Behavioral/Cognitive

Briefly Cuing Memories Leads to Suppression of Their Neural Representations

Jordan Poppenk and Kenneth A. Norman
Princeton Neuroscience Institute, Princeton University, Princeton Neuroscience Institute Building, Princeton, New Jersey 08540

Previous studies have linked partial memory activation with impaired subsequent memory retrieval (e.g., Detre et al., 2013) but have not provided an account of this phenomenon at the level of memory representations: How does partial activation change the neural pattern subsequently elicited when the memory is cued? To address this question, we conducted a functional magnetic resonance imaging (fMRI) experiment in which participants studied word-scene paired associates. Later, we weakly reactivated some memories by briefly presenting the cue word during a rapid serial visual presentation (RSVP) task; other memories were more strongly reactivated or not reactivated at all. We tested participants' memory for the paired associates before and after RSVP. Cues that were briefly presented during RSVP triggered reduced levels of scene activity on the post-RSVP memory test, relative to the other conditions. We used pattern similarity analysis to assess how representations changed as a function of the RSVP manipulation. For briefly cued pairs, we found that neural patterns elicited by the same cue on the pre- and post-RSVP tests (preA–postA; preB–postB) were less similar than neural patterns elicited by different cues (preA–postB; preB–postA). These similarity reductions were predicted by neural measures of memory activation during RSVP. Through simulation, we show that our pattern similarity results are consistent with a model in which partial memory activation triggers selective weakening of the strongest parts of the memory.

Key words: **fMRI; human memory; memory retrieval; memory weakening; multivoxel pattern analysis; nonmonotonic plasticity**

Introduction

What factors cause memory weakening? According to the nonmonotonic plasticity hypothesis (Newman and Norman, 2010), partial reactivation of memories causes weakening of those memories, whereas stronger reactivation causes strengthening (and no reactivation causes no learning). This hypothesis is supported by neurophysiological evidence showing that moderate but not high levels of depolarization lead to synaptic weakening (Artola et al., 1990; Hansel et al., 1996) and electroencephalography (EEG) and fMRI results showing impaired subsequent memory for moderately reactivated events (Newman and Norman, 2010; Detre et al., 2013) (for relevant behavioral evidence, see also Keresztes and Racsmány, 2013).

It is unclear, however, how these weakening effects arise at the level of neural representations: How does partial activation change the neural pattern elicited when the memory subsequently is cued? One possibility is that the originally activated neural pattern will simply fade in salience relative to background noise, in which case its instantiation in the brain may appear similar over time, but weaker. Another possibility, suggested by our prior neural network simulations exploring nonmonotonic plasticity (Norman et al., 2006a), is that the neural patterns elicited by cues may change qualitatively (Fig. 1). In particular, these simulations predict that partial activation of a cue will be sufficient to trigger partial reactivation of its strongest associated features, but weaker features will remain inactive. Partial reactivation of these strongly connected features will trigger weakening of connections into the features, whereas connections into weaker (inactive) features will remain intact. Collectively, these changes will result in a reversal of feature strength values, such that the features that were previously strongest are now eclipsed by other (previously weaker) features.

To address these possibilities, we had participants study word-scene pairs and compared their neural representations of scene associates before and after repeated memory reactivation. Participants were trained to ceiling levels of performance with word-scene associate pairs; we overtrained the pairs to ensure that word cues would elicit a strong representation of the associated scene. Then, as participants were scanned with fMRI, we exposed them to the word cues under conditions designed to weakly or strongly reactivate their scene associates. Critically, before and after this phase, participants performed cued visualization of the scene. This allowed us to compare visualization data across these two sessions to assess (1) whether memory strength was altered and (2) how the underlying neural representations changed.

To measure the extent to which scene memories were reactivated throughout our experiment, we trained a pattern classifier to detect scene-specific activation in participants' fMRI data. To examine how reactivation altered memory representations, we

Received Oct. 28, 2013; revised April 28, 2014; accepted May 1, 2014.

Author contributions: J.P. and K.A.N. designed research; J.P. performed research; J.P. and K.A.N. analyzed data; J.P. and K.A.N. wrote the paper.

This work was supported by a National Sciences and Engineering Research Council Post-Doctoral Fellowship to J.P. and by National Institutes of Health Grant R01 MH069456 to K.A.N. We thank M. Aly, J. Hulbert, and N. Turk-Browne for helpful comments on an earlier version of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Jordan Poppenk, Princeton Neuroscience Institute, Princeton University, Princeton Neuroscience Institute Building, Princeton, NJ 08540. E-mail: jpoppenk@princeton.edu.

DOI:10.1523/JNEUROSCI.4584-13.2014

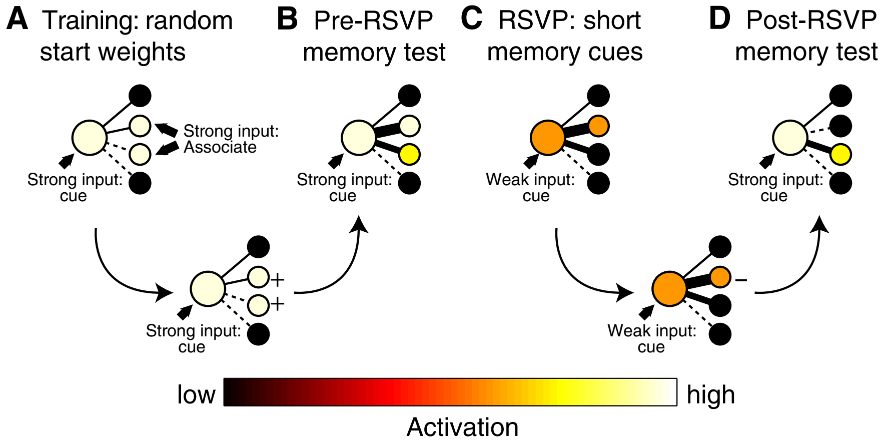

Figure 1. Predictions of the Norman et al. (2006a) neural network model. During training (***A***), both the cue and associate are presented together. Thinner lines indicate weaker connections. As a result of this strong coactivation, connections between the cue and features of the associate are strengthened (bottom row). Consequently, when the cue is presented in the pre-RSVP memory test (***B***), features of the associate are strongly activated through spreading activation. During the RSVP phase (***C***), the cue is presented only briefly, leading to partial cue activation, as well as partial reactivation of its most strongly associated features through spreading activation. As a result of this partial activation, connections between the cue and these strongly associated features are weakened; connections between the cue and other (inactive) features are unchanged (bottom row). Consequently, when the cue is presented in the post-RSVP memory test (***D***), activation preferentially spreads to features that were formerly weaker, yielding a "reversal" in the ordinal ranking of cued feature activations (in this example, the middle two features show a reversal).

assessed the similarity of neural patterns elicited by a particular memory cue before and after reactivation. We predicted that partially reactivated memories would be recalled less well (as measured by our pattern classifier) on the postreactivation memory test, relative to nonreactivated or more strongly reactivated ones. We also predicted that partially reactivated memories would show representational changes (in the pattern similarity analysis) consistent with the distinctive "reversal" pattern predicted by our computational model.

Materials and Methods

Overview

The experiment contained five main phases (Table 1; Fig. 2): paired-associate training (Phase 1), memory reactivation (Phase 4), and preactivation and postreactivation memory tests (Phases 2 and 5). In addition, a functional localizer was collected to assist with pattern classification analysis (Phase 3). To reactivate memories, we used a rapid serial visual presentation (RSVP) task in which participants monitored for target words (types of fruit) within a serial stream of nonfruit words (Potter, 1976). The stream contained occasional short (600 ms) and long (2000 ms) presentations of cue words (nonfruit nouns previously paired with scenes) that were intended to reactivate associated scene memories to differing degrees. Other cue words were omitted entirely from the RSVP task. Our hypotheses concerned the impact on the post-RSVP memory test of the RSVP reactivation manipulation (i.e., whether memories were partially reactivated by a short cue, more strongly reactivated by a long cue, or not reactivated at all during RSVP).

Participants

Sixteen right-handed volunteers participated in the experiment (five female, mean age 21.2 years). All were native English-speakers between 18 and 35 years of age with normal or corrected-to-normal vision and hearing. Participants were screened for neurological and psychological conditions and received financial remuneration. The protocol was approved by the Institutional Review Board for Human Subjects at Princeton University.

