Neural reactivation reveals mechanisms for updating memory

Brice A. Kuhl, **Wilma A. Bainbridge**, and **Marvin M. Chun**
Yale University

Abstract

Our ability to remember new information is often compromised by competition from prior learning, leading to many instances of forgetting. One of the challenges in studying why these lapses occur and how they can be prevented is that it is methodologically difficult to 'see' competition between memories as it occurs. Here, we used multi-voxel pattern analysis of human fMRI data to measure the neural reactivation of both older (competing) and newer (target) memories during individual attempts to retrieve newer memories. Of central interest was (a) whether older memories were reactivated during retrieval of newer memories, (b) how reactivation of older memories related to retrieval performance, and (c) whether neural mechanisms engaged during the encoding of newer memories were predictive of neural competition experienced during retrieval. Our results indicate that older and newer visual memories were often simultaneously reactivated in ventral temporal cortex—even when target memories were successfully retrieved. Importantly, stronger reactivation of older memories was associated with less accurate retrieval of newer memories, slower mnemonic decisions, and increased activity in anterior cingulate cortex. Finally, greater activity in the inferior frontal gyrus during the *encoding* of newer memories (memory updating) predicted lower competition in ventral temporal cortex during subsequent *retrieval*. Together, these results provide novel insight into how older memories compete with newer memories and specify neural mechanisms that allow competition to be overcome and memories to be updated.

Introduction

We are accustomed to thinking of memory lapses as situations in which we fail to bring something to mind. Yet, many of our lapses actually reflect situations in which we bring the *wrong thing to mind*. The pernicious tendency for irrelevant memories to compete with relevant memories is evident, for example, whenever we retrieve a formerly relevant password instead of the currently relevant version—that is, when we fail to *update* memory (Bjork, 1978). Given the prevalence of such failures in everyday remembering (e.g., Underwood, 1957; Slamecka and Ceraso, 1960), and their increased frequency in aging populations (Shimamura and Jurica, 1994) and those with neural injury (Shimamura et al., 1995), there is considerable interest in understanding how we successfully update our memories and why we sometimes fail.

Most typically, competition between memories is inferred from disruptions in behavioral performance—for example, reduced retrieval accuracy or slower reaction times (e.g., Anderson, 1983). However, two recent fMRI studies have illuminated neural representations of memory competition more directly by capitalizing on the sensitivity of multi-voxel pattern analysis. One study found that when older (competing) memories share a semantic category with newer (target) memories, retrieval of target memories is associated with particularly strong neural representation of the shared category (Öztekin and Badre, 2011).

Correspondence should be addressed to Brice A. Kuhl, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06520. brice.kuhl@yale.edu.

Similarly, another study found that when older and newer memories correspond to distinct visual categories, patterns of neural activity at retrieval reflect a blend of these categories (Kuhl et al., 2011). Together, these results suggest that mnemonic competition is characterized by simultaneous *neural reactivation* of target and competing memories during retrieval.

Here, we used neural reactivation as a tool for gaining novel insight into the mechanisms that support successful memory updating. We employed a paired-associate learning task in which associations between words and images changed during the experiment, requiring subjects to update their memories accordingly. To measure reactivation, we trained a pattern classifier to discriminate between image categories based on patterns of activity in ventral temporal cortex during encoding and then tested whether these category-specific patterns of activity were reactivated at retrieval (e.g., Polyn et al., 2005; Kuhl et al., 2011). Importantly, our design allowed us to separately measure reactivation of newer (target) *and* older (competing) memories.

We first assessed whether older memories were reactivated during retrieval of newer memories and, if so, whether this reactivation of older memories impaired the accuracy and reaction time with which newer memories were retrieved. Critically, we then asked how ventral temporal expressions of memory competition related to the engagement of cognitive control mechanisms both during retrieval and encoding. Specifically, given the proposed role of the anterior cingulate cortex in detecting conflict among co-active representations (Botvinick et al., 2001), we assessed whether simultaneous reactivation of older and newer memories during retrieval triggered a conflict response in anterior cingulate cortex. During encoding, we asked whether responses in the inferior frontal gyrus— considered important for resolving mnemonic competition (Badre and Wagner, 2007; Fletcher et al., 2000)— would be predictive of reduced neural competition during memory retrieval.

Materials and Methods

Subjects

Twenty-four subjects (11 female) were recruited from the Yale University community. All subjects were between 18–35 years of age (M = 21.4), right-handed, and native English speakers. Informed consent was obtained according to procedures approved by the Yale Institutional Review Board.

Materials

Stimuli consisted of 144 words (cues) and 252 pictures (associates). Words were verbs drawn from the Medical Research Council Psycholinguistic Database (http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm) with a length between 4 and 11 letters (M = 6.3) and Kucera-Francis written frequency that ranged from 1 to 311 (M = 21.8). The pictures consisted of black-and-white photographs of famous people (e.g., Tom Cruise; *faces*), famous locations (e.g., Taj Mahal; *scenes*), and common objects (e.g., wrench; *objects*). All word-picture pairings and the assignment of words and pictures to conditions were randomized for each subject.

Procedure

The experiment was divided into four phases: pre-study, study, retrieval, and post-test.

Pre-study rounds—During pre-study subjects studied word-picture pairs with the instruction to remember the pairs for memory tests that would occur during scanning (retrieval phase) and after scanning (post-test). On each trial, a word was presented directly

above a picture (face, object, or scene). Below each picture was a label naming that picture (e.g., 'Tom Cruise,' or 'Taj Mahal'; Fig. 1a). Subjects were instructed to read the label for each picture, but to primarily focus their attention on the picture. No responses were required during the pre-study rounds. Subjects completed two rounds of pre-study; one outside the scanner and one inside the scanner during the collection of a high-resolution anatomical scan. In the first pre-study round, 144 pairs were presented for 4000 ms each with a 600 ms fixation cross between trials. In the second pre-study round, all 144 pairs were re-presented for 2600 ms each with a 600 ms fixation cross between trials.

Study rounds—After the second pre-study round, subjects completed a series of 6 study-retrieval cycles, each conducted during fMRI scanning. The study rounds were similar to the pre-study rounds in that words were presented above pictures and labels (4000 ms) and subjects were instructed to try to remember the pairs for the upcoming tests; again, no responses were required during the study trials. Each of the six study rounds contained 24 pairs ($1/6^{th}$ of the cues from pre-study) and within each round some of the pairs were identical to the pairs presented during pre-study (*repeat trials*); other pairs consisted of repeated words paired with a novel picture (*change trials*) (Fig. 1a). For repeat trials, subjects were instructed to continue to try to remember the (unchanged) pair for an upcoming test. For change trials, subjects were instructed to try to remember the new (updated) pair and informed that the older association (the original picture) would no longer be relevant. Trials in the study round were separated by an 8000 ms baseline period during which a fixation cross first appeared for 800 ms, followed by the presentation of six arrows (each presented for 800 ms and followed by a fixation cross for 400 ms). The arrows were either left- or right-oriented (randomly determined) and subjects were instructed to indicate the orientation of each arrow via button press. This active baseline was intended to prevent continued rehearsal of study pairs and allow the hemodynamic response to subside.

