

Fidelity of neural reactivation reveals competition between memories

Brice A. Kuhl^{a,b,1}, Jesse Rissman^a, Marvin M. Chun^b, and Anthony D. Wagner^{a,c}

^aDepartment of Psychology and ^cNeurosciences Program, Stanford University, Stanford, CA 94305; and ^bDepartment of Psychology, Yale University, New Haven, CT 06511

Edited by Edward E. Smith, Columbia University, New York, NY, and approved February 14, 2011 (received for review November 10, 2010)

Remembering an event from the past is often complicated by the fact that our memories are cluttered with similar events. Though competition is a fundamental part of remembering, there is little evidence of how mnemonic competition is neurally represented. Here, we assessed whether competition between visual memories is captured in the relative degree to which target vs. competing memories are reactivated within the ventral occipitotemporal cortex (VOTC). To assess reactivation, we used multivoxel pattern analysis of fMRI data, quantifying the degree to which retrieval events elicited patterns of neural activity that matched those elicited during encoding. Consistent with recent evidence, we found that retrieval of visual memories was associated with robust VOTC reactivation and that the degree of reactivation scaled with behavioral expressions of target memory retrieval. Critically, competitive remembering was associated with more ambiguous patterns of VOTC reactivation, putatively reflecting simultaneous reactivation of target and competing memories. Indeed, the more weakly that target memories were reactivated, the more likely that competing memories were later remembered. Moreover, when VOTC reactivation indicated that conflict between target and competing memories was high, frontoparietal mechanisms were markedly engaged, revealing specific neural mechanisms that tracked competing mnemonic evidence. Together, these findings provide unique evidence that neural reactivation captures competition between individual memories, providing insight into how well target memories are retrieved in the present and how likely competing memories will be remembered in the future.

forgetting | pattern classification

Our ability to remember an event from the past is powerfully influenced by competition arising from memories of similar or overlapping events (1–3). For example, in searching for today's parking space, we may find ourselves standing where we parked yesterday. Though competition between memories is almost ubiquitous, and a primary reason why we forget, there is surprisingly little evidence of how competition between memories is neurally represented. In part, the lack of evidence reflects a methodological challenge of how to measure neural competition between memories. Here, we consider whether competition between memories can be measured by, and understood in terms of, the relative degree to which memories are neurally reactivated—that is, the degree to which patterns of neural activity present during event encoding are reinstated at retrieval. By this view, competitive remembering may strongly parallel competitive perception (4)—a domain that has been more extensively studied.

When competition exists between visual stimuli, responses within the ventral occipitotemporal cortex (VOTC) are strongly modulated by how attention is allocated. For example, when faces and scenes are concurrently or sequentially presented, increased activity is observed in fusiform or parahippocampal gyri according to whether faces or scenes are attended, respectively (5–9). Similarly, VOTC responses are tightly correlated with both spontaneously fluctuating perceptions (10, 11) and misperceptions (12) of visual stimuli. VOTC responses also scale with gradations in the strength of visual stimuli (13–15), thus reflecting the quality of perceptual events. Importantly, perceptual evidence emerging from the VOTC is also correlated with the engagement of fron-

toparietal structures (15–19), putatively reflecting the translation of perceptual evidence to goal-relevant behavior.

Building on the literature relating competitive perception to VOTC responses, we used a memory task in which subjects learned cue-associate pairs for which the cues were words and the associates were images of faces or scenes. Competition between individual memories was created through AB/AC learning: subjects first encoded and retrieved novel cue-associate pairs (noncompetitive, AB pairs) and subsequently encoded and retrieved overlapping pairs that contained a repeated cue paired with a novel associate (competitive, AC pairs; Fig. 1). A separate set of novel cue-associate pairs (noncompetitive, DE pairs) were not followed by overlapping pairs, thus functioning as a control condition. To allow for separation of VOTC reactivation related to target vs. competitor memories, the B and C images for a given AB/AC set were always from distinct visual categories (i.e., a face and a scene). During the critical retrieval phase, subjects were presented with word cues and attempted to covertly recall target associates, indicating by button press whether they were able to recall the specific image vs. a more general memory for the category of the image (face/scene). The scanned encoding and retrieval rounds were followed by a behavioral posttest that reassessed memory for AB and DE pairs, allowing for measurement of the impact that AC learning had on memory for previously learned AB pairs.

Reactivation was assessed via multivoxel pattern analysis (MVPA)—a technique well-suited to measuring reactivation (20–22). Accordingly, we first trained a pattern classifier to discriminate patterns of activity within the VOTC that were associated with the encoding of words paired with faces vs. scenes; we then applied the classifier to retrieval trials to estimate the degree to which VOTC responses reflected face vs. scene reactivation—that is, the fidelity with which target memories were retrieved. Importantly, because the classifier measured the relative evidence for face vs. scene reactivation, here, fidelity of reactivation refers to the relative strength of target reactivation. We predicted that competitive remembering would be associated with reactivation of both target and competitor memories, thereby reducing the fidelity of VOTC reactivation. Moreover, we predicted that measures of competitive reactivation would be related to both how well target memories were retrieved in the present and how likely competing memories would be remembered in the future. Finally, we predicted that neural evidence of competitive reactivation would correspond to the engagement of frontoparietal structures that evaluate mnemonic—and perhaps perceptual—evidence.

Results

Behavioral Results. Retrieval rounds. Responses during the scanned retrieval rounds were separated according to whether subjects

Author contributions: B.A.K., M.M.C., and A.D.W. designed research; B.A.K. performed research; B.A.K. and J.R. analyzed data; and B.A.K., J.R., M.M.C., and A.D.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: brice.kuhl@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016939108/-DCSupplemental.

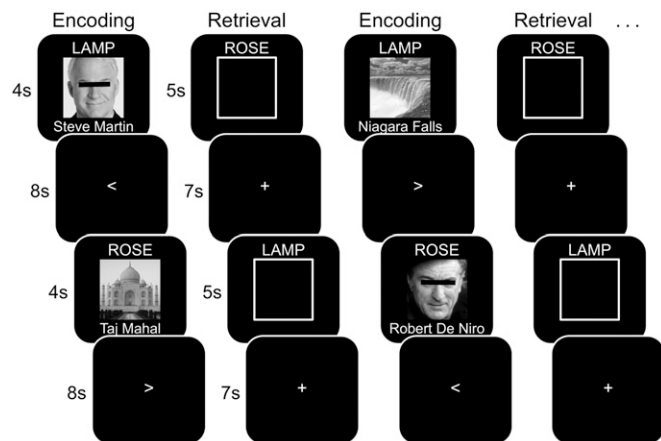


Fig. 1. Encoding and retrieval rounds. During encoding, subjects viewed nouns (cues) paired with images of well-known faces or scenes (associates). The name of each face/scene was presented below the image. During retrieval, subjects were presented with cues and instructed to recall the corresponding associate, indicating the category (face/scene) via button press. Encoding and retrieval rounds alternated for a total of seven rounds each. Half of the cues were paired with one associate (DE pairs), and half of the cues were paired with two associates (one face, one scene). If a cue was paired with two associates, the first associate (AB pair) was encoded and retrieved before encoding and retrieval of the second associate (AC pair). During retrieval, subjects were instructed to recall the most recent associate for each cue. Images of faces were not obscured in the actual experiment.

correctly indicated the category (face/scene) of the target image (hit); indicated that they did not remember the target image (don't know); or incorrectly classified the category of the target image (error). Hits were further subdivided according to whether subjects reported remembering specific or general details of the image. Because errors were relatively infrequent, they were not subdivided into specific and general errors.