Stimuli

Participants learned 30 word-scene pairings. Words were concrete, imagable nouns sampled from the MRC Psycholinguistic Database (Coltheart, 1981) (mean length = 6.3 letters; mean concreteness = 571.5; mean imagability = 561.3; mean Thorndike–Lorge verbal frequency = 241.68) and filtered to exclude nouns semantically related to rooms and fruit. Paired scenes were grayscale bedroom interiors drawn from Detre et al. (2013). Each participant received a different random pairing of words and images. Forty additional words were used as lures; these words were randomly sampled from a pool of 7000 neutral nonfruit nouns from the MRC Psycholinguistic Database (mean length = 8.4 letters; mean concreteness = 471.3; mean imagability = 483.9; mean Thorndike–Lorge verbal frequency = 191.1). Last, 28 fruit nouns were selected for use as targets during the Phase 4 RSVP task (mean length = 6.8 letters; mean concreteness = 608.3; mean imagability = 605.0; mean Thorndike–Lorge verbal frequency = 165.8).

Ten other scene, face, car, and word images were presented during a functional localizer phase. Scenes were sampled as above. Faces were unfamiliar male faces cropped to include the full face, excluding shoulders and hair. Cars were lateral views of cars oriented leftwards. Words were drawn from the word pool above and rendered in black Arial with thick white outlines. Backgrounds of face and car images were removed; face, car, and word images were then placed on top of phase-scrambled versions of the scene images to provide a complex background.

All text in the experiment was presented in black Geneva font (height = 0.8° visual angle) on a white background (with the exception of text images in the localizer, which were matched to the visual properties of other image categories). All images in the experiment were the same size (9.0° × 9.0° visual angle) and normalized with respect to their luminance using the procedure described by Detre et al. (2013).

Procedure

Phase 1: learning of stimulus materials and word-scene associates. During paired-associate training, participants were familiarized with word-scene associates. This phase took place over 2 d: on day 1, participants completed initial study of the associates, a train-to-criterion memory task, and an RSVP practice task in a behavioral testing room. On day 2, participants completed the train-to-criterion memory task and RSVP practice tasks again, but from within the fMRI scanner bore.

Initial study consisted of two passes through the set of 30 word-scene associates. All 30 pairs were presented once, then the order was randomized, and the 30 pairs were presented again. Participants were told that a memory test would follow and that, to make stronger memories, they should try to imagine the most creative, distinctive possible explanation for how each "hotel room" got its name (i.e., the cue word). Cue words were presented for 5500 ms; 1500 ms after each cue word onset, the scene image also appeared below the word. A fixation cross of 750 ms duration separated trials.

Next, participants completed a train-to-criterion memory test. Each trial incorporated three parts (Fig. 2). First, a cue word was presented for 4000 ms, during which time participants were instructed to visualize the associated scene in as much detail as possible. Next, they were asked to rate their visualization on the following scale: 1, no room-related imagery; 2, generic room with no distinguishing features; 3, room with a specific distinguishing feature; 4, room with multiple specific distinguishing features; 5, complete image. After a subjective response was entered, the associated scene image plus scenes from three other studied pairings were presented in random order from left to right. Participants had 3000 ms to select the scene associated with the presented cue word via a button press. If a correct response was entered before the deadline, green exclamation points were presented for 750 ms; otherwise, a red "X"

Table 1. Schematic of main experimental phases

Phase and purpose	Day	Location	Participant tasks
Phase 1: learning of stimulus materials and word-scene associates	1	Testing room	● Study word-scene list (twice) ● Relearn word-scene pairs to criterion ● Familiarization with lure words in an RSVP practice task
	2	Scanner (anatomical)	● Relearn word pairs to criterion ● Familiarization with lure words in an RSVP practice task
Phase 2: pre-RSVP memory test	2	Scanner (fMRI)	● Cued recall visualization ratings and multiple choice
Phase 3: category localizer	2	Scanner (fMRI)	● One-back task using four categories of images (scenes, faces, cars, words)
Phase 4: cue-exposure manipulation (RSVP)	2	Scanner (fMRI)	● Exposure of memory cues for 600 ms (short) and 2000 ms (long) durations, with other cues left out (omit)
Phase 5: post-RSVP memory test	2	Scanner (fMRI)	● Cued recall visualization ratings and multiple choice

was presented for 750 ms, followed by presentation of the cue word with the correct scene image for 4000 ms. A 5000 ms fixation cross separated each trial. Each item remained in the list until it received a correct multiple-choice response, at which point it was dropped from the study set. The order of the remaining pairs in the study set was randomly shuffled after each pass through the study set.

Participants were then given a 7.7 min "practice" version of the RSVP task they would later complete in the scanner, in which they responded to fruit words with a button press. The practice task also served to familiarize participants with 40 words to be used as lures in memory tests. These words were presented repeatedly, with the duration of each presentation sampled from a uniform distribution with limits of 300–750 ms. Six target fruit word trials were presented for 1000 ms during the task, appearing at random intervals but no sooner than 8000 ms after a previous target. Participants were given feedback on their performance at the end of the task.

Participants repeated the train-to-criterion memory test and RSVP practice task on day 2 inside the scanner bore while anatomical scans were completed. Instead of viewing a computer monitor and responding using a keyboard, participants viewed images projected into the scanner bore using an overhead mirror and responded using a MR-safe button box. Participants quickly learned the 30 paired associates to criterion levels both on day 1 (mean ± SD: 37.0 ± 7.5 trials) and on day 2 (mean ± SD: 34.3 ± 3.5 trials).

Phases 2 and 5: pre- and post-RSVP memory tests. In Phases 2 and 5, test items included all 30 studied cue words plus 20 lures that had previously been familiarized during the "practice RSVP" task. In each of the two phases, a different set of 20 lures and a different random sequence were used. A practice test item was shown at the start of both phases. On each test trial (Fig. 2), participants were presented with a hotel room name (i.e., cue word) for 5000 ms and were instructed to visualize the associated scene (if there was one). They were then presented with the text "rate visualization" for 3000 ms; during this period, they had to enter their visualization rating using the same 5-point rating scale that was used in Phase 1. Next, the multiple-choice prompt appeared; as in Phase 1, four scenes were presented, and participants had to choose which of these scenes went with the cue word. Participants had 3000 ms to enter their multiple-choice response. The multiple-choice period was followed by a fixation cross for 9000 ms. No feedback was presented, and the pace of the experiment did not vary based on participant responses. The full set of 51 trials took 17.1 min (514 volumes) to complete.

Phase 3: category localizer. The goal of this phase was to obtain a clean neural signal associated with viewing photographs of scenes, faces, cars, and words, which we later used to train a category-specific pattern classifier. To ensure that participants paid attention to the images, they performed a "one-back" task in which they pressed a button when any image appeared twice in a row. The image sequence was divided into category-specific blocks. Within each block, 10 different images from the same category (scene, face, car, or word) were presented sequentially for 900 ms each with a 100 ms interstimulus interval (Fig. 2). A random six images of the 10 were repeated within this sequence, for a total of 16 image exposures (lasting 16 s total). Each block was followed by a 10 s interblock interval. Six blocks were presented for each image category, for a total of 24 blocks. Participants demonstrated task engagement by identifying nearly all item repetitions (mean ± SD: 97.5 ± 2.4%, minimum 93.1%) with few false alarms (mean ± SD: 0.8 ± 0.8%, maximum 2.9%). This task took 10.5 min (315 volumes) to complete.

Phase 4: controlled memory reactivation in an RSVP task. The goal of Phase 4 was to repeatedly elicit controlled levels of memory retrieval using word cues. Of the 30 studied word-scene pairs, 10 pairs were assigned to the long presentation condition (which was designed to elicit the strongest reactivation), 10 pairs were assigned to the short presentation condition (which was designed to elicit weaker reactivation), and 10 pairs were omitted from this phase (so they did not undergo any reactivation). Our approach was to use a vigilance task that required full concentration, occasionally inserting memory cues into the task for short and long durations (Fig. 2). Participants pressed a button when a fruit word appeared. Randomly sampled concrete "filler" nouns were presented sequentially for a duration selected randomly from a continuous distribution between 300 and 750 ms. To discourage retrieval of filler-related memories, each filler word was used only once. Within this word stream, "event" words were inserted every ~8000 ms. Each event word was either a fruit word (1000 ms duration) or a memory cue. Cues assigned to the short presentation condition were shown for 600 ms; cues assigned to the long presentation condition were shown for 2000 ms. Each cue was presented eight times during Phase 4. To prevent attentional blink effects from influencing processing of event words, the filler word before an event word was always presented for 500 ms (for review, see Dux and Marois, 2009). Together, in each functional run, ten "long" memory cues, ten "short" cues, and six fruit words were presented. To emphasize engagement in the fruit-detection cover task, after each run, participants were given feedback about their performance on that task. Participants identified most, but not all of the fruit targets in the cover task (mean ± SD: 87.5 ± 8.9%, minimum 65%). Eight runs were completed, each lasting 7.7 min (231 volumes).

fMRI data collection

Scanning was performed using a 3 Tesla whole-body Skyra MRI system (Siemens) at Princeton University in Princeton, New Jersey. T1-weighted high-resolution MRI volumes were collected using a 3D MPRAGE pulse sequence optimized for gray-white matter segmentation, with slices collected in the AC-PC plane (176 sagittal slices; FOV = 256 mm; 256 × 256 matrix; TR = 2530 ms; TE = 3.37 ms; flip angle = 9°). All functional MRI scans were collected using T2*-weighted echo-planar image acquisition (34 axial oblique slices; FOV = 192 mm; 64 × 64 matrix; TR = 2000 ms; TE = 33.0 ms; flip angle = 71°; 2× IPAT acquisition; Siemens prospective motion correction). A T1 FLASH and fieldmap image were also collected using these parameters to assist with coregistration and to correct spatial distortions.

fMRI preprocessing

We applied retrospective motion correction (Siemens) and a despiking algorithm (3dDespike, AFNI), then coregistered data to a subject-specific T1 FLASH image and corrected for spatial distortion with a fieldmap image.

