of memory repulsion. Yet, a critical limitation 73 of these studies is that the feature dimensions along which memories move are underspecified. That is, 74 do changes in neural representations correspond to changes in the information content of memories? On 75 the one hand, neural activity pattern may become separated without any changes to underlying 76 memories. Alternatively, changes in neural activity patterns may reflect adaptive changes in memory 77 content. For example, if two vacations to the same city were associated with different weather conditions,

then weather-related information may be a salient component of corresponding memories and weatherrelated *differences* between those vacations may be exaggerated to improve memory discriminability (e.g., "That was the year it was *really cold*," vs. "That was the year it was *really hot*").

81 While it has proven difficult to translate hippocampal activity patterns to explicit feature 82 dimensions (LaRocque et al., 2013; Liang, Wagner, & Preston, 2013), feature dimensions are far more 83 accessible in (or decodable from) neocortical regions involved in memory retrieval. In particular, there is 84 rapidly growing evidence that lateral parietal cortex carries detailed information about the content of 85 retrieved memories (Chen et al., 2017; Long, Lee, & Kuhl, 2016; Xiao et al., 2017) and amplifies 86 behaviorally-relevant information (Favila, Samide, Sweigart, & Kuhl, 2018; Kuhl, Johnson, & Chun, 2013). 87 Moreover, recent studies have shown that memory representations in parietal cortex can be decomposed 88 into separable feature dimensions (Bone, Ahmad, & Buchsbaum, 2020; Favila et al., 2018; Lee, Samide, 89 Richter, & Kuhl, 2019). Thus, lateral parietal cortex may provide a unique window into how memory 90 representations are shaped by interference.

91 Here, we tested whether interference between highly similar memories triggers adaptive 92 distortions in parietal memory representations and corresponding behavioral expressions of memories. 93 Our motivating theoretical perspective was that subtle differences between similar memories are 94 prioritized and exaggerated to reduce the potential for interference. To test these ideas, we modified a 95 recent behavioral paradigm that demonstrated adaptive biases in long-term memory for objects 96 (Chanales, Tremblay-McGaw, & Kuhl, in-press). We predicted that competition between memories for 97 similar objects would trigger a memory-based exaggeration of subtle differences between those objects, 98 and that greater exaggeration would be associated with lower memory interference. Using pattern-based 99 fMRI analyses, we tested whether memory representations in lateral parietal cortex (a) preferentially 100 express features that are critical for discriminating similar objects and (b) predict feature-specific 101 distortions in behavioral expressions of memory.

102

103 Materials and Methods

### 105 Participants

106

107 Thirty-two (21 female; mean age = 23.5 years) right-handed, native English speakers from the University 108 of Oregon community participated in the experiment. Three participants were excluded from analysis (two 109 due to falling asleep inside the scanner, one due to technical error), resulting in a final set of 29 110 participants (19 female; mean age = 23.7 years) included in data analysis. Participants were screened for 111 motion during the scanned recall tasks, but no participants exceeded the exclusion criteria (mean 112 framewise displacement > 0.25) for any of the runs. The sample size was comparable to similar fMRI 113 studies in the field. All participants had normal or corrected-to-normal vision. Informed consent was 114 obtained in accordance with the University of Oregon Institutional Review Board.

115

### 116 Overview of Experimental Paradigm

117

118 We modified a paradigm from a recent behavioral study that was used to demonstrate adaptive biases in 119 long-term memory for object colors (Chanales, Tremblay-McGaw, & Kuhl, in-press). In the prior (and 120 current) study, participants learned associations between faces and object images. Critically, the objects 121 contained 'pairmates' for which the object images were identical except for their color (e.g., a blue 122 backpack and a purple backpack), and successful learning required discriminating between these 123 pairmates. In the current study, we used a two-day procedure in which participants received extensive 124 behavioral training on face-object associations on Day 1 and then returned on Day 2 for additional 125 behavioral training, followed by an fMRI session, and finally a behavioral color memory test (Fig. 1). A 126 critical feature of our design is that we held color similarity between pairmates constant (24 degrees 127 apart), but we included a competitive and non-competitive condition (Fig. 1b). In the competitive 128 condition, pairmate images corresponded to the same object category (e.g., two beanbags of slightly 129 different colors). In the non-competitive condition, pairmates corresponded to distinct object categories 130 (e.g., a pillow and a ball of slightly different colors). Thus, in both conditions the pairmates were 24 131 degrees apart in color space; but, for the competitive condition, color was the only feature dimension on 132 which the pairmates differed. In contrast, for the non-competitive condition, object category also differed

between pairmates. Thus, although color distance between pairmates was matched across conditions, color information was more important in the competitive condition. For the fMRI session, participants were shown faces, one at a time, with the only instruction being to retrieve corresponding objects as vividly as possible. An important feature of our procedure is that participants were not explicitly instructed to retrieve color information during the fMRI scans, nor had color memory been tested at any point prior to scanning. Rather, we only tested color memory after participants exited the scanner.

139

### 140 Stimuli

### 141

142 Participants learned associations between 24 object images and 24 images of white male faces. The 24 143 object images corresponded to 18 distinct object categories (e.g., beanbag, hat, umbrella, balloon) and 144 12 distinct color values. Thus, some of the 24 object images were from the same object category (e.g., 145 two beanbags) or had the same color value. The object images were generated from an image set that 146 allowed for each image's color to be rotated along a 360° color wheel (Brady, Konkle, Alvarez, & Oliva, 147 2013). To assign colors to each object, the 360° color wheel was divided into 15 evenly spaced color 148 values (0°, 24°, 48°, etc.). These 15 values were arbitrarily chosen but were fixed across participants. For 149 each participant, 6 consecutive color values were selected (randomly positioned among the set of 15 150 color values) for the competitive condition. For example, color values of 48°, 72°, 96°, 120°, 144°, and 151 168° might be selected for the competitive condition (Fig. 1b). Likewise, 6 consecutive color values were 152 selected for the non-competitive condition. The 6 values for the non-competitive condition always 'started' 153 48° after the competitive color values 'ended.' For example, if the color values for the competitive 154 condition spanned 48° to 168°, then the color values for the non-competitive condition would be 216°, 155 240°, 264°, 288°, 312°, 336° (Fig. 1b).

156

For both conditions, the 6 color values were clustered into 3 sets of consecutive color values: e.g., 48° and 72°, 96° and 120°, 144° and 168°. Each of these sets included a total of 4 object images (resulting in 12 object images for each condition). For the competitive condition, the four images in each set represented two color values (e.g., 48° and 72°) and two object categories (e.g., beanbag and jacket). For 161 example, the set might include a 48° beanbag, a 72° beanbag, a 48° jacket and a 72° jacket (Fig. 1b). 162 Object images within each set that were from the same object category (e.g., the 48° beanbag and the 163 72° beanbag) are referred to as 'pairmates.' For the non-competitive condition, the four images in each 164 set represented two color values (e.g., 216° and 240°) and four distinct object categories (Fig. 1b). 165 Although none of the object images in the non-competitive condition were from the same object category, 166 the four images in each set were also divided into pairmates, with pairmates being images from distinct 167 object categories and, as in the competitive condition, with color values 24° apart. For example, if a set in 168 the non-competitive condition included a 216° lunchbox, a 216° pillow, a 240° hat, and a 240° ball, the 169 216° lunchbox and the 240° hat might be arbitrarily designated as one set of pairmates and the 216° 170 pillow and the 240° ball as the other set of pairmates. These non-competitive pairmates functioned as a 171 critical control condition for behavioral and fMRI analyses (see fMRI Pattern Similarity Analyses, below).

172

The mapping between the 24 object images and the 24 face images was randomly determined for each
participant. All face and object images were 250 \* 250 pixels.

175

### 176 Pre-scan face-object training

177

178 Participants completed the experiment on two consecutive days (Fig. 1a). On Day 1, participants learned 179 24 face-object associations across 14 training rounds. Each training round consisted of a study phase 180 and an associative memory test phase. During study phases, participants were presented with the 24 181 face-object associations, one association at a time, in random order. Each trial started with a fixation 182 cross presented in the center of the screen (1.5 s), followed by the face-object association (3.5 s). Faces 183 were presented to the left of the objects. During the associative memory test phases, object images were 184 presented at the top of the screen with four face choices below. The four face choices always included 185 the target face (i.e., the face associated with the presented object image), the pairmate's face (i.e., the 186 face that was associated with the presented object's pairmate), and two foil faces (associated with non-187 pairmate objects). Participants were asked to select the face that was associated with the presented 188 object. After responding, participants received feedbacks indicating whether or not they were correct and

189 showing the correct face-object association for 1.5 s. Each trial in the associative memory test was self-190 paced up to a maximum of 8 s. On Day 2, participants completed 4 additional training rounds immediately 191 prior to entering the fMRI scanner. The procedure was the same as on Day 1.

