Behavioral/Cognitive

Lesser Neural Pattern Similarity across Repeated Tests Is Associated with Better Long-Term Memory Retention

Linnea Karlsson Wirebring,[1,2,4] **Carola Wiklund-Hörnqvist,**[2,4] **Johan Eriksson,**[1,2] ⬤**Micael Andersson,**[1,2] ⬤**Bert Jonsson,**[4] **and Lars Nyberg**[1,2,3]

[1]Department of Integrative Medical Biology, [2]Umeå Center for Functional Brain Imaging (UFBI), [3]Department of Radiation Sciences, and [4]Department of Psychology, Umeå University, 901 87 Umeå, Sweden

Encoding and retrieval processes enhance long-term memory performance. The efficiency of encoding processes has recently been linked to representational consistency: the reactivation of a representation that gets more specific each time an item is further studied. Here we examined the complementary hypothesis of whether the efficiency of retrieval processes also is linked to representational consistency. Alternatively, recurrent retrieval might foster representational variability—the altering or adding of underlying memory representations. Human participants studied 60 Swahili–Swedish word pairs before being scanned with fMRI the same day and 1 week later. On Day 1, participants were tested three times on each word pair, and on Day 7 each pair was tested once. A BOLD signal change in right superior parietal cortex was associated with subsequent memory on Day 1 and with successful long-term retention on Day 7. A representational similarity analysis in this parietal region revealed that beneficial recurrent retrieval was associated with representational variability, such that the pattern similarity on Day 1 was lower for retrieved words subsequently remembered compared with those subsequently forgotten. This was mirrored by a monotonically decreased BOLD signal change in dorsolateral prefrontal cortex on Day 1 as a function of repeated successful retrieval for words subsequently remembered, but not for words subsequently forgotten. This reduction in prefrontal response could reflect reduced demands on cognitive control. Collectively, the results offer novel insights into why memory retention benefits from repeated retrieval, and they suggest fundamental differences between repeated study and repeated testing.

Key words: fMRI; memory; pattern similarity; repeated testing; retrieval; RSA

Significance Statement

Repeated testing is known to produce superior long-term retention of the to-be-learned material compared with repeated encoding and other learning techniques, much because it fosters repeated memory retrieval. This study demonstrates that repeated memory retrieval might strengthen memory by inducing more differentiated or elaborated memory representations in the parietal cortex, and at the same time reducing demands on prefrontal-cortex-mediated cognitive control processes during retrieval. The findings contrast with recent demonstrations that repeated encoding induces less differentiated or elaborated memory representations. Together, this study suggests a potential neurocognitive explanation of why repeated retrieval is more beneficial for long-term retention than repeated encoding, a phenomenon known as the testing effect.

Introduction

Repetition is a fundamental principle of learning. It is well established that repeated study strengthens memory representations

and makes them resistant to forgetting (Ebbinghaus, 1885). There is also accumulating evidence that repeated testing can strengthen memory representations (Karpicke and Roediger, 2008; Roediger and Butler, 2011), and have beneficial effects relative to study at shorter (Wiklund-Hörnqvist et al., 2014) and longer retention intervals (Kornell et al., 2011). Repeated study has been demonstrated to induce durable memory retention to the degree that the same neural representations are more precisely reactivated across cumulative trials (Xue et al., 2010). Here we addressed the complementary question of what signifies repeated testing in relation to successful long-term retention.

By definition, repeated testing poses demands on recurrent retrieval processes (Jang et al., 2012). Cognitive theories suggest
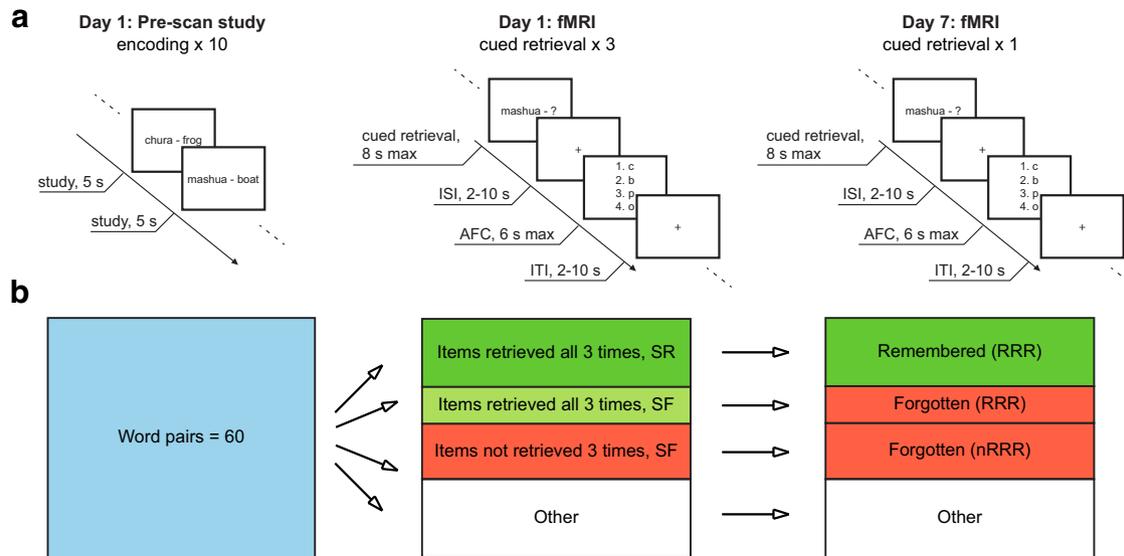
Figure 1. *a*, On Day 1, participants encoded 60 word pairs during 10 prescan encoding runs. The fMRI session on the same day consisted of three retrieval tests per word pair. One week later, Day 7, participants were scanned during a follow-up retrieval test. *b*, Study design. The figure illustrates the key conditions of interest. Of the 60 studied word pairs, items were sorted *post hoc* according to performance on both Day 1 and Day 7 to form the conditions of interest (for details, see Materials and Methods). ISI, Interstimulus interval; AFC, alternative forced choice; ITI, intertrial interval.

that recurrent retrieval affects the semantic representational space, for example, by elaborative processes that influence representations (McDaniel and Masson, 1985; Roediger and Karpicke, 2006; Carpenter, 2009). Thus, by affecting underlying representations, repeated retrieval may foster durable memory in an analogous way to that of repeated study (Xue et al., 2010). Previous studies show that successful retrieval entails reactivation of modality-specific cortices (Nyberg et al., 2000; Wheeler et al., 2000), which could serve to strengthen representational consistency during recurrent retrieval. However, there is convincing evidence that retrieval is not only a simple replay of neural ensembles engaged during learning. Rather, retrieval is an active process that may alter underlying representations (Loftus and Loftus, 1980; Nader et al., 2000; Dudai and Eisenberg, 2004) or create additional representations (Nadel and Moscovitch, 1998; Nadel and Land, 2000), which may induce representational variability. Thus, it remains unclear whether repeated retrieval serves to reactivate and strengthen consistent representations (as for repeated study) or whether it rather promotes enhanced representational variability.