Segmentation was performed in a semiautomated fashion using the Freesurfer image analysis suite, which is documented and available online (version 5.1; http://surfer.nmr.mgh.harvard.edu) with details de-

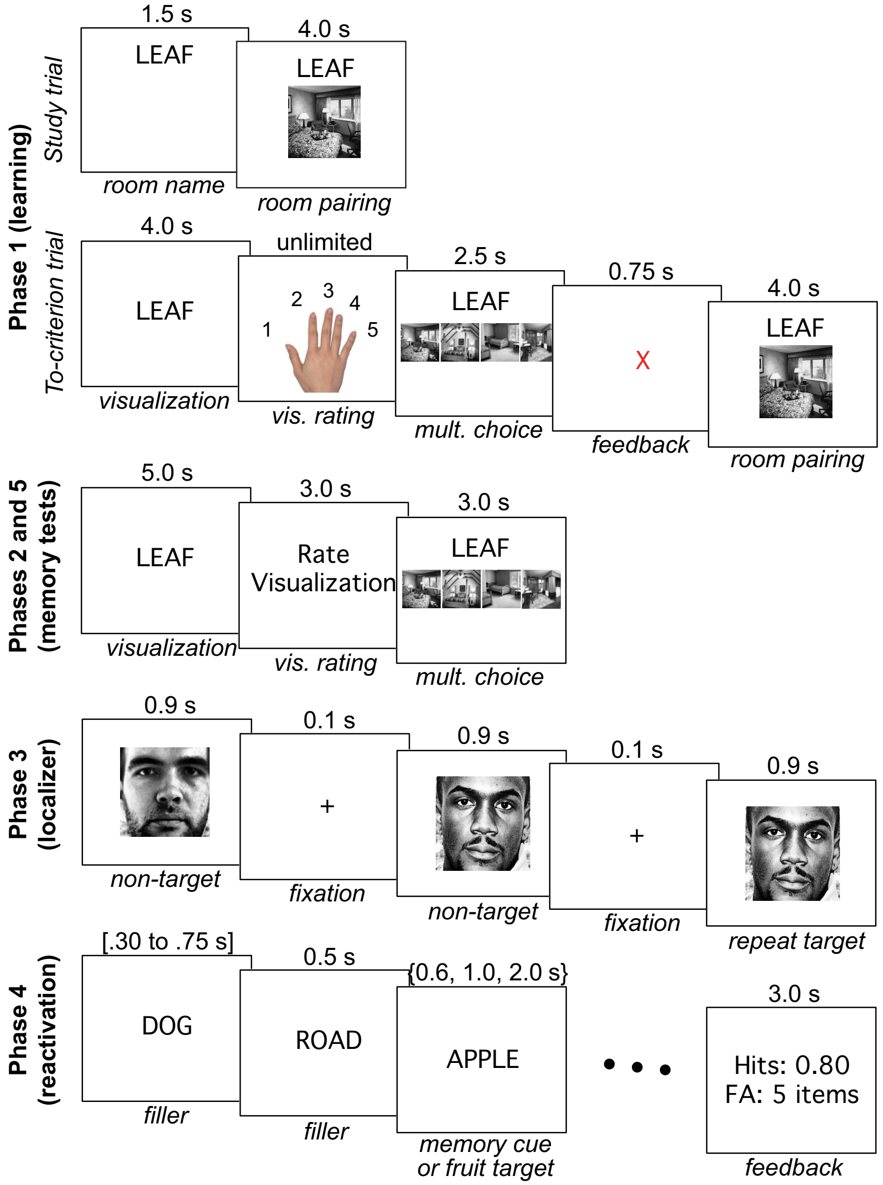

Figure 2. Trial layout for phases described in Table 1. In Phase 1 (learning phase), participants studied word-scene associate pairs and learned them to criterion. In Phases 2 and 5 (pre- and post-RSVP memory tests), we measured memory and memory-related brain activity. In Phase 4 (RSVP), we attempted to partially or fully reactivate memories by presenting memory cues at 0.6 s and 2.0 s durations while participants monitored for fruit words that appeared for 1.0 s. In Phase 3 (functional localizer phase), participants viewed images of different categories while completing a one-back task, providing brain data suitable for training a pattern classifier.

scribed previously (e.g., Fischl et al., 2004). Briefly, this processing includes removal of nonbrain tissue using a hybrid watershed/surface deformation procedure, automated Talairach transformation, intensity normalization, tessellation of the gray matter white matter boundary, automated topology correction and surface deformation following intensity gradients, parcellation of cortex into units based on gyral and sulcal structure, and creation of a variety of surface-based data, including maps of curvature and sulcal depth. Manual quality control checks were performed. We resampled Freesurfer segmentations of fusiform and parahippocampal gyri to native functional image space for use as anatomical masks.

Classifier training

Our analyses required an ongoing measure of memory activation. As all cued associates were indoor scenes, our approach was to measure evidence of scene processing in the fMRI data. To this end, we first trained a pattern classifier to be sensitive to features of Phase 3 fMRI data that distinguish between presentation of scenes versus other categories of images. Then, we used the classifier to measure the presence of scene information in other parts of the experiment. This approach has been used in numerous studies to measure the degree of memory reactivation (e.g., Polyn et al., 2005; Kuhl et al., 2011, 2012; Zeithamova et al., 2012; for a review, see Rissman and Wagner, 2012; Detre et al., 2013).

We conducted our classifier analysis in MATLAB (MathWorks) using functions from the Princeton Multi-Voxel Pattern Analysis Toolbox (Detre et al., 2006), available for download at http://www.pni.princeton.edu/mvpa (for discussion, see also Norman et al., 2006b). Classifier training was performed separately for each participant using a ridge regression algorithm, which is sensitive to graded signal information (such as might be associated with intermediate states of memory reactivation; for other applications of ridge regression algorithms to neural pattern classification, see Newman and Norman, 2010; Poppenk and Norman, 2012; Detre et al., 2013). Ridge regression learns a β for each input feature (voxel) and uses the weighted sum of voxel activation values to predict outcomes. The ridge regression algorithm optimizes each β to simultaneously minimize both the sum of the squared prediction error across the training set and also the sum of the squared β (technical details described previously; see Hoerl and Kennard, 1970; Hastie et al., 2001). A regularization parameter (λ) determines how strongly the classifier is biased toward solutions with a low sum of squared β; when this parameter is set to zero, ridge regression becomes identical to multiple linear regression. The solution found by the classifier corresponded to a β map for each regressor describing the spatial pattern that best distinguished that regressor's condition from other conditions (with regularization applied).

Following the lead of other fMRI pattern-classification studies that have tracked scene and face activity (Kuhl et al., 2011; Poppenk and Norman, 2012; Detre et al., 2013; Kim et al., 2014; Lewis-Peacock and Norman, unpublished observations), we took all gray-matter voxels from the fusiform gyrus and parahippocampal gyrus (using subject-specific masks) and fed these voxels' activation values (on a TR-by-TR basis) into the classifier (Fig. 3). The regularization parameter λ was set to 10. To label the patterns, we took the four training regressors that (respectively) described the presentation of scenes, faces, cars, and words, and shifted each by 4 s (i.e., two TRs) to accommodate hemodynamic lag effects. To assess the classifier's effectiveness at distinguishing among these image categories, we used a leave-one-block-out cross-validation procedure (Kriegeskorte et al., 2009). As there were six blocks of each image category, we left out a different block of each type (i.e., one sixth of examples) on each of six training iterations. Mean classifier accuracy across iterations was sufficiently high (mean ± SD: 0.83 ± 0.06, minimum 0.72, chance 0.25) to warrant further use of the classifier.

