

# Neural Global Pattern Similarity Underlies True and False Memories

Zhifang Ye,<sup>1,2</sup> Bi Zhu,<sup>1,2</sup> Liping Zhuang,<sup>1,2</sup> Zhonglin Lu,<sup>3</sup> Chuansheng Chen,<sup>4</sup> and Gui Xue<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China, <sup>2</sup>Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, 100875, China, <sup>3</sup>Department of Psychology, Ohio State University, Columbus, Ohio 43210, and <sup>4</sup>Department of Psychology and Social Behavior, University of California, Irvine, California 92697

The neural processes giving rise to human memory strength signals remain poorly understood. Inspired by formal computational models that posit a central role of global matching in memory strength, we tested a novel hypothesis that the strengths of both true and false memories arise from the global similarity of an item's neural activation pattern during retrieval to that of all the studied items during encoding (i.e., the encoding-retrieval neural global pattern similarity [ER-nGPS]). We revealed multiple ER-nGPS signals that carried distinct information and contributed differentially to true and false memories: Whereas the ER-nGPS in the parietal regions reflected semantic similarity and was scaled with the recognition strengths of both true and false memories, ER-nGPS in the visual cortex contributed solely to true memory. Moreover, ER-nGPS differences between the parietal and visual cortices were correlated with frontal monitoring processes. By combining computational and neuroimaging approaches, our results advance a mechanistic understanding of memory strength in recognition.

**Key words:** DRM; false memory; fMRI; global matching models; neural global pattern similarity; recognition

## Significance Statement

What neural processes give rise to memory strength signals, and lead to our conscious feelings of familiarity? Using fMRI, we found that the memory strength of a given item depends not only on how it was encoded during learning, but also on the similarity of its neural representation with other studied items. The global neural matching signal, mainly in the parietal lobule, could account for the memory strengths of both studied and unstudied items. Interestingly, a different global matching signal, originated from the visual cortex, could distinguish true from false memories. The findings reveal multiple neural mechanisms underlying the memory strengths of events registered in the brain.

## Introduction

A fundamental question of human recognition memory concerns the cognitive and neural representations and processes that give rise to the memory strength signals, which lead to our conscious feelings of familiarity and guide a variety of memory decisions. Formal computational models of memory (i.e., the

global matching models) have posited that the memory strength of a given item derives from the match (measured as similarity) between its representation and those of all the other studied items (Murdock, 1982; Gillund and Shiffrin, 1984; Hintzman, 1984; Pike, 1984). The magnitude of matching or memory strength is then subjected to a decision making process (e.g., signal detection model) to determine the response in a recognition memory task.

The global matching models provide an algorithmic explanation for why recognition memory is affected by the similarity of other studied items and why false memory occurs. According to the global matching models, false memory occurs when there is high similarity between the unstudied item and stored memories, due to the overlap in item and/or contextual information. In particular, the sum of many partial matches to memory traces can provide a strong overall match, leading to the global similarity effect (Humphreys et al., 1989; Clark and Gronlund, 1996). Consistently, unstudied category

Received Feb. 8, 2016; revised May 8, 2016; accepted May 20, 2016.

Author contributions: B.Z. and G.X. designed research; Z.Y. and L.Z. performed research; Z.Y., B.Z., and G.X. analyzed data; Z.Y., B.Z., Z.L., C.C., and G.X. wrote the paper.

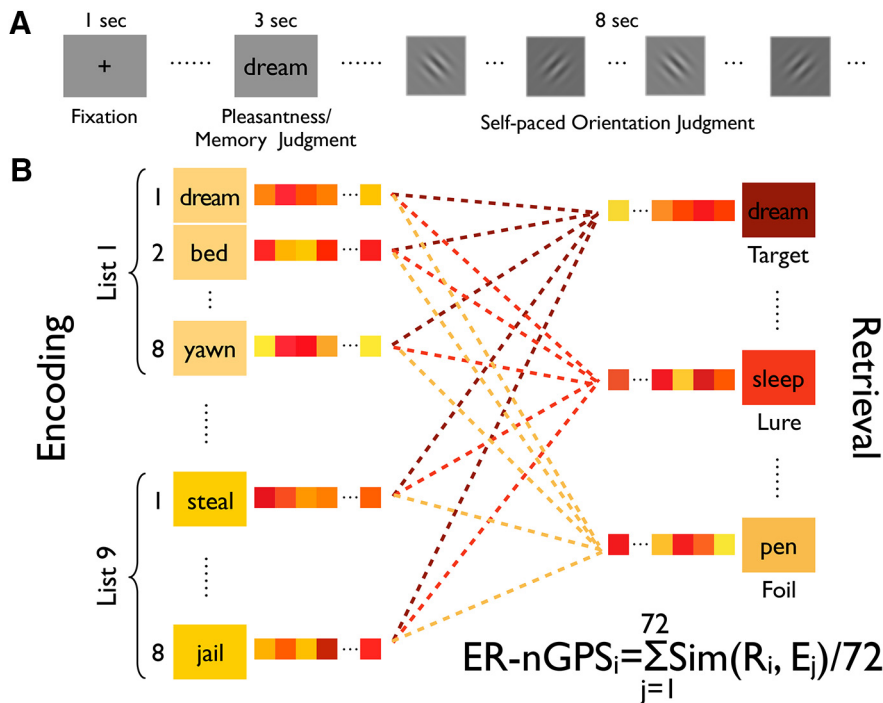
This work was supported by the National Natural Science Foundation of China 31130025, 973 Program 2014CB846102, 111 Project B07008, and NSFC Project 31571132 and 31521063. We thank Russell Poldrack, Per Sederberg, and Adam Osth for comments on earlier versions of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Gui Xue, State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China. E-mail: guixue@gmail.com.

DOI:10.1523/JNEUROSCI.0425-16.2016

Copyright © 2016 the authors 0270-6474/16/366792-11\$15.00/0



**Figure 1.** Schematic depiction of the experimental design and data analysis procedure. **A**, Experimental design. The slow event-related design (each trial lasting 12 s) was used for both encoding and retrieval phases of the task. Each trial started with a 1 s fixation point, followed by a Chinese word (e.g., “Dream”) that was presented on the screen for 3 s. Three seconds after the onset of the word, participants were asked to perform a perceptual orientation judgment task for 8 s. A self-paced procedure was used to make this task engaging. During the encoding phase, participants were asked to make a pleasantness judgment on each word (i.e., from 1 = “Very unpleasant” to 4 = “Very pleasant”) by pressing 1 of 4 buttons. Over 3 scanning runs, 72 words (9 lists of 8 words) were studied. During the retrieval phase, participants were asked to make a memory judgment on each word (i.e., from 1 = “Definitely new” to 4 = “Definitely old”) by pressing 1 of 4 buttons. **B**, A depiction of the neural global pattern similarity measure between encoding and retrieval. The global pattern similarity between encoding and retrieval (ER-nGPS) was calculated by averaging the Fisher’s *z*-scores reflecting neural activation pattern similarity of each test item with the neural activation pattern of all studied items ( $n = 72$ ).

prototypes show high false-alarm rates (Posner and Keele, 1970; Bransford and Franks, 1971), and both the hit and false-alarm rates increase as the number of similar items in a list is increased (Hintzman, 1988).

The neural implementation of the global matching mechanism is barely understood. According to the global matching models, one would predict that memory strength signal is scaled to the similarity between a test item’s neural activation pattern during retrieval and that of all studied items during encoding (i.e., the encoding-retrieval neural global pattern similarity [ER-nGPS]). In partial support of this hypothesis, emerging studies found that the similarity between the neural activation pattern of a studied item during encoding to that of all other studied items (i.e., nGPS), was predictive of subsequent memory (LaRocque et al., 2013; Davis et al., 2014a; Lu et al., 2015). Nevertheless, because the neural responses at the retrieval phase were not recorded, these studies could not examine ER-nGPS and its potential role in true and false memory.

Moreover, the global matching models need to deal with important differences between true and false memories, both at the behavioral and neural levels. At the behavioral level, the global matching models suggest either a cubing mechanism (i.e., the MINERVA 2) (Hintzman, 1984) or a multiplicative cue mechanism (i.e., the SAM model) (Gillund and Shiffrin, 1984) that would increase the weight of an exact match (true memory) and reduce the weight of multiple partial matches

(false memory). These mechanisms provide a good explanation of false recall (Kimball et al., 2007) and the effects of stimulus duration, number of exemplars, and associative strength on true and false memory (Arndt and Hirshman, 1998), but have difficulty explaining the effect of stimulus duration and repetition on the variance of memory (Ratcliff et al., 1990). At the neural level, neuroimaging studies found that, compared with false memory, true memory is associated with increased activity in sensory cortices (Schacter et al., 1996; Abe et al., 2008; Atkins and Reuter-Lorenz, 2011; Dennis et al., 2012), but reduced activity in the bilateral prefrontal cortices (Kensinger and Schacter, 2006; Kubota et al., 2006; Garoff-Eaton et al., 2007; Kim and Cabeza, 2007). These neural differences have yet to be integrated into the global matching models.