Overall, subjects were fairly successful at retrieving target images: specific hit = 53.2% (mean); general hit = 23.5%; don't know = 14.7%; error = 6.4%; no response = 2.1% (Fig. 2). Critically, competition significantly impacted retrieval success, as evidenced by a robust reduction in the rate of specific hits ($t_{17} = 5.10, P < 0.001$; Fig. 2). Though the reduction in the rate of specific hits necessarily corresponded to a higher rate of responses in the remaining bins—and thus nonindependence of additional tests—competition increased general hits ($t_{17} = 2.22, P < 0.05$), don't knows ($t_{17} = 2.23, P < 0.05$), and errors ($t_{17} = 2.38, P < 0.05$). The rate of specific hits, general hits, don't knows, and errors did not differ for faces vs. scenes (all P 's > 0.4).

Posttest. Recall performance at posttest probed memory for AB and DE pairs from the scanned encoding/retrieval rounds. Posttest trials were coded as successful (recalled) if subjects retrieved at least some detail of the target image beyond the category. As with

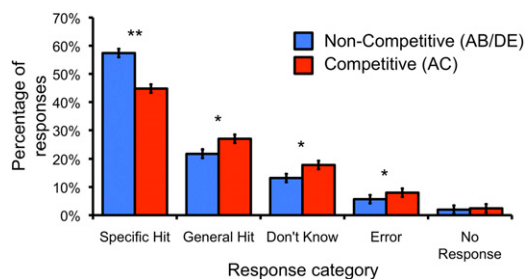


Fig. 2. Behavioral accuracy during retrieval rounds as a function of retrieval competition. Error bars indicate within-subject SEM.

the retrieval rounds, competition negatively impacted performance at posttest, evidenced by a lower rate of recalled AB pairs than DE pairs (mean = 52.4% vs. mean = 57.2%; $t_{17} = 2.32$, $P < 0.05$; *SI Results*).

fMRI Results. Category-selective encoding. Univariate analysis of the encoding data revealed robust category-sensitive neural responses within VOTC (Fig. S1), including greater responses to face trials within fusiform gyrus and greater responses to scene trials in parahippocampal cortex, confirming category-sensitivity within the VOTC. To next permit quantification of the fidelity of cortical reactivation at retrieval, we trained a pattern classifier to discriminate face vs. scene encoding trials using the encoding data from all voxels within an anatomically defined VOTC mask (SI Methods and Fig. S2A). MVPA of encoding trials confirmed highly robust sensitivity to face- vs. scene-related encoding responses in the VOTC (SI Results).

Cortical reactivation at retrieval. To assess reactivation at retrieval, pattern classification of retrieval trials was performed using the encoding data as the training set and the retrieval data as the testing set. Because the training set was restricted to the encoding data, classification of retrieval data could only succeed to the extent that neural responses that differentiated faces vs. scenes at encoding were reactivated at retrieval.

Considering all retrieval trials, irrespective of the subject's response or condition, trial-by-trial classification accuracy of the target associate's perceptual category (i.e., whether a retrieved item was a face vs. scene) was well above chance (mean = 66.6%, $t_{17} = 7.70$, $P < 0.001$), providing strong evidence that encoding-related activity within VOTC was reactivated at retrieval. This cued-recall result builds on prior evidence of neural reactivation of episodic memories revealed by MVPA, including during free recall (22), source memory (21), and item recognition (20). Notably, classification accuracy scaled with the number of voxels included in the VOTC mask (Fig. S2B), consistent with the idea that the representation of visual stimuli is distributed across the VOTC (23).

We next considered classifier performance in several additional ways. First, by computing classifier performance at each volume [i.e., repetition time (TR)-by-TR classification], we confirmed that classification accuracy generally conformed to the pattern of a hemodynamic response function, with peak accuracy achieved 4–8 s poststimulus onset (Fig. S2C). Second, the distribution of classifier evidence revealed that trials varied considerably in the strength of evidence for target reactivation (Fig. S2D). Finally, this distribution of evidence allowed for generation of receiver operating characteristic (ROC) curves (Fig. S2E); the area under these curves (AUC) provides another useful index of classifier performance (24, 25).

Cortical reactivation and retrieval strength. Having confirmed the approach, our first fundamental objective was to determine whether cortical reactivation scales with retrieval performance. Supporting this possibility, classification accuracy for specific hits, general hits, don't know, and error trials revealed a significant main effect of retrieval performance ($F_{3,51} = 21.70$, $P < 0.001$; Fig. 3*A–D*). Though classification accuracy for don't know trials did not differ from chance ($t_{17} = 0.27$, $P = 0.79$), classification accuracy for general hit trials was significantly above chance ($t_{17} = 6.48$, $P < 0.001$) and significantly greater than accuracy for don't know trials ($t_{17} = 3.45$, $P < 0.005$). Moreover, as the specificity of retrieval increased, so did the fidelity of cortical reactivation: classification accuracy for specific hit trials was well above chance ($t_{17} = 10.19$, $P < 0.001$) and significantly greater than accuracy for general hit trials ($t_{17} = 2.80$, $P < 0.05$). Finally, classification accuracy for error trials was numerically, but not significantly, below chance ($t_{17} = -1.53$, $P < 0.14$). Measures of AUC mirrored these results (Fig. 3*E*).