192

### 193 Scanned perception and cued recall tasks

194

195 During fMRI scanning, participants completed 6 consecutive rounds of a perception task and 6 196 consecutive rounds of a cued recall task (each round corresponded to a separate fMRI scan). The order 197 of the perception and cued recall tasks was counterbalanced across participants. In the perception task, 198 each trial presented one of the 24 object images in the center of the screen for 0.5 s followed by a fixation 199 cross for 3.5 s. A black cross was embedded within the object images at a random location on 25% of 200 trials and participants were instructed to make a button press whenever they detected a black cross. In 201 each perception round, each object image was presented twice, in block randomized order. Participants 202 were instructed to remain centrally-fixated, on a white fixation cross, throughout each perception run. 203 Each perception round contained a 10 s null trial (fixation cross only) at the beginning and end of each 204 scan and 12 null trials (4 s each) randomly distributed throughout the run. Here, we do not consider data 205 from the perception task because (a) our primary hypotheses related to participants' memories for the 206 object images and (b) subtle color differences between were more to detect in the scanner environment.

207

208 In the cued recall task, each trial started with one of the 24 face images presented at the center of the 209 screen for 0.5 s, followed by a blank screen for 2.5 s, and then a question mark for 1 s. Participants were 210 instructed to recall the object image that was associated with the presented face as vividly as possible 211 and to hold the image in mind throughout the trial. Participants were instructed to rate the vividness of 212 their memories ('vivid' or 'not vivid') via a button box response when the question mark appeared. The 213 question mark was followed by a fixation cross for 2 s before next trial began. Responses were recorded 214 during the trial and during the 2 s fixation cross between trials. Together, the intertrial interval was 6 s. All 215 face-object associations were tested twice in each retrieval round, in block randomized order. Each

retrieval round contained a 10 s null trial (fixation cross only) at the beginning and end of each scan and

217 12 null trials (4 s each) randomly distributed throughout the run.

218

### 219 Post-scan behavioral tests

220

221 After participants completed the perception and cued recall tasks, they exited the scanner and completed 222 five rounds of the color memory test. During the color memory test, each trial began with one of the 24 223 face images presented on the left side of the screen and the corresponding object image presented on 224 the right of the screen. Importantly, the object image was initially in grey scale. Participants were 225 instructed to move a cursor along a color wheel (Fig. 1a, c) to adjust the color of the object to the 226 remembered color value. Participants clicked the mouse to record their response and then moved on to 227 the next trial. Each face-object association was tested once per round and the task was self-paced. After 228 completing the five color memory test rounds, participants completed two final rounds of the associative 229 memory test-the same task they completed during the training rounds on Day 1 and just prior to fMRI 230 scanning. The sole purpose of the post-scan associative memory test was to motivate participants to 231 maintain their effort and memory accuracy throughout the fMRI session as the post-scan associative 232 memory test was used to determine a monetary bonus for participants (a fact which participants were 233 made aware of prior to the fMRI scan).

234

### 235 Measuring color memory bias

236

The post scan color memory test was used to measure participants' color memory for each object image. However, rather than focusing on the accuracy of recall, we were critically interested in recall bias. Bias was measured in two ways. The first measure—*mean signed distance*—was computed by first averaging the responses across the 5 color memory test trials for each object image. The difference between the mean response and the actual color value for a given object image reflects the color memory distance for that object image. Critically, if the mean response was biased *away* from the color of the pairmate object (Fig. 1c), the distance measure was positively signed; if the mean response was biased *toward* the color 244 of the pairmate object (Fig. 1c), the distance measure was negatively signed. By averaging the signed 245 distance measure across the 12 object images within each condition, the mean signed distance was 246 computed for each condition (competitive, non-competitive) and for each participant. The second 247 measure—percentage of away responses—was computed by ignoring the distance between participants' 248 responses and the actual color values and instead simply computing the percentage of responses that 249 were biased away from the color of the pairmate object. It is important to note that this measure was 250 computed at the trial level. Thus, for a given object image, if a participant recalled the object's color 'away 251 from' the pairmate on 4 out of the 5 test trials for that object image, the percentage of away responses for 252 that object image would be 80%. Although we did not expect (or observe) notable differences between 253 the two measures (mean signed distance and percentage of away responses), the percentage of away 254 responses addressed the concern that any observed effects for the mean signed distance measure were 255 driven by a few extreme responses.

256

### 257 fMRI data acquisition

258

259 Imaging data were collected on a Siemens 3 T Skyra scanner at the Robert and Beverly Lewis Center for 260 NeuroImaging at the University of Oregon. Functional data were acquired using a T2\*-weighted multiband 261 EPI sequence with whole-brain coverage (repetition time = 2 s, echo time = 36 ms, flip angle =  $90^{\circ}$ , 262 multiband acceleration factor = 3, inplane acceleration factor = 2, 72 slices, 1.7 × 1.7 × 1.7 × mm voxels) 263 and a 32-channel head coil. Note that due to an a priori decision to focus on visual and parietal cortical 264 areas, we used a high-resolution protocol that fully covered visual/parietal regions but only partially 265 covered frontal cortex. Each perception scan (6 total) consisted of 130 total volumes. Each retrieval scan 266 (6 total) consisted of 190 total volumes. Oblique axial slices were aligned parallel to the plane defined by 267 the anterior and posterior commissures. A whole-brain T1-weighted MPRAGE 3D anatomical volume (1 x 268 1 x 1 mm voxels) was also collected.

269

270 fMRI data preprocessing

272 fMRI data preprocessing was performed using fMRIPrep 1.3.1 (Esteban et al., 2019). The T1-weighted 273 (T1w) image was corrected for intensity non-uniformity with N4BiasFieldCorrection (Tustison et al., 2010) 274 and skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0) with OASIS30ANTs as the target template. 275 Brain surfaces were reconstructed using recon-all from FreeSurfer 6.0.1 (Dale, Fischl, & Sereno, 1999). 276 Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov, Evans, 277 McKinstry, Almli, & Collins, 2009) was performed through nonlinear registration with antsRegistration 278 (ANTs 2.2.0). For the functional data, susceptibility distortion corrections were estimated using 3dQwarp 279 (Cox & Hyde, 1997). The BOLD reference was then co-registered to the T1w reference by bbregister 280 (FreeSurfer) using boundary-based registration with nine degrees of freedom (Greve & Fischl, 2009). 281 Head-motion parameters were estimated by mcflirt from FSL 5.0.9 (Jenkinson, Bannister, Brady, & Smith, 282 2002). Slice-time correction was done by 3dTshift from AFNI 20160207 (Cox & Hyde, 1997). Functional 283 data were smoothed with a 1.7 mm FWHM Gaussian kernel and high pass filtered at 0.01Hz. Smoothing 284 and filtering were done with the Nipype pipeline tool (Gorgolewski et al., 2011).

285

Response estimates were obtained for each trial (one regressor per trial, 4 s duration) in each cued recall run using the "least-squares separate" method (Mumford, Turner, Ashby, & Poldrack, 2012). With this method, each item was estimated in a separate GLM as a separate regressor while all remaining items were modeled together with another regressor. The six movement parameters and framewise displacement were included in each GLM as confound regressors. This resulted in *t* maps that were used for the pattern similarity analysis. Given that all analyses averaged data across multiple trials—mitigating the influence of any one trial—we did not perform any data exclusion for outliers at the trial level.

293

### 294 Regions of interest

295

fMRI analyses were conducted using a set of visual and parietal regions of interest (ROIs) that were identical to those used by Favila, Samide, Sweigart, & Kuhl (2018) to measure object and color representations during memory recall. While our primary focus was on the parietal ROIs, we anticipated that visual regions might also reflect feature-specific information during memory retrieval. For low level 300 visual regions, we combined bilateral V1v and V1d as V1 and combined bilateral LO1 and LO2 as LO 301 based on Wang, Mruczek, Arcaro, & Kastner (2014). For high level visual regions, we generated a VTC 302 ROI by combining bilateral fusiform gyrus, collateral sulcus, and lateral occipitotemporal sulcus derived 303 from the output of Freesurfer segmentation routines. For lateral parietal cortex, we referenced Yeo et al. 304 (2011)'s 17-network resting state atlas. The parietal nodes from Network 12 and 13 (subcomponents of 305 the frontoparietal control network) are referred to as dorsal lateral intraparietal sulcus (dLatIPS) and 306 ventral lateral intraparietal sulcus (vLatIPS), respectively. For the parietal node of Network 5 (dorsal 307 attention network), we separated it along the intraparietal sulcus to create a dorsal region we refer to as 308 posterior intraparietal sulcus (pIPS) and a ventral region we refer to as ventral IPS (vIPS) (Sestieri et al., 309 2017) . The vertices in lateral occipital cortex were eliminated in these two regions. The parietal nodes of 310 Networks 15–17 (subcomponents of the default mode network) were combined into a region we refer to 311 as angular gyrus (AnG).

312

For post hoc analyses, we generated medial temporal and hippocampus subfield ROIs using ASHS (Yushkevich et al., 2015). We selected bilateral CA1, subiculum, entorhinal cortex, and parahippocampal cortex. We combined CA2, CA3 and dentate gyrus into a single ROI (CA23DG) and combined BA35 and BA36 into a perirhinal cortex ROI.