Classifier evidence as a dependent measure

We next used the classifier to obtain a temporal "read-out" of memory reactivation during Phases 2 and 5 (the pre- and post-RSVP memory

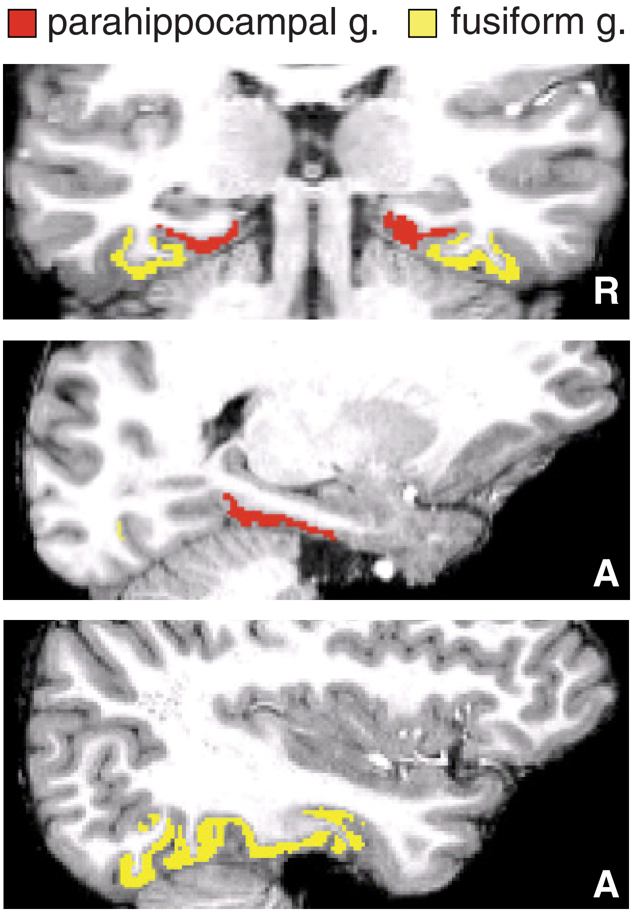

Figure 3. Example anatomical masks. Because we required a classifier that could distinguish scenes from other categories (including faces), we applied the classifier to voxels from regions of ventral temporal cortex (fusiform gyrus and parahippocampal gyrus) that have been implicated in category-specific processing (Kuhl et al., 2011; Detre et al., 2013). By contrast, similarity analysis focused on within-category information (scene exemplars), so we limited our mask to the parahippocampal gyrus, which has previously been implicated in scene-specific processing (e.g., Epstein and Kanwisher, 1998). Anatomical masks were automatically derived in native subject space using Freesurfer software. The extent of these regions is illustrated within one example subject.

tests) and Phase 4 (the RSVP task). For these analyses, we trained the classifier using all of the data from the localizer phase, and we used the trained classifier to measure evidence of scene and face activity in each functional volume from Phases 2, 4, and 5. We used the difference in scene and face evidence as our estimate of memory reactivation to focus our estimate on category-specific activity (this approach has yielded greater sensitivity in other studies relative to just looking at scene activity; e.g., Detre et al., 2013). The result was a TR-by-TR (i.e., one 2 s functional scan a time) time series of memory reactivation. We parsed this into events that began with cue onset (i.e., TR0) and ended with the subsequent cue onset. To ensure measurement of evoked signals (as opposed to low-frequency state-based signals), we normalized each event by subtracting the value at TR0 from all TRs. To restrict statistical comparisons and account for hemodynamic lag, we focused on values between TR2 (4 s after onset) and the TR marking the end of the trial. For our analyses of Phase 2 and Phase 5 classifier evidence, we measured the change in average classifier evidence from Phase 2 to Phase 5 for each condition, and then we compared these "change in classifier evidence" scores across conditions. This last step (comparing across conditions) is crucial. If we just looked at the short-cue condition by itself and observed a pre-to-post decrease in classifier evidence, that decrease could be caused by (1) an effect that is selective to that condition or (2) a nonselective effect (e.g., people might be generally more tired during the post-test, in which case classifier evidence might be lower for all three conditions; also, scanner drift could cause a global decrease in classifier evidence). Comparing "change in classifier evidence" across conditions controls for these nonselective effects, thereby making it possible to determine whether the RSVP phase differentially reduces classifier evidence in the short-cue condition (as is predicted by our theory).

To validate the classifier's ability to detect memory retrieval, we performed the following sanity checks: In the visualization task, we checked for more scene activation for target cues than lures (Fig. 4*A*) and for a positive correlation between scene activation and subjective visualization ratings (Fig. 4*B*); in the RSVP task, we checked for more scene activation for long than short cues (Fig. 5*A*).

Pattern similarity analysis

Our pattern classification analysis allowed us to quantify the amount of scene activation elicited by a memory cue. By contrast, we used pattern similarity analysis to assess whether our reactivation manipulation caused any qualitative changes in the underlying neural representations. To conduct this analysis, we had to extract vectors for each item characterizing the item's neural representation (1) before the RSVP reactivation phase and (2) after the RSVP reactivation phase. All of the studied associates were scenes; based on prior work implicating the parahippocampal region in scene processing (Epstein and Kanwisher, 1998), we limited this pattern similarity analysis to voxels from the parahippocampal gyrus (Fig. 3); a qualitatively similar pattern of results was observed when using the combined parahippocampal and fusiform gyrus mask used in our classifier analysis. Because TR2-TR4 were most sensitive to reactivation in our pattern classification analysis, we averaged across those volumes for each event in each memory test; then we reshaped parahippocampal gyrus voxels from the mean volume into a vector (for other examples of this strategy, see Moore et al., 2013; Kim et al., 2014). To evaluate the similarity of neural representations over time, we correlated each item's post-RSVP (Phase 5) feature vector with that same item's pre-RSVP (Phase 2) feature vector; we will refer to this measure as same-item similarity. We also correlated each item's feature vector from the post-RSVP memory test with the feature vectors for all of the noncorresponding items from the pre-RSVP memory test from the same condition (e.g., if the item was from the short-cue condition, we would correlate its post-RSVP feature vector with the pre-RSVP feature vectors for all of the other short-cue items). We will refer to this measure as different-item similarity. In addition to computing same-item similarity and different-item similarity for each condition, we also correlated (across individual items) same-item similarity with the average scene classifier evidence triggered by that item during the RSVP phase. This allowed us to test whether the amount of change in an item's neural representation (indexed by same-item similarity) was related to the degree of reactivation of that memory during RSVP (as indexed by scene classifier evidence).

Significance testing

To provide a random-effects statistical test of condition-level differences in our pattern classifier, pattern similarity, and behavioral measures, we computed these measures at the single-subject level. Group-level pairwise comparisons of condition means were then conducted using a nonparametric bootstrapping analysis. For each score or time point, pairwise differences between condition means across participants were calculated. These computations were repeated 1000 times, each time drawing 16 samples with replacement from the group of 16 participants. The SD of differences provided an SE estimate for each comparison. We divided the overall mean difference by the difference bootstrap SE to obtain a bootstrap ratio (BSR), which approximately corresponds to a *z* statistic. This use of bootstrap statistics allowed us to maximize statistical power when working with smaller samples and to avoid making normality assumptions associated with parametric statistics (Efron and Tibshirani, 1986; McIntosh and Mišić, 2013). We set our significance threshold at an absolute BSR value of 1.96 (corresponding to a ~95% confidence interval). When relating pattern similarity to classifier measures of memory activation from the RSVP phase, we computed correlations within subjects and then compared the distribution of within-subject correlation coefficients against zero by using the same bootstrap resampling approach described above.

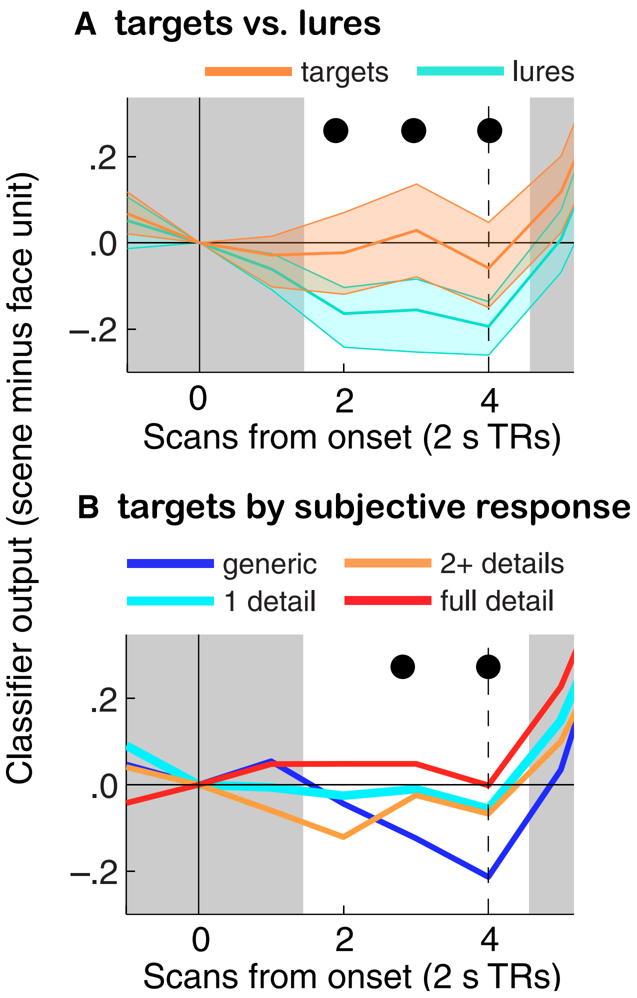

Figure 4. Linking scene visualization to classifier evidence. In the pre-RSVP memory test (Phase 2), participants were presented with a cue (onset marked with a solid vertical line), performed cued visualization of scene images, then selected the correct image from four alternatives (onset marked with a dashed vertical line). ***A***, Classifier evidence is shown separately for cue words with scene associates (targets) and familiar words without scene associates (lures). ***B***, Classifier evidence is shown separately for targets as a function of the amount of self-reported visualization detail. In both, the large increase in classifier evidence of scene processing approximately TR5 corresponds to the visual presentation of scenes during the multiple-choice part of the trial. Ribbons around each time course in ***A*** denote 95% bootstrap confidence intervals; these are absent in ***B*** to preserve legibility. Black dots represent a significant difference (***A***; see also Table 3) or significant correlation between visualization ratings and classifier evidence (***B***).