Mnemonic competition and cortical reactivation. Our second fundamental objective was to determine whether and how competition impacts the reactivation of target memories during retrieval. To this end, we first computed mean classification accuracy separately for noncompetitive DE/AB trials and for competition-

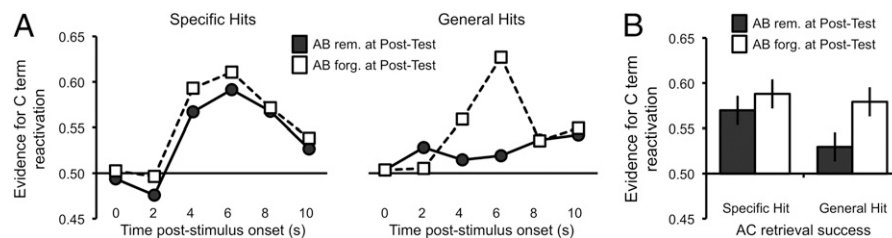


Fig. 5. Relationship between fidelity of reactivation during AC retrieval and subsequent memory for AB pairs. (A) Mean classifier evidence for C term reactivation during AC retrieval as a function of acquisition volume (2 s per volume), behavioral measures of AC retrieval success (specific hit vs. general hit), and subsequent memory for corresponding AB pairs at posttest (remembered vs. forgotten). (B) Mean classifier evidence for C term reactivation during AC retrieval as a function of AC retrieval success (specific hits vs. general hits) and subsequent memory for corresponding AB pairs at posttest [recalled vs. forgotten (don't know or error)]. Error bars indicate within-subject SEM.

memberings of new associations corresponded to better subsequent memory for previously encoded, competing associations.

Frontoparietal mechanisms and low-fidelity reactivation. The preceding analyses indicate that low-fidelity remembering of AC pairs is positively associated with later memory for AB pairs. One interpretation of this relationship is that past associations (B terms) are reactivated during the retrieval of newer associations (AC pairs), and this coactivation of B and C terms yields low-fidelity reactivation (i.e., ambiguous classifier evidence) but ultimately benefits AB retention. By this account, low-fidelity AC trials can be characterized in terms of high, but conflicting, mnemonic evidence. Alternatively, it is possible that when AC pairs are weakly reactivated, memory for AB pairs is less likely to be disrupted, thus accounting for the inverse relationship between AC reactivation and subsequent AB memory. By this account, low-fidelity AC trials can be characterized in terms of low, but not necessarily conflicting, evidence.

To differentiate between these competing accounts, we first identified frontoparietal regions that positively tracked retrieval evidence, and then assessed how responses in these regions related to the fidelity of AB vs. AC reactivation. Specifically, we performed a conjunction analysis of two independent contrasts: (i) a contrast to identify regions that tracked behavioral expressions of mnemonic evidence, and (ii) a contrast to identify regions that were differentially related to the fidelity of reactivation across AB vs. AC trials. To first identify regions that tracked mnemonic evidence, we contrasted specific hit trials against general hit and don't know trials. Importantly, we restricted this contrast to only DE pairs, so as to obtain a contrast that was independent of the critical AB/AC trials (Table S1). To next identify regions that were differentially engaged in relation to the fidelity of reactivation across AB vs. AC retrieval events, we first separated AB and AC trials according to the fidelity of classifier-based evidence for VOTC reactivation. AB trials were sorted into three bins of equal size: low-, medium-, and high-fidelity reactivation, and likewise for AC trials. It should be emphasized that these bins only represented the relative strength of target evidence; strong reactivation of both face and scene representations would correspond to low-fidelity reactivation because of the lack of evidence selectively favoring the target category. Using these binned data, we then tested for a voxel-level interaction between the fidelity of reactivation (high vs. low) and pair type (AB vs. AC), with the prediction being that AC trials associated with low-fidelity reactivation would elicit relatively high engagement of regions that tracked mnemonic evidence.

A conjunction analysis of the two contrasts (each thresholded at $P < 0.005$) revealed overlapping activation in several regions, including the dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex, and lateral and medial parietal cortex (Fig. 6A and Fig. S3A). Critically, these regions were characterized by relatively high activation during low-fidelity AC retrieval events (Fig. 6B). Indeed, a direct contrast of low-fidelity AC retrieval events vs. low-fidelity AB retrieval events revealed highly similar frontoparietal regions (Table S2). These data are particularly striking considering that, following our trial binning procedure,

low-fidelity AC retrieval events were associated with even lower-fidelity reactivation (mean = 0.41) than low-fidelity AB retrieval events (mean = 0.43; $P < 0.01$). High-fidelity AB (mean = 0.70) vs. AC (mean = 0.70) trials did not differ ($P = 0.73$). Thus, whereas low-fidelity AC retrieval events were associated with weaker classifier-based evidence that target memories were reactivated, they were associated with elevated frontoparietal responses, suggesting that a relatively high amount of episodic information was nonetheless retrieved. Together, these data strongly favor the interpretation that low-fidelity reactivation during AC retrieval reflected robust but nonselective retrieval—that is, reactivation of both target and competitor memories.

Discussion

Emerging evidence indicates that neural reactivation is a fundamental component of event remembering (26). The present results provide unique evidence that reactivation reflects not only how successfully memories are retrieved but also how competition impacts remembering—namely, that competitive remembering of visual memories is associated with lower-fidelity

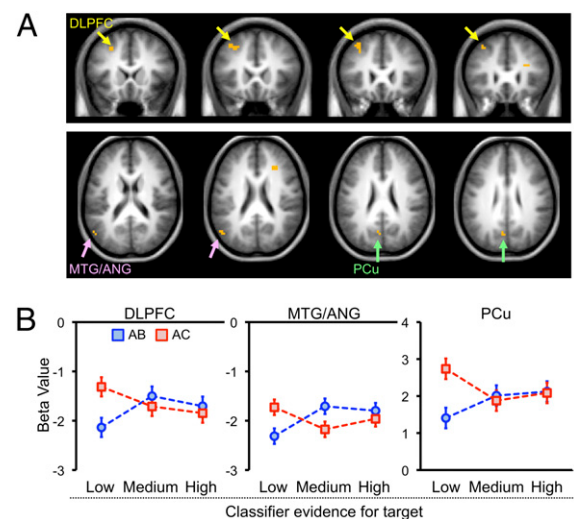


Fig. 6. Frontoparietal regions that tracked mnemonic evidence and fidelity of reactivation. (A) Conjunction analysis: regions that tracked retrieval of specific event details (specific hit trials > general hit/don't know trials; DE pairs only; $P < 0.005$) and displayed an interaction between fidelity of reactivation (low vs. high) and trial type (AB vs. AC; $P < 0.005$). DLPFC, dorsolateral prefrontal cortex; MTG, middle temporal gyrus; ANG, angular gyrus; PCu, precuneus. For complete set of implicated regions, see Fig. S3. (B) Responses in several regions of interest drawn from the conjunction analysis revealed that the interaction between fidelity of reactivation and trial type was characterized by marked activation for AC trials associated with low-fidelity reactivation. Error bars indicate within-subject SEM.

reactivation within VOTC. Critically, these decreases in the fidelity of target reactivation were predictive of subsequent competitor memory. These data provide a striking link between neural expressions of memory competition and the corresponding consequences for future remembering.