317

### 318 fMRI Pattern similarity analyses

319

320 Pattern similarity analyses were used to measure the similarity of fMRI activity patterns for various pairs 321 of object images during the cued recall task. To calculate pattern similarity, we first computed the mean 322 activity pattern for each of the 24 recalled objects by averaging t maps for odd runs and even runs 323 separately. Pearson correlations were then computed between the mean t map of odd runs and even 324 runs. All the correlations were z-transformed (Fisher's z) before subsequent analyses. All analyses were 325 performed in the participant's native T1w space and were done separately for each ROI. Pattern similarity 326 analyses focused on three specific correlations within each 'set' of 4 object images (see Fig. 1b and 327 Stimuli for explanation of 'sets'): (1) 'Pairmate correlations' (see Stimuli for definition of pairmates), (2)

<sup>328</sup> 'Same-color correlations,' which refer to correlations between object images from different object <sup>329</sup> categories but with identical color values (Fig. 1b), and (3) 'Baseline correlations,' which refer to object <sup>330</sup> images from different object categories and different color values (24 degrees apart; Fig 1b). Again, it is <sup>331</sup> important to emphasize that all pattern similarity analyses were performed within the sets of 4 object <sup>332</sup> images and, critically, the same correlations were applied for the competitive and non-competitive <sup>333</sup> conditions.

334

### 335 Neural representation of color information

336

337 To test whether representation of color information was stronger in the competitive condition than the 338 non-competitive condition, we first obtained (for each condition, ROI, and participant) the mean 'Same-339 color correlation' and the mean 'Baseline correlation.' Both of these correlations reflect correlations 340 between object images from different object categories (Fig. 1b), but the same-color correlation reflects 341 images with identical color values whereas the baseline correlation reflects images with a 24° difference 342 in color. Thus, the difference between these measures (same-color - baseline) isolates color-related 343 similarity. Of critical interest was whether this color-related similarity was stronger in the competitive 344 condition than the non-competitive condition. Critically, color similarity was objectively identical across 345 conditions, but we predicted stronger color representation in the competitive condition owing to its greater 346 diagnostic value in the competitive condition. It is important to note that the inclusion of a separate 347 baseline correlation for each condition (competitive, non-competitive) controlled for potential global 348 similarity differences between conditions (i.e., that correlations among all pairs of object images might be 349 higher in one condition vs. the other).

350

### 351 Neural similarity between pairmates

352

To test whether similarity between pairmates was stronger in the competitive condition than the noncompetitive condition, we first obtained (for each condition, ROI, and participant) the mean 'Pairmate correlation' and the mean 'Baseline correlation.' For the competitive condition, pairmate correlations 356 reflect object images from the same object category but with a 24° difference in color (Fig. 1b). For the 357 non-competitive condition, pairmate correlations reflect object images from different object categories, 358 again with a 24° difference in color (Fig. 1b). Thus, pairmate similarity was objectively greater in the 359 competitive condition than the non-competitive condition. For both conditions, the baseline correlations 360 reflect object images from different object categories and with a 24° difference in color. Thus, the 361 difference between these measures (pairmate - baseline) was intended to isolate object-related similarity 362 (specifically for the competitive condition). As with the color information analysis, the condition-specific 363 baseline correlations controlled for potential global similarity differences between conditions.

364

### 365 Neural measures of pairmate similarity predict color memory repulsion

366

367 To test whether similarity between vIPS representations of pairmates during competitive recall predicted 368 the degree to which there was repulsion of color memories (as measured in the post-scan color memory 369 test), we first computed the mean signed color memory distance for the two objects in each set of 370 pairmates. This yielded a single value representing the distance between a given set of pairmates, with 371 greater distance reflecting greater repulsion. Next, for vIPS we computed dissimilarity between each set 372 of pairmates, as defined by: 1 - the Pairmate correlation. (Note: for this analysis we used dissimilarity, as 373 opposed to similarity, simply for ease of interpretation). Thus, for each participant and for each condition 374 (competitive, non-competitive), this resulted in 6 values representing color memory distance between 375 each set of pairmates and 6 values representing vIPS dissimilarity between each set of pairmates. We 376 then performed a Spearman correlation between these two measures. For each condition, one-sample t-377 tests were performed on the participants' z-transformed Spearman's r<sub>s</sub> values to test whether the mean 378 correlation between color memory distance and vIPS dissimilarity differed from 0. For comparison, similar 379 analyses were also performed for other ROIs (Table 1).

380

To better visualize the relationship between color memory distance and vIPS dissimilarity, for each participant the 6 pairmates in the competitive condition were divided into three bins (2 pairmates per bin) based on vIPS pairmate dissimilarity (low, medium, high). We then computed the mean signed color 384 memory distance (from the post-scan color memory test) and the mean associative memory accuracy 385 (from the pre-scan associative memory test) for each of these bins. One-way ANOVA was used to test 386 whether mean signed distance and/or mean associative memory accuracy varied as a function of vIPS 387 dissimilarity bin. Finally, we performed a multilevel mediation analysis to test whether color memory 388 mediated the relationship between vIPS pairmate dissimilarity and associative memory accuracy. This 389 analysis was performed by obtaining, for each participant, the mean color memory distance, vIPS 390 dissimilarity, and associative memory performance for each of the 6 pairmates in each condition. 391 Mediation analyses included a random intercept for each participant, but random slopes were not 392 included due to the small number of data points per condition/participant.

393

### 394 Statistical analysis

395

Statistical analyses were performed using R version 3.6.3. All *t*-tests were two-tailed, with  $\alpha$  = 0.05. All repeated measures ANOVAs were computed with the afex package using Type III sums of squares. Effect sizes for *t*-tests were estimated using the effsize package. Multilevel mediation analyses were computed using the mediation package. Multilevel models were built using the Ime4 package. All error bars in the figures represent S.E.M.

401

### 402 RESULTS

403

### 404 Associative Memory Performance

405

Participants completed three separate sessions that tested memory for object-face associations (14 rounds on Day 1; 4 rounds before scanning on Day 2; 2 rounds after scanning on Day 2; Fig. 1a). Participants showed improved accuracy across test rounds in the Day 1 session, from a mean of 56.9% (SD = 12.8%) on round 1 to a mean of 95.5% (SD = 4.8%) on round 14 (main effect of test round:  $F_{5.56}$ ,  $_{155.73} = 91.29$ , p < 0.0001,  $\eta^2 = 0.55$ ). Accuracy did not vary by test round for either of the Day 2 sessions (Day 2 pre-scan:  $F_{2.77, 77.63} = 1.63$ , p = 0.194,  $\eta^2 = 0.01$ ; Day 2 post-scan:  $F_{1, 28} = 0.14$ , p = 0.713,  $\eta^2 =$  412 0.0009). Critically, accuracy was lower in the competitive condition than the non-competitive condition for 413 each of the sessions (Day 1:  $F_{1,28}$  = 115.89, p < 0.0001,  $\eta^2$  = 0.29; Day 2 pre-scan:  $F_{1,28}$  = 21.8,1 p < 0.0001,  $\eta^2$ 0.0001,  $\eta^2 = 0.15$ ; Day 2 post-scan:  $F_{1, 28} = 22.25$ , p < 0.0001,  $\eta^2 = 0.20$ ; Fig. 2a). For subsequent 414 415 analyses, we focused on associative memory performance from the Day 2 pre-scan session (an a priori 416 decision; see Methods). Notably, for the Day 2 pre-scan session, lower accuracy in the competitive 417 condition (M = 93.2%, SD = 6.9%) than the non-competitive condition (M = 98.9%, SD = 2.1%) was 418 driven by an increased rate of selecting faces that were associated with the pairmate image (competitive 419 condition: M = 6.0%, SD = 6.6%; non-competitive condition: M = 0.2%, SD = 0.6%;  $t_{28} = 4.74$ , p < 0.0001, 420 95% CI = [0.03 0.08], Cohen's d = 1.16, paired t-test; Fig. 2a). The rate of other errors did not differ in the 421 competitive vs. non-competitive conditions (competitive: M = 0.8%, SD = 1.4%; non-competitive: M =422 0.98%, SD = 1.6%; t<sub>28</sub> = -0.18, p = 0.861, 95% CI = [-0.01 0.01], Cohen's d = -0.04, paired t-test). Thus, 423 as intended, the competitive condition specifically increased interference between pairmate images.

424

### 425 Color Memory Bias

426

Immediately after the fMRI session, participants completed a color memory test. Color memory was indexed in two ways: (1) using a continuous, signed measure of distance, in degrees, between the reported and actual color; positive values indicate a bias *away from* the competing memory and negative values indicate a bias *toward* the competing memory, and (2) using a categorical measure of the percentage of responses that were biased away from the competing memory (see Methods for details of each measure). We refer to these two measures as the *signed distance* and *percentage of away responses*, respectively.