Here, event-related fMRI was used in conjunction with representational similarity analysis (RSA; Kriegeskorte et al., 2008; Xue et al., 2010) to delineate brain responses characterizing successful long-term retrieval effects. Subjects were scanned twice (Days 1 and 7), where Day 1 involved three (Xue et al., 2010) repeated tests of Swahili–Swedish word pairs that had been practiced before scanning, and Day 7 involved one test. Memory performance on Day 7 defined the following two critical item classes: (1) items successfully retrieved three times on Day 1 and subsequently remembered on Day 7 (RRR_SR); and (2) items successfully retrieved three times on Day 1 but subsequently forgotten (RRR_SF) on Day 7. First, RRR_SR was contrasted with RRR_SF on Day 7, which held Day 1 retrieval success history constant, to examine the actual neural correlates of long-term retention after repeated retrieval. We hypothesized that this contrast would reveal differential fMRI responses most notably in lateral temporal cortex (Patterson et al., 2007; Binder et al., 2009) and parietal cortex (Wagner et al., 2005; Binder et al., 2009). Thereafter, changes in BOLD signal intensity and pattern similarity across

the three tests on Day 1 were examined in relation to successful retention 1 week later.

Materials and Methods

Study overview and procedure

Figure 1*a* presents a schematic overview of the study. In an initial encoding session before scanning, 60 word pairs were presented on a computer screen one by one in random order for 5 s each. Participants were instructed to memorize each word pair. Based on the results from a pilot study, 10 consecutive study runs were included to ensure that a sufficient number of items were going to be retrieved with high confidence during scanning. The Swahili–Swedish word pairs were adopted from studies by Pyc and Rawson (2009) and Nelson and Dunlosky (1994), 40 of which were the same as those found in the study by Eriksson et al. (2011). The additional 20 items were translated by the authors of this report. E-prime version 2.0 (Psychology Software Tools) was used to control presentation and responses during the experiment.

During a scanning session on the same day as the initial encoding (Day 1), each participant performed three repeated tests on each word pair. They received the Swahili word as a probe and were asked to recall the Swedish counterpart. The word pairs were randomly divided for each participant in three sessions of 20 word pairs each (60 trials per session). The lag between each repetition of a word ranged from 4 to 44 items. The Swahili word was shown for a maximum of 8 s (Fig. 1*a*). Within this time, participants were asked to respond by pressing a four-button keypad with their right hand fingers to indicate whether they had recalled a Swedish word they (1) "Knew was correct" (index finger), (2) "Believed was correct" (middle finger), or (3) "Did not retrieve a word" (ring finger). Next, a jittered (2–10 s) crosshair appeared on the screen. Participants were asked to choose among four alternatives for the second letter in the word they had just retrieved, using the right-hand fingers, within 6 s. They were instructed to guess the second letter if they had responded "did not retrieve a word." The second letter was used instead of the first letters to discourage participants from relying on simpler recognition strategies, such as first-letter cueing. This was followed by a crosshair for 2–10 s before the next word was shown. The position of the correct letter relative to the lures was systematically varied such that the target appeared equally often in each of the four possible positions. The scan session ended with structural imaging. In total, the scanning session lasted for ~2 h.

One week after the initial fMRI session, all participants returned for an additional fMRI session (Day 7). Participants were again tested on the

Table 1. Summary table of the main fMRI data analyses

Step	Purpose	Analysis	Data	Items compared	N	Trials on average over subjects
1	Long-term retention after repeated retrieval	One-sample t test	Day 7	RRR_SR vs RRR_SF	18	16.8 vs 9.4
2	Subsequent memory during repeated retrieval	t contrast (main effect)	Day 1	RRR_SR vs RRR_SF	18	50.4 vs 28.2
3	Repeated retrieval as a function of subsequent memory	F contrast (interaction)	Day 1	RRR (SR/SF) × repetition (1, 2, and 3)	15	16.4 vs 10.7 × 3 repetitions
4	Repeated retrieval as a function of subsequent memory	Representational similarity analysis	Day 1 (voxels identified on Day 7)	RRR_SR vs RRR_SF (correlated over repetitions 1, 2, and 3)	15	16.4 vs 10.7 × 3 repetitions

word pairs, but only once, using the same procedure as in the initial scanning session (Fig. 1a).

Following the scanning session on Day 7, all subjects completed a postscan confirmatory test as well as a few other tests that are not reported here. On the postscan confirmatory test, the participants were asked to fill in the Swedish word of those they classified as "know is correct" in the scanner. They were presented with a list of the Swahili words, and they were asked to write down the Swedish counterpart. No postscan confirmatory tests were given after scanning on Day 1 to prevent interference of recognition.

Participants
Twenty-three neurologically healthy subjects (mean age, 24.7 years; age range, 19–36 years; 9 males) participated in the study. They were all right handed by self-report, and they all had normal or corrected-to-normal vision. All were Swedish natives, and no participant reported prior experience with the Swahili language. Written informed consent was obtained in accordance with the Helsinki declaration, and the study was approved by the Regional Ethics Committee at Umeå University. Participants received 700 Swedish kronor for participation. Data from five participants were discarded from all analyses as they completely lacked events in one of the conditions of interest (see below). Data from an additional three participants were discarded from two of the main imaging analyses (Table 1) as they had fewer than five items per repetition in each condition of interest, impeding stable model estimates.