The present findings point to a strong parallel between visual perception and visual remembering (4). First, although competition impacted the fidelity of reactivation, VOTC reactivation was clearly modulated by retrieval goals. That is, the majority of competitive retrieval trials were associated with VOTC responses that favored the target category (Fig. 4*B*). This finding parallels evidence that, when competing visual stimuli are presented, VOTC responses track attended stimuli (5–9). Thus, both during competitive perception and competitive remembering, patterns of responses within VOTC provide insight into the representations that are favored via perceptual or mnemonic selection. Second, we observed a strong relationship between behavioral measures of retrieval specificity—that is, the strength of retrieved information—and the fidelity of VOTC reactivation (Fig. 3). Thus, the degree to which relevant VOTC structures are engaged corresponds to the strength of visual evidence, whether that evidence comes from external inputs that drive current perception (15, 19) or episodic remembering. Together, these findings reveal that, like visual perception, visual remembering is intimately related to the evoked patterns of activation across VOTC.

A primary goal of the present study was to characterize how competition impacts neural reactivation. Does competition reduce the fidelity of VOTC reactivation? If so, are such reductions due to reinstatement of competing memories, and are these reductions predictive of later memory outcomes? Our findings suggest that retrieval cues associated with competing images elicited lower-fidelity reactivation of target categories, relative to retrieval cues associated with a single image. This finding is consistent with the observed behavioral costs associated with competition—namely, a reduction in the rate with which specific event details were retrieved. Interestingly, we also observed a trend for lower-fidelity reactivation during competitive retrieval even when behavioral accuracy was matched, suggesting that VOTC responses may reflect costs that are not otherwise apparent in behavior.

Given that reductions in the fidelity of AC reactivation putatively reflect interference from corresponding AB pairs, it is notable that the fidelity of AC reactivation did not significantly correlate with the fidelity of corresponding AB reactivation. To the extent that AB pairs compete with AC pairs, a negative correlation would have been predicted. Though the data were numerically in this direction (the correlation coefficient was negative in 12 of 18 subjects, and the group mean was negative), we may have lacked sufficient power for this subtle analysis. In particular, AC reactivation should be a product of both the degree to which AC pairs are successfully encoded, which is unaccounted for in this analysis, and the degree to which AB pairs interfere. As this finding is ultimately inconclusive, this issue is worth future consideration. It is of note, however, that we did observe a positive correlation between the fidelity of AB reactivation and prefrontal engagement during corresponding AC retrieval (*SI Results*), consistent with the idea that the strength of AB associations does influence AC retrieval.

Perhaps the most compelling aspect of VOTC responses in the present study is that variance in the fidelity of AC reactivation was predictive of future remembering of competing memories. The lower the evidence for target memory reactivation—and, therefore, the stronger the evidence for competitor reactivation—the more likely that competing memories were later remembered. Critically, this relationship was observed even when only considering trials that both the subject and the classifier categorized correctly. In other words, even when targets were successfully retrieved and VOTC reactivation was biased toward target representations, there was nonetheless meaningful variance in the fidelity of reactivation that reflected the influence of competing memories. These data indicate that retrieval success was graded, and these gradations were diagnostic of future remembering.

Though we primarily focus on this variance in relation to competitive dynamics, there was also variance in the fidelity of reactivation for noncompetitive retrieval trials (Fig. 4*A*). Indeed, variance in the fidelity of AB/DE reactivation was predictive of subsequent memory for these pairs (Fig. *S4*). Thus, AB memory at posttest was a function of both how strongly AB pairs were initially reactivated and how weakly AC pairs were subsequently reactivated. Collectively, the present findings provide evidence for a relationship between distributed patterns of neural reactivation and subsequent mnemonic outcomes. As such, these results are highly relevant to a growing literature that considers the powerful ways in which current acts of retrieval can influence future remembering (27–29).

Consideration of frontoparietal responses during retrieval provided strong evidence that competing items were reactivated during target retrieval, thereby reducing the fidelity of classifier-based evidence for reactivation. Specifically, low-fidelity AC retrieval events disproportionately engaged frontoparietal regions that tracked mnemonic evidence (Fig. 6). This relationship raises two important questions. First, why was reactivation of B terms during AC retrieval associated with better subsequent AB memory? On the one hand, our findings are surprising in light of reconsolidation theory, which posits that reactivation renders memories susceptible to disruption (30, 31). On the other hand, the benefits of AB reactivation during AC retrieval are consistent with evidence documenting the powerful benefits of event retrieval for subsequent memory retention (29), as well as with event integration theories, which posit that the reinstatement of older associations during processing of newer associations can lead to the direct binding of past and present in memory (32, 33). Similarly, behavioral evidence indicates that integration can reduce forgetting of otherwise competing associations (34). Indeed, the relationship observed here strongly parallels recent evidence that reactivation of competing memories during encoding can protect competing memories against forgetting (35) (*SI Discussion*). However, seemingly inconsistent with a strong role of integration, behavioral evidence indicated that recall of AB and AC pairs was conditionally independent ($\chi^2 = 0.085$, $df = 1$, $P = 0.77$; *SI Discussion*). A more subtle but theoretically well-articulated possibility is that reactivation may either increase or decrease the likelihood of forgetting depending on how strongly competing memories are reactivated. In other words, memory disruption may be a nonmonotonic function of reactivation strength (36–38). Understanding the situations in which reactivation is beneficial vs. disruptive represents an important avenue for future investigation.

A second question is whether the frontoparietal regions that tracked mnemonic evidence are convergent with those engaged during visual perception. In particular, we observed a relationship between classifier evidence and responses within a subregion of the left DLPFC, extending from the superior frontal sulcus to the middle frontal gyrus. Notably, a highly similar DLPFC subregion has been argued to play a fundamental role in perceptual decision-making (15, 17; cf. ref. 39). For example, when viewing noise-degraded face vs. scene images, DLPFC activation increases as a function of the strength of visual evidence, regardless of category, whereas category-sensitive VOTC responses increase according to the strength of category-preferred visual stimuli (15). Moreover, DLPFC responses directly scale with both the strength of VOTC responses and subjects' behavioral performance (15).

The present data are strikingly consistent with the data implicating DLPFC in perceptual decision-making, as our data similarly reveal that DLPFC activation scaled with both behavioral performance (i.e., DLPFC was modulated by retrieval success) and VOTC responses (i.e., DLPFC activation was related to classifier evidence; *Table S3*). Importantly, follow-up analyses confirmed the anatomical consistency of the present DLPFC foci with the DLPFC region previously implicated in perceptual decision-making (*Table S3*). The strong convergence across these distinct domains suggests a commonality in the operations that DLPFC performs during visual perception and visual remembering—namely, these collective findings indicate that DLPFC is engaged in relation to the fidelity or strength of

responses within VOTC, regardless of whether these responses are driven by current visual input or memory-based reactivation of visual events. These functional interactions between the prefrontal cortex and the VOTC putatively enable perceptual or mnemonic evidence to be translated to goal-relevant behavioral responses (15, 17, 18). Importantly, these interactions are thought to be fundamental to the implementation of cognitive control in service of competitive remembering (*SI Discussion*).