For the competitive condition, mean signed distance was significantly greater than 0 (5.09 ± 4.69, mean ± SD;  $t_{28} = 5.84$ , p = 0.000003, 95% CI = [3.30 6.87], Cohen's d = 1.08, one-sample *t*-test; Fig. 2b), indicating that participants' color memory was systematically biased away from the color of the pairmate. In contrast, for the non-competitive condition—where the only difference was that pairmates were not from the same object category—signed distance did not differ from 0 (-0.39 ± 7.08;  $t_{28} = -0.29$ , p = 0.771, 95% CI = [-3.08 2.31], Cohen's d = -0.05, one-sample *t*-test). Signed distance was significantly greater 440 (i.e., a stronger bias away from the pairmate) in the competitive condition compared to the non-441 competitive condition ( $t_{28} = 2.90$ , p = 0.007, 95% CI = [1.61 9.34], Cohen's d = 0.92, paired *t*-test). These 442 data clearly demonstrate that similarity between images triggered the color memory bias.

443 The pattern of data was identical when considering the percentage of away responses. Namely, 444 the percentage of away responses was significantly greater than 50% for the competitive condition (61.4 445  $\pm$  3.6%;  $t_{28}$  = 4.49, p = 0.0001, 95% CI = [56.2% 66.6%], Cohen's d = 0.83, one-sample t-test; Fig. 2c), but 446 not for the non-competitive condition (46.5 ± 14%;  $t_{28}$  = -1.35, p = 0.189, 95% CI = [41.2% 51.8%], 447 Cohen's d = -0.25, one-sample t-test). The difference between the two conditions was also significant ( $t_{28}$ 448 = 3.58, p = 0.001, 95% CI = [0.06 0.23], Cohen's d = 1.08, paired *t*-test). While the percentage of away 449 responses does not contain information about the magnitude of the bias in color memory, it rules out the 450 possibility that the effects observed with the signed distance measure were driven by a minority of trials 451 with very high bias.

452

### 453 Relationship between associative memory and color memory bias

454

### 455 A key component of our theoretical framework is that exaggerating the color distance (in memory) 456 between similar objects plays an adaptive role in reducing memory interference. To test this idea, we 457 correlated each participant's associative memory performance (from the Day 2 pre-scan session) with 458 their color memory performance. For the competitive condition, mean associative memory performance was positively correlated with mean signed distance (r = 0.50, $t_{26} = 2.91$ , p = 0.007, 95% CI = [0.15 0.73], 459 460 one outlier excluded for associative memory performance < 3 SD below mean; Fig. 2d), consistent with 461 the idea that stronger color memory repulsion (i.e., a bias in color memory away from the pairmate) 462 supports lower associative memory interference. For the non-competitive condition, this correlation was 463 not significant ( $r = -0.31 t_{26} = -1.63$ , p = 0.114, 95% CI = [-0.61 0.08], one outlier excluded for signed 464 distance > 3 SD above the mean). Thus, a bias in color memory away from the pairmate was not 465 beneficial if the pairmate was not similar to (competitive with) the target. An identical pattern of data was 466 observed when considering the percentage of away responses as an index of color memory. Namely, for 467 the competitive condition there was a positive correlation between associative memory performance and

the mean percentage of away responses (r = 0.42,  $t_{26} = 2.39$ , p = 0.025, 95% CI = [0.06 0.69], one outlier excluded for associative memory performance < 3 SD below mean) and no significant correlation for the non-competitive condition (r = -0.37,  $t_{27} = -2.05$ , p = 0.050, 95% CI = [-0.65 -0.002]).

471

### 472 Neural representation of color information during recall

473

474 The key design feature of the competitive condition was that color information was critical for 475 discriminating between pairmates. Specifically, in the competitive condition the only difference between 476 pairmates was a 24-degree color difference. This contrasts with the non-competitive condition where 477 pairmates differed in color (again 24 degrees) and object category. Because color information was 478 therefore more important in the competitive condition, we predicted that representation of color 479 information during the scanned recall trials would be relatively stronger in the competitive condition than 480 the non-competitive condition. Notably, participants' only instruction on the recall trials was to bring each 481 stimulus to mind as vividly as possible (mean percentage of vivid responses = 95.42%, SD = 5.43%). 482 Participants were not explicitly oriented to color information nor had participants' memory for color been 483 tested in any way to that point in the experiment.

484 To test for representation of color information, we computed the mean correlation of activity 485 patterns evoked during recall of non-pairmate stimuli that shared an identical color value (e.g., red bean 486 bag and red jacket; 'same-color' comparison, see Fig. 1b) and subtracted from this value the mean 487 correlation between non-pairmate stimuli that were 24 degrees apart in color space (e.g., red bean bag 488 and brown jacket; 'baseline' comparison, see Fig. 1b). Thus, the difference between these two measures 489 (same-color - baseline) provided an index of color information. We then compared this index across the 490 competitive and non-competitive trials. Critically, in terms of physical properties of the stimuli, the 491 comparison between the competitive and non-competitive trials was perfectly matched: there was no 492 objectively greater similarity between the stimuli included in this analysis in the competitive condition 493 compared to the non-competitive condition-there was only a difference in the importance of the 494 information.

For this and subsequent fMRI analyses we used a set of visual and parietal regions of interest (ROIs) previously described in Favila et al. (2018) (see Methods; Fig. 3a). Critically, these ROIs were previously shown to contain color and object feature representations during a memory recall task very similar to the current study. The set of ROIs included three visual ROIs (V1, LO, VTC) and five lateral parietal ROIs (pIPS, dLatIPS, vLatIPS, AnG, vIPS).

500 An ANOVA with factors of condition (competitive, non-competitive) and ROI (all eight ROIs) 501 revealed a significant main effect of condition, with relatively stronger color information in the competitive 502 condition than the non-competitive condition ( $F_{1,28} = 5.03$ , p = 0.033,  $\eta^2 = 0.04$ ). Neither the main effect of 503 ROI nor the condition x ROI interaction were significant (ROI:  $F_{4.55, 127.36} = 0.12$ , p = 0.984,  $\eta^2 < 0.001$ ; condition x ROI:  $F_{4,10, 114,92} = 0.78$ , p = 0.542,  $\eta^2 = 0.008$ ). Considering individual ROIs, only LO and vIPS 504 505 exhibited significantly stronger color representation in the competitive than non-competitive condition (LO: 506 t<sub>28</sub> = 2.27, p = 0.031, 95% CI = [0.002 0.03], Cohen's d = 0.69; vIPS: t<sub>28</sub> = 2.67, p = 0.012, 95% CI = 507 [0.004 0.03], Cohen's d = 0.63; paired t-tests, uncorrected; Fig 3b). Thus, as predicted, the greater 508 relevance of color information in the competitive condition resulted in stronger representation of color 509 information during recall, despite the fact that participants had not been explicitly oriented to color 510 information in any way by this point of the experiment (the critical behavioral test of color memory 511 occurred after fMRI scanning).

512 Post-hoc analyses of medial temporal and hippocampal ROIs (see Methods) did not reveal 513 stronger color representation in the competitive than non-competitive condition for any of the ROIs (|t|'s < 514 1.66, *p*'s > 0.109).

515

### 516 Neural similarity between pairmates during recall

517

We next tested whether neural similarity between pairmate stimuli was greater in the competitive than non-competitive condition. In terms of physical stimulus properties, pairmates were, of course, more similar in the competitive condition (e.g., two bean bags 24 degrees apart in color space) than in the noncompetitive condition (e.g., a pillow and a ball 24 degrees apart in color space). Thus, based on stimulus properties alone, fMRI pattern similarity between pairmates should be greater in the competitive condition than the non-competitive condition. To measure pairmate similarity we computed the mean correlation between pairmate stimuli ('pairmate' comparison, see Fig. 1b) and subtracted from this value the mean correlation between non-pairmate stimuli that were also 24 degrees apart in color space ('baseline' comparison, see Fig. 1b). The difference between these two values (pairmate – baseline) yielded an index of pairmate similarity which was then compared across the competitive and non-competitive conditions.

529 Although pairmate similarity was numerically greater in the competitive than non-competitive 530 condition across each of the eight ROIs, an ANOVA with factors of ROI and condition did not reveal a significant main effect of condition ( $F_{1,28}$  = 2.30, p = 0.140,  $\eta^2$  = 0.016). The main effect of ROI and the 531 532 condition x ROI interaction were also not significant (ROI:  $F_{4.57, 127.90} = 0.68$ , p = 0.626,  $\eta^2 = 0.006$ ; 533 condition x ROI:  $F_{3.82, 106.85} = 0.58$ , p = 0.670,  $\eta^2 = 0.006$ ). However, there was a significant effect of 534 condition, corrected for multiple comparisons (Bonferroni corrected), in vIPS, with greater pattern 535 similarity in the competitive than non-competitive conditions ( $t_{28} = 3.12$ , p = 0.004, 95% CI = [0.005 0.02], 536 Cohen's d = 0.70, paired t-test; Fig. 3c). Notably, as described above (Fig. 3b), vIPS also exhibited 537 significantly stronger color representation in the competitive than non-competitive condition. Moreover, 538 vIPS also exhibited significant object and color representations during a recall task in a prior study (Favila 539 et al., 2018). Thus, across two independent studies, we have consistently observed feature 540 representations in this ROI during memory recall.