Retrieval rounds—After each study round subjects completed a retrieval round which tested their memory for the 24 pairs presented during the immediately preceding study round. On each trial, subjects were presented with a word above a rectangular outline (giving the appearance of an 'empty box'; Fig. 1a). Subjects were instructed to try to recall the picture that had been presented with each word in the immediately preceding study round. For repeat trials, this corresponded to a pair that had been encoded a total of three times (twice during pre-study, once during study). For change trials this corresponded to a pair that had only been encoded once (during study) but that overlapped with a competing (original) pair that was encoded twice (during pre-study). During the retrieval rounds, subjects were instructed to try to recall target images in as much detail as possible, but were only asked to indicate, via button press, the category of the target image. Thus, subjects made one of four responses via button press: 'face,' 'object,' 'scene,' or 'don't know.' Subjects were trained on the mapping of response to button box key before entering the scanner and reminded of the mapping before each retrieval round, but the mapping was not displayed on screen during retrieval trials. Each retrieval trial lasted 5000 ms; during the first 4000 ms, the rectangular outline presented below each word appeared in white; during the final 1000 ms, the outline changed to red, indicating that the trial was about to end and that subjects should make an immediate response. Subjects were not instructed to respond quickly—only to be sure to respond before the trial ended. Before subjects began the experiment, it was emphasized that: (a) throughout the experiment, they would always be asked to retrieve the *most recent* associate for each cue, whether the cue had been paired with multiple associates or not; (b) if, during retrieval, they indicated the category of the original associate for a word that had subsequently been paired with a new associate, this would be scored as 'incorrect,' and (c) while they would only be asked to indicate the category of the target image during the scanned retrieval rounds they would later be asked

(at post-test) to recall each of the images in more detail and they would receive monetary bonuses based on their performance.

Post-test—Following the last retrieval round, subjects exited the scanner and completed the post-test which re-tested their memory for the most recent associate for each of the 144 cue words. Post-test trials were similar to retrieval trials except that subjects were asked to respond aloud on each trial. Subjects were instructed that if they were able to recall the specific label corresponding to the target image (e.g., 'Tom Cruise'), they should provide this label; if they were unable to recall the specific label, they were instructed to provide any information about the image that they could recall (e.g., 'face,' or 'male'). Subjects received a maximum of $15 in bonus money depending on their recall accuracy at post-test. Subjects were informed at the beginning of the experiment that they would have the opportunity to earn bonus money based on their performance at post-test and, critically, that bonus money would only be based on how successfully they remembered the *most recent* associate for each cue word. Subjects were explicitly informed that if they mistakenly recalled the original associate, instead of the most recent associate, at post-test, they would not earn a monetary bonus for that trial. Because the post-test primarily served to motivate subjects to learn and recall in detail the new (updated) associations during fMRI scanning, data from this phase are not considered here.

Design

The 144 pairs presented during the pre-study round corresponded to different experimental conditions according to the category of the original associate (face, object, scene), whether the pair would later repeat vs. change during the study rounds, and, for pairs that would change, the category of the new associate. For example, a word paired with a face at pre-study might be paired with a scene during study ('F-S condition'); a word paired with an object during pre-study might be paired with a different object during study ('O-O'); or a word paired with a scene at pre-study might be paired with the same scene at study ('S-repeat'). This resulted in 12 total conditions (F-F, F-S, F-O, F-repeat, S-S, S-F, S-O, S-repeat, O-O, O-F, O-S, O-repeat); 75% of the conditions thus involved a change in the associate; the remaining 25% involved a repeated associate. Each of the 12 conditions contained 12 cue-associate sets. Two sets from each of the 12 conditions appeared in each study/retrieval round which, along with pseudo-randomization within each round, ensured a comparable average serial position across conditions. Data from the 12 conditions were collapsed into the following higher-level conditions: repeat trials (F-repeat + S-repeat + O-repeat), within-category change trials (F-F, S-S, O-O), and between-category change trials (F-S, F-O, S-F, S-O, O-F, O-S). Note: behavioral performance during the retrieval rounds for within-category change trials is not considered here because subjects' responses on these trials were ambiguous (i.e., the target and competitor corresponded to the same category); within-category change trials were primarily included so that, for change trials, the new associate was equally likely to belong to each of the three categories, thereby reducing subjects' ability to predict or infer the new category.

fMRI Methods

Data acquisition—Imaging data were collected on a 3T Siemens Trio scanner at the Anlyan Center at Yale University using a twelve-channel head coil. Prior to the functional imaging, two T1-weighted anatomical scans were collected (in-plane and high-resolution 3D). Functional data were collected using a T2*-weighted gradient EPI sequence; TR = 2000 ms, TE = 25 ms, flip angle = 90°, 34 axial-oblique slices, 224-mm FOV ($3.5 \times 3.5 \times 4$ mm). A total of twelve functional scans were collected (6 study rounds, 6 retrieval rounds). Each scan consisted of 149 volumes; the first five volumes were discarded to allow for T1 equilibration.

fMRI data analysis—fMRI data preprocessing and univariate analyses were conducted using SPM8 (Wellcome Department of Cognitive Neurology, London). Images were first corrected for slice timing and head motion. The high-resolution anatomical image was then co-registered to the functional images and segmented into gray matter, white matter, and cerebrospinal fluid. The segmented gray matter image was stripped of remaining skull, normalized to a gray matter Montreal Neurological Institute template, and then used for normalization of functional images. Functional images were resampled to 3-mm cubic voxels and smoothed with a Gaussian kernel (8 mm FWHM).

Data were analyzed under the assumptions of the general linear model (GLM). Trials were modeled using a canonical hemodynamic response function and its first-order temporal derivative; scan session was treated as a covariate. Linear contrasts were used to obtain subject-specific estimates for effects of interest, which were then entered into second-level, random-effects analyses using one-sample *t* tests against a contrast value of zero at each voxel. For targeted anatomical region of interest analyses, the Anatomical Automatic Labeling (AAL) atlas was used (http://www.cyceron.fr/web/aal__anatomical_automatic_labeling.html).

Multi-voxel pattern analysis—Neural reactivation was assessed by first training a pattern classifier to discriminate between the three visual categories of images (faces, objects, scenes). The classifier was then applied to the retrieval phase to test for reactivation of category-level information. All fMRI data used for classification analyses were preprocessed in the same way that data for univariate analyses were (including normalization and smoothing); additionally, fMRI data used for classification analyses were high-pass filtered (0.01 Hz), detrended, and z- scored within scan. After relevant trials and corresponding volumes had been selected, data were z-scored again, first across voxels within each volume, and then across all volumes corresponding to the encoding phase and, separately, across all volumes corresponding to the retrieval phase (thus, mean response for each voxel within each phase = 0). The decision to z-score across voxels and across time was made *a priori* and was based on analyses applied to a separate data set—thus, these steps did not necessarily optimize classifier performance in the present study. Classification analyses were performed using penalized logistic regression with L2-norm regularization (penalty parameter = 100). Voxel inclusion was restricted using anatomical regions of interest generated from the AAL atlas. Specifically, a VOTC mask was created that consisted of the union of the left and right masks labeled as fusiform, parahippocampal, and inferior temporal. The mask consisted of 5,789 total voxels. No additional feature selection was performed. Our use of a VOTC mask was motivated by (i) prior work demonstrating visual category selectivity within VOTC; (ii) our interest in characterizing mnemonic reactivation specifically within higher-level visual areas, and (iii) potential benefits to classification accuracy associated with reducing the number of features (voxels).