Results

Overview

Our analyses focused on how the RSVP (Phase 4) cue exposure manipulation (short, long, or no cue presentation) affected participants' memories for word-scene pairs. fMRI data confirmed that, within the RSVP stream, long cue exposures elicited more reactivation than short ones. Also, by comparing fMRI data from the pre-RSVP (Phase 2) memory test and the post-RSVP (Phase 5) memory test, we found that short cue exposures led to (1) weaker memory retrieval overall and (2) changes in neural representations (from pre-RSVP to post-RSVP) that fit with our computational model of learning (Norman et al., 2006a). We also found that changes in the neural representations of short-cue items were correlated with how strongly the memory reactivated during the RSVP phase.

Behavioral results

As noted above, behavioral performance was normed to be at ceiling. This was done (1) to ensure that word cues would elicit strong recollection of the associated scene during the pre-RSVP (Phase 2) memory test, thereby allowing us to clearly measure the pattern of neural activity associated with the memory; and (2) to ensure that the word cues would elicit scene retrieval during the RSVP (Phase 4) procedure. We made this design choice with the understanding that it would limit our ability to see effects of the RSVP cue exposure manipulation on behavioral memory measures. Table 2 summarizes behavioral data from the three cue exposure conditions. In general, our analyses focused on changes from Phase 2 (pre-RSVP) to Phase 5 (post-RSVP), which allowed each item to serve as its own baseline; we examined whether these pre-to-post changes in behavior varied as a function of condition (short, long, omit; no significant differences were found within the pre-RSVP test). We expected that memory weakening effects would be largest in the short condition, compared with the omit and long conditions; thus, any behavioral measure sensitive to this effect (despite the presence of near-ceiling levels of performance) would be expected to show a larger pre-to-post decrease in performance in the short condition compared with the omit and long conditions. None of our pre-to-post behavioral measures showed this pattern of effects. The only significant condition-wise difference related to reaction times was in the multiple-choice test: In the omit condition, participants showed a numerical decrease in reaction times from the pre-RSVP test to the post-RSVP test (indicating improved performance), whereas in the short condition participants showed a numerical increase in reaction times from the pre-RSVP test to the post-RSVP test. The change from pre to post was significantly different between the short and omit conditions (BSR = 4.04, $p < 0.001$). However, the change from pre to post was not significantly greater in the short than long condition (BSR = 0.69, p = not significant).

Validation of our classifier measure of memory reactivation

Because, under the conditions necessary to perform the current experiment, overt responses were nondiscriminative of memory signal, it was necessary to use covert memory measures that were in fact sensitive to differences in memory strength across items and conditions. To this end, we relied on our classifier measures to examine variation in strength of memory recall. As a first validation step, we performed cross-validation within the localizer run and obtained a high accuracy rate (see Materials and Methods). Next, to validate the use of our classifier as a measure of associative memory retrieval, we examined whether classifier measures of scene activation tracked memory retrieval strength in the pre-RSVP memory test (Phase 2). We found significantly more classifier evidence for memory cues than lures at TR2-TR5, corresponding to 4 to 10 s after cue onset (Fig. 4*A*; Table 3). Additionally, classifier evidence at TR3 and TR4 was positively correlated with participants' scene visualization ratings for target items (Fig. 4*B*; Table 3; no such correlation was present for multiple-choice reaction time). These results indicate that the classifier measure successfully tracked associative memory retrieval during the cued visualization task in our experiment, and they converge with other studies that have observed a relationship between classifier activity at test and behavioral indices of memory retrieval strength (e.g., Johnson et al., 2009; Kuhl et al., 2011; Gordon et al., 2014).

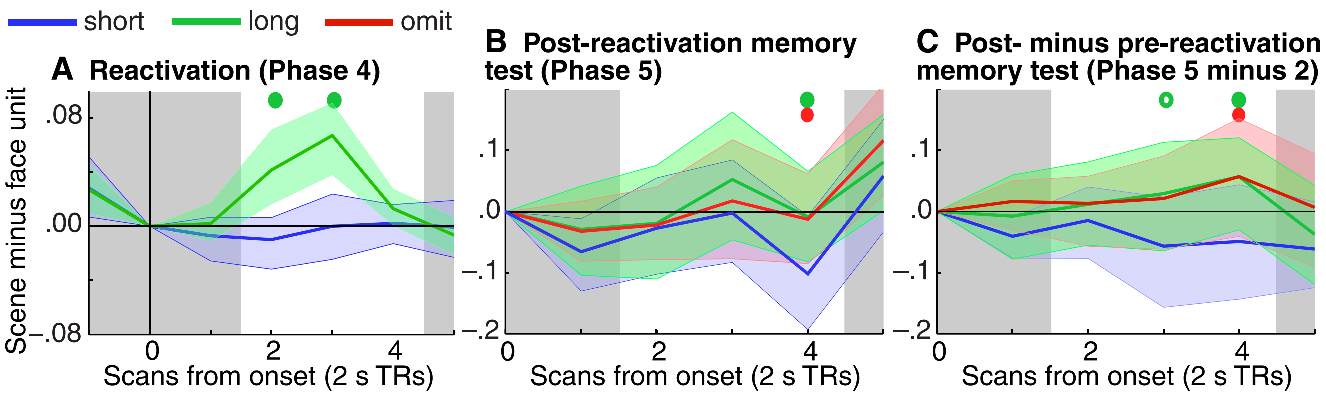

Figure 5. Impact of Phase 4 exposure duration on memory retrieval. ***A***, In Phase 4, we exposed memory cues for 600 ms (short), 2000 ms (long), or not at all (omit; absent from panel). Long items were associated with more classifier evidence of memory reactivation. ***B***, Classifier evidence of scene processing during the visualization period of the Phase 5 post-RSVP memory test. ***C***, The same data after subtracting out classifier evidence from the Phase 2 pre-RSVP test. In both cases, short items were associated with less classifier evidence of scene processing than long or omit items. In all graphs, word cue onset occurred at TR0. Ribbons around each time course denote 95% bootstrap confidence intervals. Green and red dots represent a significant difference between the short condition and long or omit conditions, respectively (see also Table 3). Open dots represent trend-level ($p < 0.1$) significance.

Table 2. Behavioral measures of the impact of our RSVP manipulation on memory (SD in parentheses)

Condition	Mean visualization rating post-RSVP	Change in visualization rating pre-to-post-RSVP	Mean multiple choice accuracy post-RSVP	Change in multiple choice accuracy pre-to-post RSVP	Median multiple choice RT post-RSVP (ms)	Change in median multiple choice RT pre-to-post RSVP (ms)
Short	3.82 (0.42)	0.12 (.18)	0.89 (0.22)	−0.01 (0.09)	1461 (238)	30 (184)
Long	3.85 (0.43)	0.19 (.36)	0.88 (0.25)	0.00 (0.11)	1506 (268)	−12 (265)
Omit	3.79 (0.46)	0.08 (.36)	0.87 (0.24)	−0.01 (0.09)	1354 (229)	−157 (181)
Lures	1.21 (0.46)	NA	NA	NA	NA	NA

NA, Not applicable.

Table 3. Statistical comparisons within fMRI time series (bootstrap ratio)

			TR		
Phase	Statistical test	Figure	2	3	4
2	Targets minus lures	4	5.33***	4.70***	3.49***
2	Positive correlation with visual rating (one-way test)	4	1.21	1.89*	2.36**
4	Long minus short items	5	2.65**	3.34***	0.83
5	Long minus short items	5	0.26	1.70†	2.34*
5	Omit minus short items	5	0.22	0.53	3.05**
5, 2	Long minus short items (post-RSVP minus pre-RSVP)	5	0.90	1.92†	2.17*
5, 2	Omit minus short items (post-RSVP minus pre-RSVP)	5	0.58	1.51	2.35*

Significant correlations: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; trends: †$p < 0.1$.