541 Post-hoc analyses of medial temporal and hippocampal ROIs (see Methods) did not reveal 542 greater pairmate similarity in the competitive than non-competitive condition for any of the ROIs (|t|'s < 543 1.42, *p*'s > 0.168).

544

### 545 Neural measures of pairmate similarity predict color memory bias

546

Results from the preceding analysis revealed greater similarity in vIPS representations of pairmates in the competitive condition than the non-competitive condition. While this measure of neural similarity reflects the greater physical similarity between pairmates in the competitive condition than the non-competitive condition, the key finding from our behavioral results is that there is an adaptive benefit to *reducing*  551 similarity (in memory) between pairmates in the competitive condition. This raises the question of whether 552 similarity between vIPS representations of pairmates during competitive recall predicted the degree to 553 which there was repulsion of color memories (as measured in the post-scan color memory test). To test 554 this, for each condition (competitive, non-competitive) we correlated fMRI measures of pairmate 555 dissimilarity (1 - pattern similarity) with behavioral measures of mean signed color memory distance. This 556 analysis was performed within participant (i.e., at the level of individual pairmates). Given that each 557 condition only corresponded to 6 pairmates per participant, Spearman rank correlation was used in order 558 to reduce the influence of any one data point. Correlation coefficients were then z-transformed, yielding a 559 single z-transformed value for each condition and participant.

560 For the competitive condition, the mean correlation between pairmate dissimilarity in vIPS during 561 recall and mean signed color memory distance was significantly positive (vIPS:  $t_{28} = 3.75$ , p = 0.0008, 562 95% CI = [0.34 1.14], Cohen's d = 0.70, one-sample t-test; Fig. 4a). In other words, the more dissimilar 563 vIPS activity patterns were when recalling pairmates, the greater the color memory repulsion effect for 564 those pairmates. There was no correlation between pairmate dissimilarity in vIPS and signed color memory distance for the non-competitive condition ( $t_{28} = 0.78$ , p = 0.443, 95% CI = [-0.22 0.49], Cohen's 565 566 d = 0.14; Fig. 4a) and the difference between the competitive and non-competitive conditions was 567 significant ( $t_{28} = 2.39$ , p = 0.024, 95% CI = [0.09 1.12], Cohen's d = 0.61, paired *t*-test). Significant positive 568 relationships were also observed when pairmate dissimilarity was measured from pIPS, VTC, and 569 vLatIPS—again, only for the competitive condition (see table 1).

570 As a complementary analysis—and to better visualize the results in vIPS—we binned pairmates, 571 for each participant, based on vIPS dissimilarity (competitive condition only). We generated three bins per 572 participant: low, medium, and high pairmate dissimilarity. We then computed the mean signed color 573 memory distance for each of these bins. A one-way ANOVA revealed a significant main effect of pairmate 574 dissimilarity in vIPS on mean signed color memory distance (Fig. 4b;  $F_{1.75, 48.90} = 4.95$ , p = 0.014,  $\eta^2 =$ 575 0.062), with greater dissimilarity between vIPS representations associated with greater distance in 576 remembered color values (i.e., greater repulsion). We also computed mean accuracy on the associative 577 memory test for these same vIPS dissimilarity bins in order to more directly test whether vIPS dissimilarity 578 was associated with lower interference. Indeed, we again found a significant main effect of bin (F1.78, 49.87

579 = 4.52, p = 0.019,  $\eta^2 = 0.068$ ), with behavioral accuracy increasing as a function of pairmate dissimilarity 580 in vIPS. Finally, a mediation analysis performed at the level of individual pairmates (see Methods) 581 revealed that the relationship between vIPS dissimilarity and associative memory accuracy was 582 significantly mediated by signed color memory distance ( $\beta = 0.12$ ,  $CI = [0.02 \ 0.23]$ , p = 0.016, 1000 583 bootstrapped samples), consistent with the interpretation that vIPS dissimilarity reflected the degree of 584 color memory repulsion, which in turn was associated with better associative memory accuracy (lower 585 interference).

586

### 587 DISCUSSION

588

Here, we show that competition between similar memories triggers biases in their neural representations and corresponding behavioral expressions. Specifically, we demonstrate that subtle, diagnostic differences between events were exaggerated in long-term memory and that this exaggeration reduced interference. Critically, these behavioral expressions of memory distortion were predicted by adaptive, feature-specific changes to memory representations in parietal cortex.

594 Our behavioral paradigm was designed to isolate the effect that competition had on color memory. 595 Specifically, the competitive and non-competitive conditions had perfectly matched structures, with 596 equivalent color distances between pairmates in both conditions (Fig. 1b). The only difference was that 597 pairmates in the competitive condition were from the same object category. As intended, this increased 598 the number of interference-related errors, particularly during early stages of learning (Fig. 2a). The 599 increase in interference-related errors is consistent with a long history of behavioral studies of memory 600 interference (Anderson & Spellman, 1995; Mensink & Raaijmakers, 1988; Wixted, 2004). Our critical 601 question, however, was whether competition distorted memory for object features that were otherwise 602 successfully remembered. Results from the color memory post-test revealed a robust bias in color 603 memory in the competitive condition-that is, participants exaggerated the distance between pairmates-604 but no systematic bias in the non-competitive condition. We refer to the bias in the competitive condition 605 as a repulsion effect in order to emphasize that the bias was triggered by the representational proximity of

competing memories (Bae & Luck, 2017; Chanales et al., 2017, in-press; Golomb, 2015), just as spatial
proximity of like-poled magnets triggers magnetic repulsion.

608 It is important to emphasize that the repulsion effect is distinct from-in fact, opposite to-an 609 interference effect. That is, interference-related errors should lead participants to occasionally recall the 610 color of the competing object—an error that would produce a bias in color memory toward the pairmate 611 (Fig. 1c, d). Here, we did not test color memory until the very end of the experiment, so as to avoid 612 explicitly orienting participants to color information prior to (or during) the fMRI session, but our 613 speculation is that the repulsion effect only emerged after extensive practice and as interference errors 614 subsided (Chanales et al., in-press). In this sense, the repulsion effect can be thought of as an aftereffect 615 of initial memory interference. Although repulsion reflects a form of memory error, our findings indicate 616 that it is an adaptive error: participants who exhibited a stronger repulsion effect also exhibited fewer 617 interference-related errors (Fig. 2d). To the extent that objective similarity between stimuli is a root cause 618 of memory interference (Osgood, 1949), then exaggerating the difference between stimuli in memory is a 619 potentially powerful means for reducing interference (Chanales et al., in-press; Favila et al., 2016; Hulbert 620 & Norman, 2015).

621 Our fMRI analyses, which measured neural activity patterns as participants recalled object 622 images, provided a unique means for covertly probing the qualities of participants' memories. These 623 analyses revealed two forms of adaptive memory representations in parietal cortex. First, despite the fact 624 that participants were not instructed to think about or report objects' colors during these recall trials, we 625 observed stronger color information-across the full set of visual and parietal ROIs, and in vIPS 626 specifically-during competitive than non-competitive recall trials. The stronger representation of color 627 information during competitive trials can be viewed as an adaptive response to competition in that color 628 information was the only (or diagnostic) feature dimension for discriminating pairmates in the competitive 629 condition.

630 Second, although pairmate similarity in vIPS was stronger during competitive than non-631 competitive recall trials (indicating that vIPS was sensitive to object similarity; Fig. 3c), we found that 632 greater *dissimilarity* between vIPS pairmate representations during competitive recall trials was 633 associated with greater color memory repulsion and less memory interference. In other words, minimizing the overlap of neural representations of pairmates was an adaptive response to competition. This relationship was observed within participants, at the level of individual pairmates, but it is important to emphasize that these measures were temporally offset: vIPS pattern similarity was measured during recall trials in the scanner (with the only instruction being to recall objects as vividly as possible) whereas behavioral expressions of color memory were only tested after scanning was completed. This again makes the point that color information—in this case the subtle *difference* in pairmate colors—was a salient component of activity patterns in vIPS during competitive recall.

641 Importantly, when our two main fMRI findings are taken together, they indicate that an adaptive 642 response to competition involved an increase in similarity between stimuli that shared a diagnostic feature 643 value (i.e., objects of the same color) but a decrease in similarity between stimuli that had subtly different 644 values for a diagnostic feature (i.e., pairmates, which had slightly different colors). This indicates that 645 avoiding memory interference does not necessarily require a global reduction in similarity to all other 646 memories (LaRocque et al., 2013), but instead may be accomplished by more targeted changes in 647 representational structure that emphasize relevant similarities as well as important differences between 648 events that are stored in memory. Critically, this idea is distinct from-if not fundamentally incompatible 649 with-the traditional, and dominant view that interference is avoided through the orthogonalization of 650 memory representations (Colgin et al., 2008; Yassa & Stark, 2011). Specifically, whereas 651 orthogonalization emphasizes an initial encoding of new memories as independent from existing 652 memories, our findings instead emphasize that the representation of a given memory is highly dependent 653 on representations of other memories (Hulbert & Norman, 2015).