The present study aimed to address these issues with fMRI and an adapted Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995). To examine the ER-nGPS, research participants were scanned during both the encoding and retrieval phases of the memory task. We hypothesized that the memory strength for both true and false memories should be scaled to the ER-nGPS. We further predicted that both common and differential ER-nGPS signals underlay true and false memories.

## Materials and Methods

**Participants.** Participants were 35 healthy college students (19 males; mean age = 23.0 ± 1.9 years, range 19–27 years), who had normal or corrected-to-normal vision, and no history of psychiatric or neurological diseases. Written consent was obtained from each participant after a full explanation of the study procedure. The study was approved by the Institutional Review Board at Beijing Normal University.

**Experimental materials.** Nine word lists, each containing 12 words that describe one theme, were used in this study. They were translated and adapted from materials used in Roediger and McDermott (1995). For example, one list included words “dream,” “awake,” “bed,” “doze,” “yawn,” “snore,” “drowsy,” “blanket,” “sleep,” “rest,” “tired,” and “pillow” (Fig. 1A). Of the 12 words in each list, 8 were studied (only 4 of them would be tested, i.e., targets) and 4 were used as critical lures (see below). In addition, 36 semantically unrelated words were used as foils in the recognition test.

**fMRI task.** Laying supine on the scanner bed, participants viewed visual stimuli back-projected onto a screen through a mirror attached onto the head coil. Foam pads were used to minimize head motion. Stimulus presentation and timing of all stimuli were achieved using MATLAB (The MathWorks) and Psychtoolbox (www.psychtoolbox.org) on a Windows PC.

During the encoding phase, participants were explicitly instructed to intentionally memorize each word presented on the screen and were told that a recognition test would be conducted after the encoding session. They were also warned that there would be some unstudied words that were semantically related to the studied words in the following recognition test. Participants studied 72 words (i.e., 9 word lists) in the scanner

over three scanning runs, each containing three word lists. Before the start of each word list, there was a 2 s visual cue (i.e., “List 1”). Each word was presented only once. The order of the 8 studied words within each word list and that of the 9 word lists was counterbalanced across participants. A slow event-related design (12 s for each trial) was used in this study (Fig. 1). Each trial started with a 1 s fixation point, followed by a Chinese word that was presented on the screen for 3 s. To help the participants to remember these words, participants were asked to make a pleasantness judgment on the word by pressing 1 of 4 buttons (1 = “Very unpleasant,” 2 = “Mildly unpleasant,” 3 = “Mildly pleasant,” 4 = “Very pleasant”). Three seconds after the onset of the word, participants were asked to perform a perceptual judgment task for 8 s, which was included to prevent the participants from further processing the studied words. A Gabor image tilting 45° to the left or the right was randomly selected and presented on the screen, and participants were asked to identify the orientation of the Gabor by pressing one of two buttons. Participants were asked to respond as quickly and accurately as possible. The next trial started 100 ms after participants responded.

After studying all 9 lists, participants were given a 2-back working memory task for 10 min before they took the recognition test. The working memory task served as a distractor task between the encoding and retrieval phases and allowed time for an anatomic MRI scan. During the retrieval phase, participants were asked to judge whether they had studied the words earlier by pressing 1 of 4 buttons (1 = “Definitely new,” 2 = “Probably new,” 3 = “Probably old,” 4 = “Definitely old”). These confidence responses were used to index memory strength. The use of right versus left hand for old versus new response was counterbalanced across participants. In total, 108 words (36 target words, 36 critical lures, and 36 foils) were presented over three scanning sessions, and the order was pseudorandomized. Following the procedure used in previous studies (Gerard et al., 1991; Duverne et al., 2009), three unstudied foil words were placed at the beginning of each test run. The same slow event-related design (12 s for each trial) as in the study phase was used for the retrieval phase.

**Postscan semantic similarity rating.** Immediately after the scan, a semantic similarity rating task was given to the participants to assess pairwise semantic similarity of the tested items. The large number of words (108 in total) made it very difficult to do complete pairwise ratings (5778 pairs in total). Therefore, participants were asked only to rate the similarity between 8 studied words and 4 critical lures within each word list because previous studies suggested that the false memory effect was mainly contributed by the within-list similarity. There were 594 word pairs in total (i.e., 66 pairs/list  $\times$  9 lists). Within each list, the 66 pairs were randomly mixed together. For each pair, the participants were asked to judge semantic similarity between the two words using a 6 point scale, with 1 = “Very weak semantic association” and 6 = “Very strong semantic association.” The stimulus remained on the screen until a response was made. There was no time limit for responses on this rating task.

**Behavioral data analysis.** Paired-samples *t* test was used to examine the differences in the endorsement rates of target, lure, and foil items and associated reaction times in the recognition test. To measure the effect of semantic similarity on recognition memory, Pearson correlation was calculated between the memory scores in the recognition test and semantic global similarity ratings (i.e., sGS), which were obtained by averaging the semantic similarity of each tested item to all studied words within a list. These analyses were conducted separately for targets and lures.

**fMRI data acquisition and preprocessing.** Imaging data were acquired on a 3.0 T Siemens MRI scanner in the MRI Center at Beijing Normal University. A single-shot T2\*-weighted gradient-echo, EPI sequence was used for functional imaging acquisition with the following parameters: TR/TE/ $\theta$  = 2000 ms/30 ms/90°, FOV = 192  $\times$  192 mm, matrix = 64  $\times$  64, and slice thickness = 3.3 mm. Forty-one contiguous axial slices parallel to the AC-PC line were obtained to cover the whole cerebrum and partial cerebellum. Anatomical MRI was acquired using a T1-weighted, 3D, gradient-echo pulse-sequence (MPRAGE). The parameters for this sequence were as follows: TR/TE/ $\theta$  = 2530 ms/3.39 ms/7°, FOV = 256  $\times$  256 mm, matrix = 256  $\times$  256, and slice thickness = 1 mm. A total of 144 sagittal slices were acquired to provide high-resolution structural images of the whole brain.

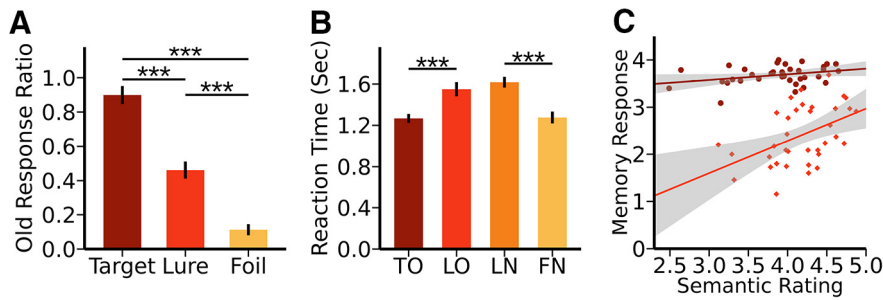
Image preprocessing and statistical analysis were performed using FEAT (FMRIB Expert Analysis Tool) version 6.00, part of the FSL (FMRIB software library, version 5.0.9, www.fmrib.ox.ac.uk/fsl). The first 3 volumes before the task were automatically discarded by the scanner to allow for T1 equilibrium. The remaining images were then realigned to correct for head movements. Translational movement parameters never exceeded 1 voxel in any direction for any participant or session. Data were spatially smoothed using a 5 mm FWHM Gaussian kernel. The data were filtered in the temporal domain using a nonlinear high-pass filter with a 90 s cutoff. A two-step registration procedure was used whereby EPI images were first registered to the MPRAGE structural image and then into the standard MNI space using affine transformations (Jenkinson and Smith, 2001). Registration from MPRAGE structural image to the standard space was further refined using FNIRT nonlinear registration (Andersson et al., 2007a, b).

**Univariate activation analysis.** The GLM within the FILM module of FSL was used to model the data. During the retrieval phase, four types of trials were modeled: targets judged as old (TO), lures judged as old (LO), lures judged as new (LN), and foils judged as new (FN). The targets judged as new and foils judged as old were rare and were included as two separate nuisance variables. The incorrect trials of the perceptual orientation task were coded as an additional nuisance variable, whereas the correct orientation trials were not coded and thus were treated as an implicit baseline. To control for the differences in reaction time, reaction times for all items were included as one parametric regressor. Three contrasts (TO vs FN, LO vs LN, and TO vs LO) were defined to examine the effect of true memory, false memory, and their direct comparison, respectively. The contrast between activations in LN and FN was used to examine the region(s) involved in cognitive control. A higher-level analysis based on a fixed-effects model created cross-run contrasts for each participant. These contrasts were then used for group analysis with a random-effects model, using FLAME (FMRIB’s Local Analysis of Mixed Effect) Stage 1 (Beckmann et al., 2003; Woolrich et al., 2004). Unless otherwise noted, group images were thresholded using cluster detection statistics, with a height threshold of  $z > 2.3$  and a cluster probability of  $p < 0.05$ , corrected for whole-brain multiple comparisons using Gaussian Random Field Theory. The same threshold was used for both univariate and ER-nGPS analysis.