Behavioral data analysis
Item classification. Items stated as "know is correct" and given the correct second letter were considered as "retrieved items" with high confidence, whereas items stated as "did not retrieve a word" regardless of the second letter were considered as "forgotten items." The items were *post hoc* sorted as follows (Fig. 1b): Day 1: (1) RRR_SR, items correctly retrieved with high confidence three times on Day 1 and subsequently recalled at Day 7; (2) RRR_SF, items correctly retrieved with high confidence three times on Day 1 but subsequently forgotten at Day 7; (3) unsuccessfully RRR_SF (nRRR_SF), items not correctly retrieved with high confidence three times on Day 1 and subsequently forgotten at Day 7 [this category included all subsequently forgotten items on Day 7 that were not RRR_SF, except items that were (1) stated as "believe is correct" and given the correct second letter at least once Day 1, and (2) items that were stated as "know is correct" but given the incorrect second letter at least once Day 1, as both of these item categories presumably involved retrieval as well]; and (4) other, all items on Day 1 that were not included in any of the three categories above; and Day 7: (1) RRR_SR, items correctly retrieved with high confidence on Day 7 that were previously correctly retrieved with high confidence three times on Day 1; (2) RRR_SF, items forgotten on Day 7 that were previously correctly retrieved with high confidence three times on Day 1; (3) nRRR_SF, items forgotten on Day 7 that were not correctly retrieved with high confidence three times on Day 1; and (4) other, Day 7 items not included in any of the three categories above.

Response times. For each participant, the median response time (RT) was separately computed for the conditions RRR_SR and RRR_SF, for each of the three repetitions during Day 1. These individual median RTs were entered into a repeated-measurement ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated. Therefore, the degrees of freedom were corrected using Greenhouse–Geisser esti-

mates of sphericity. The corrected degrees of freedom were rounded up to the nearest integer.

fMRI data acquisition and preprocessing
All images were acquired using a 3.0 T whole-body MRI system (MR 750, GE Medical Systems) equipped with a 32 channels head coil. T2*-weighted images were obtained with a single-shot GE-EPI sequence used for BOLD imaging. The sequence had the following parameters: echo time, 30 ms; repetition time, 2000 ms; flip angle, 90°; FOV, 248 × 248 mm; matrix, 96 × 96; and slice thickness, 3.4 mm (37 slices acquired). Ten dummy scans were collected before data collection to allow for equilibration of the fMRI signal. High-resolution T1-weighted structural images were obtained for each participant. Stimuli were presented on a computer screen seen by the participant through a mirror attached to the head coil. Stimulus presentation was handled by a PC running E-prime version 2.0 (Psychology Software Tools) and Lumitouch fMRI optical response keypads (Photon Control) were used to collect responses. To reduce the scanner noise, headphones and earplugs were used, and cushions in the coil reduced head movement.

Functional data were preprocessed and analyzed in SPM8 (The Wellcome Department of Cognitive Neurology, Institute of Neurology, University College London, London, UK), with a batch function in an in-house program (DataZ). All images were corrected for slice timing, realigned with unwarp to correct for head movements, spatially normalized to MNI space, and smoothed (8 mm FWHM Gaussian filter kernel). Statistical analyses were calculated on the smoothed data with a high-pass filter (128 s cutoff period) to remove low-frequency noise.

fMRI data analyses
Day 7 BOLD level analysis. The point of departure for the fMRI analyses was to compare Day 7 data for items that were successfully retrieved three times (RRR) on Day 1, and then either subsequently remembered (RRR_SR) or subsequently forgotten (RRR_SF) on Day 7. This analysis identified the neural correlates underlying successful long-term retention after repeated retrieval during the acquisition phase. On Day 7, we also compared RRR_SF with nRRR_SF words to examine whether these items differed in terms of brain activity on Day 7. Thus, the statistical model of the Day 7 data consisted of three effects of interest (RRR_SR, RRR_SF, and nRRR_SF) and two effects of no interest: words not included in RRR_SR, RRR_SF, or nRRR_SF (other); and the second letter alternative forced-choice event (Fig. 1a, alternative forced choice). The six movement parameters were included as covariates. Button presses during the Swahili event were used as the onset, and the duration was set to zero. All regressors except the movement parameters were convolved with a hemodynamic response function. In the first-level analysis, model estimates were made for each participant. The model estimations from each individual were taken into a second level one-sample t test with a statistical threshold of $p < 0.001$ (uncorrected) at the voxel level, and $p < 0.05$ (FWE corrected) at the cluster level using a method based on Gaussian random field theory (Worsley et al., 1999; Hayasaka et al., 2004).

Day 1 BOLD level analysis. Next, Day 1 data were analyzed to examine how changes in BOLD signal intensity and pattern similarity across the three tests reflected successful retention 1 week later.

Button presses during the Swahili event were used as the onset, and the duration was set to zero. The model of Day 1 data consisted of nine effects of interest (RRR_SR repetition 1, RRR_SR repetition 2, RRR_SR repetition 3, RRR_SF repetition 1, RRR_SF repetition 2, RRR_SF repetition 3,

and nRRR_SF repetition 1, nRRR_SF repetition 2, and nRRR_SF repetition 3) and four effects of no interest (other repetition 1, other repetition 2, other repetition 3, and alternative forced choice event). The six movement parameters were included as covariates. All regressors except the movement parameters were convolved with a hemodynamic response function. In the first-level analysis, model estimations were made for each participant. Model component estimations (e.g., RRR_SR repetition 1) from each individual were taken into a second-level repeated-measurement ANOVA (3×3: RRR_SR, RRR_SF, nRRR_SF × three repetitions) investigating the main effect of subsequent memory as well as the interaction between subsequent memory and retrieval repetition with a statistical threshold of $p < 0.001$ (uncorrected) at the voxel level, and $p < 0.05$ (FWE-corrected) at the cluster level using a method based on Gaussian random field theory (Worsley et al., 1999; Hayasaka et al., 2004). From the interaction analysis, we report results only for clusters where there was no significant difference between RRR_SR and RRR_SF at the first retrieval repetition of a word ($p > 0.1$) followed by monotonic increases or decreases as a function of repetition.