654 Our fMRI findings also add to a growing body of evidence that implicates parietal cortex in 655 actively representing content during memory retrieval (Kuhl & Chun, 2014; Lee & Kuhl, 2016; Lee et al., 656 2019; Rugg & King, 2018; Sestieri, Shulman, & Corbetta, 2017). Of most direct relevance, in a recent 657 study we found that vIPS (a ventral subregion of parietal cortex) actively represents color and object 658 category information during memory recall (Favila et al., 2018). However, this prior study focused on 659 decoding the objective properties of recalled stimuli and did not test whether competition influenced or 660 distorted these representations, nor did it establish a link between vIPS representations and behavioral 661 expressions of memory. The current findings provide unique evidence that representations within this

same vIPS subregion reflect subtle distortions in how events are remembered that are *dissociable from the objective properties of the event.* More generally, our findings highlight the behavioral relevance and
detailed nature of memory representations in parietal cortex.

665 While our findings provide strong evidence that representations in parietal cortex reflect the 666 influence that competition had on memory representations, it is not necessarily the case that parietal 667 cortex was the source of this influence. Rather, competition between memories is thought to induce 668 targeted plasticity in the hippocampus (Norman, Newman, & Detre, 2007; Ritvo, Turk-Browne, & Norman, 669 2019). In fact, hippocampal representations have been shown to specifically exaggerate differences 670 between highly similar stimuli (Ballard et al., 2019; Chanales et al., 2017; Dimsdale-Zucker et al., 2018; 671 Favila et al., 2016; Hulbert & Norman, 2015; Schapiro et al., 2012; Schlichting et al., 2015). However, 672 these exaggerations in hippocampal activity patterns have generally been observed during memory 673 encoding or perception (Ballard et al., 2019; Chanales et al., 2017; Dimsdale-Zucker et al., 2018; Favila 674 et al., 2016; Hulbert & Norman, 2015; Schapiro et al., 2012; Schlichting et al., 2015), as opposed to 675 memory recall, and they have not been translated to explicit feature spaces. Indeed, attempts to translate 676 hippocampal activity patterns to explicit feature dimensions or categories have tended to be unsuccessful 677 (LaRocque et al., 2013; Liang et al., 2013). In post hoc analyses, we did not find any evidence that 678 competition influenced feature representations in the hippocampus or medial temporal lobe ROIs. That 679 said, one notable aspect of our study is that each object was retrieved from memory many times before 680 fMRI scanning began. Given that repeated retrieval has specifically been shown to hasten the transfer of 681 representations to parietal cortex (Brodt et al., 2018, 2016), this raises the question of whether the 682 observed findings in parietal cortex were dependent on repeated retrieval. For example, it is possible that 683 competition induces exaggerated representations that are initially expressed in the hippocampus but 684 ultimately transformed, via retrieval, into stable representations in parietal cortex (Favila, Lee, & Kuhl, 685 2020). While the current study cannot address this question, it represents an interesting avenue for future 686 research.

In summary, our findings provide unique evidence that memory-based representations in parietal
 cortex exhibit adaptive, feature-specific changes in response to competition and that these changes in
 parietal representations predict distortions in behavioral expressions of memory. More generally, our

- 690 findings provide unique evidence in support of the perspective that memory distortions are an adaptive
- 691 component of the memory system (Schacter, Guerin, & St. Jacques, 2011).

| 693 | Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory        |
|-----|---|
| 694 | retrieval as a model case. Psychological Review. US: American Psychological Association.                      |
| 695 | https://doi.org/10.1037/0033-295X.102.1.68  |
| 696 | Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. Attention,      |
| 697 | Perception, and Psychophysics, 79(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8                    |
| 698 | Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports               |
| 699 | reinforcement learning. Nature Communications, 10(1), 1073. https://doi.org/10.1038/s41467-019-               |
| 700 | 08998-1   |
| 701 | Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. Journal of    |
| 702 | Experimental Psychology, 58(2), 97–105. https://doi.org/10.1037/h0047507                                      |
| 703 | Bone, M. B., Ahmad, F., & Buchsbaum, B. R. (2020). Feature-specific neural reactivation during episodic       |
| 704 | memory. Nature Communications, 11(1), 1945. https://doi.org/10.1038/s41467-020-15763-2                        |
| 705 | Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as       |
| 706 | bound units: Independent forgetting of different object details from visual memory. Journal of                |
| 707 | Experimental Psychology: General, 142(3), 791–808. https://doi.org/10.1037/a0029649                           |
| 708 | Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., & Schönauer, M. (2018). Fast track to the neocortex: A |
| 709 | memory engram in the posterior parietal cortex. Science, 362(6418), 1045 LP – 1048.                           |
| 710 | https://doi.org/10.1126/science.aau2528   |
| 711 | Brodt, S., Pöhlchen, D., Flanagin, V. L., Glasauer, S., Gais, S., & Schönauer, M. (2016). Rapid and           |
| 712 | independent memory formation in the parietal cortex. Proceedings of the National Academy of                   |
| 713 | Sciences, 113(46), 13251 LP – 13256. https://doi.org/10.1073/pnas.1605719113                                  |
| 714 | Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers     |
| 715 | Repulsion of Hippocampal Representations. Current Biology, 27(15), 2307-2317.e5.                              |
| 716 | https://doi.org/10.1016/j.cub.2017.06.057   |
| 717 | Chanales, A. J. H., Tremblay-McGaw, A. G., & Kuhl, B. A. (in-press). Adaptive repulsion of long-term          |
| 718 | memory representations is triggered by event similarity. Psychological Science.                               |
| 719 | https://doi.org/10.1101/2020.01.14.900381   |

| 720 | Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories   |
|-----|--|
| 721 | reveal shared structure in neural activity across individuals. Nature Neuroscience, 20(1).               |
| 722 | https://doi.org/10.1038/nn.4450  |
| 723 | Colgin, L. L., Moser, E. I., & Moser, MB. (2008). Understanding memory through hippocampal               |
| 724 | remapping. Trends in Neurosciences, 31(9), 469-477.  |
| 725 | https://doi.org/https://doi.org/10.1016/j.tins.2008.06.008   |
| 726 | Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. NMR in     |
| 727 | Biomedicine, 10(4-5), 171–178. https://doi.org/10.1002/(SICI)1099-                                       |
| 728 | 1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L   |
| 729 | Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and    |
| 730 | Surface Reconstruction. NeuroImage, 9(2), 179–194.   |
| 731 | https://doi.org/https://doi.org/10.1006/nimg.1998.0395   |
| 732 | Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and   |
| 733 | CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal           |
| 734 | subfields. Nature Communications, 9(1). https://doi.org/10.1038/s41467-017-02752-1                       |
| 735 | Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Gorgolewski, K. |
| 736 | J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods, 16(1),          |
| 737 | 111–116. https://doi.org/10.1038/s41592-018-0235-4   |
| 738 | Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern        |
| 739 | differentiation prevents interference during subsequent learning. Nature Communications, 7(1),           |
| 740 | 11066. https://doi.org/10.1038/ncomms11066   |
| 741 | Favila, S. E., Lee, H., & Kuhl, B. A. (2020). Transforming the Concept of Memory Reactivation. Trends in |
| 742 | Neurosciences. https://doi.org/https://doi.org/10.1016/j.tins.2020.09.006                                |
| 743 | Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal representations of stimulus   |
| 744 | features are amplified during memory retrieval and flexibly aligned with top-down goals. Journal of      |
| 745 | Neuroscience, 38(36), 0564–18. https://doi.org/10.1523/JNEUROSCI.0564-18.2018                            |
| 746 | Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear  |
| 747 | average age-appropriate brain templates from birth to adulthood. Neurolmage, 47, S102.                   |

|                         | 748 | https://doi.org/https://doi.org/10.1016/S1053-8119(09)70884-5   |
|-------------------------|-----|---|
|                         | 749 | Golomb, J. D. (2015). Divided spatial attention and feature-mixing errors. Attention, Perception, &       |
| ـب                      | 750 | Psychophysics, 77(8), 2562–2569. https://doi.org/10.3758/s13414-015-0951-0                                |
| Q                       | 751 | Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Ghosh, S. (2011).        |
|                         | 752 | Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Pyt              |
| C                       | 753 | . Frontiers in Neuroinformatics . Retrieved from  |
| S                       | 754 | https://www.frontiersin.org/article/10.3389/fninf.2011.00013  |
| $\square$               | 755 | Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based         |
|                         | 756 | registration. NeuroImage, 48(1), 63-72.   |
| J                       | 757 | https://doi.org/https://doi.org/10.1016/j.neuroimage.2009.06.060  |
| $\geq$                  | 758 | Hulbert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing        |
|                         | 759 | Memories Following Interleaved Study and Retrieval Practice. Cerebral Cortex, 25(10), 3994–40             |
| O                       | 760 | https://doi.org/10.1093/cercor/bhu284   |
|                         | 761 | Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and     |
| $\overline{\mathbf{O}}$ | 762 | Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage, 17(2), 825-84             |
|                         | 763 | https://doi.org/https://doi.org/10.1006/nimg.2002.1132  |
| Ö                       | 764 | Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted      |
| ŏ                       | 765 | memories. Journal of Neuroscience, 37(8), 2022–2031.  |
|                         | 766 | Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in     |
|                         | 767 | Lateral Parietal Cortex. Journal of Neuroscience, 34(23), 8051-8060.                                      |
| Ö                       | 768 | https://doi.org/10.1523/jneurosci.4328-13.2014  |
| S                       | 769 | Kuhl, B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable Neural Mechanisms for Goal-Directed        |
| 0                       | 770 | Versus Incidental Memory Reactivation. Journal of Neuroscience, 33(41), 16099–16109.                      |
|                         | 771 | https://doi.org/10.1523/jneurosci.0207-13.2013  |
|                         | 772 | LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Glob |
|                         | 773 | Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Mem                |
| 2                       | 774 | Journal of Neuroscience, 33(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013                |
|                         | 775 | Lee, H., & Kuhl, B. A. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in     |
|                         |     |   |