**Single-item response estimation.** GLM was used to compute the  $\beta$ -map for each of the 72 unique word stimuli during encoding and 108 words during retrieval. In this single-trial model, the presentation of each stimulus was modeled as an impulse and convolved with a canonical hemodynamic response function (double gamma) (Mumford et al., 2012). The least-square single method was used to obtain reliable estimates of single trial responses. The  $\beta$  values of each stimulus were used to calculate neural global similarity and then used for further statistical analysis (see below).

**Neural global pattern similarity between encoding and retrieval (ER-nGPS).** The searchlight method was used to locate brain regions whose neural global pattern similarity between encoding and retrieval was associated with memory strength (Kriegeskorte et al., 2006). For each voxel, signals were extracted from the cubic ROI containing 125 surrounding voxels. For each tested item, paired-wise Pearson correlation was calculated between the activation patterns of this item during retrieval with the activation pattern of all studied words during encoding. These similarity scores were transformed into Fisher’s *z*-scores and then averaged to generate the ER-nGPS value. The ER-nGPS for each type of trials (TO, LO, LN, and FN) were then separately averaged and contrasted at individual participant level. A random-effects model was used for group analysis. Because no first-level variance was available, an ordinary least square model was used.

**ROI analysis.** We defined five ROIs based on the whole-brain searchlight results, including the left inferior frontal gyrus (LIFG), left inferior parietal lobule (LIPL), left superior parietal lobule (LSPL), left ventral lateral occipital complex (LvLOC), and right ventral lateral occipital complex (RvLOC). They were defined by including all the voxels in each cluster showing suprathreshold activation for the contrast between TO and FN. The mean ER-nGPS, activation level, and variance for each trial were then extracted and correlated with other measures, including the



**Figure 2.** Behavioral results during retrieval and after the scan. **A**, The endorsement (judged as old during retrieval) rate of target, lure, and foil items. **B**, The reaction time of TO, LO, LN, and FN items. Error bars indicate within-participant SE. **C**, Scatter plot and regression lines indicate the relationship between global semantic similarity for each retrieval item with all studied items and memory performance for target (shown in dark red) and lure items (shown in light red).  $***p < 0.001$ .

pattern similarity and mean activation level in other regions. To examine the ER-nGPS in the medial temporal lobe (MTL), four ROIs were defined based on the Harvard-Oxford probabilistic atlas (threshold at 25% probability), namely, the bilateral parahippocampus gyrus and hippocampus.

**Mixed-effects model.** Mixed-effects modeling is a powerful statistical tool that offers many advantages over conventional *t* test, regression, and ANOVA in sophisticated fMRI designs (Mumford and Poldrack, 2007; Ward et al., 2013). It is especially useful when the number of trials differs by condition and/or across participants, as was the case in this study, in which participants remembered varying numbers of words by condition. In this study, the mixed-effects model was implemented with lme4 in R (Bates and Bolker, 2012). We used the likelihood ratio test to compare the models (with vs without the predictor) to determine the effect of the predictor.

**Mediation analysis.** We performed the mediation effect test to examine whether the ER-nGPS mediated the semantic similarity effect on memory strength. Mixed-effects models were used to test the relationship between (1) semantic similarity and memory strength ( $Y = a_1 + b_1X + \varepsilon_1$ ); (2) semantic similarity and ER-nGPS ( $M = a_2 + b_2X + \varepsilon_2$ ); and (3) semantic similarity and memory strength with a mediator ( $Y = a_3 + b_3X + bM + \varepsilon_3$ ). In the equations, *Y* is the dependent variable, *X* is the predictor, and *M* is the mediator. The indirect effect was estimated as  $b_2 \times b_3$ . We used distribution-of-the-product method to compute confidence intervals, which has been shown to be more accurate when the sample size is small (MacKinnon et al., 2002).

## Results

### Behavioral results

The mean endorsement rate (judged as old items, scored 3 or 4) of targets, lures, and foils were 90%, 46%, and 11%, respectively (Fig. 2A), suggesting that participants were generally accurate in their true memory, but also showed a fair amount of false memory. Paired-samples *t* tests revealed that the endorsement rate was significantly higher for targets than for lures ( $t_{(34)} = 11.22, p < 0.001$ ), and for lures than for foils ( $t_{(34)} = 13.21, p < 0.001$ ). For reaction times, paired-samples *t* tests showed that LO responses were slower than TO responses ( $t_{(34)} = 8.18, p < 0.001$ ) and LN responses were slower than FN responses ( $t_{(34)} = 10.84, p < 0.001$ ), but there was no difference between the reaction times of TO and FN ( $t_{(34)} = -0.25, p = 0.803$ ), or between the reaction times of LO and LN ( $t_{(34)} = -1.51, p = 0.142$ ) (Fig. 2B).

Next, we examined whether memory strength was affected by semantic similarity. After the recognition memory test, participants were asked to rate, within each list, the semantic similarity between each tested item (targets and lures) and each studied item. We calculated the sGS of each tested item by averaging its semantic similarity ratings with all studied items in that list. As shown in Figure 2C, the sGS was significantly correlated with the memory strength of lures ( $t_{(34)} = 3.07, p = 0.004, \beta = 0.69$ ). This

correlation was marginally significant for target items ( $t_{(34)} = 1.96, p = 0.058, \beta = 0.12$ ), probably due to the ceiling effect as most target items had high memory strength. Direct comparison of the slopes of the regression models suggested that semantic similarity had a stronger effect on lures than on targets ( $t_{(68)} = 2.45, p = 0.017$ ).

### fMRI results

#### ER-nGPS was associated with memory strength

We developed a method to examine whether the memory strength of true and false memories was associated with the ER-nGPS (Fig. 1B). Formally, the ER-nGPS was

calculated by averaging Fisher's *z*-scores of neural activation pattern similarity (Pearson *r*) between each item (*i*) during retrieval (R) with all other items (*j*) during encoding (E, from 1 to *n*):  $ER-nGPS = \sum_{j=1}^n sim(R_i, E_j) / n$ . We first examined whether the ER-nGPS was associated with true memory strength (targets reported as old [3, 4] vs foils reported as new [1, 2]). A whole-brain searchlight analysis showed that, consistent with the global matching hypothesis, items with greater memory strengths (i.e., TO) showed greater ER-nGPS than items with lower memory strengths (i.e., FN) in distributed brain regions, including the LIFG (MNI:  $-48, 26, 8, Z = 4.74$ ), LIPL (MNI:  $-56, -48, 48, Z = 5.46$ ), LSPL (MNI:  $-30, -52, 38, Z = 5.13$ ), LvLOC (MNI:  $-42, -78, -10, Z = 5.06$ ), and RvLOC (MNI:  $36, -78, -6, Z = 4.87$ ) (Fig. 3; Table 1).

Because the TO items might have greater perceptual or conceptual overlap with encoding items than might FN items, we did a linear mixed-model analysis to directly test the quantitative association between ER-nGPS and memory strength, using all items (targets, lures, and foils). We found that the ER-nGPS increased with true memory strength (1–4) in LIFG ( $\chi^2_{(1)} = 39.96, p < 0.001$ ), LIPL ( $\chi^2_{(1)} = 65.86, p < 0.001$ ), LSPL ( $\chi^2_{(1)} = 48.40, p < 0.001$ ), LvLOC ( $\chi^2 = 31.10, p < 0.001$ ), and RvLOC ( $\chi^2_{(1)} = 29.83, p < 0.001$ ). These results suggested a strong association between ER-nGPS and memory strength.

In the above mixed-effect analysis, the effect might have been due to differences between the TO and FN trials. A more robust test of the association between ER-nGPS and memory strength is to correlate memory strength with the ER-nGPS within each category (i.e., targets, lures, and foils). However, because memory strength was very high for targets and very low for foils, only lures had enough variations in memory strength and hence the power for this analysis. Therefore, we examined whether the ER-nGPS in the above regions was associated with false memory by comparing LO and LN items, both of which were not studied but showed different subjective memory strengths. Indeed, we found that the frontoparietal regions, including the LIFG ( $t_{(34)} = 2.39, p = 0.022$ ), LIPL ( $t_{(34)} = 2.32, p = 0.027$ ), and LSPL ( $t_{(34)} = 2.14, p = 0.040$ ), showed stronger ER-nGPS for LO than LN items, but the LvLOC ( $t_{(34)} = 1.72, p = 0.095$ ) and RvLOC ( $t_{(34)} = -0.19, p = 0.847$ ) did not (Fig. 3). A linear mixed-model analysis suggested that, for lures, the ER-nGPS increased with memory strength in the LIFG ( $\chi^2_{(1)} = 9.18, p = 0.002$ ), LIPL ( $\chi^2_{(1)} = 6.92, p = 0.008$ ), and the LSPL ( $\chi^2_{(1)} = 5.61, p = 0.018$ ), but not in the LvLOC ( $\chi^2_{(1)} = 2.89, p = 0.089$ ) or RvLOC ( $\chi^2_{(1)} = 0.004, p = 0.950$ ). These results suggest that the ER-nGPS in the frontoparietal regions was associated with the strength of both true and

false memories, whereas the ER-nGPS in the visual cortex was only associated with the strength of true memory.