Classifier results: detecting memory reactivation during RSVP

We also used our classifier to measure whether our cue exposure manipulation was effective at modulating memory reactivation during RSVP (Phase 4). As retrieval may be triggered automatically by relevant memory cues (for discussion, see Moscovitch, 2008), we predicted that participants would retrieve scene associates of memory cues (despite their irrelevance to the RSVP task) and that amount of retrieval would be influenced by the duration of memory cue exposure (600 ms vs 2000 ms). Consistent with this prediction, long memory cue presentations were associated with greater scene activation than short ones (Fig. 5*A*; Table 3). Also, although we could not compare classifier signal in the short condition against a closely matched baseline, two observations suggested that signal was present in that condition. First, although "fruit" cues differed in various ways from those in the short condition (including longer exposure duration: 1000 ms), they had no scene associates, so no scene activity was predicted. Consistent with this idea, more scene activation was found for short than fruit cues at TR4 (BSR = 2.60, $p < 0.01$). Second, as we will discuss later, a relationship was found between fluctuations in the short-condition classifier signal and pattern similarity.

Classifier results: partial reactivation reduces later reactivation

We next examined the impact of the RSVP phase on retrieval of scene associates in the post-RSVP (Phase 5) memory test, as measured by the classifier. For each condition (short cue/long cue/omitted cue) and each TR, we computed the change in classifier evidence of memory activation from Phase 2 (pre-RSVP) to Phase 5 (post-RSVP). At TR4, a significant difference was found between the short-cue condition and the long-cue and omitted-cue conditions; specifically, the post-minus-pre difference in classifier evidence was more negative (i.e., classifier evidence decreased more) in the short-cue condition than the long-cue and omitted-cue conditions (Fig. 5*C*; Table 3). This same pattern of significance was observed when post-RSVP test data were analyzed independently of pre-RSVP test data: At TR4, there was less classifier evidence elicited by short-cue items than by long-cue and omitted-cue items (Fig. 5*B*; Table 3).

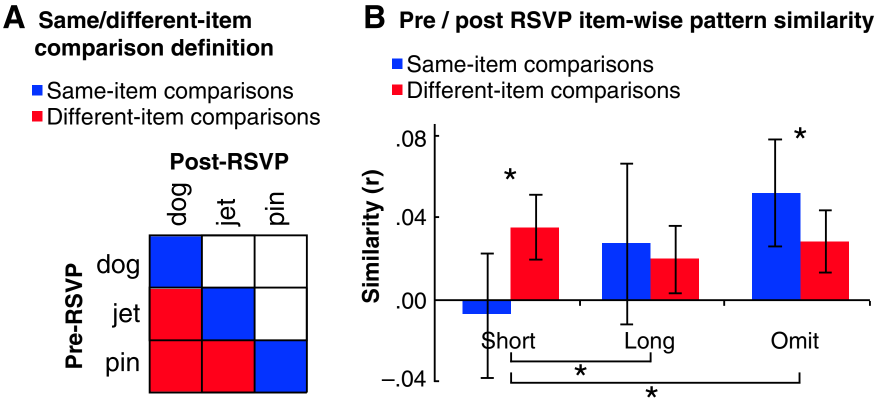

Figure 6. Similarity of representations before and after RSVP. Responses of parahippocampal gyrus voxels in the pre- and post-RSVP memory tests were extracted for each cue presentation and sorted by condition; these patterns were used to compute a correlation matrix for each condition (i.e., short, long, and omit) showing the similarity of each pre-RSVP pattern to each post-RSVP pattern within that condition (***A***). On-diagonal cells of this matrix (blue) are described here as same-item comparisons; off-diagonal cells (red) are described here as different-item comparisons. ***B***, Bar plot describes mean similarity of multivoxel patterns before and after the RSVP phase, with separate condition means for same-item comparisons and different-item comparisons. Same-item comparison similarity was significantly greater than different-item comparison similarity for omit items, and long items showed the same pattern numerically. By contrast, short items showed the opposite pattern. Error bars indicate 95% random effects confidence intervals. *$p < 0.05$.

Pattern similarity: partial reactivation inhibits neural representations

In our classifier analyses, we assessed overall memory activation, whereas in our pattern similarity analysis we assessed changes in the structure of individual memory representations. To do this, we correlated the brain's response to each memory cue as sampled during Phases 2 and 5 (i.e., during pre- and post-RSVP memory tests). For each item, it was possible to compare the same item across time and also different items across time. To the extent that memory cues trigger reinstatement of associated items in memory, a basic prediction is that correlation coefficients should be higher for "same" than "different" comparisons. Consistent with this prediction, "same" comparisons in the omit condition yielded significantly higher similarity values than "different" comparisons in that condition (BSR = 1.98, $p < 0.05$; Fig. 6*B*). In the long condition, this difference was numerically in the same direction but not significant (BSR = 0.45, p = not significant). By contrast, in the short condition, there was a reversal of this pattern, such that "same" comparisons yielded significantly lower similarity values than "different" comparisons (BSR = 2.25, $p < 0.05$). Putting these results together, there was an interaction of same versus different similarity between the short and long conditions (BSR = 2.84, $p < 0.005$) and between the short and omit conditions (BSR = 2.71, $p < 0.01$). The interaction of same versus different similarity between the long and omit conditions was not significant (BSR = 0.17, p = not significant).

We next attempted to directly relate pattern similarity to the degree of memory reactivation during RSVP (Phase 4), as measured by the pattern classifier. If the changes in pattern similarity in the short-cue condition (relative to the omit condition) were driven by reactivation during RSVP, then the amount of reactivation in that phase should negatively predict the amount of pattern similarity for a given cue-scene pair (i.e., more reactivation should lead to less pattern similarity for that pair). For each item, the dependent variable was the level of scene classifier evidence elicited by that item during TR2-TR4, averaging across the eight RSVP reactivation events per item; the independent variable was the corresponding similarity value for the item. We found a reliable negative relationship between RSVP reactivation in the short condition and similarity, $r = -0.073$, BSR = -3.41, $p < 0.001$. For completeness, we also looked at the relationship between RSVP reactivation in the long condition and similarity; the relationship in the long condition was not reliable, $r = -0.017$, BSR = -0.41, p = not significant. The lack of a reliable correlation in this condition fits with the other null effects that were observed in the long condition (i.e., no significant change in classifier activation relative to omit, and no significant change in same vs different pattern similarity relative to omit).

Simulation: pattern-similarity results reflect weakening of strongest features

To understand which representational changes could have produced this pattern of similarity, we created mock patterns, subjected them to different transformation algorithms that could plausibly be associated with memory weakening, and correlated the resulting patterns with their initial state. We focused on five possible mechanisms: As noted in the Introduction, our prior neural network modeling work predicts that memories in the short-cue condition will show differential weakening of their strongest features, inducing a reversal in the ordinal ranking of activation strength among features (Norman et al., 2006a); we will call this possibility reversal. Other possibilities are that features will be weakened at random; that weakening reflects the introduction of noise, such that memories in the short-cue condition will show a reduced signal-to-noise ratio; or that memories will become more schematic (i.e., showing a greater influence of shared vs item-specific features). Last, recent work suggests that univariate changes can contaminate similarity measures (Davis et al., 2014), so we examined whether such a confound could account for the observed pattern of results.

Each simulation consisted of 100 patterns containing 1000 features that were randomly and independently assigned a value s sampled from a uniform distribution between 0 and 1, where s corresponds to signal strength. To simulate the fact that our scenes contained some features that are shared across rooms and some features that are idiosyncratic to each room, 500 of these random features were shared across all patterns and 500 were idiosyncratic (i.e., generated randomly for each pattern). To simulate reversal, we set s to 0 for all features exceeding a threshold of i; this transformation had the effect of selectively weakening the strongest features. To simulate random weakening, we set s to 0 for i random features. To simulate changes in the signal-to-noise ratio, we added zero-mean Gaussian noise, adjusting the variance of this noise. To simulate schematization, in which the idiosyncratic part of patterns becomes less prominent than the shared part, we specifically downscaled the amplitude of idiosyncratic features. Finally, to simulate the effect of adding a univariate "blob" of activation that is shared across all patterns, we set s to 4 for i random features across all patterns.

Upon reducing the threshold in our reversal simulation, "same" similarity fell below "different" similarity (Fig. 7), as observed in our fMRI results. This was the only simulation in which this pattern emerged. Randomly weakening features, adding

Gaussian noise, or adding a blob of univariate activation to patterns, caused both "same" and "different" similarity to asymptote toward zero, never reversing their order. When patterns were "schematized," "same" similarity decreased because patterns were distorted, whereas "different" similarity increased because the portion of the variance accounted for by the shared part of the pattern grew; however, reversal never occurred.