### https://doi.org/https://doi.org/10.1016/S1053-8119(09)70884-5 nb, J. D. (2015). Divided spatial attention and feature-mixing errors. Attention, Perception, & Psychophysics, 77(8), 2562-2569. https://doi.org/10.3758/s13414-015-0951-0 olewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Ghosh, S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python . Frontiers in Neuroinformatics . Retrieved from https://www.frontiersin.org/article/10.3389/fninf.2011.00013 e, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. NeuroImage, 48(1), 63-72. https://doi.org/https://doi.org/10.1016/j.neuroimage.2009.06.060 ert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. Cerebral Cortex, 25(10), 3994-4008. https://doi.org/10.1093/cercor/bhu284 nson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage, 17(2), 825-841. https://doi.org/https://doi.org/10.1006/nimg.2002.1132 G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted memories. Journal of Neuroscience, 37(8), 2022-2031. B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. Journal of Neuroscience, 34(23), 8051-8060. https://doi.org/10.1523/jneurosci.4328-13.2014 B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable Neural Mechanisms for Goal-Directed Versus Incidental Memory Reactivation. Journal of Neuroscience, 33(41), 16099–16109. https://doi.org/10.1523/jneurosci.0207-13.2013 cque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. Journal of Neuroscience, 33(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013

|                         | 776 | Lateral Parietal Cortex. Journal of Neuroscience, 36(22), 6069–6082.                              |
|-------------------------|-----|---|
|                         | 777 | https://doi.org/10.1523/JNEUROSCI.4286-15.2016  |
| ـب                      | 778 | Lee, H., Samide, R., Richter, F. R., & Kuhl, B. A. (2019). Decomposing Parietal Memory Reactive   |
| Q                       | 779 | Predict Consequences of Remembering. Cerebral Cortex, 29(8), 3305–3318.                           |
|                         | 780 | https://doi.org/10.1093/cercor/bhy200   |
| C                       | 781 | Liang, J. C., Wagner, A. D., & Preston, A. R. (2013). Content Representation in the Human Med     |
| S                       | 782 | Temporal Lobe. Cerebral Cortex, 23(1), 80–96. https://doi.org/10.1093/cercor/bhr379               |
| $\square$               | 783 | Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal Mismatch Signals Are Modulated by th      |
|                         | 784 | of Neural Predictions and Their Similarity to Outcomes. Journal of Neuroscience, 36(50),          |
| $\overline{\mathbf{D}}$ | 785 | 12687.  |
| $\geq$                  | 786 | Mensink, GJ., & Raaijmakers, J. G. (1988). A model for interference and forgetting. Psycholog     |
|                         | 787 | Review, 95(4), 434–455. https://doi.org/10.1037/0033-295X.95.4.434                                |
| Ō                       | 788 | Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activ    |
| Ð                       | 789 | event-related designs for multivoxel pattern classification analyses. NeuroImage, 59(3), 26       |
| $\mathbf{O}\mathbf{f}$  | 790 | https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.08.076                                  |
|                         | 791 | Norman, K. A., Newman, E. L., & Detre, G. (2007). A Neural Network Model of Retrieval-Induce      |
| Ö                       | 792 | Forgetting. Psychological Review, 114(4), 887–953. https://doi.org/10.1037/0033-295X.11           |
| Ö                       | 793 | O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and rec   |
|                         | 794 | Avoiding a trade-off. Hippocampus, 4(6), 661-682. https://doi.org/10.1002/hipo.45004060           |
|                         | 795 | Osgood, C. E. (1949). The similarity paradox in human learning: a resolution. Psychological Re    |
| Ö                       | 796 | American Psychological Association. https://doi.org/10.1037/h0057488                              |
| S                       | 797 | Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Men     |
| 0                       | 798 | Retrieval Drives Learning. Trends in Cognitive Sciences, 23(9), 726–742.                          |
|                         | 799 | https://doi.org/10.1016/j.tics.2019.06.007  |
|                         | 800 | Rugg, M. D., & King, D. R. (2018). Ventral lateral parietal cortex and episodic memory retrieval. |
|                         | 801 | 107, 238-250. https://doi.org/https://doi.org/10.1016/j.cortex.2017.07.012                        |
| 2                       | 802 | Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive pers  |
|                         | 803 | Trends in Cognitive Sciences, 15(10), 467–474. https://doi.org/10.1016/j.tics.2011.08.004         |
|                         |     |   |

| 776 | Lateral Parietal Cortex. Journal of Neuroscience, 36(22), 6069-6082.                                      |
|-----|---|
| 777 | https://doi.org/10.1523/JNEUROSCI.4286-15.2016  |
| 778 | Lee, H., Samide, R., Richter, F. R., & Kuhl, B. A. (2019). Decomposing Parietal Memory Reactivation to    |
| 779 | Predict Consequences of Remembering. Cerebral Cortex, 29(8), 3305–3318.                                   |
| 780 | https://doi.org/10.1093/cercor/bhy200   |
| 781 | Liang, J. C., Wagner, A. D., & Preston, A. R. (2013). Content Representation in the Human Medial          |
| 782 | Temporal Lobe. Cerebral Cortex, 23(1), 80–96. https://doi.org/10.1093/cercor/bhr379                       |
| 783 | Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal Mismatch Signals Are Modulated by the Strength    |
| 784 | of Neural Predictions and Their Similarity to Outcomes. Journal of Neuroscience, 36(50), 12677-           |
| 785 | 12687.  |
| 786 | Mensink, GJ., & Raaijmakers, J. G. (1988). A model for interference and forgetting. Psychological         |
| 787 | <i>Review</i> , <i>95</i> (4), 434–455. https://doi.org/10.1037/0033-295X.95.4.434                        |
| 788 | Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in    |
| 789 | event-related designs for multivoxel pattern classification analyses. NeuroImage, 59(3), 2636–2643.       |
| 790 | https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.08.076  |
| 791 | Norman, K. A., Newman, E. L., & Detre, G. (2007). A Neural Network Model of Retrieval-Induced             |
| 792 | Forgetting. Psychological Review, 114(4), 887–953. https://doi.org/10.1037/0033-295X.114.4.887            |
| 793 | O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall:       |
| 794 | Avoiding a trade-off. Hippocampus, 4(6), 661–682. https://doi.org/10.1002/hipo.450040605                  |
| 795 | Osgood, C. E. (1949). The similarity paradox in human learning: a resolution. Psychological Review. US:   |
| 796 | American Psychological Association. https://doi.org/10.1037/h0057488                                      |
| 797 | Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory          |
| 798 | Retrieval Drives Learning. Trends in Cognitive Sciences, 23(9), 726–742.                                  |
| 799 | https://doi.org/10.1016/j.tics.2019.06.007  |
| 800 | Rugg, M. D., & King, D. R. (2018). Ventral lateral parietal cortex and episodic memory retrieval. Cortex, |
| 801 | 107, 238–250. https://doi.org/https://doi.org/10.1016/j.cortex.2017.07.012                                |
| 802 | Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective.  |

| 804 | Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the    |
|-----|---|
| 805 | human medial temporal lobe based on temporal regularities. Current Biology, 22(17), 1622–1627.            |
| 806 | https://doi.org/10.1016/j.cub.2012.06.056   |
| 807 | Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes    |
| 808 | reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex.        |
| 809 | Nature Communications, 6(1), 8151. https://doi.org/10.1038/ncomms9151                                     |
| 810 | Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal     |
| 811 | cortex to episodic memory. Nature Reviews Neuroscience, 18(3), 183–192.                                   |
| 812 | https://doi.org/10.1038/nrn.2017.6  |
| 813 | Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory.       |
| 814 | Hippocampus, 4(3), 374–391. https://doi.org/10.1002/hipo.450040319  |
| 815 | Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). |
| 816 | N4ITK: Improved N3 Bias Correction. IEEE Transactions on Medical Imaging, 29(6), 1310–1320.               |
| 817 | https://doi.org/10.1109/TMI.2010.2046908  |
| 818 | Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2014). Probabilistic Maps of Visual Topography |
| 819 | in Human Cortex. Cerebral Cortex, 25(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277                |
| 820 | Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. Annual Review of Psychology,         |
| 821 | 55(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555                                    |
| 822 | Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern       |
| 823 | Reinstatement during Episodic Memory Retrieval. The Journal of Neuroscience, 37(11), 2986–2998.           |
| 824 | https://doi.org/10.1523/JNEUROSCI.2324-16.2017  |
| 825 | Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. Trends in Neurosciences,   |
| 826 | 34(10), 515–525. https://doi.org/https://doi.org/10.1016/j.tins.2011.06.006                               |
| 827 | Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Buckner, R. L.   |
| 828 | (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity.     |
| 829 | Journal of Neurophysiology, 106(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011                      |
| 830 | Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, SL., Gertje, E. C., Wolk, D. A. (2015).         |