Collectively, the present results constitute unique evidence relating patterns of neural reactivation elicited during competitive remembering to the quality of information retrieved in the present and to memory outcomes experienced in the future. These results parallel findings from studies of visual perception, indicating that VOTC structures are modulated by current mnemonic goals while reflecting the costs associated with mnemonic competition. Importantly, our results also point to overlap in frontoparietal mechanisms that operate upon perceived vs. remembered representations. More broadly, our results reveal that reactivation provides unique insight into understanding competitive dynamics between memories, and is central to both the experience and consequences of episodic remembering.

Methods

Procedure. The experiment was comprised of four phases: encoding, retrieval, a face/scene localizer task (not considered here), and a posttest. All phases except the posttest were conducted during fMRI scanning. Seven encoding rounds and seven retrieval rounds occurred in alternation. During encoding rounds, subjects studied words (cues) paired with either faces or scenes (associates) for 4 s each. Subjects encoded a total of 48 pairs in each of three conditions (AB, AC, DE). AB and DE pairs appeared in encoding rounds 1–6 (eight pairs per condition per round); AC pairs appeared in rounds 2–7 (eight pairs per round). AB pairs and corresponding AC pairs always appeared in encoding rounds n and $n + 1$, respectively. Encoding trials were separated by

an 8-s active baseline period (Fig. 1). During retrieval rounds, subjects' memory was tested for each pair from the immediately preceding encoding round. Subjects were presented with cues and instructed to covertly recall the corresponding associate. Each trial lasted 5 s, and subjects indicated their retrieval success by making one of five responses using a five-key button box: (i) don't know, (ii) face-specific, indicating that they remembered the specific image and that it was a face, (iii) face-general, indicating that they had a nonspecific memory of a face, (iv) scene-specific, and (v) scene-general. Retrieval trials were separated by a 7-s baseline during which a fixation cross was presented. After exiting the scanner, subjects completed the posttest. For each posttest trial (5 s), subjects were presented with a cue along with the instruction to retrieve either the face or the scene that was previously studied with that cue. For cues that had been associated with more than one image, subjects were always prompted to retrieve the first associate (B term). Subjects responded aloud during posttest. Each trial was followed by a 1-s fixation cross. For additional details of each phase, see *SI Methods*.

fMRI Data Analysis. fMRI scanning was conducted at the Lucas Center at Stanford University on a 3.0T GE Signa MRI system (GE Medical Systems). Functional images were obtained using a T2*-weighted 2D gradient echo spiral-in/out pulse sequence; TR = 2 s; echo time (TE) = 30 ms; flip angle = 75°; 30 slices, $3.4 \times 3.4 \times 4$ mm; axial oblique sequential acquisition. Encoding rounds corresponded to seven functional scans (940 volumes total), and likewise for retrieval rounds. Image preprocessing and data analysis were performed using SPM5 (Wellcome Department of Cognitive Neurology, London). Pattern classification analyses were conducted using the Princeton Multi-Voxel Pattern Analysis Toolbox (<http://code.google.com/p/princeton-mvpa-toolbox/>) and custom code implemented in MATLAB (MathWorks).

ACKNOWLEDGMENTS. This work was supported by National Institute of Mental Health Grants 5R01-MH080309 and 5R01-MH076932 (to A.D.W.); R01-EY014193 and P30-EY000785 (to M.M.C.); and EY019624-02 (to B.A.K.).

- Anderson JR (1974) Retrieval of propositional information from long-term memory. *Cogn Psychol* 6:451–474.
- Anderson MC (2003) Rethinking interference theory: Executive control and the mechanisms of forgetting. *J Mem Lang* 49:415–445.
- Mensink G, Raaijmakers JG (1988) A model for interference and forgetting. *Psychol Rev* 95:434–455.
- Anderson MC, Spellman BA (1995) On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychol Rev* 102:68–100.
- Dudukovic NM, Preston AR, Archie JJ, Glover GH, Wagner AD (2011) High-resolution fMRI reveals match enhancement and attentional modulation in the human medial temporal lobe. *J Cogn Neurosci* 23:670–682.
- Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M (2005) Top-down enhancement and suppression of the magnitude and speed of neural activity. *J Cogn Neurosci* 17:507–517.
- O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. *Nature* 401:584–587.
- Yi D-J, Kelley TA, Marois R, Chun MM (2006) Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. *Brain Res* 1080:53–62.
- Yi D-J, Chun MM (2005) Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. *J Neurosci* 25:3593–3600.
- Lumer ED, Rees G (1999) Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proc Natl Acad Sci USA* 96:1669–1673.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21:753–759.
- Summerfield C, Egner T, Mangels J, Hirsch J (2006) Mistaking a house for a face: Neural correlates of misperception in healthy humans. *Cereb Cortex* 16:500–508.
- Bar M, et al. (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529–535.
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3: 837–843.
- Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for perceptual decision-making in the human brain. *Nature* 431:859–862.
- Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci* 5:10–16.
- Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG (2006) Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proc Natl Acad Sci USA* 103:10023–10028.
- Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci* 2:176–185.
- Ploran EJ, et al. (2007) Evidence accumulation and the moment of recognition: Dissociating perceptual recognition processes using fMRI. *J Neurosci* 27:11912–11924.
- Johnson JD, McDuff SGR, Rugg MD, Norman KA (2009) Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron* 63:697–708.
- McDuff SGR, Frankel HC, Norman KA (2009) Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *J Neurosci* 29:508–516.
- Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* 310:1963–1966.
- Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Bradley AP (1997) The use of area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 30:1145–1159.
- Rissman J, Greely HT, Wagner AD (2010) Detecting individual memories through the neural decoding of memory states and past experience. *Proc Natl Acad Sci USA* 107: 9849–9854.
- Danker JF, Anderson JR (2010) The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychol Bull* 136:87–102.
- 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.
- Dudukovic NM, Dubrow S, Wagner AD (2009) Attention during memory retrieval enhances future remembering. *Mem Cognit* 37:953–961.
- Roediger HL, Karpicke JD (2006) Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol Sci* 17:249–255.
- Dudai Y (2004) The neurobiology of consolidations, or, how stable is the engram? *Annu Rev Psychol* 55:51–86.
- Nader K (2003) Memory traces unbound. *Trends Neurosci* 26:65–72.
- Hupbach A, Gomez R, Hardt O, Nadel L (2007) Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learn Mem* 14:47–53.
- Shohamy D, Wagner AD (2008) Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* 60:378–389.
- Anderson MC, McCulloch JC (1999) Integration as a general boundary condition on retrieval-induced forgetting. *J Exp Psychol Learn Mem Cogn* 25:608–629.
- Kuhl BA, Shah AT, DuBrow S, Wagner AD (2010) Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat Neurosci* 13: 501–506.
- Newman EL, Norman KA (2010) Moderate excitation leads to weakening of perceptual representations. *Cereb Cortex* 20:2760–2770.
- Norman KA, Newman EL, Detre G (2007) A neural network model of retrieval-induced forgetting. *Psychol Rev* 114:887–953.
- Anderson MC, Levy BJ (2010) On the relationship between interference and inhibition in cognition. *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork*, ed Benjamin AS (Psychology Press, London), pp 107–132.
- Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. *J Neurosci* 29:8675–8687.