Three separate classifiers were generated, each trained on a different pairing of the image categories (i.e., face vs. scene, scene vs. object, face vs. object). For the training data, encoding trials were grouped according to the category of the currently encoded image, regardless of condition. For example, the face category corresponded to encoding trials in the conditions: F-repeat, F-F, O-F, and S-F. Thus, each image category corresponded to 48 trials used for classifier training. While each of the classifiers was trained on study trials for only two categories, each classifier was tested on all trials within the retrieval phase. For the critical between-category change trials, one category functioned as the target category, one as the competitor category, and one as a baseline category. For example, for the condition O-F, where subjects would be attempting to retrieve a face, the target category = face, competitor = object, and baseline = scene. For repeat and within-category change trials, one category functioned as the target, and the remaining two categories functioned as baseline.

Classifier performance was assessed by considering the probability that the classifier assigned to the target, baseline, and (if relevant) competitor category for each trial. Specifically, the binary logistic regression classifier used here generated a continuous value ranging from 0 to 1 for each trial; this value represented the probability of category A membership (and, inversely, the probability of category B membership). To compute, for example, the probability that a trial corresponded to the face category, the probability of face membership was summed across the face vs. scene and face vs. object classifiers; this value was then divided by three which resulted in a number with a possible range of 0 to 2/3, with 1/3 representing chance performance. Note: the probabilities assigned to the three categories necessarily summed to 1. We refer to the trial-specific classifier probabilities as *classifier evidence*. The use of three separate classifiers (as opposed to a multinomial classifier) allowed us to assess classifier performance for each pair of conditions—that is, to be sure that each pair was discriminable. Thus, while the data that are presented here average across these classifiers, having access to classifier performance for each pairwise comparison was considered useful.

Because each encoding and retrieval trial corresponded to several fMRI volumes, averaging was performed across volumes, prior to training or testing the classifier, so that statistical analysis of classifier performance could be assessed at the level of individual trials. For encoding trials, averaging was performed across the 3rd and 4th TRs (corresponding to 4–8 s after stimulus onset). For retrieval trials, a weighted average was performed across the 3rd–6th TRs (corresponding to 4–12 s after stimulus onset; weights = [.35 .35 .15 .15]). A wider temporal window was used for retrieval trials because of variability in subjects' reaction times at retrieval; the weighted average was used based on data from a prior study (Kuhl et al., 2011), which indicated that information value was highest for TRs 3–4. While all statistical analyses were based on data averaged across volumes, we also computed TR-by-TR (volume-by-volume) classification accuracy for the purpose of visualizing classifier performance across time and confirming that classification accuracy generally corresponded to the shape of a hemodynamic response function. In such cases, the classifier training data consisted of data averaged across the 3rd and 4th TRs of encoding trials and the classifier was then separately applied (tested) at each TR (1st–6th) of retrieval trials. Importantly, whenever statistical tests involved collapsing data across conditions, classifier performance was first averaged within-condition and then across conditions, ensuring that any imbalances in the number of trials within a condition (as might occur when considering classifier performance as a function of retrieval accuracy) were not confounded with differences in classifier performance across image categories.

Results

Behavioral performance

Across all retrieval trials, subjects were generally successful at recalling the visual category of the target (most recent) image (M = 63.5%). Subjects were less successful at recalling the target category for between-category change trials (M = 57.3%) than repeat trials (M = 71.8%) (Fig. 1b). For between-category change trials, when subjects indicated a category that did not correspond to the target, it was more likely to correspond to the category of the original, competing image (M = 19.7%) than the other, 'baseline' category (M = 8.8%; $t_{(23)}$ = 6.15, p < .001) (Fig. 1b).

fMRI results

Target reactivation during retrieval—Considering all retrieval trials, regardless of behavioral response, evidence for target reactivation was highly robust, as classifier evidence for the target category was significantly greater than evidence for the baseline

category/categories for repeat trials ($t_{(23)} = 5.37$, $p < .001$), within-category change trials ($t_{(23)} = 6.03$, $p < .001$), and between-category change trials ($t_{(23)} = 6.56$, $p < .001$; Fig. 2a). These data indicate that patterns of neural activity elicited during target encoding were robustly reactivated during target retrieval (e.g., Polyn et al., 2005; Lewis-Peacock and Postle, 2008; McDuff et al., 2009; Johnson et al., 2009; Kuhl et al., 2011).

Competitor reactivation during retrieval—Central to our aims, we next considered whether there was evidence for reactivation of competing (older) memories. Indeed, for the between-category change trials, evidence for the competitor category was significantly greater than evidence for the baseline category ($t_{(23)} = 2.95$, $p < .01$; Fig. 2a), providing clear evidence that competing memories were reactivated during attempted target retrieval. Competitor evidence was, however, significantly weaker than target evidence ($t_{(23)} = 2.83$, $p < .01$; Fig. 2a), indicating that neural reactivation was, on average, biased toward target (newer) memories (consistent with subjects' goals). As we consider below, subjects varied considerably in the degree to which they reactivated competing memories during target retrieval.

Reactivation and retrieval success

<u>**Within-subject analysis:**</u> The preceding analyses provide evidence that both newer (target) and older (competing) memories were reactivated during attempted target retrieval. We next considered how target and competitor reactivation related to behavioral measures of retrieval success. In particular, we asked whether target and competitor reactivation differed across between-category change trials for which the target was successfully retrieved vs. those trials for which the competitor was (inappropriately) retrieved—hereinafter, *intrusions*. First, considering only the trials associated with successful target retrieval, classifier evidence for the target category was well above evidence for the baseline category ($t_{(23)} = 8.83$, $p < .001$; Fig. 2b) and competitor category ($t_{(23)} = 4.82$, $p < .001$). Notably, however, classifier evidence for the competitor category was still greater than evidence for the baseline category ($t_{(23)} = 2.29$, $p < .05$; Fig. 2b), providing striking evidence that competing memories were reactivated in ventral temporal cortex even when target memories were successfully retrieved.