In another analysis, we examined whether the ER-nGPS for targets and lures was contributed mainly by the within-list similarity. We calculated the ER-nGPS for targets and lures using only the within-list items (8 in total). This analysis revealed very similar results as reported above, suggesting that within-list similarity contributed to memory strength.

#### No memory strength signal in the MTL

Previous studies have suggested that the nGPS in the MTL could predict subsequent memory (LaRocque et al., 2013; Davis et al., 2014a). However, our whole-brain searchlight analysis found that the ER-nGPS in the MTL was not related to memory strength, even with a liberal threshold ( $z = 2.3$ , uncorrected). We further examined this issue by anatomically defining four ROIs in the MTL based on the Harvard-Oxford template, namely, the bilateral parahippocampal gyrus and hippocampus. The comparison of ER-nGPS in these ROIs only revealed a marginally significant difference between TO and FN in the left hippocampus ( $p = 0.07$ ) (Fig. 4).

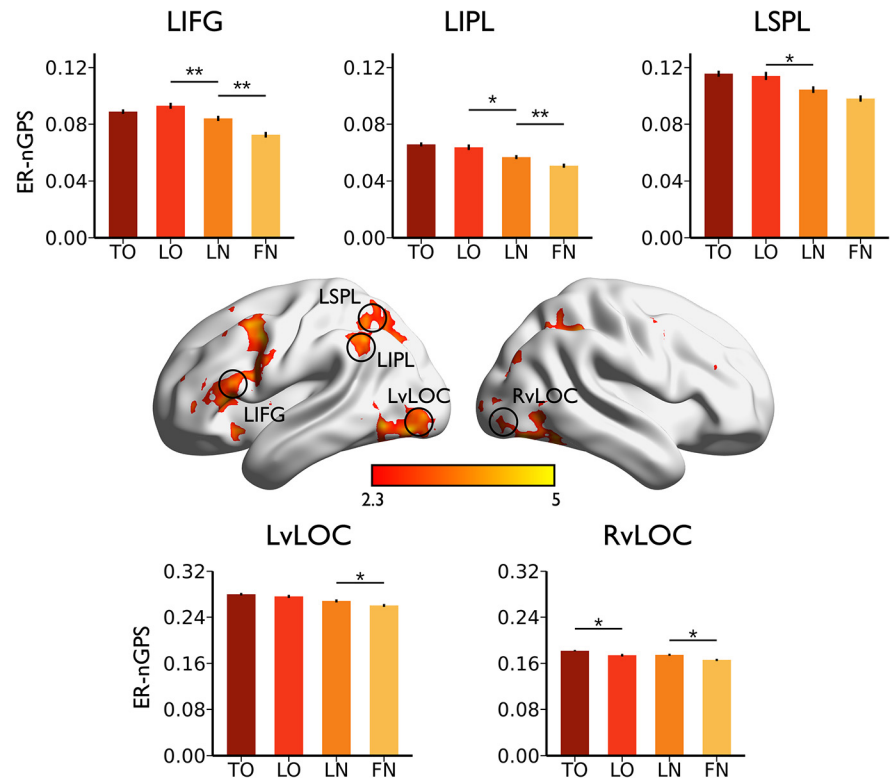
#### Parietal ER-nGPS mediated the effect of semantic similarity on false memory

The above analysis suggests that both semantic similarity (sGS) and ER-nGPS were associated with memory strength, and the effect of sGS was stronger for lures than for targets. We further evaluated whether the ER-nGPS indeed reflected semantic similarity and thus mediated the effect of sGS on memory strength. Focusing on the brain regions where the ER-nGPS was associated with memory strength, a linear mixed-effect model revealed that, for lures, the ER-nGPS was associated with the sGS in the LIPL ( $\chi^2_{(1)} = 4.14$ ,  $p = 0.042$ ) and the LSPL ( $\chi^2_{(1)} = 3.86$ ,  $p = 0.050$ ), but not in the LIFG ( $\chi^2_{(1)} = 1.70$ ,  $p = 0.193$ ). Mediation analysis shows the ER-nGPS in the LIPL partially mediated the effect of sGS on memory strength (indirect effect = 0.007,  $p = 0.05$ ) (Fig. 5), whereas the ER-nGPS in the LSPL did not ( $p = 0.11$ ). It is worth noting that the univariate activation level and the sGS were not associated in either of these two ROIs (all  $p$  values  $> 0.28$ ) and that activation levels in these regions did not mediate the sGS effect in memory. These results suggest that the ER-nGPS is more sensitive to the content of episodic representation than is the univariate activation level.

Consistent with the limited role of semantic similarity in memory strengths of target items, the association between sGS and ER-nGPS for targets was not significant in the LIFG ( $\chi^2_{(1)} = 2.20$ ,  $p = 0.138$ ), LIPL ( $\chi^2_{(1)} = 0.001$ ,  $p = 0.970$ ), or the LSPL ( $\chi^2_{(1)} = 0.72$ ,  $p = 0.396$ ).

#### ER-nGPS differentiated true and false memories

Could ER-nGPS distinguish true memory from false memory? Focusing on the regions where the ER-nGPS was associated with memory strength, the RvLOC ( $t_{(34)} = 2.58$ ,  $p = 0.014$ ) showed greater ER-nGPS for TO than LO items (Fig. 3). An

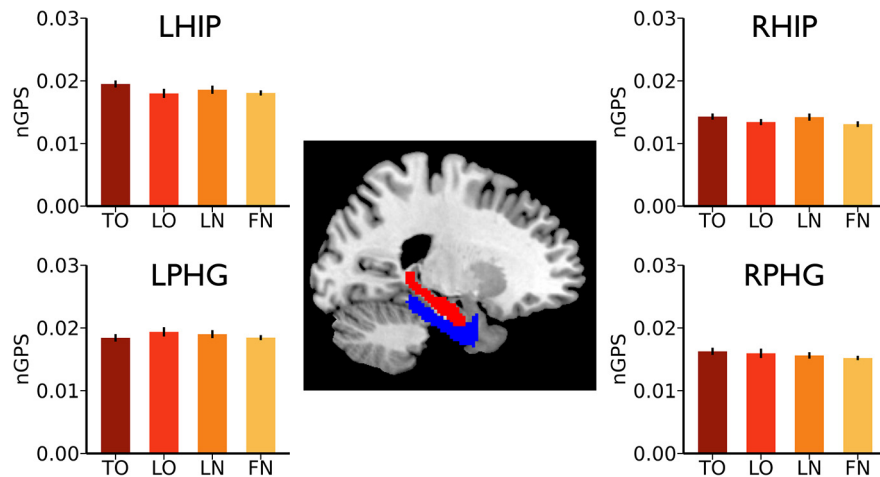


**Figure 3.** Brain regions with significant associations between ER-nGPS and memory strength. Greater ER-nGPS for TO than FN items, thresholded at  $z > 2.3$  (whole-brain corrected), were rendered onto a population-averaged surface atlas (Xia et al., 2013) (see also Table 1). Bar graphs of ER-nGPS, as a function of memory status, are shown for the LIFG, LIPL, LSPL, LvLOC, and RvLOC. Error bars indicate within-participant SE. \* $p < 0.05$ , \*\* $p < 0.01$ .

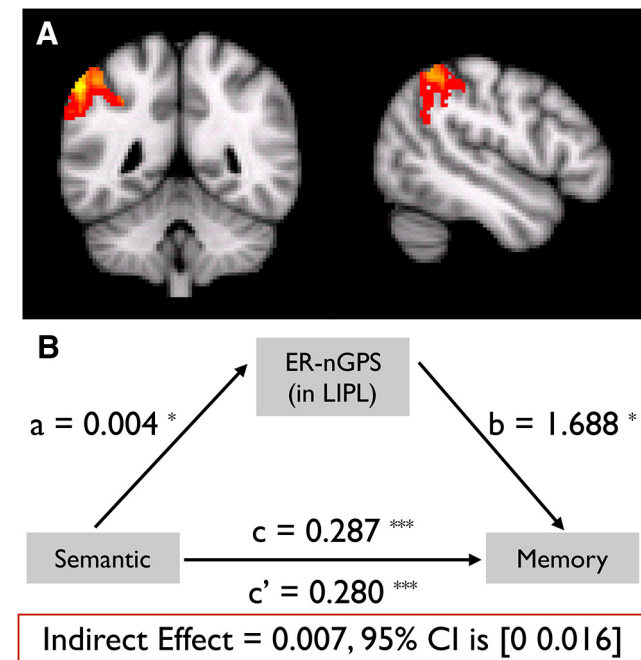
**Table 1. Regions whose ER-nGPS was greater for TO than FN items**

Region	Coordinates			Z
	x	y	z	
Left inferior frontal gyrus (LIFG)	-48	26	8	4.74
Left middle frontal gyrus (LMFG)	-42	12	50	4.97
Right middle frontal gyrus (RMFG)	50	26	28	3.99
Left inferior parietal lobule (LIPL)	-56	-48	48	5.46
Right inferior parietal lobule (RIPL)	42	-52	46	4.37
Left superior parietal lobule (LSPL)	-30	-52	38	5.13
Right superior parietal lobule (RSPL)	42	-42	46	4.24
Left dorsal lateral occipital complex (LdLOC)	22	-64	38	5.15
Right dorsal lateral occipital complex (RdLOC)	32	-76	18	4.59
Left ventral lateral occipital complex (LvLOC)	-42	-78	-10	5.06
Right ventral lateral occipital complex (RvLOC)	40	-74	-16	4.71
Left inferior temporal gyrus (LITG)	-42	-62	-6	4.42
Right inferior temporal gyrus (RITG)	48	-58	-14	4.17
Left middle temporal gyrus (LMTG)	-52	-58	4	3.60
Right middle temporal gyrus (RMTG)	64	-46	2	3.31
Left occipital fusiform gyrus	-34	-80	-10	4.78
Left temporal fusiform gyrus	-30	-60	-18	4.7
Right occipital fusiform gyrus	30	-80	-10	4.93
Right temporal fusiform gyrus	44	-56	-14	4.40
Left precuneus	-14	-76	40	4.16
Right precuneus	18	-56	12	4.13

additional whole-brain comparison showed that a cluster located in the right intracalcarine cortex extending to the right lingual gyrus (MNI: 28, -66, 2,  $Z = 3.28$ ) and a small cluster in the right superior parietal lobule (MNI: 26, -44, 66,  $Z = 3.00$ ) also showed greater ER-nGPS for TO than LO items (Fig. 6).