Supporting Information

Kuhl et al. 10.1073/pnas.1016939108

SI Results

Recall Accuracy at Posttest. Posttest trials were coded as recalled if subjects recalled at least some detail of the target image beyond the category (face vs. scene) with which they were cued. The reason for this liberal coding scheme was because subjects often provided detailed descriptions of the images without actually producing the specific label that corresponded to the image. Notably, in cases where details were recalled without the specific label, it was often not clear that the visual memories were any weaker than when the specific verbal label was correctly recalled. For example, for the image of Robert DeNiro, subjects may have recalled “actor, mafia guy,” or for the Acropolis of Athens, subjects may have recalled “ancient ruins.” Likewise, in some cases subjects described details specific to the photographs that were used (e.g., “building with interesting shadows”). Overall, the most frequent response type at posttest was successful recall of the target associate (recalled: mean = 54.8%). Of these trials, 63.1% corresponded to recall of the specific verbal label, and the remaining 36.9% corresponded to recall of details beyond the category but without the specific label. The next most frequent response type were trials on which subjects indicated they did not remember the target (don’t know: mean = 39.6%), and finally trials on which subjects retrieved an incorrect associate (error: mean = 5.6%). AB pairs yielded more don’t know responses than DE pairs (mean = 41.9% vs. mean = 37.3%; $t_{17} = 3.18$, $P < 0.01$); AB and DE pairs yielded a similar rate of errors (mean = 5.7% vs. mean = 5.6%; $t_{17} = 0.12$, $P = 0.91$). Note, however, that comparison of don’t know and error responses for AB and DE pairs were nonindependent of the primary comparison of recall rate.

Importantly, target associates reported as specific hits during the initial retrieval phase were much better remembered at posttest than were those reported as general hits. Indeed, the majority (mean = 77.1%) of associates reported as specific hits at initial retrieval were successfully remembered at posttest, whereas subsequent retrieval occurred much less frequently for general hits (mean = 34.4%; $t_{17} = 13.69$, $P < 0.001$). Thus, overt recall performance at posttest validated the covert recall data collected during the scanned retrieval phases, because specific and general hit trials from the retrieval phase robustly differed in terms of retrieval success at posttest.

Classification of Encoding Trials. To confirm that the pattern classifier was able to discriminate face- and scene-related neural activity during the encoding rounds, we used a cross-validation technique, training and testing the classifier on subsets of the encoding data. Classification accuracy for encoding trials was near ceiling (mean = 98.3%), documenting the separability of face- and scene-related cortical patterns during encoding. Though classification accuracy did not differ for competitive (AC; 98.0%) vs. noncompetitive (AB/DE; 98.5%) encoding trials ($t_{17} = 0.72$, $P = 0.48$), a more-sensitive continuous measure of classifier evidence (*SI Methods*) revealed that competitive encoding trials were characterized by modestly weaker classifier evidence for the target visual category, relative to noncompetitive encoding trials (mean = 0.874 vs. mean = 0.884; $t_{17} = 2.22$, $P = 0.04$). It should be emphasized that competitive encoding trials were equivalent to noncompetitive trials in terms of the visual input—they differed only in terms of their mnemonic history. Thus, though competitive encoding trials produced VOTC responses that could be classified with high accuracy, there was nonetheless

subtle evidence that past associations impacted VOTC responses during present encoding.

Univariate Analysis of AC Retrieval Events as a Function of Prior AB Reactivation. Behavioral and neural data from the present study suggest that B terms interfere with AC retrieval events. One consequence of this interference should be increased demands on cognitive control mechanisms during AC retrieval. To test whether the strength of AB associations impacted frontoparietal responses during AC retrieval, we created an additional univariate model in which, for each AC retrieval event, a regressor was included representing the fidelity with which corresponding B terms had been previously reactivated. In other words, if a B term had been strongly reactivated before AC encoding, did AC retrieval events display any evidence of this proactive interference? As with analyses in the main text, we separated AC retrieval events into high, medium, and low bins, reflecting the fidelity of corresponding AB reactivation; we then contrasted the high and low bins to test for regions displaying greater activation during AC retrieval when corresponding AB pairs had been reactivated with high fidelity. At a standard threshold ($P < 0.001$, five-voxel extent threshold) we observed a single cluster in the left midventrolateral prefrontal cortex (*pars triangularis*, Brodmann’s area 45; MNI coordinates: $x = -51$, $y = 27$, $z = 9$). A highly consistent region of the left midventrolateral prefrontal cortex has previously been implicated in resolving proactive interference in working memory (1, 2) and during task switching (3), consistent with the idea that this region was engaged during AC retrieval in support of resolving interference from prior associations (i.e., AB pairs). Left mid-VLPFC activation has also been observed in situations where weak episodic memories are retrieved amid competition from more dominant memories (4, 5), again consistent with a role for this region in resolving mnemonic interference.

SI Discussion

Integration Across Associations. In the AB/AC paradigm used here, recall of AC pairs is putatively impaired because B terms interfere with retrieval of C terms. We suggest that this interference is reflected in the coactivation of B and C terms during AC retrieval. However, simultaneous reactivation of B and C terms may reflect integration across AB and AC associations. That is, during the encoding of AC pairs, B terms may have been reactivated (6) and integrated into the new (AC) associations (7). During AC retrieval, both terms would then be reactivated as part of an integrated set. We next consider several lines of evidence that potentially address the role of integration in the present study.

First, VOTC responses during encoding were at least consistent with the possibility that B terms were reactivated while AC pairs were encoded, as evidence for the target (presented) category was modestly lower for competitive (AC) than noncompetitive (AB/DE) events (*SI Results*). However, though reactivation of B terms during AC encoding would putatively be necessary for integration to occur, it would not, on its own, require that integration take place. Thus, though not inconsistent with an integration account, these encoding data do not compel such an account.