Discussion

We obtained behavioral and neural evidence that representations of memory associates were suppressed following brief cuing, relative to conditions where cues were presented for a longer duration or not presented. Using a classifier trained to be sensitive to scene information in fMRI data, we found that brief cuing during the RSVP phase led to lower levels of scene activation on a later memory test. Using pattern similarity analysis, we found that neural representations of briefly cued memories bore less resemblance to their own initial state than to the initial states of other patterns. Moreover, this reduction in similarity in the short-cue condition was directly predicted by the extent to which scene associates were reactivated during the RSVP phase (as estimated by our classifier).

Our simulations showed these pattern similarity results were consistent with a "reversal" dynamic whereby the strongest features drop out of the pattern, leaving previously weaker ones to become dominant. We also showed that these results were inconsistent with other hypotheses about how representations might be altered (e.g., weakening of randomly selected features). As described in the Introduction, a reversal dynamic is predicted by our neural network model of learning (Norman et al., 2006a). In the model, initially stronger features suffer more weakening because they partially activate in response to brief cues, triggering weakening of connections into these features; initially weaker features do not activate and thus do not suffer weakening. The executive control theory proposed by Anderson (2003) makes a similar prediction: According to this theory, greater activation of irrelevant features triggers greater executive control and thus more inhibition.

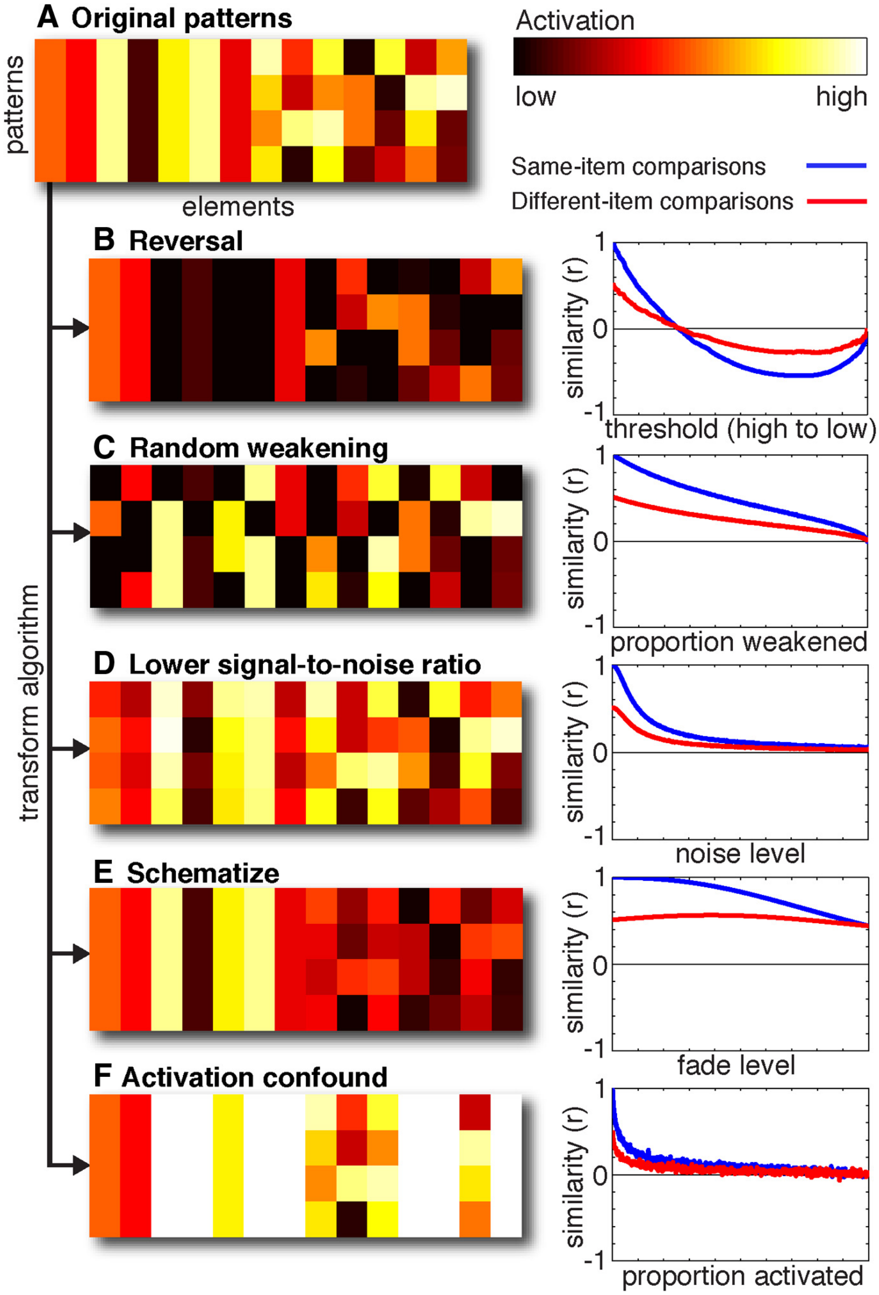

Figure 7. Impact of different simulated pattern transformations on pattern similarity. To understand how various mechanisms acting on neural representations would impact similarity scores, we generated starting patterns that featured shared and idiosyncratic parts (***A***). We then applied different transformation algorithms (***B–F***) and measured the impact on similarity (plotted on right). The *y*-intercept corresponds to a null transformation and is equivalent across plots. The reversal transformation (***B***) was the only one to generate the "similarity reversal" pattern ("same" similarity < "different" similarity) that was observed in the short-cue condition (Fig. 6).

Effects of short versus long RSVP cues

In addition to predicting that partial activation causes weakening, our theory also predicts that strong activation should cause strengthening (Norman et al., 2006a). Past research has shown that, under some circumstances, memory retrieval can elicit activation strong enough to cause further strengthening (Detre et al., 2013). Based on this, one might have expected long RSVP cues to cause strengthening, but we did not obtain any significant differences for long cues versus omitted cues. One possible explanation is that, because participants were distracted by the fruit-detection task, memories triggered by long RSVP cues may have straddled the moderate-activity zone (associated with weakening) and the strong-activity zone (associated with strengthening), instead of falling clearly into the strong-activity zone; strengthening and weakening may have, on average, cancelled each other out.

We had no way of knowing *a priori* which exposure durations would lead to moderate activation (causing weakening) and which exposure durations would lead to strong activation (causing strengthening). Accordingly, we used conditions designed to trigger a range of activation values, hoping that these sampled a wide enough range to delineate the distinctive "u" shape of the curve predicted by the nonmonotonic plasticity hypothesis (i.e., with increasing activation, we should first see weakening and then strengthening). We appear to have succeeded: The RSVP manipulation had a more negative effect on our classifier measure of memory recall for the 600 ms (short-cue) condition than for the 0 ms (omitted-cue) or 2000 ms (long-cue) condition. Likewise, the distinctive pattern changes our model predicts after memory weakening (same-item similarity < different-item similarity) were more evident for the short-exposure condition than the omitted-cue or long-exposure condition. Our only "miss" was that we incorrectly predicted 2000 ms exposures would be adequate to cause strengthening. Regardless, 2000 ms exposures were still long enough to mitigate the weakening effects associated with 600 ms exposures, allowing us to demonstrate nonmonotonic effects of exposure time on learning.

Relationship to prior work

Our results complement recent work from our laboratory showing that partial activation of memories reduces subsequent accessibility of those memories (Newman and Norman, 2010; Detre et al., 2013; Kim et al., 2014; Lewis-Peacock and Norman, unpublished observations). These studies used a variety of methods to elicit partial activation: For example, participants in Detre et al. (2013) learned word-scene paired associates like those used here; later, participants were given cue words but asked to not think of the studied associate. Scene activation during these "no think" trials, as measured using an fMRI pattern classifier, showed that moderate (but not high or low) levels of retrieval were associated with subsequently impaired recall of scene associates.

The present study makes three main novel contributions: First, we replicated memory-weakening effects using novel means of eliciting partial activation (i.e., brief cues embedded in an RSVP stream). Second, although previous studies have revealed behavioral evidence of memory weakening arising from partial memory reactivation, ours is the first to show neural evidence of the same (via our classifier measure). Third, and most importantly, this study examined how partial activation qualitatively changes the neural pattern elicited by subsequent retrieval; the studies listed above only measured quantitative changes in memory strength.

Are neural pattern changes attributable to formation of new associations?

As discussed, our preferred explanation of our results is that briefly cued memories were weakened. However, we should consider the possibility that impaired memory was caused by formation of additional, interfering associations during RSVP. When cues were briefly presented during RSVP, they may have accrued new associations with whatever the participant was thinking (via Hebbian plasticity); later, when cues were presented again during the Phase 5 test, these new associations may have competed with the correct associate, impairing its retrieval. This could occur even without successful cued recall during RSVP; the only requirement is cue activation (so they can be linked to new associates; for a similar account of forgetting in the think-no think paradigm, see Tomlinson et al., 2009).