While intrusions were fairly common (~1/5th of all responses), the rate of these errors varied considerably across subjects. In order to assess how neural reactivation related to intrusions, we considered only those subjects (N = 12) that committed at least one of these errors for each type of between-category change trial (e.g., target = face, competitor = scene; target = face, competitor = object; etc.). For this subset of subjects, intrusions were associated with highly robust classifier evidence for the competitor category (Fig. 2b), which was significantly greater than evidence for the target ($t_{(23)} = 4.49$, $p < .005$) or baseline category ($t_{(23)} = 3.02$, $p < .05$). Notably, classifier evidence for the target category on these intrusion trials did not differ from evidence for the baseline category ($t_{(23)} = .28$, $p = .78$). Thus, intrusions were characterized by a clear dominance of competitor reactivation and the absence of target reactivation. For the successful retrieval trials, the data from this subset of subjects were comparable to the data from all subjects: there was significantly greater evidence for the target category relative to baseline ($t_{(23)} = 4.37$, $p < .005$), and greater evidence for the competitor category relative to the baseline category ($t_{(23)} = 2.27$, $p < .05$). Overall, for this subset of subjects, the difference in target vs. competitor reactivation as a function of retrieval success was reflected in a highly significant interaction of category (target, baseline, competitor) by retrieval outcome (target successfully retrieved vs. intrusion) ($F_{(2,22)} = 11.46$, $p < .001$; Fig. 2b).

Between-subject analysis: The preceding results indicate that neural reactivation of target (newer) vs. competitor (older) memories predicted whether the target would be successfully retrieved or whether the competitor would intrude. We next asked whether target vs. competitor reactivation was related to between-subject differences in the rate with which older memories intruded (as reflected in behavioral responses). As a first step, we median-split subjects into two groups according to the proportion of their between-category retrieval trials that corresponded to intrusions. We then considered the distribution of classifier-based evidence for target and competitor reactivation across all between-category change trials (i.e., collapsing across all behavioral responses), separately for each group.

Subjects with the lowest rate of intrusions (intrusions M = 11.2%; successful retrieval M = 66.4%) were characterized by distributions of target and competitor evidence that were clearly separated (Fig. 2c), with mean classifier evidence for the target category well above evidence for the baseline ($p < .001$) and competitor categories ($p < .005$); moreover, competitor evidence did not significantly differ from baseline ($p = .15$), indicating that, on the whole, these subjects were effective in reactivating relevant memories. In contrast, subjects with the highest rate of intrusions (intrusions M = 28.1%; successful retrieval M = 48.2%), were characterized by distributions of target and competitor evidence that were largely un-differentiated (Fig. 2c), with mean evidence for both the target and competitor categories greater than baseline (p's $< .05$) and no significant difference between evidence for target versus competitor reactivation ($p = .53$). A direct comparison of high- vs. low-intrusion subjects revealed that low-intrusion subjects were associated with less overlapping distributions of target versus competitor reactivation than high-intrusion subjects ($p < .05$).

It is important to emphasize that these between-subject analyses reflect averages across all trials (i.e., collapsing across behavioral accuracy) and therefore do not provide information about differences in selectivity on individual trials. In other words, these results are largely an extension of, and complement to, the within-subject analyses in that they indicate that variance across subjects in the proportion of successful retrieval versus intrusion trials was related to each subject's overall mean strength of target versus competitor reactivation. A more specific question is whether high-intrusion subjects were associated with reduced selectivity of reactivation even when behavioral accuracy was matched—that is, when retrieval was ultimately successful. A between-subject correlation relating the rate of intrusions to the selectivity of reactivation (target–competitor) revealed a marginally significant negative correlation ($r = -.37$, $p = .07$), indicating that subjects with higher rates of intrusions tended to display less selective reactivation even when they successfully retrieved the target.

Target vs. competitor reactivation and reaction time—Although competitor reactivation was most robust when intrusions occurred, it is notable that there was clear evidence of competitor reactivation even when targets were successfully retrieved (Fig. 2). An important question is whether this reactivation of competing memories during successful target retrieval was associated with a cost to retrieval. To address this question, we tested whether stronger competitor reactivation (on successful retrieval trials) was associated with slower reaction time during retrieval. Accordingly, for each between-category change condition (e.g., F-O, F-S, etc.) and each subject, we calculated the correlation coefficient reflecting the relationship between classifier evidence for competitor reactivation and behavioral measures of reaction time. This analysis was restricted to successful retrieval trials only. (Note: three subjects were excluded from this analysis due to an insufficient number of trials in one of the relevant conditions). Each correlation coefficient was z-transformed (Fischer's z) and the resulting z scores were then averaged across conditions within each subject in order to obtain a single value representing the relationship between competitor reactivation and reaction time for that subject. The mean z scores for each

subject were then compared to a test value of 0 to determine whether the correlation, across subjects, was significantly different from 0. Indeed, the correlation was significantly greater than 0 (mean $z = .14$, $t_{(20)} = 3.65$, $p < .005$), indicating that the reactivation of competing memories during successful retrieval trials was associated with slower reaction time. Interestingly, the strength of target reactivation was not, on its own, significantly correlated with reaction time (mean $z = -.09$, $t_{(20)} = -1.27$, $p = .22$), though the strength of the relationship between competitor reactivation and reaction time was not significantly stronger than that between target reactivation and reaction time ($p = .45$).

Neural mechanisms that track competitive reactivation—Complementing the prior analysis, we next asked whether there were specific neural mechanisms that displayed increased activity when ventral temporal measures of reactivation indicated that older memories were competing with newer memories. That is, were there regions that showed greater activity during retrieval when target reactivation was relatively *weak* compared to competitor reactivation? To address this, we conducted a univariate analysis that tested for voxels that displayed a negative linear relationship with classifier-based evidence for target–competitor reactivation at retrieval. We thus constructed a GLM for which each between-category change trial was associated with a regressor representing the difference in target vs. competitor reactivation (target–competitor). Each condition was separately modeled to prevent potential differences across categories from contributing to any observed relationship. The results of this analysis revealed a cluster within anterior cingulate cortex (Fig. 3) whose activation at retrieval was negatively correlated with the relative strength of target vs. competitor reactivation (for complete results, see Table 1). In other words, activity in the anterior cingulate cortex was high when the target memories were weak, relative to competing memories.

Prefrontal engagement during memory updating biases subsequent reactivation—The preceding results indicate that retrieval competition was strongly reflected in measures of neural reactivation. We next sought to determine whether these neural expressions of memory competition during *retrieval* were predicted by neural responses observed during the *encoding* of newer memories (i.e., updating). In other words, were neural responses during the encoding of newer memories related to the selectivity with which those memories were later reactivated in ventral temporal cortex? We approached this question by first identifying regions of prefrontal cortex that showed increased encoding activity when mnemonic associations changed and updating was required (change trials) relative to when associations remained unchanged (repeat trials). We then assessed how responses within prefrontal cortex during memory updating related to later neural expressions of memory competition.

<u>Response to change trials:</u> A univariate contrast of encoding activity for change trials (between-category + within-category) vs. repeat trials revealed a cluster in the left inferior frontal gyrus, including portions of *pars opercularis*, *pars triangularis*, and *pars orbitalis* (for complete results, see Table 2).