**Figure 4.** ER-nGPS in the MTL was not correlated with memory strength. The ROIs for the hippocampus (HIP, red) and parahippocampal gyrus (PHG, blue) were anatomically defined based on the Harvard-Oxford probabilistic atlas. No significant difference between conditions was found in any of the four ROIs, except for a marginal effect in the left HIP for TO versus FN items ( $p = 0.07$ ).



**Figure 5.** ER-nGPS in the LIPL mediated the relationship between sGS and memory response. **A**, The LIPL ROI’s ER-nGPS was associated with memory strength (TO > FN). **B**, Mediation analysis showed that the positive association between sGS and memory strengths for lures was partially mediated by the ER-nGPS in the LIPL.  $^{***}p < 0.001$ ,  $^*p < 0.05$ .

*Testing the cubing hypothesis with a nonlinear similarity gradient function*

Following the SAM model (Gillund and Shiffrin, 1984) and a previous neuroimaging study (LaRocque et al., 2013), the above analyses used a linear similarity gradient function. To differentiate true and false memory, the MINERVA 2 (Hintzman, 1984) has posited a cubing mechanism (nonlinear similarity gradient function, such as a third power function). Similarly, the exponential similarity gradient function has been used in a neuroimaging study (Davis et al., 2014a). To test the cubing hypothesis, we did two additional analyses using the exponential and third power similarity gradient function, respectively. Consistent with a previous observation (Davis et al., 2014a), these nonlinear functions

generated results that were very similar to those from the linear function (Table 2), suggesting that the choice of similarity gradient function did not change our findings.

*The prefrontal cortex was involved in cognitive control*

To identify the regions involved in cognitive control, we compared the activations for correctly rejected lures (LN) and foils judged as new (FN). This analysis revealed strong activation in a large cluster in the left lateral prefrontal cortex (MNI: -52, 32, 16,  $Z = 3.87$ ) (Fig. 7A) and a small cluster in the medial prefrontal cortex (MNI: -2, 20, 56,  $Z = 4.81$ ). The MFPC has been implicated in conflict processing (Carter et al., 1998; MacDonald et al., 2000; Alexander and Brown, 2010), whereas the LPFC has been implicated in cognitive control and conflict resolution (MacDonald et al., 2000). Focusing on the

LPFC, we found stronger activation for TO than FN items ( $t_{(34)} = 4.78$ ,  $p < 0.001$ ), but no difference between LO and LN items ( $t_{(34)} = 0.84$ ,  $p = .404$ ) or between TO and LO items ( $t_{(34)} = -1.11$ ,  $p = 0.27$ ) (Fig. 7B).

Could a strong cognitive control process help reduce the ER-nGPS signals of the lures and reduce false memory? If this hypothesis were true, we would observe a negative association between LPFC activity and the ER-nGPS in the parietal lobule. Contrary to this hypothesis, we found a strong positive association between LPFC’s activation and ER-nGPS in the LIPL for lure items ( $\chi^2_{(1)} = 411.29$ ,  $p < 0.001$ ). After controlling for semantic similarity (sGS), this positive association was greatly reduced but still significant ( $\chi^2_{(1)} = 4.49$ ,  $p = 0.034$ ).

*ER-nGPS discrepancy in the parietal and visual cortices was associated with cognitive control*

The above analysis suggested that ER-nGPS signals in the LIPL were strong for both true and false memories, whereas the signals in the visual cortex were stronger for true memory than for false memory. We hence speculated that the LPFC’s activation was associated with the discrepancy of the ER-nGPS in the LIPL and the visual cortex. Consistently, a mixed-effect model analysis revealed a strong positive association between the ER-nGPS difference (LIPL - RvLOC) and the left LPFC activation ( $\chi^2_{(1)} = 97.22$ ,  $p < 0.001$ ) (Fig. 7C,D). The same relationship was found when using the ER-nGPS in the intracalcarine cortex cluster ( $\chi^2_{(1)} = 60.88$ ,  $p < 0.001$ ). These results suggest that the cognitive control process might result from a discrepancy between the ER-nGPS in the parietal and visual cortices.

*Controlling for the effects of univariate activation levels and their variances*

As Figure 8 shows, the univariate activation levels in distributed brain regions were also associated with memory performance. Furthermore, linear mixed model showed moderate-to-high correlations between univariate activation levels and ER-nGPS in all ROIs for both true memory ( $r = 0.41$  to  $0.73$ , all  $p$  values  $< 0.0001$ ) and false memory ( $r = 0.40$  to  $0.69$ , all  $p$  values  $< 0.0001$ ). Therefore, to confirm the effect of ER-nGPS on memory, we needed to control for the univariate activation levels. A mixed-effect regression model revealed that after controlling

for univariate activation level, ER-nGPS was still a significant predictor of true memory strength in all ROIs (LIFG,  $\chi^2_{(1)} = 13.12$ ,  $p < 0.001$ ; LIPL,  $\chi^2_{(1)} = 37.36$ ,  $p < 0.001$ ; LSPL,  $\chi^2_{(1)} = 10.19$ ,  $p = 0.001$ ; LvLOC,  $\chi^2_{(1)} = 5.57$ ,  $p = 0.018$ ; RvLOC,  $\chi^2_{(1)} = 18.07$ ,  $p < 0.001$ ), and was a significant predictor for the strength of false memory in the LIPL ( $\chi^2_{(1)} = 6.87$ ,  $p = 0.009$ ) and a marginally significant predictor in the LIFG ( $\chi^2_{(1)} = 3.33$ ,  $p = 0.068$ ).

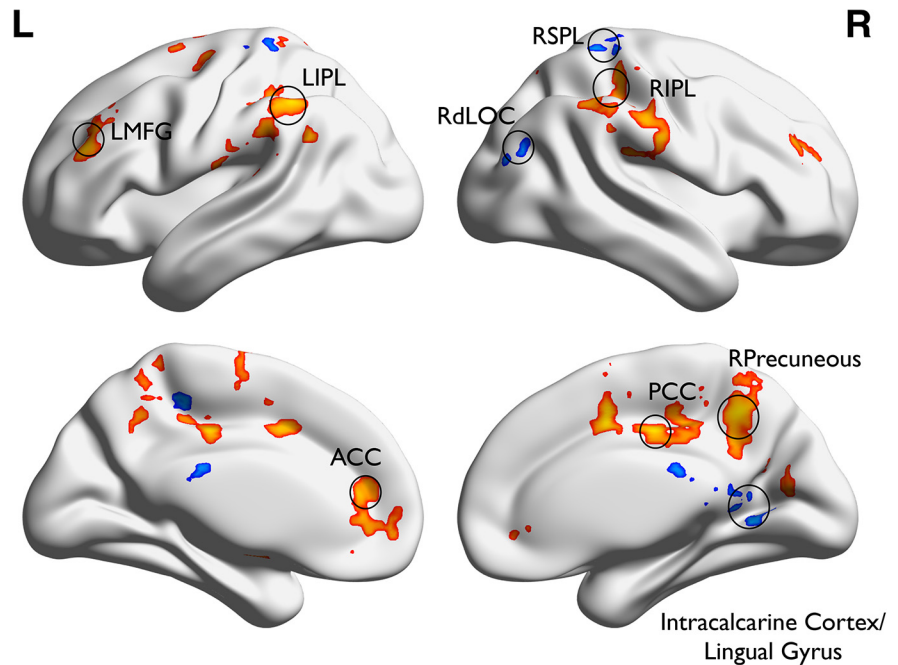
A previous simulation study suggests that the variance of activation level across voxels could also affect pattern similarity estimation (Davis et al., 2014b). To examine this effect, we calculated the variance in each ROI for each type of trials. Only the variance for TO was slightly larger than that for FN in the LIFG ( $t_{(34)} = 2.07$ ,  $p = 0.046$ ), LSPL ( $t_{(34)} = 2.48$ ,  $p = 0.018$ ), and LvLOC ( $t_{(34)} = 2.03$ ,  $p = 0.050$ ). No other difference was significant. After controlling for both the activation level and the variance using a linear mixed effect model, we found that the ER-nGPS still predicted true memory strength in all ROIs (LIFG,  $\chi^2_{(1)} = 13.88$ ,  $p < 0.001$ ; LIPL,  $\chi^2_{(1)} = 40.08$ ,  $p < 0.001$ ; LSPL,  $\chi^2_{(1)} = 10.66$ ,  $p = 0.001$ ; LvLOC,  $\chi^2_{(1)} = 7.25$ ,  $p = 0.007$ ; RvLOC,  $\chi^2_{(1)} = 20.24$ ,  $p < 0.001$ ), and false memory strength in LIFG ( $\chi^2_{(1)} = 3.91$ ,  $p = 0.048$ ) and LIPL ( $\chi^2_{(1)} = 8.14$ ,  $p = 0.004$ ).