Day 1 representational similarity analysis. For the RSA (Kriegeskorte et al., 2008; Xue et al., 2010), we constructed a β-series model with one regressor per trial and re-estimated the model on normalized but nonsmoothed data from Day 1. We then extracted the β-values for each item, for each repetition, and for each voxel in one functionally defined region of interest (ROI; see below). We next correlated (Pearson's correlation coefficient) the β-values between items and between repetitions in the three main conditions (RRR_SR, RRR_SF, and nRRR_SF) to investigate the possible item-specific effects on pattern similarity as a function of repeated retrieval. As regions of interest for the RSA were regions that would potentially prove to be of significance for retrieval success both at Day 1 and Day 7 (i.e., regions with significant activity differences between RRR_SR and RRR_SF both at Day 1 and Day 7). Importantly, we used regions of interest from Day 7 to define the voxels in which we performed a similarity analysis on Day 1. Correlations averaged over items and repetitions were then statistically compared with a paired t test to investigate whether mean within-item correlations over subjects were higher for RRR_SR than RRR_SF, or vice versa. For an overview of the main fMRI data analyses, see Table 1.

Control analyses: representational similarity. To establish reference regions with which to compare the size of the correlations in our regions of interest, we performed representational similarity analysis in two regions for which we had no hypotheses in relation to the beneficial effect of retrieval repetition (regions anatomically defined by the *Automated Anatomical Labeling* atlas; Tzourio-Mazoyer et al., 2002). We expected the correlations in bilateral amygdala (AMG) to be small for this emotionally neutral material, whereas we expected high correlations in bilateral inferior occipital cortex (OCC) owing to the fact that the exact same word was repeated on every repetition.

To ascertain that the pattern similarity results would not be driven by potential confounding variables such as item repetition lag, response times, and the number of trials in the regressor (Xue et al., 2010), correlations were performed between pattern similarity and these three variables for each of the different item categories (RRR_SR and RRR_SF).

To investigate whether differences in BOLD amplitude affected the pattern similarity correlations, the pattern similarity analyses were reperformed in the same regions of interest after removing the voxels that were active in the contrast RRR_SR versus RRR_SF Day 1 under a lenient threshold ($p < 0.005$; for a similar procedure, see Xue et al., 2010). To control for signal amplitude when investigating pattern similarity is a common procedure (Xue et al., 2010, 2013; Staresina et al., 2012; Ritchey et al., 2013; LaRocque et al., 2013), and it has been shown through simulations that signal amplitude might drive or obscure the correlations (LaRocque et al., 2013) due to, for example, increasing the signal-to-noise ratio for stimuli in a category.

Finally, to investigate whether any additional brain regions than those identified in the Day 7 and Day 1 contrasts would show differences in pattern similarity between RRR_SR and RRR_SF, we implemented a whole-brain representational similarity analysis based on a searchlight approach (Gilbert et al., 2012; Xue et al., 2013; Copara et al., 2014). We generated sphere-based ROIs with a radius of 2.5 mm (81 voxels) cen-

tered at each voxel of the brain. For each such ROI, we performed identical correlational analyses as we did in the original ROI analysis. Based on the β-series model with one regressor per trial on normalized but nonsmoothed data from Day 1, we extracted the β-values for each item, for each repetition, for each voxel (in the ROI), and for each searchlight ROI, and then correlated (Pearson's correlation coefficient) the β-values between items and repetitions in the three main conditions (RRR_SR, RRR_SF, and nRRR_SF). Correlations, averaged over items and repetitions, were then statistically compared with paired t tests to investigate differences between RRR_SF and RRR_SR in mean within-item correlations over subjects with a statistical threshold of $p < 0.001$ (uncorrected) at the voxel level, and $p < 0.05$ (FEW corrected) at the cluster level using a method based on Gaussian random field theory (Worsley et al., 1999; Hayasaka et al., 2004).

Results

Memory performance and response times
The validity of responses judged as "know is correct" and given the correct second letter in the scanner (i.e., the items retrieved with high confidence) was supported by a high correlation between these responses and the postscan confirmatory test performance ($r = 0.99, p < 0.001$).

On Day 1, on average, 33.0 of the 60 encoded word pairs were RRR. Of these RRR words, 52% were later remembered with high confidence (RRR_SR), and 28% were forgotten (RRR_SF) on Day 7. The rest of the RRR words were remembered with low confidence on Day 7 and/or given the incorrect second letter on Day 7. Moreover, on Day 1, on average, 13.9 words were nRRR_SF.

An ANOVA on the response-time data from Day 1 revealed a significant ($F_{(1,23)} = 86.5; p < 0.001$) main effect of repetition (1, 2, 3) for both RRR_SR words (2393, 1841, 1690 ms) and RRR_SF words (2831, 2071, 1845 ms). This effect confirmed a speed up of the response times as a function of repeated retrieval. In addition, there was a significant ($F_{(1,17)} = 18.2; p = 0.001$) main effect of subsequent memory (RRR_SR, RRR_SF), with faster response times for subsequently remembered words (average response times across repetitions: RRR_SR, 1975 ms; RRR_SF, 2249 ms).

Thus, although many words were correctly retrieved on all three tests on Day 1, some of them were going to be subsequently forgotten on Day 7. Moreover, the response time data indicated differences on Day 1 in the ease of access to words that would subsequently be remembered compared with forgotten.

Day 7 BOLD signal changes for successfully remembered word pairs
We compared Day 7 fMRI BOLD activity for RRR_SR versus RRR_SF words. Critically, both RRR_SR and RRR_SF words had been successfully retrieved three times on Day 1. As expected, a one-sample t test revealed significantly higher activity for RRR_SR compared with RRR_SF in bilateral posterior parietal cortex (Wagner et al., 2005; Binder et al., 2009) and lateral temporal cortex (Patterson et al., 2007; Binder et al., 2009), but also in several additional brain regions, including the right hippocampus and frontal cortex (Fig. 2a, Table 2). Conversely, contrasting RRR_SF with RRR_SR or RRR_SF with nRRR_SF at Day 7 yielded no clusters above the statistical threshold.