<u>Subsequent reactivation:</u> To test whether neural responses during memory updating were related to subsequent neural reactivation we ran a univariate analysis that tested for a linear relationship between neural responses observed during the *encoding* of change trials and classifier-based evidence of target vs. competitor reactivation at *retrieval*. By relating neural responses at encoding to subsequent neural expressions of memory, this analysis paralleled 'subsequent memory analyses' that relate neural activity at encoding to subsequent behavioral outcomes (e.g., remember vs. forget) (for reviews, see Paller and Wagner, 2002; Blumenfeld and Ranganath, 2007; Kim, 2011). We thus constructed a GLM for which every

encoding trial was associated with a regressor representing the *subsequently measured* difference in target vs. competitor reactivation (target–competitor). This analysis revealed two clusters within prefrontal cortex, both within the left inferior frontal gyrus (Table 3). Comparing the results of this analysis to the contrast of change > repeat trials (Table 2) revealed overlapping activation in left inferior frontal gyrus (Fig. 4c). Thus, left inferior frontal gyrus (a) displayed increased activity when mnemonic associations changed, relative to when they repeated, and (b) when associations changed, activity in this region was predictive of the degree to which newer (target) memories were later reactivated in ventral temporal cortex, relative to older (competing) memories.

Complementing the whole brain analysis, and motivated by prior evidence that left inferior frontal gyrus contributes to the resolution of interference (Badre and Wagner, 2007), we conducted a second, more focused region of interest (ROI) analysis that specifically probed responses within three subregions of left inferior frontal gyrus: *pars opercularis, pars triangularis*, and *pars orbitalis*. To generate regions of interest, we used anatomical masks for each subregion (see Materials and Methods) and selected all voxels within each mask that were identified from the contrast of change vs. repeat trials (Fig. 4a). We then compared the beta values generated from the subsequent reactivation GLM for each ROI (averaging across all voxels within an ROI). For each subregion, the observed beta value was significantly greater than 0 (*opercularis:* $t_{(23)} = 2.50$, $p < .05$; *triangularis:* $t_{(23)} = 2.61$, $p < .05$; *orbitalis:* $t_{(23)} = 2.43$, $p < .05$; Fig. 4b) and the magnitude of the relationship did not significantly differ across subregions ($F<1$).

While the preceding result indicates that responses within the inferior frontal gyrus during memory updating were predictive of subsequent neural reactivation, it is possible that this relationship simply reflected, or mirrored, a relationship between inferior frontal gyrus activation during encoding and behavioral success at retrieval. Thus, we next asked whether encoding responses in the inferior frontal gyrus predicted gradations in the selectivity of reactivation *above and beyond* what was attributable to behavioral accuracy. To this end, we constructed a second GLM that tested for subsequent reactivation, this time restricting analysis to only those trials for which the target was (behaviorally) successfully retrieved. Again, we considered the relationship between encoding and subsequent neural reactivation by comparing beta values for each of the three left inferior frontal gyrus ROIs. For this analysis, two subjects were excluded as outliers due to beta values for one or more of the subregions that corresponded to a z score greater than 2.5 (one subject) or less than 2.5 (one subject). (Note: exclusion of these two subjects from the preceding analyses did not qualitatively change any results). Among the three subregions, the relationship between encoding activity and subsequent reactivation remained significant only for *pars triangularis* (*opercularis:* $t_{(21)} = -.23$, $p = .82$; *triangularis:* $t_{(21)} = 2.33$, $p < .05$; *orbitalis:* $t_{(21)} = 1.12$, $p = .28$; Fig. 4b). The selectivity of this effect to *pars triangularis* was reflected in a main effect of subregion ($F_{(1,21)} = 3.60$, $p < .05$). Thus, *pars triangularis* was predictive of gradations in target vs. competitor reactivation above and beyond what was attributable to retrieval accuracy.

Discussion

Many of our lapses in memory reflect instances where outdated or otherwise irrelevant memories come to mind. Specifying how competition between relevant and irrelevant memories is experienced and how it may be overcome remains a fundamental objective in memory research but is handicapped by difficulty in measuring—or 'seeing'—memory competition as it occurs. Here we show that when older visual memories competed with newer memories, both memories were often simultaneously reactivated in high-level visual cortical areas. Importantly, these measures of neural reactivation were robustly related to the

costs that are a hallmark of memory competition—increases in retrieval errors and slowing of reaction time—confirming their relevance to the experience of retrieval competition. Moreover, increased competition within ventral cortex was associated with greater engagement of anterior cingulate cortex, consistent with a role for this region in tracking retrieval conflict. Strikingly, we also found that responses within the left inferior frontal gyrus during the encoding of newer memories were positively correlated with the selectivity with which newer memories were later reactivated.

Competitive reactivation of memories

While previous studies have indicated that mnemonic competition impacts neural reactivation at retrieval (Kuhl et al., 2011; Öztekin and Badre, 2011) these studies have not *separately* measured reactivation of target and competing memories and therefore could not directly establish whether, or to what degree, competing memories were neurally reactivated. Here, we provide clear and novel evidence that, at least within the domain of visual remembering, competing memories are often simultaneously reactivated and, critically, these neural measures of memory competition are related to trial-by-trial costs associated with retrieval competition.

The negative influence that competition has on memory is typically evidenced by either a lower probability of retrieval success or slower reaction time (e.g., Anderson, 1983). Here, we found that neural expressions of competition were related to both of these costs. With respect to unsuccessful retrieval trials, we were particularly interested in cases where older memories intruded and were erroneously reported. On these trials, reactivation of older memories was robust, but there was no detectable reactivation of newer memories. The failure to reactivate newer memories may have been partly attributable to poor encoding of newer items, but it is also probable that at least some of these failures reflected interference that occurred at retrieval.

When retrieval cues elicited initial and robust reactivation of competing memories, this may have interfered with and reduced the probability of target reactivation (Smith, 1971). Indeed, it has been suggested that many everyday instances of forgetting may be attributable to prepotent but irrelevant memories coming to mind and blocking access to target memories (Roediger and Neely, 1982; Anderson and Neely, 1996; Schacter, 1999). Thus, intrusions were not characterized by a failure to select among active memories—e.g., a source monitoring error (Johnson et al., 1993)—but a failure to reactivate target memories at all. In contrast, because successful retrieval trials were characterized by simultaneous target and competitor reactivation, these trials likely involved successful source monitoring. Of course, at least in some cases, successful retrieval may have involved robust target reactivation that 'blocked' competitor reactivation, resulting in a successful low-competition retrieval event. Importantly, by presenting a methodological approach for measuring and quantifying the degree to which relevant and irrelevant memories come to mind during retrieval, the present study suggests potentially exciting applications to populations that are characterized by memory deficits. For example, among older adults—a group that is particularly sensitive to memory interference (Shimamura and Jurica, 1994)—it may be asked whether retrieval events are more likely to be associated with a failure to select among active memories or a greater probability that one representation dominates (blocks) the other?

While intrusions represent a particularly salient cost associated with competition, the influence of competition was also evident in more subtle ways. Namely, even when newer memories were successfully retrieved, we found clear evidence that older memories were reactivated and that stronger reactivation of older memories was associated with slower reaction time. These results suggest that neural reactivation reflected in patterns of ventral

temporal activity was fundamentally related to the mnemonic decisions that subjects reached.