Finally, univariate analysis revealed greater activation for true memory (TO) than false memory (LO) in the left MFG, bilateral IPL, precuneus, and anterior and posterior cingulate cortex (Fig. 6). However, these regions did not overlap with those showing differences in ER-nGPS between true and false memories.

## Discussion

Inspired by formal computational models of recognition memory, the current study aimed at leveraging the fine activation pattern measured by fMRI to provide a mechanistic understanding of the neural signals underlying recognition memory strength (i.e., familiarity). We tested a novel hypothesis that the neural global pattern similarity between encoding and retrieval (ER-nGPS) was associated with subjective strength of recognition memory (either true or false memory). By adapting the DRM paradigm, we observed a significant amount of false memory, although the 40% false memory rate was lower than that found in the original DRM paradigm (Roediger and McDermott, 1995). This lower rate was partially due to the shorter word lists and a smaller number of lists, as well as the use of prewarning (McCabe and Smith, 2002). In addition, the use of visual presentation of the stimuli during encoding, compared with the standard auditory presentation (Roediger and McDermott, 1995), could have also reduced false recognition of visually presented test words (Smith and Hunt, 1998).

Consistent with the global matching hypothesis, we found that the ER-nGPS in the frontal and parietal cortices scaled with the strength of subjective feelings of familiarity of both true and false memories. The current study and many previous studies have also found that univariate activation in these regions was sensitive to the memory strength of true and false memories



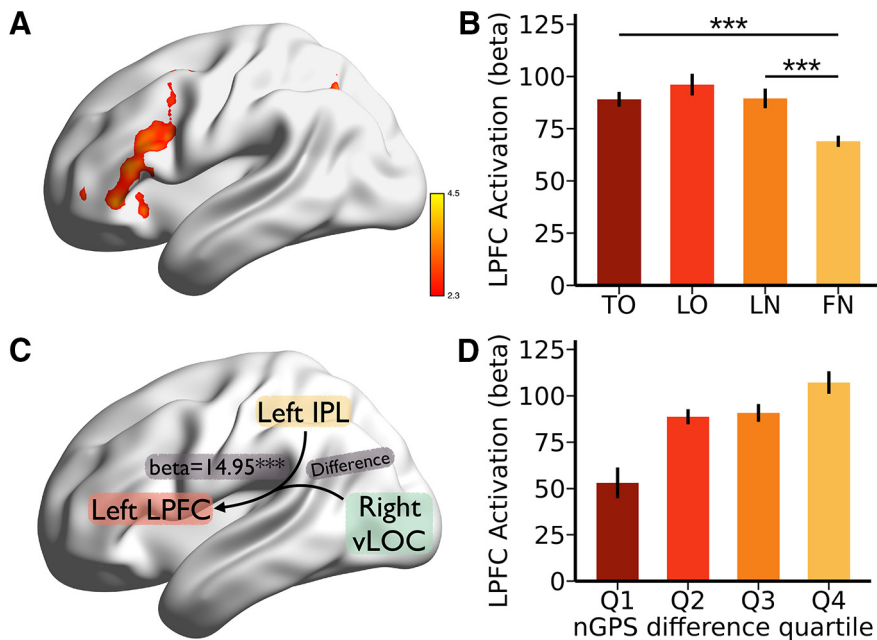
**Figure 6.** Neural differences between true and false memories. Whole-brain searchlight analysis results for regions that showed greater ER-nGPS (shown in blue) or activation level (shown in red) for TO items than LO items. The differences were thresholded at  $z > 2.3$  (whole-brain corrected) and rendered onto a population-averaged surface atlas (Xia et al., 2013). There was no overlap between the univariate and multivariate effects.

**Table 2. Statistical comparison of the ER-nGPS between conditions, using exponential and third power functions as the similarity gradient functions**

ROI	TO versus FN		LO versus LN		TO versus LO		LN versus FN	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Exponential similarity gradient function								
LIFG	6.29	0.000	3.02	0.005	−1.33	0.193	3.42	0.002
LIPL	6.89	0.000	2.50	0.017	1.02	0.317	2.87	0.007
LSPL	5.64	0.000	0.63	0.028	0.48	0.632	1.81	0.079
LvLOC	5.13	0.000	1.66	0.107	1.11	0.275	2.14	0.039
RvLOC	5.67	0.000	−0.16	0.871	2.63	0.013	2.91	0.006
Third power similarity gradient function								
LIFG	6.04	0.000	3.24	0.003	−1.49	0.144	3.56	0.001
LIPL	6.23	0.000	2.31	0.027	1.23	0.226	2.79	0.009
LSPL	4.49	0.000	1.92	0.063	0.59	0.556	2.05	0.048
LvLOC	4.94	0.000	1.62	0.115	0.63	0.533	2.23	0.032
RvLOC	4.81	0.000	0.10	0.924	2.44	0.020	2.47	0.019

(Schacter et al., 1996; Cabeza et al., 2001; Garoff-Eaton et al., 2007; Stark et al., 2010). However, the ER-nGPS made unique contributions to memory strength after controlling for univariate activation level and variance. Furthermore, only the ER-nGPS in the parietal lobe was correlated with semantic similarity and mediated the effect of semantic similarity on memory strength of lure items. Together, these results provide clear support to the global matching hypothesis.

More importantly, the current study found multiple ER-nGPS signals that contributed differentially to true and false memories. Specifically, the ER-nGPS in the frontoparietal region was correlated with the memory strength of both true and false memories, whereas the ER-nGPS in the visual cortex was only associated with true but not false memory. Furthermore, the ER-nGPS in the early visual cortex was stronger for true memory than for false memory, and was not correlated with semantic

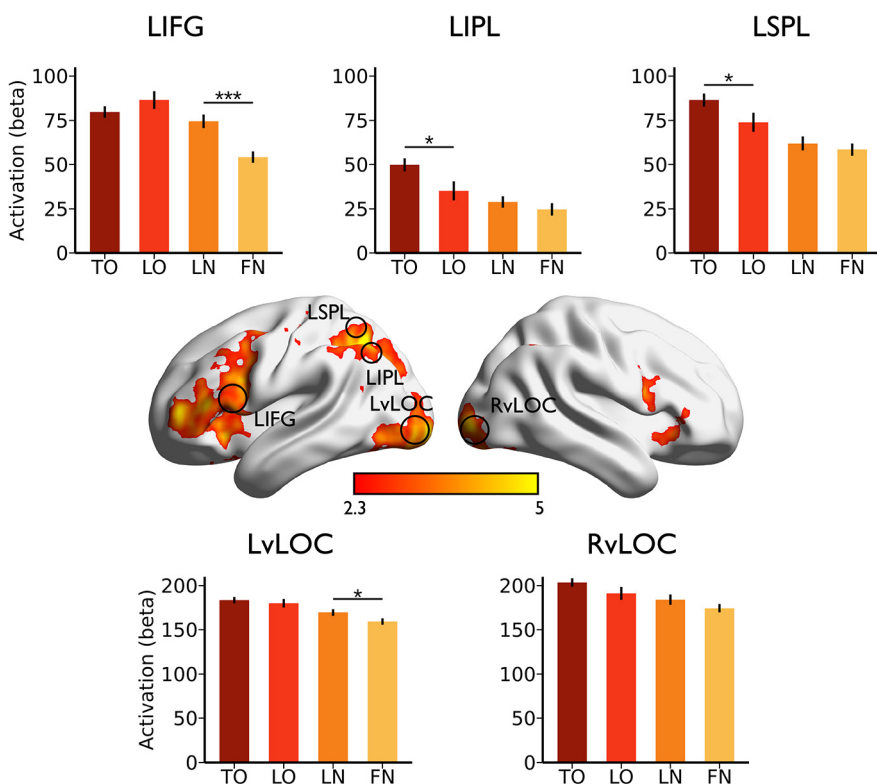


**Figure 7.** The ER-nGPS difference in the LIPL and RvLOC was associated with left LPFC activation level. **A**, The LPFC showed greater activation level for LN than FN, reflecting the cognitive control process. The differences were thresholded at  $z > 2.3$  (whole-brain corrected), and rendered onto a population-averaged surface atlas (Xia et al., 2013). **B**, Bar graph of the activation level in the LPFC by condition. **C**, An illustration of the relationship between LIPL/RvLOC’s ER-nGPS and LPFC’s activation level. The ER-nGPS difference (LIPL – RvLOC) was associated with the LPFC activation level across trials. **D**, LPFC’s activation level by ER-nGPS difference quartile. \*\*\* $p < 0.001$ .

similarity. Together, these results suggest that ER-nGPS in the visual and frontoparietal cortices reflects distinct contents of memory.