Day 1 BOLD signal changes in relation to successful long-term retention
The analysis of Day 1 fMRI BOLD activity as a function of subsequent memory on Day 7 (i.e., a t contrast between Day 1 RRR_SR and RRR_SF across repetitions; Fig. 1b) revealed a difference in a right posterior parietal cluster (MNI x, y, z coordi-
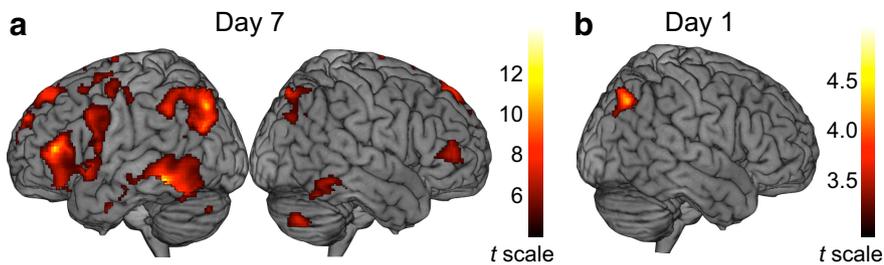
Figure 2. Day 7 and Day 1 BOLD responses reflecting long-term retention due to repeated testing. ***a***, Day 7 BOLD signal changes for successfully remembered word pairs (RRR_SR vs RRR_SF) includes bilateral posterior parietal cortex, lateral temporal cortex, and inferior and superior prefrontal cortex. ***b***, Day 1 BOLD signal changes in relation to successful long-term retention (RRR_SR vs RRR_SF) in right posterior parietal cortex.

nates, 34, −68, 42; $t = 4.31$, $k = 325$). This cluster overlapped with the right parietal region that was identified on Day 7 (Fig. 2a,b, Table 2). Thus, despite the fact that the RRR_SR and RRR_SF item classes were behaviorally identical on Day 1 (successful retrieval on all three consecutive tests), the right parietal response was diagnostic of the difference in long-term retention on Day 7 (elevated for RRR_SR words). The parietal locus of the effect is in line with recent imaging work on the testing effect (i.e., testing is more beneficial for long-term retention of a material than rereading, for a comprehensive review, see Roediger and Butler, 2011). Most, if not all, of the relevant imaging studies reported parietal activity differences (Eriksson et al., 2011; Nelson et al., 2013; Wing et al., 2013; van den Broek et al., 2013; Keresztes et al., 2014; Liu et al., 2014; Vestergren and Nyberg, 2014), and some studies demonstrated a positive relation between right posterior parietal activity during retrieval and subsequent retrieval performance (van den Broek et al., 2013; Liu et al., 2014).

In addition, the contrast revealed a difference in a medial prefrontal cluster (x, y, z, 0, 60, 14; $t = 3.90$; $k = 396$). Some previous work has implicated medial frontal cortex in successful long-term memory. For example, medial prefrontal cortex has been implicated in systems-level consolidation of memories in humans (Takashima et al., 2007; Sterpenich et al., 2009; van Kesteren et al., 2010) and in rodents (Frankland and Bontempi, 2005; Restivo et al., 2009; Lehmann and McNamara, 2011). Also, Eriksson et al. (2011) demonstrated a relationship between long-term retention after successful repeated retrieval and medial prefrontal cortex activity (in dorsal medial PFC/ACC).

The reversed contrast, testing whether any brain regions were more strongly recruited on Day 1 for words that would be forgotten on Day 7, did not reveal any significant effects.

Day 1 BOLD signal repetition effects in relation to successful long-term retention
A three (repetition, 1, 2, 3) by two (subsequent memory; i.e., RRR_SR, RRR_SF) ANOVA was used to examine whether a BOLD signal change as a function of the three repeated tests related in a systematic way to subsequent memory (i.e., this analysis tested for a repetition by subsequent-memory interaction). This analysis yielded no clusters above the statistical threshold. Under a more liberal threshold, the strongest significant effect was observed in left dorsolateral prefrontal cortex (DLPFC; x, y, z, −34, 34, 22; $F = 17.01$; $k = 106$; $p = 0.0001$; Fig. 3a), which reflected a more marked BOLD signal reduction across tests for RRR_SR words than for RRR_SF words (Fig. 3b).

Thus, although the words in both of these item classes were successfully retrieved on all three consecutive tests, the change in

DLPFC responses as a function of repeated tests differed depending on long-term retention on Day 7. This finding, although it should be considered preliminary as it did not pass the chosen statistical threshold, replicates previous findings of decreases in prefrontal activity as a function of retrieval repetitions (Kuhl et al., 2007). The DLPFC is associated with processes related to retrieval–selection mechanisms, and the resolution of response conflicts between different memory associations (Badre and Wagner, 2004). Presumably, the need for such control processes is reduced with every retrieval repetition, as the material gets more and more rehearsed.

Repeated retrieval in relation to pattern similarity
The observed Day 7 BOLD signal differences between remembered and forgotten words in right posterior parietal cortex (Fig. 2a) was mirrored by the finding that already at Day 1, the response in the same region was diagnostic of the difference in long-term retention on Day 7 (Fig. 2b). Interestingly, the difference in parietal BOLD signals is one of the key findings in related studies on the mnemonic benefits of retrieval (Eriksson et al., 2011; Nelson et al., 2013; Wing et al., 2013; van den Broek et al., 2013; Keresztes et al., 2014; Liu et al., 2014; Vestergren and Nyberg, 2014). This region was therefore chosen as the region of interest for the representational similarity analyses (i.e., analyses of whether repeated retrieval on Day 1 altered neural pattern similarity). The voxels in right superior parietal cortex at Day 7, which is important for successful long-term retention after repeated retrieval (Table 2, cluster 3), formed the basis for the representational similarity analysis of Day 1 data.

A paired t test revealed significantly lower item-specific pattern similarity across repeated retrieval on Day 1 for RRR_SR words compared with RRR_SF words in this right parietal region (Fig. 4 SR vs SF white bars; $t_{(14)} = -2.36$, $p = 0.033$). Notably, the highest item-specific similarity was observed for words that were nRRR_SF; Fig. 4, nR gray bar), implying that repeated retrieval per se might be associated with low pattern similarity in this region.

Control analyses: pattern similarity
As reference regions for the pattern similarity analysis, we computed pattern similarity in bilateral inferior OCC, where we expected relatively higher correlations due to the visual presentation modality, and in bilateral AMG, where we expected relatively lower correlations due to the nonemotional nature of the stimulus materials. These expectations were confirmed (mean SR_OCC = 0.28, mean SF_OCC = 0.27, $t_{(14)} = 0.64$, $p = 0.54$; and mean SR_AMG = 0.03, mean SF_AMG = 0.04, $t_{(14)} = -0.66$, $p = 0.52$).

Neither repetition lags ($p = 0.52$ vs $p = 0.36$) and mean response times ($p = 0.63$ vs $p = 0.34$), nor the number of trials in the regressors ($p = 0.11$ vs $p = 0.93$) were significantly correlated with pattern similarity for RRR_SR and RRR_SF items, respectively, thus ruling out potential confounds highlighted previously (Xue et al., 2010).