To the extent that information represented in ventral temporal cortex served as the basis for mnemonic decisions, other neural mechanisms should track or covary with this information. In particular, when ventral temporal reactivation indicates high conflict between target and competing memories, control mechanisms should be recruited to help resolve competition. It has been suggested that anterior cingulate cortex may play an initial role in detecting mnemonic competition and thereby triggering the engagement of other prefrontal control mechanisms (Norman et al., 2007). Indeed, anterior cingulate cortex activation has previously been found to correlate with behavioral measures of memory competition (Kuhl et al., 2007; Kuhl et al., 2008). Here, we found novel evidence that anterior cingulate cortex activation during retrieval was correlated with neural measures of competition between memories. Thus, complementing our finding that ventral temporal expressions of memory competition were related to behavioral costs, this relationship indicates that ventral temporal expressions of memory competition also engaged a neural mechanism that putatively detects conflict and triggers other cognitive control processes (Botvinick et al., 2001).

Biased reactivation

While the present results indicate that older memories were often reactivated during the retrieval of newer memories, on average, reactivation was biased in favor of newer memories, consistent with subjects' retrieval goals. The general bias of reactivation in favor of newer memories likely reflects a combination of two factors. First, during retrieval, control processes putatively favor those mnemonic representations that match retrieval goals (i.e., to retrieve the newer association for each retrieval cue). This biasing likely plays out across time within a given retrieval event, with initial reactivation of both target and competitor memories giving way to a later bias toward target reactivation (Norman et al., 2007). While the temporal resolution of fMRI data precludes testing this temporal hypothesis in the present study, the idea that ventral temporal representations are biased as a function of retrieval goals is consistent with biased competition theories (e.g., Miller and Cohen, 2001) and supports a general parallel between memory retrieval and selective attention (Anderson and Spellman, 1995; Chun and Johnson, 2011; Kuhl et al., 2011).

The second way in which reactivation at retrieval may be biased is through the influence of encoding mechanisms. Here, we found that encoding responses within left inferior frontal gyrus were predictive of subsequent target vs. competitor reactivation during retrieval. Notably, the *pars triangularis* subregion of left inferior frontal gyrus was predictive of gradations in target vs. competitor reactivation beyond what was attributable to behavioral measures of retrieval success, indicating a particular relevance of processing in this area to the selectivity of later reactivation. This novel observation is consistent with evidence that this region plays a fundamental role in resolving mnemonic interference (Badre and Wagner, 2007). For example, *pars triangularis* has been implicated in resolving competition from older and irrelevant information across the domains of working memory (e.g., Jonides et al., 1998; Badre and Wagner, 2005; Feredoes et al., 2007; Nee et al., 2007) (for review, see Jonides and Nee, 2006), episodic memory (e.g., Dolan and Fletcher, 1997; Fletcher et al., 2000; Henson et al., 2002; Sohn et al., 2003), and task switching (Badre and Wagner, 2006). Thus, in the present study, left inferior frontal gyrus was potentially recruited in situations where older memories came to mind during the encoding of newer memories (Kuhl et al., 2010).

In addition to resolving interference, left inferior frontal gyrus has also repeatedly been implicated in successful encoding of episodic memories (Paller and Wagner, 2002; Blumenfeld and Ranganath, 2007; Kim, 2011). Going beyond this literature, the present

association between left inferior frontal gyrus activation at encoding and the selectivity of subsequent reactivation provides novel evidence for a relationship between *neural* mechanisms engaged at encoding and *neural* expressions of remembering. However, an important question is whether this relationship is better characterized as an 'encoding success' effect or a 'resolution of interference' effect? In other words, did left inferior frontal gyrus contribute to memory updating by encoding target memories or by keeping competing memories out of mind? Ultimately, both mechanisms may be relevant to memory updating (Bjork, 1978) and, from a biased competition perspective, they may reduce to a single mechanism (Miller and Cohen, 2001). Indeed, it has been argued that left inferior frontal gyrus contributes to episodic encoding by biasing memory toward relevant representations and away from irrelevant representations, or 'sculpting the response space' (Fletcher et al., 2000).

Summary

This study provides novel evidence characterizing how competition between memories is experienced and overcome. By applying multi-voxel pattern analysis to measure neural reactivation of memories, we found that older visual memories are often reactivated during retrieval of newer memories and that this neural competition between memories is fundamentally related to whether retrieval fails or succeeds. Finally, we identified specific neural mechanisms that (a) detect neural expressions of mnemonic conflict during retrieval and (b) influence the degree of neural competition at retrieval through their contribution during memory encoding.

Acknowledgments

References

Anderson JR. A spreading activation theory of memory. J Verb Learn Verb Behav. 1983; 22:261–295.

Anderson MC, Neely JH. Interference and inhibition in memory retrieval. Memory. 1996; 22:237–313.

Anderson MC, Spellman BA. On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. Psychol Rev. 1995; 102:68–100. [PubMed: 7878163]

Badre D, Wagner AD. Frontal lobe mechanisms that resolve proactive interference. Cereb Cortex. 2005; 15:2003–2012. [PubMed: 15788702]

Badre D, Wagner AD. Computational and neurobiological mechanisms underlying cognitive flexibility. Proc Natl Acad Sci USA. 2006; 103:7186–7191. [PubMed: 16632612]

Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007; 45:2883–2901. [PubMed: 17675110]

Bjork, RA. The updating of human memory. In: Bower, GH., editor. The psychology of learning and motivation. New York: Academic Press; 1978. p. 235-259.

Blumenfeld RS, Ranganath C. Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. Neuroscientist. 2007; 13:280–291. [PubMed: 17519370]

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychol Rev. 2001; 108:624–652. [PubMed: 11488380]

Chun MM, Johnson MK. Memory: Enduring traces of perceptual and reflective attention. Neuron. 2011; 72:520–535. [PubMed: 22099456]

Dolan RJ, Fletcher PC. Dissociating prefrontal and hippocampal function in episodic memory encoding. Nature. 1997; 388:582–585. [PubMed: 9252188]

Feredoes E, Tononi G, Postle BR. Direct evidence for a prefrontal contribution to the control of proactive interference in verbal working memory. Proc Natl Acad Sci USA. 2006; 103:19530–19534. [PubMed: 17151200]

Fletcher PC, Shallice T, Dolan RJ. "Sculpting the response space"--an account of left prefrontal activation at encoding. Neuroimage. 2000; 12:404–417. [PubMed: 10988034]

Henson RN, Shallice T, Josephs O, Dolan RJ. Functional magnetic resonance imaging of proactive interference during spoken cued recall. Neuroimage. 2002; 17:543–558. [PubMed: 12377133]

Johnson JD, McDuff SG, Rugg MD, Norman KA. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron. 2009; 63:697–708. [PubMed: 19755111]

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol Bull. 1993; 114:3–28. [PubMed: 8346328]

Jonides J, Nee DE. Brain mechanisms of proactive interference in working memory. Neuroscience. 2006; 139:181–193. [PubMed: 16337090]