Second, our behavioral results, at first pass, suggest that integration was unlikely to play a major role, in light of evidence that integration powerfully reduces interference-related forgetting (8), whereas the behavioral costs of interference observed here were quite robust. However, a more subtle possibility is that integration of B and C terms did take place, eliciting simulta-

neous reactivation of B and C terms, but memory for C terms was impaired because relevant contextual information concerning the temporal order of the associations—i.e., which item was more recent—was not maintained. Importantly, forgetting of this type, related to forgetting the relevant temporal source (9), is qualitatively different from a failure to reactivate the target memory or outright forgetting of C terms. Thus, the presence of interference-related forgetting is also inconclusive with respect to establishing the role of integration.

Finally, and most informatively, we considered whether recall of B and C terms was independent. If B and C terms are directly integrated, memory for corresponding B and C terms should be positively correlated. However, we did not observe a relationship between recall of AC pairs during the scanned retrieval rounds and recall of AB pairs at posttest. Specifically, memory for AB pairs at posttest was similar if corresponding AC pairs were specific hits (mean = 52.5%) vs. other (general hits, don't know, or error; mean = 53.8%). Similarly, the percentage of specific hits during AC retrieval was similar if corresponding AB pairs were recalled (mean = 43.7%) vs. other (don't know or error; mean = 45.0%). Analysis of conditional independence (Mantel-Haenszel test) revealed no violation of independence ($\chi^2 = 0.085$, $df = 1$, $P = 0.77$). Though this null relationship argues against a strong integration account of the present data, the present study was not specifically designed to test this possibility, and the relationship between reactivation and integration is worthy of future study.

Reactivation During Competitive Encoding. Kuhl et al. (6) recently found that during new learning, older, competing memories are reactivated, thereby protecting them against forgetting. The relationship between competitive encoding and reduced forgetting of past events was mediated by the hippocampus and reflected in cortical and subcortical reactivation of contextual features. Although the present study differs in several regards, including materials and analysis approach, the two studies share several parallel findings. First, here we also observed a link between the hippocampus and cortical reactivation, as recall of specific event details was associated with both robust VOTC reactivation and marked hippocampal engagement (Table S1). Second, both of these studies indicate that reactivation of previously encoded, competing events during ongoing mnemonic processing (either encoding or retrieval) is associated with reduced competitor forgetting. Third, although we primarily focus on reactivation at retrieval in the present study, there was also evidence suggesting that competing memories were reactivated during encoding as well. Namely, relative to noncompetitive encoding, competitive encoding trials were associated with a subtle decrease in classifier-based evidence for the target (perceived) category (SI Results), indicating that overlap in mnemonic associations impacted VOTC responses during present encoding—putatively because previously encoded memories were reactivated. Indeed, pooling data across both studies, we observed a robust relationship between individual differences in hippocampal activation during competitive encoding and protection against competitor forgetting (Fig. S5). Specifically, greater hippocampal activation during competitive encoding was associated with less forgetting of competing events. Together, these data provide converging evidence that (i) the hippocampus supports cortical reactivation of past events and (ii) that reactivation of past events—whether it occurs during present encoding or retrieval—can protect against forgetting of events past.

Cognitive Control and Competitive Remembering. The present results reveal a relationship between frontoparietal BOLD responses and the fidelity of memory reactivation within the VOTC. These findings relate to prior work that has demonstrated that frontally mediated cognitive control mechanisms are engaged

during competitive remembering (10). For example, Kuhl et al. (10) have shown that during competitive retrieval, responses in anterior cingulate cortex and right inferior frontal gyrus decrease in direct relation to the weakening of competing memories; these regions also display heightened responses when weak memories are retrieved in the face of more dominant memories (4). In the present study, competition was putatively maximal during AC retrieval events that were associated with low-fidelity VOTC reactivation. Indeed, direct interrogation of anterior cingulate and right inferior frontal regions of interest revealed marked activation during low-fidelity AC retrieval events (Fig. S3 B and C). Thus, complementing prior evidence that these regions are engaged in relation to the competitive status of retrieval targets, here we observed a relationship between the engagement of these regions and neural expressions of mnemonic competition within posterior cortical sites. These results provide evidence of a relationship between prefrontal regions that putatively support cognitive control operations and posterior cortical expressions of competitive visual remembering.

SI Methods

Participants. Subjects consisted of 18 (10 female), right-handed, native English speakers between the ages of 18 and 27 y (mean = 22 y). Informed consent was obtained in accordance with procedures approved by the Stanford University Institutional Review Board. Subjects were paid \$20/h for their participation.

Materials. Stimuli consisted of 96 nouns, 72 images of faces, and 72 images of scenes. Nouns were drawn from the Medical Research Council Psycholinguistic Database (http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm) and ranged in length from four to eight letters (mean = 5.4), with a Kucera–Francis written frequency of at least 5 (mean = 20.7), and a concreteness rating of at least 500 (mean = 600). Faces consisted of black-and-white photographs of well-known male and female actors and musicians (e.g., Robert DeNiro). Faces included hair and varied in emotional expression, but were cropped such that other body parts were not visible. Scenes consisted of black-and-white photographs of well-known locations, including natural landscapes (e.g., Niagara Falls) and manmade structures (e.g., Empire State Building). All face/scene images were 225 × 225 pixels, with a resolution of 150 pixels/inch. An additional eight nouns, four faces, and four scenes were used as filler items (all with similar properties to the nonfiller stimuli). All stimuli were randomly assigned to conditions for each subject.

Procedures. Encoding rounds. Encoding trials (4 s) consisted of a single noun (cue) presented directly above either a face or scene (associate), with the name of the associate (e.g., Robert De Niro) presented beneath the image (Fig. 1). Subjects were instructed to try to remember the association between each cue and associate such that they would later be able to retrieve the associate when presented with the cue. No overt response was required during encoding trials.

During the first encoding round, all cues and associates were novel. During encoding rounds 2–7, some cues were repeated but paired with novel associates. In total, half of all cues were paired with two associates (competitive condition) and half were paired with one associate (noncompetitive condition). Critically, for cues in the competitive condition, if the first associate was studied during encoding round n (AB trials; A = cue, B = associate), the second associate was studied during encoding round $n + 1$ (AC trials). Thus, AB trials were evenly distributed across encoding rounds 1–6, whereas AC trials were evenly distributed across encoding rounds 2–7. Importantly, for cues in the competitive condition, one associate was always a face and one associate was always a scene (i.e., the B and C terms were always from distinct categories). Cues in the noncompetitive condition were paired

with novel associates and evenly distributed across encoding rounds 1–6 (DE trials). Thus, the distinction between AB and DE pairs was entirely based on whether an overlapping (competing) pair was subsequently learned. In summary, the first encoding round contained AB and DE trials (pseudorandomly intermixed); encoding rounds 2–6 contained AB, AC, and DE trials (pseudorandomly intermixed); and the last encoding round contained AC trials as well as an equal number of filler trials consisting of novel cues and novel associates (pseudorandomly intermixed). In total, there were 48 encoding trials in each condition (AB, AC, DE). Half of the trials in each condition contained face associates; half contained scene associates. The number of face/scene associates was also balanced within each encoding round.