Our study provides a novel way to link neural activity to memory strength. Existing fMRI studies using the subsequent memory paradigm have consistently revealed that the activation level (Wagner et al., 1998; Kim, 2011) and/or the distributed pattern (Xue et al., 2010; Ward et al., 2013) for a given item are associated with the subsequent memory of that item. The implicit assumption underlying those studies is that the activation for each studied item mainly reflects the “strength” gained for that item. In contrast, the global matching models posit that a test item could gain certain degree of “strength” from each of the studied items, which is determined by the similarity between the tested item and the studied items (Gillund and Shiffrin, 1984). Using neural pattern similarity analysis, our data suggest that it might be possible to estimate and sum up the “strength” of the test item obtained from each studied item. It should be noted that, although the choices of similarity gradient function in computational models are based on different mechanisms and have significant impacts on model performance, their effect on neuroimaging data seems to be subtle as shown in both a previous study (Davis et al., 2014a) and the current study, probably due to the overall low and limited range of neural similarity as a result of the noisy nature of the fMRI data.

Our results also help to elucidate the function of the parietal lobule, whose activation has been consistently observed in successful retrieval (Wagner et al., 2005; Cabeza et al., 2008). Mechanistic accounts of the role of the inferior parietal lobule in memory have emphasized either general processes, such as attention on internal mnemonic representations (Cabeza et al., 2008) and accumulation of mnemonic evidence (Wagner et al., 2005), or specific processes, such as representations of retrieved content in an “output buffer” (Vilberg and Rugg, 2012) and binding information from other cortical inputs (Shimamura, 2011). Using representational similarity analysis to probe the content of information representation, recent studies have provided evidence to support the content-specific episodic representation (Xue et al., 2013; Kuhl and Chun, 2014). We found that ER-nGPS in the LPC underlay both true and false memories and that the strength of the ER-nGPS was correlated with the semantic similarity of



**Figure 8.** Brain regions whose activation level was associated with memory strength. Greater activation for TO than FN items, thresholded at  $z > 2.3$  (whole-brain corrected), is rendered onto a population-averaged surface atlas. Bar graphs of activation, as a function of memory status, are plotted for the LIFG, LIPL, LSPL, LvLOC, and RvLOC. Error bar indicates within-participant SE. \* $p < 0.05$ , \*\*\* $p < 0.001$ .



lures. These results provide further support for the role of the LPC in representing the content of episodic retrieval. Its correlation with semantic similarity is consistent with the DRM paradigm and the role of the LPC in semantic processing (Binder et al., 2009; Seghier, 2013). Unlike a previous study, which revealed differentiated functions of the SPL, IPL, and intraparietal sulcus (Hutchinson et al., 2014), the similar pattern of results in the SPL and IPL obtained in the present study might be due to the limited spatial resolution of fMRI and the use of the searchlight method and group-level ROIs.

Consistent with existing neuroimaging studies (Kensinger and Schacter, 2006; Kubota et al., 2006; Garoff-Eaton et al., 2007; Kim and Cabeza, 2007), we also found more frontal activation for lures than for targets and foils as predicted by the activation/monitoring model (Roediger et al., 2001). Furthermore, our results suggest that the high ER-nGPS signal from the parietal cortex but low ER-nGPS signal from the visual cortex might have triggered the monitoring processes, as well as the slower reaction time for lures. This would provide a mechanism for the brain to detect and correct potential false memory. Essentially, this cognitive control mechanism could also be expanded easily to explain how task and instruction manipulations could modulate relative contributions of different sources of global matching signals to mnemonic decisions.

The present study failed to find strong association between MTL ER-nGPS signal and memory strength. The MTL plays an important role in pattern completion during retrieval (Leutgeb et al., 2007; Bakker et al., 2008), leading to reactivation of cortical patterns from initial encoding (Wheeler et al., 2000; Johnson et al., 2009; Manning et al., 2011; Staresina et al., 2012; Kuhl and Chun, 2014). Several possible reasons could account for the absence of MTL global similarity signals. First, whereas previous studies examined global similarity among encoding trials, the current study examined global similarity between retrieval and encoding. Second, the MTL contains multiple subregions that are involved in distinct functions and may represent distinct types of information (O'Reilly and McClelland, 1994; Bakker et al., 2008). High-resolution fMRI in combination of individualized ROI analysis might be required to elucidate the finer functional dissociations (LaRocque et al., 2013). Third, the nature of sparse representation in the hippocampal regions would make it hard to probe content/category-level representations (Quiroga et al., 2008), due to the limited spatial resolution of fMRI. Consistently, content-specific representation has not been reliably observed in the hippocampus (LaRocque et al., 2013; Liang et al., 2013). Finally, the current experimental task might rely more on semantic information in other cortical regions rather than episodic/contextual information encoded in the MTL.

Several important questions need to be addressed in future research. First, future studies should examine the nature of multiple global matching signals found in the current study and whether they could account for the various false memory effects, including the modality effect (Smith and Hunt, 1998; Gallo et al., 2001), stimulus duration and repetition effect (Seamon et al., 2002), and developmental effect (Dennis et al., 2007, 2014). Second, future studies should examine the list strength and list length effects within the global matching framework (Shiffrin et al., 1990). The multiple-trace models posit that repetitions create multiple independent traces, increasing global matching. This account, however, would suggest a similar effect of list length and list strength on memory, which has been contradicted by empirical findings (Ratcliff et al., 1990). To address this issue, Shiffrin and Steyvers (1997) proposed a differentiation hypothesis, which

posits that repetition enhances the memory representation of single items, and therefore reduces item noise and the interitem similarity. Consistent with this idea, neuroimaging studies have found that greater neural pattern similarity across repetitions (i.e., self-similarity) was associated with better memory (Xue et al., 2010, 2013; Ward et al., 2013; Lu et al., 2015). Future studies need to examine how list strength enhances self-similarity, global matching, and memory strength.

In conclusion, by integrating multivariate representation similarity methods with formal computational models of memory, our study provides important results connecting behavioral, neural, and computational accounts of recognition memory strength. Multiple global matching signals in distributed brain regions were found to carry distinct information and to contribute differentially to true and false memories. These results provide strong neural evidence for the global matching models and highlight the importance of considering multiple global matching signals. More broadly, these results suggest that the neural measures inspired by formal computational models could provide important data to test and improve existing models, and could contribute to a deeper mechanistic understanding of the representation and processes underlying human recognition memory.

## References

- Abe N, Okuda J, Suzuki M, Sasaki H, Matsuda T, Mori E, Tsukada M, Fujii T (2008) Neural correlates of true memory, false memory, and deception. *Cereb Cortex* 18:2811–2819. [CrossRef Medline](#)
- Alexander WH, Brown JW (2010) Computational models of performance monitoring and cognitive control. *Top Cogn Sci* 2:658–677. [CrossRef Medline](#)
- Andersson J, Jenkinson M, Smith S (2007a) Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. <http://www.fmrib.ox.ac.uk/analysis/techrep>.
- Andersson J, Jenkinson M, Smith S (2007b) Non-linear optimisation. FMRIB technical report TR07JA1. <http://www.fmrib.ox.ac.uk/analysis/techrep>.
- Arndt J, Hirshman E (1998) True and false recognition in MINERVA2: explanations from a global matching perspective. *J Mem Lang* 39:371–391. [CrossRef](#)
- Atkins AS, Reuter-Lorenz PA (2011) Neural mechanisms of semantic interference and false recognition in short-term memory. *Neuroimage* 56:1726–1734. [CrossRef Medline](#)
- Bakker A, Kirwan CB, Miller M, Stark CE (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319:1640–1642. [CrossRef Medline](#)
- Bates D, Bolker B (2012) lme4. 0: Linear mixed-effects models using Eigen and R syntax. *J Stat Softw* 65:1–68. [CrossRef Medline](#)
- Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in fMRI. *Neuroimage* 20:1052–1063. [CrossRef Medline](#)
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796. [CrossRef Medline](#)
- Bransford JD, Franks JJ (1971) The abstraction of linguistic ideas. *Cogn Psychol* 2:331–350. [CrossRef](#)
- Cabeza R, Rao SM, Wagner AD, Mayer AR, Schacter DL (2001) Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc Natl Acad Sci U S A* 98:4805–4810. [CrossRef Medline](#)
- Cabeza R, Ciaramelli E, Olson IR, Moscovitch M (2008) The parietal cortex and episodic memory: an attentional account. *Nat Rev Neurosci* 9:613–625. [CrossRef Medline](#)
- Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280:747–749. [CrossRef Medline](#)
- Clark SE, Gronlund SD (1996) Global matching models of recognition memory: how the models match the data. *Psychon Bull Rev* 3:37–60. [CrossRef Medline](#)