Jonides J, Smith EE, Marshuetz C, Koeppe RA, Reuter-Lorenz PA. Inhibition in verbal working memory revealed by brain activation. Proc Natl Acad Sci USA. 1998; 95:8410–8413. [PubMed: 9653200]

Kim H. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. NeuroImage. 2011; 54:2446–2461. [PubMed: 20869446]

Kuhl BA, Dudukovic NM, Kahn I, Wagner AD. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat Neurosci. 2007; 10:908–914. [PubMed: 17558403]

Kuhl BA, Kahn I, Dudukovic NM, Wagner AD. Overcoming suppression in order to remember: Contributions from anterior cingulate and ventrolateral prefrontal cortex. Cogn Affect Behav Neurosci. 2008; 8:211–221. [PubMed: 18589510]

Kuhl BA, Rissman J, Chun MM, Wagner AD. Fidelity of neural reactivation reveals competition between memories. Proc Natl Acad Sci USA. 2011; 108:5903–5908. [PubMed: 21436044]

Kuhl BA, Shah AT, DuBrow S, Wagner AD. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat Neurosci. 2010; 13:501–506. [PubMed: 20190745]

Lewis-Peacock JA, Postle BR. Temporary activation of long-term memory supports working memory. J Neurosci. 2008; 28:8765–8771. [PubMed: 18753378]

McDuff SG, Frankel HC, Norman KA. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. J Neurosci. 2009; 29:508–516. [PubMed: 19144851]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001; 24:167–202. [PubMed: 11283309]

Nee DE, Jonides J, Berman MG. Neural mechanisms of proactive interference-resolution. NeuroImage. 2007; 38:740–751. [PubMed: 17904389]

Norman KA, Newman EL, Detre G. A neural network model of retrieval-induced forgetting. Psychol Rev. 2007; 114:887–953. [PubMed: 17907868]

Öztekin I, Badre D. Distributed patterns of brain activity that lead to forgetting. Front Hum Neurosci. 2011; 5:1–8. [PubMed: 21283556]

Paller KA, Wagner AD. Observing the transformation of experience into memory. Trends Cogn Sci. 2002; 6:93–102. [PubMed: 15866193]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. Science. 2005; 310:1963–1966. [PubMed: 16373577]

Roediger HL, Neely JH. Retrieval blocks in episodic and semantic memory. Can J Psychol. 1982; 36:213–242.

Schacter DL. The seven sins of memory. Insights from psychology and cognitive neuroscience. Am Psychol. 1999; 54:182–203. [PubMed: 10199218]

Shimamura AP, Jurica PJ. Memory interference effects and aging: Findings from a test of frontal lobe function. Neuropsychology. 1994; 8:408–412.

Shimamura AP, Jurica PJ, Mangels JA, Gershberg FB, Knight RT. Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. J Cogn Neurosci. 1995; 7:144–152.

Slamecka NJ, Ceraso J. Retroactive and proactive inhibition of verbal learning. Psychol Bull. 1960; 57:449–475.

Smith AD. Output interference and organized recall from long-term memory1. J Verb Learn Verb Behav. 1971; 10:400–408.

Sohn MH, Goode A, Stenger VA, Carter CS, Anderson JR. Competition and representation during memory retrieval: roles of the prefrontal cortex and the posterior parietal cortex. Proc Natl Acad Sci USA. 2003; 100:7412–7417. [PubMed: 12773617]

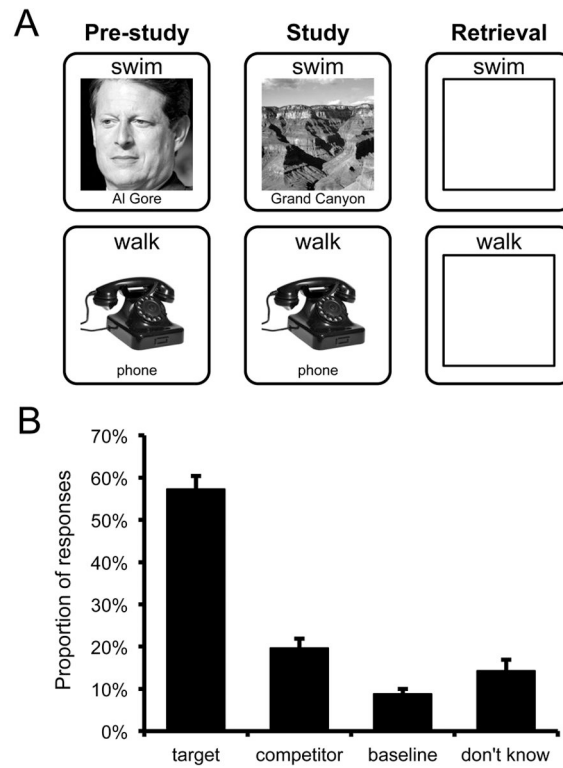Underwood BJ. Interference and forgetting. Psychol Rev. 1957; 64:49–60. [PubMed: 13408394]