Importantly, this account fails to explain our pattern-similarity results. If short cues were linked to other information during RSVP (e.g., fruit words), this would introduce random noise to the pattern similarity analysis (akin to Fig. 7*D*). This would reduce both "same" and "different" pattern similarity but would not induce the reversal (different-similarity > same-similarity) observed in our data. Even if short cues became linked to scenes activated by long cues earlier in the RSVP stream, no reversal would occur (for short-cue items, "different" similarity describes similarity to other short-cue items, so increased similarity to long-cue items would not boost "different" similarity). Finally, as noted above, same-cue pattern similarity in the short-cue condition was predicted by evoked scene activity during RSVP (i.e., changes relative to a pretrial baseline). This fits with the idea that changes in pattern similarity were caused by retrieval of associated scene memories, as opposed to incorporation of new information from the RSVP phase. For these reasons, it is unlikely that retroactive interference caused our pattern similarity results.

Key design features

Our design incorporated several features important for providing insight into neural pattern change. The first of these was the inclusion of both pre- and post-RSVP cued recall memory tests; this pre-post design allowed us to measure changes in specific representations as a function of memory reactivation. An important goal for future research is to apply a similar pre-post design to other paradigms that have been used to examine memory inhibition effects: for example, the think/no-think paradigm (Anderson and Green, 2001) and the retrieval-induced forgetting paradigm (Anderson et al., 1994). Our working hypothesis is that the pattern-similarity reversal observed here (same-item similarity < different-item similarity) reflects a general neural signature of memory suppression that will generalize to other paradigms known to induce forgetting.

Although no one has yet attempted a pre-post design with the think/no-think paradigm, Gagnepain et al. (2014) ran a pattern similarity analyses on data collected during the "think" and "no think" trials. Specifically, they took a snapshot of each item's pattern and computed the pairwise similarities of all item patterns. Like us, they ran simulations to see which models best fit the data. The models they considered encompassed a wide range of hypotheses about effects of the no-think procedure, including random suppression of features (akin to "random weakening" in Fig. 7*C*) and targeted suppression of the strongest features (akin to "reversal" in Fig. 7*B*). As in our study, the model positing targeted suppression of the strongest features best explained the data. The results of Gagnepain et al. (2014) show targeted suppression can be observed without a pre-post design; however, we contend that a pre-post design is preferable because it yields a qualitative signature of suppression (same-item similarity < different-item similarity). By contrast, Gagnepain et al. (2014) had to rely on quantitative differences in model fit.

Another key design feature was our use of cued recall instead of recognition. If we had instead used a recognition memory test (i.e., with participants directly viewing word-scene associates), detailed bottom-up information arising from repeated scene presentations in Phases 2 and 5 likely would have anchored associated neural representations, keeping them similar. In this situation, we would have been unlikely to observe a similarity reversal whereby (in the short-cue condition) "same" similarity was lower than "different" similarity. By contrast, in cued recall, there are fewer bottom-up constraints on the representation (the cue is repeated, but participants are free to imagine different things in response), thereby making it easier to detect effects of learning on memory representations.

In conclusion, we examined the impact of brief memory cue exposures on subsequent memory for word-scene paired associates. Our pattern similarity analyses revealed a similarity reversal, such that neural patterns in the short-cue condition became less similar to their own original state than that of other items. This was not observed for items in the long-cue or omitted-cue condition. On an item-by-item basis, the degree to which an item's pattern changed from pre- to post-RSVP was correlated with the amount of scene reactivation elicited by the brief cue during the RSVP phase. Together, these results support the hypothesis that partial memory activation in response to brief cues induced qualitative representational changes responsible for subsequent changes in the accessibility of scene memory associates. Overall, these results converge with prior findings showing that partial activation leads to forgetting, and they provide an initial glimpse into how weakened memory representations are altered.

References

Anderson MC (2003) Rethinking interference theory: executive control and the mechanisms of forgetting. J Mem Lang 49:415– 445. CrossRef

Anderson MC, Green C (2001) Suppressing unwanted memories by executive control. Nature 410:366 –369. CrossRef Medline

Anderson MC, Bjork RA, Bjork EL (1994) Remembering can cause forgetting: retrieval dynamics in long-term memory. J Exp Psychol Learn Mem Cogn 20:1063–1087. CrossRef Medline

Artola A, Bröcher S, Singer W (1990) Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature 347:69 –72. CrossRef Medline

Coltheart M (1981) The MRC psycholinguistic database. Q J Exp Psychol 33:497–505. CrossRef

Davis T, Larocque KF, Mumford J, Norman KA, Wagner AD, Poldrack RA (2014) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage. Advance online publication. Retrieved April 21, 2014. CrossRef Medline

Detre GJ, Natarajan A, Gershman SJ, Norman KA (2013) Moderate levels of activation lead to forgetting in the think/no-think paradigm. Neuropsychologia 51:2371–2388. CrossRef Medline

Detre GJ, Polyn SM, Moore CD, Natu VS, Singer BD, Cohen JD, Haxby JV, Norman KA (2006) The Multi-Voxel Pattern Analysis (MVPA) toolbox. Presented at the Organization for Human Brain Mapping Annual Meeting, Florence, Italy, June. Poster 50.

Dux PE, Marois R (2009) The attentional blink: a review of data and theory. Atten Percept Psychophys 71:1683–1700. CrossRef Medline

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54 – 75. CrossRef

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598 – 601. CrossRef Medline

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004) Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11–22. CrossRef Medline

Gagnepain P, Henson RN, Anderson MC (2014) Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. Proc Natl Acad Sci U S A 111:E1310 –E1319. CrossRef Medline

Gordon AM, Rissman J, Kiani R, Wagner AD (2014) Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cereb Cortex. Advance online publication. Retrieved Aug. 6, 2013. Medline

Hansel C, Artola A, Singer W (1996) Different threshold levels of postsynaptic $[Ca^{2+}]$i have to be reached to induce LTP and LTD in neocortical pyramidal cells. J Physiol Paris 90:317–319. CrossRef Medline

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning. New York: Springer.

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 55– 67.

Johnson JD, McDuff SG, Rugg MD, Norman KA (2009) Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron 63:697–708. CrossRef Medline

Keresztes A, Racsmány M (2013) Interference resolution in retrieval-induced forgetting: behavioral evidence for a nonmonotonic relationship between interference and forgetting. Mem Cognit 41:511–518. CrossRef Medline

Kim G, Lewis-Peacock JA, Norman KA, Turk-Browne NB (2014) Pruning of memories due to context-based prediction error. Proc Natl Acad Sci U S A. In press.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540. CrossRef Medline

Kuhl BA, Rissman J, Chun MM, Wagner AD (2011) Fidelity of neural reactivation reveals competition between memories. Proc Natl Acad Sci U S A 108:5903–5908. CrossRef Medline

Kuhl BA, Rissman J, Wagner AD (2012) Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia 50:458 – 469. CrossRef Medline

McIntosh AR, Mišić B (2013) Multivariate statistical analyses for neuroimaging data. Annu Rev Psychol 64:499 –525. CrossRef Medline

Moore KS, Yi DJ, Chun M (2013) The effect of attention on repetition suppression and multivoxel pattern similarity. J Cogn Neurosci 25:1305– 1314. CrossRef Medline

Moscovitch M (2008) The hippocampus as a "stupid," domain-specific module: implications for theories of recent and remote memory, and of imagination. Can J Exp Psychol 62:62–79. CrossRef Medline

Newman EL, Norman KA (2010) Moderate excitation leads to weakening of perceptual representations. Cereb Cortex 20:2760 –2770. CrossRef Medline

Norman KA, Newman E, Detre G, Polyn S (2006a) How inhibitory oscillations can train neural networks and punish competitors. Neural Comput 18:1577–1610. CrossRef Medline

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006b) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424 – 430. CrossRef Medline

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963– 1966. CrossRef Medline

Poppenk J, Norman KA (2012) Mechanisms supporting superior source memory for familiar items: a multi-voxel pattern analysis study. Neuropsychologia 50:3015–3026. CrossRef Medline

Potter MC (1976) Short-term conceptual memory for pictures. J Exp Psychol Hum Learn 2:509 –522. CrossRef Medline

Rissman J, Wagner AD (2012) Distributed representations in memory: insights from functional brain imaging. Annu Rev Psychol 63:101–128. CrossRef Medline

Tomlinson TD, Huber DE, Rieth CA, Davelaar EJ (2009) An interference account of cue-independent forgetting in the no-think paradigm. Proc Natl Acad Sci U S A 106:15588 –15593. CrossRef Medline

Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75:168 –179. CrossRef Medline