Removing the voxels from the right parietal region that displayed a main effect in BOLD amplitude of RRR_SR versus RRR_SF on Day 1 did not affect the results. That is, the difference between subsequently remembered and forgotten words with re-

Table 2. Day 7: Regional responses characterizing long-term retention after repeated retrieval [words successfully repeatedly retrieved and recalled at the 1 week follow-up (RRR_SR) versus words successfully repeatedly retrieved and forgotten (RRR_SF)]

Cluster	Local maxima	Hemisphere	BA	MNI coordinates			z-score	Voxels (k)
				x	y	z		
1	Inferior temporal	L	20	−54	−42	−12	6.38	24,398
	Inferior frontal triangularis	L	45	−44	38	12	6.04	
	Intraparietal sulcus	L	7	−26	−74	42	5.91	
	Putamen	L		−28	6	2	5.31	
	Superior frontal	L	9	−10	40	50	5.18	
2	Fusiform gyrus	R	20	42	−16	−24	5.21	490
	Amygdala	R	34	30	0	−20	4.17	
	Hippocampus	R	34	20	−2	−18	3.57	
	Parahippocampal gyrus	R	37	36	−32	−14	3.39	
3	Superior parietal	R	7	32	−76	52	5.07	1161
	Middle occipital	R	19	34	−74	34	4.35	
	Angular gyrus	R	7	44	−64	54	4.03	
4	Cerebellum	R		38	−70	−42	4.78	1341
	Inferior temporal	R	37	62	−46	−14	4.49	
	Middle temporal	R	21	66	−36	0	3.54	
5	Middle frontal	R	46	48	46	6	4.57	562
	Inferior frontal triangularis	R	45	44	28	12	3.40	

Up to 5 local maxima are reported for each cluster. BA, Brodmann area; L, left; R, right. MNI coordinates were determined using SPM8. z-values are at the peak voxel. Voxel, $p < 0.001$ (uncorrected); cluster, $p < 0.05$ (FWE).

spect to pattern similarity persisted and did not depend on the presence of these voxels (SR vs SF: $t_{(14)} = −2.30$; $p = 0.037$).

For completeness, we also computed pattern similarity in the second region identified in the analysis of Day 1 fMRI BOLD activity contrasting RRR_SR versus RRR_SF (the medial prefrontal cluster: $x, y, z, 0, 60, 14$; $t = 3.90$; $k = 396$), even though it was not similarly mirrored by an overlapping cluster at Day 7, as was the right parietal cluster. Overall, correlations in this region were small, and there was no significant difference between RRR_SR and RRR_SF words (mean SR = 0.05; mean SF = 0.06; $t_{(14)} = −0.94$; $p = 0.363$).

Additionally, the searchlight-based whole-brain analysis replicated lower item-specific pattern similarity across repeated retrieval on Day 1 for subsequently remembered RRR_SR compared with RRR_SF items in the right parietal cluster ($x, y, z, 44, −76, 38$; $t = −2.42$; $k = 7$; $p = 0.015$). However, this analysis yielded no clusters above the statistical threshold.

Discussion

The present study provides novel neuroimaging findings on the mnemonic benefits of repeated retrieval. One week following initial repeated successful retrieval, we found that remembered items elicited higher activity compared with forgotten items in brain regions previously implicated in retrieval success (Spaniol et al., 2009), notably left lateral temporal cortex and bilateral posterior parietal cortices (Martin, 2007; Patterson et al., 2007; Binder et al., 2009; Binder and Desai, 2011; Whitney et al., 2011). Critically, on Day 1, repeatedly retrieved items that were later going to be recalled relative to items that were later going to be forgotten evoked a higher BOLD signal in the same right posterior parietal cluster, which was associated with retrieval success on Day 7.
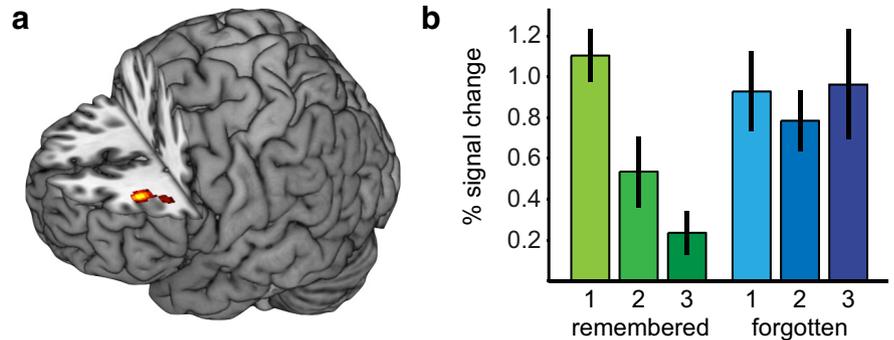

Figure 3. Day 1 BOLD signal repetition effects in relation to successful long-term retention. *a*, A repetition × subsequent memory effect was evident in left DLPFC ($x, y, z, −34, 34, 22$; $F = 17.01$; $k = 106$; $p = 0.0001$). *b*, This effect was characterized by marked reductions in the BOLD response for RRR_SR items compared with RRR_SF items. Error bars denote ±1 SE.
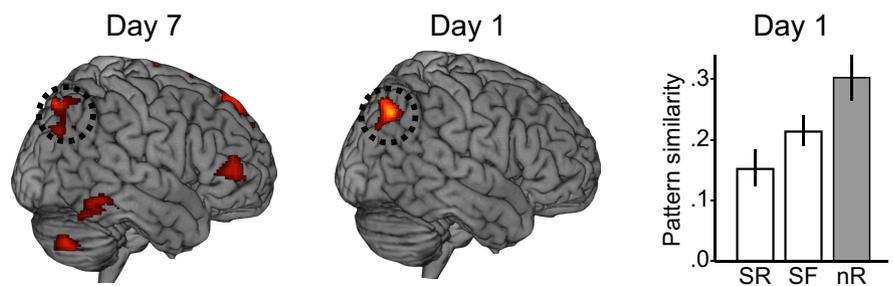

Figure 4. Repeated retrieval in relation to pattern similarity. In the right posterior parietal region of interest, within-item pattern similarity on Day 1 [as measured with Pearson's correlation (r)] was lower for items that were SR compared with SF. nR = nRRR_SF. Error bars denote ±1 SE.