Figure 1.
(*a*) Task diagram. During pre-study, subjects encoded initial word-image pairings. Images were drawn from one of three categories: faces, objects, scenes. All pairings were encoded twice during pre-study. During study, all words were presented again and were either paired with a new image (change trials) or the same image as pre-study (repeat trials). At retrieval, subjects were presented with words and attempted to retrieve the most recent (or only) image paired with that word. Subjects responded by indicating the category of the image (face, object, scene) or by responding "don't know" via button press. (*b*) Behavioral performance at retrieval for change trials where the original and newer images were from different categories (between-category change trials). Subjects were most likely to indicate the category of the newer (target) image, but were more likely to indicate the category of the older (competitor) image than a category never paired with that word (baseline). Error bars reflect standard error of the mean.
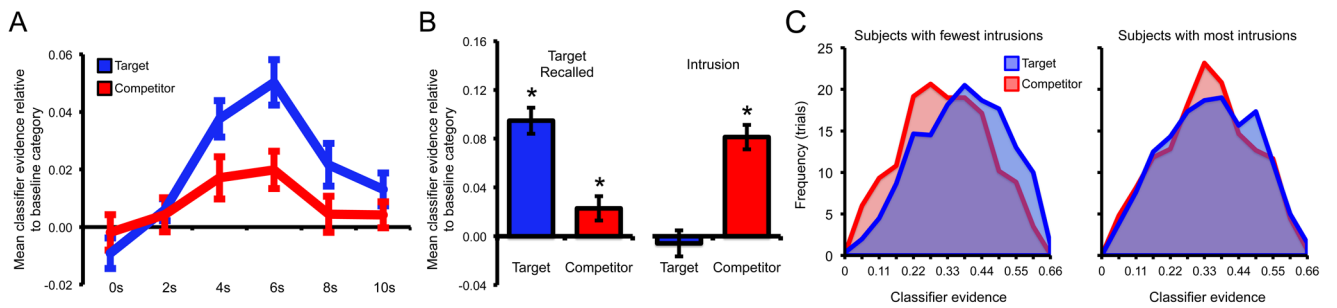
Figure 2.
Target vs. competitor reactivation within ventral temporal cortex for between-category change trials at retrieval. (**a**) Volume-by-volume classifier evidence (TR = 2s) for all trials, regardless of behavioral accuracy. Evidence for both the target and competitor categories was greater than evidence for the baseline category; evidence for the target category was also greater than evidence for the competitor category. (**b**) Left panel: successful target retrieval was associated with above-baseline classifier evidence for both the target and competitor memory. Right panel: intrusions (erroneous retrieval of competing memories) were associated with above-baseline evidence for the competitor memory, but not the target memory [note: for right panel, data are based on a sub-set of subjects (N = 12) for which a sufficient number of intrusions occurred]. (**c**) Histograms representing the frequency of classifier evidence across all retrieval trials (collapsed across behavioral accuracy) for bins of width = .055. Left panel: subjects with the lowest rate of intrusions displayed distinct distributions of target and competitor evidence. Right panel: subjects with the highest rate of intrusions displayed overlapping distributions of target and competitor evidence. Error bars reflect standard error of the mean.
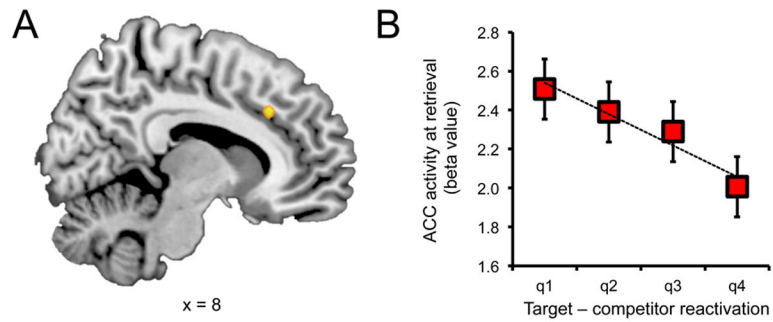
Figure 3.
Anterior cingulate cortex and retrieval competition. (***a***) Trial-by-trial variance in target vs. competitor evidence was negatively correlated with activation in anterior cingulate cortex. (***b***) Activation in anterior cingulate cortex as a function of target vs. competitor evidence (broken down by quartiles); activation in anterior cingulate cortex was relatively high when target–competitor evidence was low and relatively low when target–competitor evidence was high. Note: these data simply allow for visualization of the relationship referred to in (***a***), and do not constitute an independent analysis. Error bars reflect within-subject standard error.
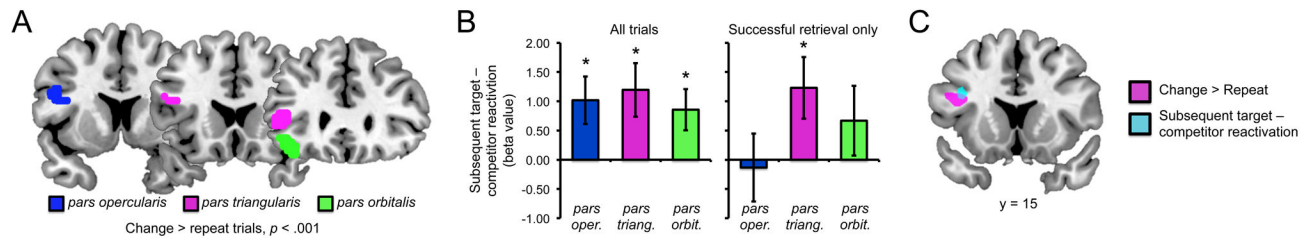
Figure 4.
Left inferior frontal gyrus contributions to encoding. (**a**) A contrast of change > repeat trials revealed clusters of activation within three anatomically defined left inferior frontal gyrus ROIs: *pars opercularis, pars triangularis*, and *pars orbitalis* (p < .001, 5 voxel extent threshold). (**b**) Left panel: activity in the left inferior frontal gyrus ROIs during the encoding of between-category change trials was predictive of subsequent target–competitor reactivation at retrieval, as reflected by positive beta values. Right panel: activity in *pars triangularis* was uniquely predictive of gradations in subsequent target–competitor reactivation when only considering trials that were subsequently remembered. (**c**) Direct comparison of the whole-brain contrast of change vs. repeat encoding trials (*p* < .001; see Table 2) and the analysis of subsequent target–competitor reactivation (*p* < .001; see Table 3) revealed a single area of overlap in frontal cortex: left inferior frontal gyrus (*pars triangularis* and *pars opercularis*). Error bars reflect standard error of the mean.

Table 1

Retrieval activity negatively correlated with target–competitor reactivation (between category change trials only)

Region	MNI coordinates			Cluster size	Peak Z
	x	y	z		
R. temporal pole	36	8	−26	6	3.66
R. anterior cingulate cortex	9	26	34	8	3.51
L. Insula	−30	−25	22	6	3.43

$p < .001$, uncorrected; 5 voxel extent threshold

Table 2

Encoding of change trials versus repeat trials

Region	MNI coordinates			Cluster size	Peak Z
	x	y	z		
R. inf./mid./sup. occipital gyrus, fusiform gyrus, middle temporal gyrus	42	−82	16	680	5.78
L. inf./mid./sup. occipital gyrus, middle temporal gyrus	−39	−70	22	350	4.32
L. fusiform gyrus, inferior temporal gyrus	−42	−49	−14	221	4.31
L. inferior frontal gyrus (*pars orbitalis, triangularis, opercularis*)	−39	29	−17	186	4.22
R. inferior frontal gyrus (*pars triangularis*)	51	32	4	27	3.69
R. medial frontal gyrus	9	62	19	11	3.55
R. precentral/postcentral gyrus	63	−7	37	26	3.55
L. medial frontal gyrus	−6	62	19	5	3.27

p < .001, uncorrected; 5 voxel extent threshold

Table 3

Encoding activity positively correlated with target–competitor reactivation at retrieval (between-category change trials only)

Region	MNI coordinates			Cluster size	Peak Z
	x	y	z		
R. cingulate gyrus	15	−31	28	19	3.83
R. middle occipital gyrus	39	−85	22	15	3.82
L. brainstem/cerebellum	−9	−31	−17	96	3.75
R. parahippocampal gyrus, brainstem, hippocampus	15	−28	−11	53	3.68
R. brainstem	18	−28	−29	7	3.66
L. inferior frontal gyrus (*pars orbitalis*)	−30	26	−20	6	3.65
R. cerebellum	9	−64	−38	53	3.59
L. inferior/middle occipital gyrus	−24	−88	−2	70	3.54
R. fusiform gyrus	24	−79	−11	32	3.54
L. fusiform gyrus	−27	−61	−11	28	3.53
R. fusiform/parahippocampal gyrus	33	−43	−20	27	3.50
L. inferior frontal gyrus (*pars opercularis, triangularis*)	−39	14	25	20	3.37
L. intraparietal sulcus	−27	−70	43	12	3.29

$p < .001$, uncorrected; 5 voxel extent threshold; activations in white matter excluded