- Davis T, Xue G, Love BC, Preston AR, Poldrack RA (2014a) Global neural pattern similarity as a common basis for categorization and recognition memory. *J Neurosci* 34:7472–7484. [CrossRef Medline](#)
- Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA (2014b) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage* 97:271–283. [CrossRef Medline](#)
- Deese J (1959) On the prediction of occurrence of particular verbal intrusions in immediate recall. *J Exp Psychol* 58:17–22. [CrossRef Medline](#)
- Dennis NA, Kim H, Cabeza R (2007) Effects of aging on true and false memory formation: an fMRI study. *Neuropsychologia* 45:3157–3166. [CrossRef Medline](#)
- Dennis NA, Bowman CR, Vandekar SN (2012) True and phantom recollection: an fMRI investigation of similar and distinct neural correlates and connectivity. *Neuroimage* 59:2982–2993. [CrossRef Medline](#)
- Dennis NA, Bowman CR, Peterson KM (2014) Age-related differences in the neural correlates mediating false recollection. *Neurobiol Aging* 35:395–407. [CrossRef Medline](#)
- Duvernois S, Motamedinia S, Rugg MD (2009) The relationship between aging, performance, and the neural correlates of successful memory encoding. *Cereb Cortex* 19:733–744. [CrossRef Medline](#)
- Gallo DA, McDermott KB, Percer JM, Roediger HL 3rd (2001) Modality effects in false recall and false recognition. *J Exp Psychol Learn Mem Cogn* 27:339–353. [CrossRef Medline](#)
- Garoff-Eaton RJ, Kensinger EA, Schacter DL (2007) The neural correlates of conceptual and perceptual false recognition. *Learn Mem* 14:684–692. [CrossRef Medline](#)
- Gerard L, Zacks RT, Hasher L, Radvansky GA (1991) Age deficits in retrieval: the fan effect. *J Gerontol* 46:P131–P136. [CrossRef Medline](#)
- Gillund G, Shiffrin RM (1984) A retrieval model for both recognition and recall. *Psychol Rev* 91:1–67. [CrossRef Medline](#)
- Hintzman DL (1984) MINERVA 2: a simulation model of human memory. *Behav Res Methods Instrum Comput* 16:96–101. [CrossRef](#)
- Hintzman DL (1988) Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol Rev* 95:528. [CrossRef](#)
- Humphreys MS, Pike R, Bain JD, Tehan G (1989) Global matching: a comparison of the SAM, Minerva II, Matrix, and TODAM models. *J Math Psychol* 33:36–67. [CrossRef](#)
- Hutchinson JB, Uncapher MR, Weiner KS, Bressler DW, Silver MA, Preston AR, Wagner AD (2014) Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cereb Cortex* 24:49–66. [CrossRef Medline](#)
- Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–156. [CrossRef Medline](#)
- Johnson JD, McDuff SG, Rugg MD, Norman KA (2009) Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron* 63:697–708. [CrossRef Medline](#)
- Kensinger EA, Schacter DL (2006) Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *J Neurosci* 26:2564–2570. [CrossRef Medline](#)
- Kim H (2011) Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *Neuroimage* 54:2446–2461. [CrossRef Medline](#)
- Kim H, Cabeza R (2007) Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *J Neurosci* 27:12190–12197. [CrossRef Medline](#)
- Kimball DR, Smith TA, Kahana MJ (2007) The fSAM model of false recall. *Psychol Rev* 114:954–993. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Kubota Y, Toichi M, Shimizu M, Mason RA, Findling RL, Yamamoto K, Calabrese JR (2006) Prefrontal hemodynamic activity predicts false memory: a near-infrared spectroscopy study. *Neuroimage* 31:1783–1789. [CrossRef Medline](#)
- Kuhl BA, Chun MM (2014) Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J Neurosci* 34:8051–8060. [CrossRef Medline](#)
- LaRocque KF, Smith ME, Carr VA, Witthoft N, Grill-Spector K, Wagner AD (2013) Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J Neurosci* 33:5466–5474. [CrossRef Medline](#)
- Leutgeb JK, Leutgeb S, Moser MB, Moser EI (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315:961–966. [CrossRef Medline](#)
- Liang JC, Wagner AD, Preston AR (2013) Content representation in the human medial temporal lobe. *Cereb Cortex* 23:80–96. [CrossRef Medline](#)
- Lu Y, Wang C, Chen C, Xue G (2015) Spatiotemporal neural pattern similarity supports episodic memory. *Curr Biol* 25:780–785. [CrossRef Medline](#)
- MacDonald AW 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288:1835–1838. [CrossRef Medline](#)
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V (2002) A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* 7:83–104. [CrossRef Medline](#)
- Manning JR, Polyn SM, Baltuch GH, Litt B, Kahana MJ (2011) Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc Natl Acad Sci U S A* 108:12893–12897. [CrossRef Medline](#)
- McCabe DP, Smith AD (2002) The effect of warnings on false memories in young and older adults. *Mem Cognit* 30:1065–1077. [CrossRef Medline](#)
- Mumford JA, Poldrack RA (2007) Modeling group fMRI data. *Soc Cogn Affect Neurosci* 2:251–257. [CrossRef Medline](#)
- Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59:2636–2643. [CrossRef Medline](#)
- Murdock BB (1982) A theory for the storage and retrieval of item and associative information. *Psychol Rev* 89:609–626. [CrossRef](#)
- O'Reilly RC, McClelland JL (1994) Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus* 4:661–682. [CrossRef Medline](#)
- Pike R (1984) Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychol Rev* 91:281. [CrossRef](#)
- Posner MI, Keele SW (1970) Retention of abstract ideas. *J Exp Psychol* 83:304–308. [CrossRef](#)
- Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87–91. [CrossRef Medline](#)
- Ratcliff R, Clark SE, Shiffrin RM (1990) List-strength effect: I. Data and discussion. *J Exp Psychol Learn Mem Cogn* 16:163–178. [CrossRef Medline](#)
- Roediger HL, McDermott KB (1995) Creating false memories: remembering words not presented in lists. *J Exp Psychol Learn Mem Cognit* 21:803–814. [CrossRef](#)
- Roediger HL 3rd, Watson JM, McDermott KB, Gallo DA (2001) Factors that determine false recall: a multiple regression analysis. *Psychon Bull Rev* 8:385–407. [CrossRef Medline](#)
- Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, Roediger HL 3rd (1996) Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron* 17:267–274. [CrossRef Medline](#)
- Seamon JG, Luo CR, Schwartz MA, Jones KJ, Lee DM, Jones SJ (2002) Repetition can have similar or different effects on accurate and false recognition. *J Mem Lang* 46:323–340. [CrossRef](#)
- Seghier ML (2013) The angular gyrus multiple functions and multiple subdivisions. *Neuroscientist* 19:43–61. [CrossRef Medline](#)
- Shiffrin RM, Steyvers M (1997) A model for recognition memory: REM—retrieving effectively from memory. *Psychon Bull Rev* 4:145–166. [CrossRef Medline](#)
- Shiffrin RM, Ratcliff R, Clark SE (1990) List-strength effect: II. Theoretical mechanisms. *J Exp Psychol Learn Mem Cognit* 16:179. [CrossRef](#)
- Shimamura AP (2011) Episodic retrieval and the cortical binding of relational activity. *Cogn Affect Behav Neurosci* 11:277–291. [CrossRef Medline](#)
- Smith RE, Hunt RR (1998) Presentation modality affects false memory. *Psychon Bull Rev* 5:710–715. [CrossRef](#)
- Staresina BP, Henson RN, Kriegeskorte N, Alink A (2012) Episodic reinstatement in the medial temporal lobe. *J Neurosci* 32:18150–18156. [CrossRef Medline](#)
- Stark CE, Okado Y, Loftus EF (2010) Imaging the reconstruction of true and

- false memories using sensory reactivation and the misinformation paradigms. *Learn Mem* 17:485–488. [CrossRef Medline](#)
- Vilberg KL, Rugg MD (2012) The neural correlates of recollection: transient versus sustained fMRI effects. *J Neurosci* 32:15679–15687. [CrossRef Medline](#)
- Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL (1998) Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281:1188–1191. [CrossRef Medline](#)
- Wagner AD, Shannon BJ, Kahn I, Buckner RL (2005) Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci* 9:445–453. [CrossRef Medline](#)
- Ward EJ, Chun MM, Kuhl BA (2013) Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *J Neurosci* 33:14749–14757. [CrossRef Medline](#)
- Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci U S A* 97:11125–11129. [CrossRef Medline](#)
- Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage* 21:1732–1747. [CrossRef Medline](#)
- Xia M, Wang J, He Y (2013) BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One* 8:e68910. [CrossRef Medline](#)
- Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA (2010) Greater neural pattern similarity across repetitions is associated with better memory. *Science* 330:97–101. [CrossRef Medline](#)
- Xue G, Dong Q, Chen C, Lu ZL, Mumford JA, Poldrack RA (2013) Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. *Cereb Cortex* 23:1562–1571. [CrossRef Medline](#)