This difference at Day 1 was seen despite identical behavioral performance (i.e., both item categories were successfully retrieved, with high confidence, three times on Day 1). These observations from Days 1 and 7 point to an important role for parietal cortex in relation to the long-term beneficial effect of repeated retrieval. The representational similarity analysis provided insights into the basis for the parietal effect: repeated retrieval induced representational variability rather than representational consistency. That is, the pattern similarity in the right posterior parietal region was lower for words repeatedly retrieved on Day 1 and subsequently recalled on Day 7,

compared with words repeatedly retrieved but subsequently forgotten. This difference remained after controlling for a number of factors, including BOLD amplitude differences.

A key role for the parietal cortex in learning and retrieving word pairs may reflect a role for this brain region in the representation of semantic memories, by acting as a convergence zone where information is bound together to form schematic semantic representations (Geschwind, 1965; Binder et al., 2009; Binder and Desai, 2011; Bonner et al., 2013). The parietal cortex is tightly connected with several modality-specific zones (Bonner et al., 2013), which makes it a strong candidate region for semantic integration of modality-specific information in the service of long-term memory. The finding that lower pattern similarity in the parietal cortex is predictive of long-term retention after repeated retrieval is in line with neurocognitive notions of retrieval as a process that can create additional representations (Nadel and Moscovitch, 1998; Nadel and Land, 2000). In fact, our observations might resonate with the bulk of evidence highlighting pattern separation as a key mechanism of learning, whereby similar representations get stored separately from each other to enable flexible use of differentiated representations of a stimulus (for extensive review, see Yassa and Stark, 2011). Pattern separation is a mechanism commonly ascribed to episodic memory, and particularly often to the hippocampus (Yassa and Stark, 2011). Interestingly, using pattern similarity analysis, two recent studies have demonstrated that pattern separation in the hippocampus can predict subsequent memory after encoding (LaRocque et al., 2013) and is related to more accurate spatiotemporal memory (Copara et al., 2014). Evidence for pattern separation have also been observed in the cortex (Rotshtein et al., 2005).

Moreover, our finding that lower pattern similarity in the parietal cortex is predictive of long-term retention after repeated retrieval might also be in line with neurocognitive notions of retrieval as a process that can alter underlying representations (Loftus and Loftus, 1980; Nader et al., 2000; Dudai and Eisenberg, 2004). This would then imply that repeated retrieval enhances the formation of more variable or elaborate semantic representations over the course of multiple rounds of successful retrieval.

Our observations stand in contrast to the results previously reported by Xue et al. (2010, 2013) concerning repeated encoding, where it was demonstrated that subsequent memory after repeated encoding was associated with greater neural pattern similarity, most robustly in parietal cortex (i.e., angular gyrus; Xue et al., 2013; for similar results albeit in other brain regions and for fear memory and scene recognition, respectively, see also Visser et al., 2013; Ward et al., 2013).

These seemingly contradictory results for repeated encoding compared with repeated retrieval might offer insight into the neurocognitive mechanisms underlying the testing effect (Roediger and Butler, 2011). For recurrent encoding, greater neural pattern similarity might be related to successful long-term retention because of the benefits of providing a more consistent input to the medial temporal lobes (Xue et al., 2013), potentially simplifying the interplay between pattern separation and pattern completion processes (Yassa and Stark, 2011). For recurrent retrieval, however, lesser neural pattern similarity might be related to successful long-term retention because of the benefits of creating more variable or elaborated cue–target associations. Indeed, the notion that elaborated cue–target associations gives retrieval its mnemonic benefits compared with encoding has been put forward as a theoretical explanation for the testing effect (McDaniel and Masson, 1985; Carpenter and Delosh, 2006; Roediger and Karpicke, 2006; see also Karpicke and Roediger, 2008;

Carpenter, 2009, 2011; Roediger and Butler, 2011). Tentatively, the neurocognitive mechanism we found to characterize successful long-term retention after repeated retrieval might be considered more beneficial for promoting durable learning than the mechanism proposed for repeated encoding, as evidenced by numerous demonstrations of the testing effect (Roediger and Karpicke, 2006; Roediger and Butler, 2011).

In conclusion, we demonstrate that recurrent retrieval processes may boost long-term memory retention via representational variability, as indicated by an elevated BOLD signal and low pattern similarity in parietal cortex for words repeatedly retrieved and subsequently remembered. These results are in the opposite direction of what has been found previously for recurrent encoding processes, and is potentially a sign that semantic elaboration of memory representations is an important factor in promoting durable learning following repeated retrieval. This study adds to the growing body of evidence suggesting that recurrent retrieval should be endorsed as an effective learning method, perhaps together with task instructions encouraging semantic elaboration during retrieval.

References

Badre D, Wagner AD (2004) Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron 41:473–487. CrossRef Medline

Binder JR, Desai RH (2011) The neurobiology of semantic memory. Trends Cogn Sci 15:527–536. 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

Bonner MF, Peelle JE, Cook PA, Grossman M (2013) Heteromodal conceptual processing in the angular gyrus. Neuroimage 71:175–186. CrossRef Medline

Carpenter SK, DeLosh EL (2006) Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. Mem Cognit 34:268–276. CrossRef Medline

Carpenter SK (2009) Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. J Exp Psychol Learn Mem Cogn 35:1563–1569. CrossRef Medline

Carpenter SK (2011) Semantic information activated during retrieval contributes to later retention: support for the mediator effectiveness hypothesis of the testing effect. J Exp Psychol Learn Mem Cogn 37:1547–1552. CrossRef Medline

Copara MS, Hassan AS, Kyle CT, Libby LA, Ranganath C, Ekstrom AD (2014) Complementary roles of human hippocampal subregions during retrieval of spatiotemporal context. J Neurosci 34:6834–6842. CrossRef Medline

Dudai Y, Eisenberg M (2004) Rites of passage of the engram: reconsolidation and the lingering consolidation hypothesis. Neuron 44:93–100. CrossRef Medline

Ebbinghaus H (1885) Memory: a contribution to experimental psychology. New York: Teachers College, Columbia University.