831 Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal

- 832 cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36(1), 258–287.
- 833 https://doi.org/10.1002/hbm.22627

Table 1. Summary of key statistical analyses. Color representation analyses refer to paired-samples t-tests comparing color similarity effects (see Methods) for the competitive vs. non-competitive conditions. Pairmate similarity analyses refer to paired-samples t-tests comparing pairmate similarity effects (see Methods) for the competitive vs. non-competitive conditions. The relation to mean signed distance refers to one-sample t-tests comparing z-transformed correlations between fMRI pairmate dissimilarity and mean signed color memory distance to a test statistic of 0 (no relationship). Results from individual visual and parietal ROIs are presented in separate rows. Note: \* p < .05, uncorrected; \*\* p < .05, Bonferroni corrected; \*\*\* *p* < .01, Bonferroni corrected. 

|         | Color representation   |        | Pairmate similarity    |         | Relation to mean signed distance |           |                        |       |
|---------|------------------------|--------|------------------------|---------|----------------------------------|-----------|------------------------|-------|
| ROI     |                        |        |                        |         | Competitive                      |           | Non-competitive        |       |
|         | <b>t</b> <sub>28</sub> | p      | <b>t</b> <sub>28</sub> | p       | <b>t</b> <sub>28</sub>           | р         | <b>t</b> <sub>28</sub> | p     |
| V1      | 1.22                   | 0.232  | 0.89                   | 0.382   | 0.82                             | 0.417     | -0.34                  | 0.734 |
| LO      | 2.27                   | 0.031* | 1.71                   | 0.098   | 1.34                             | 0.190     | -0.75                  | 0.458 |
| VTC     | 1.16                   | 0.257  | 0.45                   | 0.653   | 2.13                             | 0.042*    | 0.59                   | 0.558 |
| pIPS    | 1.85                   | 0.075  | 0.84                   | 0.409   | 3.08                             | 0.005**   | 1.08                   | 0.289 |
| dLatIPS | 1.68                   | 0.104  | 0.73                   | 0.472   | 1.50                             | 0.145     | 0.65                   | 0.520 |
| vLatIPS | 1.69                   | 0.101  | 0.52                   | 0.609   | 2.92                             | 0.007**   | -1.89                  | 0.069 |
| AnG     | 0.57                   | 0.573  | 0.36                   | 0.720   | 0.75                             | 0.462     | -0.72                  | 0.475 |
| vIPS    | 2.67                   | 0.012* | 3.12                   | 0.004** | 3.75                             | 0.0008*** | 0.78                   | 0.443 |

904

905

907

855 Fig. 1. Experimental Design and Procedure. a. Overview of paradigm. On Day 1, participants 856 completed 14 Study and Associative Memory Test rounds. During Study, participants were shown object-857 face pairs and during Associative Memory Test, participants were shown an object and selected the 858 corresponding face from a set of four choices. The set of four choices included the target face along with 859 the face associated with the object's pairmate. On Day 2, participants completed four additional Study 860 and Associative Memory Test rounds before entering the fMRI scanner. During scanning, participants 861 completed a Cued Recall task during which face images were shown and participants recalled the 862 corresponding image and indicated, by button press, the vividness of their recall. After exiting the scanner, 863 participants completed a Color Memory Test during which a face image was shown alongside a greyscale 864 version of the corresponding object. Participants used a continuous color wheel to indicate their memory 865 for the object's color. Finally participants completed 2 more Associative Memory Test rounds. b, Sample 866 structure of object stimuli. For both the competitive and non-competitive conditions, pairmate stimuli were 867 24 degrees apart in color space. For the competitive condition, pairmates were from the same object 868 category; for the non-competitive condition, pairmates were from distinct categories. For both conditions, 869 some objects had identical colors (Same-color). fMRI pattern similarity for Pairmate and Same-color 870 comparisons were compared against a Baseline comparison of stimuli that were from different object 871 categories and 24 degrees apart in color space. c,d, Responses on the color memory test were used to 872 categorize memory for each object's color as being biased toward or away from the color of the 873 competing object (c) and to measure the signed distance, in degrees, between participants' responses 874 and the true color of the target (d). 875

876 Fig. 2. Behavioral results. a, Associative memory performance across the experiment. The overall error 877 rate (pairmate error + other error) was higher in the competitive condition than the non-competitive 878 condition for each of the associative memory test sessions (Day 1, Day 2 pre-scan, Day 2 post-scan (not 879 shown); all p's < 0.0001). Subsequent analyses focused on associative memory performance from the 880 Day 2 pre-scan session. For the Day 2 pre-scan session, participants were significantly more likely to 881 select faces that were associated with the pairmate image (pairmate error) in the competitive condition (M 882 = 6.0%, SD = 6.6%) compared to the non-competitive condition (M = 0.2%, SD = 0.6%; p < 0.0001), 883 confirming that similarity between pairmates was a source of interference. b, Signed distance of 884 responses in the color memory test. For the competitive condition, mean signed distance was significantly 885 greater than 0 (p = 0.000003), reflecting a bias away from the color of the pairmate object (repulsion). 886 Signed distance did not differ from 0 in the non-competitive condition (p = 0.771). The difference between 887 the competitive and non-competitive conditions was also significant (p = 0.007). c, Percentage of away 888 responses in the color memory test. The percentage of color memory responses 'away from' the color of 889 the pairmate object was significantly greater than 50% for the competitive condition (p = 0.0001), but not 890 for the non-competitive condition (p = 0.189). The difference between the competitive and non-891 competitive conditions was also significant (p = 0.001). **d**, Relationship between associative memory 892 accuracy and mean signed color memory distance. For the competitive condition, participants with 893 greater mean signed color memory distance (greater repulsion) exhibited better associative memory 894 accuracy [r = 0.50, p = 0.007, one outlier (red dot) excluded for associative memory performance < 3 SD 895 below mean]. Notes: colored dots reflect data from individual participants. Error bars reflect +/- S.E.M.; \*\*\* 896 p < 0.001, \*\* p < 0.01 897

898 Fig. 3. Neural feature representations as a function of memory competition. a, Anatomical ROIs 899 visualized on the Freesurfer average cortical surface. b, Color information as a function of memory 900 competition. Color information was defined as the fMRI pattern similarity between pairs of same-color 901 objects relative to pattern similarity between baseline pairs of objects (see Fig. 1b). Color information was 902 significantly stronger in the competitive than non-competitive condition (i.e., values greater than 0) across 903 the set of ROIs as a whole and in LO and vIPS individually (p's < .05). c, Pairmate similarity as a function of memory competition. Pairmate similarity was defined as the fMRI pattern similarity between pairmate objects relative to pattern similarity between baseline pairs of objects. Only vIPS showed significantly 906 greater pairmate similarity in the competitive than non-competitive conditions (p = 0.004). Error bars reflect +/- S.E.M.; \*\* p < 0.01, \* p < 0.05 908

Fig. 4. Neural measures of pairmate (dis)similarity predict color memory bias in vIPS. a, Mean 909 910 correlation between vIPS pairmate dissimilarity during recall and mean signed color memory distance. 911 Correlations were performed within participant and correlation coefficients were z-transformed. For the 912 competitive condition, the mean correlation was significantly positive (p = 0.004), indicating that greater 913 pairmate dissimilarity in vIPS was associated with a stronger bias to remember pairmates' colors as away 914 from each other. There was no correlation between vIPS pairmate dissimilarity and signed color memory 915 distance for the non-competitive condition (p = 0.566). **b**, Relationship between vIPS pairmate 916 dissimilarity (binned into low, medium, high groups) and mean signed color memory distance (purple) and 917 associative memory accuracy (teal). Mean signed color memory distance and associative memory 918 accuracy each significantly varied as a function of vIPS dissimilarity (p's < .05), with greater vIPS 919 dissimilarity associated with greater mean signed color memory distance and higher associative memory 920 accuracy. \*\*\* p < 0.001, \* p < 0.05

# **JNeurosci Accepted Manuscript**

а



1 168

# **JNeurosci Accepted Manuscript**





# **JNeurosci Accepted Manuscript**