Eriksson J, Kalpouzos G, Nyberg L (2011) Rewiring the brain with repeated retrieval: a parametric fMRI study of the testing effect. Neurosci Lett 505:36–40. CrossRef Medline

Frankland PW, Bontempi B (2005) The organization of recent and remote memories. Nat Rev Neurosci 6:119–130. CrossRef Medline

Geschwind N (1965) Disconnexion syndromes in animals and man. I. Brain 88:237–294. CrossRef Medline

Gilbert SJ, Armbruster DJ, Panagiotidi M (2012) Similarity between brain activity at encoding and retrieval predicts successful realization of delayed intentions. J Cogn Neurosci 24:93–105. CrossRef Medline

Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE (2004) Nonstationary cluster-size inference with random field and permutation methods. Neuroimage 22:676–687. CrossRef Medline

Jang Y, Wixted JT, Pecher D, Zeelenberg R, Huber DE (2012) Decomposing the interaction between retention interval and study/test practice: the role of retrievability. Q J Exp Psychol 65:962–975. CrossRef

Karpicke JD, Roediger HL 3rd (2008) The critical importance of retrieval for learning. Science 319:966–968. CrossRef Medline

Keresztes A, Kaiser D, Kovács G, Racsmány M (2014) Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. Cereb Cortex 24:3025–3035. CrossRef Medline

Kornell N, Bjork RA, Garcia MA (2011) Why tests appear to prevent forgetting: a distribution-based bifurcation model. J Mem Lang 65:85–97. CrossRef

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

Kuhl BA, Dudukovic NM, Kahn I, Wagner AD (2007) Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat Neurosci 10:908–914. 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

Lehmann H, McNamara KC (2011) Repeatedly reactivated memories become more resistant to hippocampal damage. Learn Mem 18:132–135. CrossRef Medline

Liu XL, Liang P, Li K, Reder LM (2014) Uncovering the neural mechanisms underlying learning from tests. PLoS One 9:e92025. CrossRef Medline

Loftus EF, Loftus GR (1980) On the permanence of stored information in the human brain. Am Psychol 35:409–420. CrossRef Medline

Martin A (2007) The representation of object concepts in the brain. Annu Rev Psychol 58:25–45. CrossRef Medline

McDaniel MA, Masson MEJ (1985) Altering memory representations through retrieval. J Exp Psychol Learn Mem Cogn 11:371–385. CrossRef

Nadel L, Land C (2000) Commentary-reconsolidation: memory traces revisited. Nat Rev Neurosci 1:209–212. CrossRef Medline

Nadel L, Moscovitch M (1998) Hippocampal contributions to cortical plasticity. Neuropharmacology 37:431–439. CrossRef Medline

Nader K, Schafe GE, LeDoux JE (2000) The labile nature of consolidation theory. Nat Rev Neurosci 1:216–219. CrossRef Medline

Nelson SM, Arnold KM, Gilmore AW, McDermott KB (2013) Neural signatures of test-potentiated learning in parietal cortex. J Neurosci 33:11754–11762. CrossRef Medline

Nelson TO, Dunlosky J (1994) Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. Memory 2:325–335. CrossRef Medline

Nyberg L, Habib R, McIntosh AR, Tulving E (2000) Reactivation of encoding-related brain activity during memory retrieval. Proc Natl Acad Sci U S A 97:11120–11124. CrossRef Medline

Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci 8:976–987. CrossRef Medline

Pyc MA, Rawson KA (2009) Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? J Mem Lang 60:437–447. CrossRef

Restivo L, Vetere G, Bontempi B, Ammassari-Teule M (2009) The formation of recent and remote memory is associated with time-dependent formation of dendritic spines in the hippocampus and anterior cingulate cortex. J Neurosci 29:8206–8214. CrossRef Medline

Ritchey M, Wing EA, LaBar KS, Cabeza R (2013) Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb Cortex 23:2818–2828. CrossRef Medline

Roediger HL, Karpicke JD (2006) The power of testing memory. Perspect Psychol Sci 1:181–210. CrossRef

Roediger HL 3rd, Butler AC (2011) The critical role of retrieval practice in long-term retention. Trends Cogn Sci 15:20–27. CrossRef Medline

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci 8:107–113. CrossRef Medline

Spaniol J, Davidson PS, Kim AS, Han H, Moscovitch M, Grady CL (2009) Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. Neuropsychologia 47:1765–1779. CrossRef Medline

Staresina BP, Henson RN, Kriegeskorte N, Alink A (2012) Episodic reinstatement in the medial temporal lobe. J Neurosci 32:18150–18156. CrossRef Medline

Sterpenich V, Albouy G, Darsaud A, Schmidt C, Vandewalle G, Dang Vu TT, Desseilles M, Phillips C, Degueldre C, Balteau E, Collette F, Luxen A, Maquet P (2009) Sleep promotes the neural reorganization of remote emotional memory. J Neurosci 29:5143–5152. CrossRef Medline

Takashima A, Nieuwenhuis IL, Rijpkema M, Petersson KM, Jensen O, Fernández G (2007) Memory trace stabilization leads to large-scale changes in the retrieval network: a functional MRI study on associative memory. Learn Mem 14:472–479. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289. CrossRef Medline

van den Broek GS, Takashima A, Segers E, Fernández G, Verhoeven L (2013) Neural correlates of testing effects in vocabulary learning. Neuroimage 78:94–102. CrossRef Medline

van Kesteren MT, Fernández G, Norris DG, Hermans EJ (2010) Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. Proc Natl Acad Sci U S A 107:7550–7555. CrossRef Medline

Vestergren P, Nyberg L (2014) Testing alters brain activity during subsequent restudy: evidence for test-potentiated encoding. Trends Neurosci Educ 3:69–80. CrossRef

Visser RM, Scholte HS, Beemsterboer T, Kindt M (2013) Neural pattern similarity predicts long-term fear memory. Nat Neurosci 16:388–390. 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 cortices. Proc Natl Acad Sci U S A 97:11125–11129. CrossRef Medline

Whitney C, Jefferies E, Kircher T (2011) Heterogeneity of the left temporal lobe in semantic representation and control: priming multiple versus single meanings of ambiguous words. Cereb Cortex 21:831–844. CrossRef Medline

Wiklund-Hörnqvist C, Jonsson B, Nyberg L (2014) Strengthening concept learning by repeated testing. Scand J Psychol 55:10–16. CrossRef Medline

Wing EA, Marsh EJ, Cabeza R (2013) Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. Neuropsychologia 51:2360–2370. CrossRef Medline

Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC (1999) Detecting changes in nonisotropic images. Hum Brain Mapp 8:98–101. 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

Yassa MA, Stark CE (2011) Pattern separation in the hippocampus. Trends Neurosci 34:515–525. CrossRef Medline