Each encoding trial was followed by an 8-s baseline period (Fig. 1) beginning with the presentation of a fixation cross (800 ms), followed by six arrows (800 ms each). Each arrow was followed by a brief fixation cross (400 ms). Subjects were instructed to indicate the direction (left/right pointing) of each arrow via a button box held in their right hand (11). Arrow orientation was randomly determined.

Retrieval rounds. Retrieval rounds probed subjects' memory for each of the pairs—and only those pairs—that were encoded in the immediately preceding encoding round. Each trial (5 s) consisted of a single cue presented above a square (equal in size to associate images). The interior of the square was black (matching the background screen color), thereby giving the impression of an empty box. The outline of the square was white for the first 4 s of the trial, then changed to red for 1 s, indicating that the trial was about to end. Subjects were instructed to covertly recall the associate that was presented with each cue in the immediately preceding study round. Subjects were made aware that, in some cases, a cue would be paired with more than one associate (AB, AC pairs), but that a cue would never be paired with more than one associate within a single encoding round. In such cases, subjects were instructed to always retrieve the associate from the immediately preceding encoding round—that is, the most recent associate. Subjects were not explicitly told that when a cue was paired with two associates, the associates would always be from different categories (i.e., one face, one scene).

Because memory for AB pairs was assessed during the retrieval round before corresponding AC pairs were encoded, both AB and DE retrieval trials were noncompetitive (as during encoding). Thus, for AB and DE trials, the B and E terms represented the retrieval targets, with no relevant competitors for these trials. For AC trials, the C term represented the retrieval target, and the previously encoded B term represented the competitor.

Subjects indicated their retrieval success by making one of five responses via the button box: (i) “don't know” indicated they could not remember anything about the associate; (ii) “face-specific” indicated they successfully recalled the specific associate and it was a face; (iii) “face-general” indicated they recalled that the associate was a face but could not recall the specific image; and likewise for (iv) scene-specific and (v) scene-general. For all subjects, don't know responses corresponded to the subject's right thumb; the assignment of the remaining four responses was counterbalanced across subjects. Subjects could make their response at any point during the trial; no emphasis was placed on responding quickly.

Retrieval trials were followed by a 7-s baseline period during which a fixation cross was presented; no responses were required during this period.

Posttest. After exiting the scanner, memory for all AB and DE pairs was assessed again in a single posttest, allowing for measurement of the consequence that AC encoding/retrieval had on memory for AB pairs, relative to DE pairs. Notably, the posttest differed from retrieval rounds in that it involved overt retrieval of the associate images, thus allowing for validation of subjects' covert responses collected during the scanned retrieval rounds.

On each trial (5 s), subjects were presented with a cue, above a square, as during retrieval. However, within each box was the word “face” or “scene,” which oriented subjects to the category of the associate that they were to retrieve. Subjects were instructed that for each trial, they should retrieve the associate that they previously studied with the presented cue and that matched the indicated category cue (face/scene). Subjects were explicitly told that if a cue had been paired with more than one associate, here they would always be cued to retrieve the first associate. Subjects were instructed to respond aloud and to indicate the name of the associate image as it appeared during the encoding rounds, but, if they could not remember the name of the image, they should indicate any additional details that they did remember (e.g., male or female, manmade structure or natural scene, etc.). Because each posttest trial included a cue indicating the category of the target associate (face/scene), if subjects simply indicated the category of the associate (i.e., repeated aloud the category cue), this was not recorded as an actual response. Thus, whereas memory for the target category was an acceptable response during the retrieval rounds, this was not the case for the posttest.

AB and DE pairs were pseudorandomly intermixed during the posttest, equating for average testing position of each condition. Each trial was followed by a 1-s fixation cross.

fMRI Data Analysis. Preprocessing. Functional data were corrected for slice timing and head motion. Structural images were coregistered to functional images and segmented into gray matter, white matter, and cerebrospinal fluid. Gray matter images were stripped of remaining skull and normalized to a gray matter Montreal Neurological Institute template. Normalized gray matter images were used for normalization of the structural and functional images. Images were resampled to 3-mm cubic voxels and smoothed with a Gaussian kernel (8 mm FWHM).

Univariate analyses. 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. Encoding and retrieval data were modeled separately with scan session (round) treated as a covariate (details of each GLM are included below). Linear contrasts were used to obtain subject-specific estimates for each effect of interest, which were then entered into a second-level, random-effects analysis using a one-sample *t* test against a contrast value of zero at each voxel. Unless otherwise noted, a threshold of $P < 0.001$, uncorrected, with a five-voxel extent threshold, was used for group-level contrasts. All contrast maps were overlaid on a mean anatomical image. Unless otherwise noted, region-of-interest (ROI) analyses were performed by extracting beta values from all significantly active voxels within a 6-mm radius of local maxima.

GLM for encoding data. A single GLM was generated for analysis of the encoding data. The model included regressors for six conditions, representing the two visual categories (faces vs. scenes) and three pair types (DE, AB, and AC pairs). A seventh regressor represented filler pairs that were included in the experiment.

GLM for retrieval data. Three GLMs were generated for analysis of the retrieval data. First, a model was generated that represented two levels of retrieval success (specific hits vs. general hits/don't knows) and the three pair types (DE, AB, AC), plus a regressor representing all excluded and filler trials. This model was used to identify regions of interest that were sensitive to detailed episodic retrieval.

Second, a model was generated that included regressors representing the visual category of the target image (face or scene), the pair type (DE, AB, AC), and the strength of classifier-based evidence for the target image (low, medium, high). Importantly, the classifier evidence bins were separately generated within each condition. For example, the 24 face-AB trials were sorted according to the strength of target evidence; the lowest eight trials constituted

the low bin, the middle eight trials the medium bin, and the highest eight trials the high bin; likewise for the remaining five conditions (scene-AB, face-AC, scene-AC, face-DE, and scene-DE). A separate regressor was included to represent filler trials.

Finally, a third model was generated specifically for AC trials that included regressors representing the visual category of the target image (face or scene) and the strength of classifier-based evidence for target reactivation during the preceding AB retrieval trials. That is, AC retrieval events were modeled in terms of how strongly corresponding B terms had previously been reactivated. As above, responses were binned into three groups (low, medium, high). Separate regressors represented filler trials and AB/DE trials.

Multivoxel pattern analyses. All fMRI data used for classification analyses were high-pass filtered (0.01 Hz), detrended, and z-scored (mean response for each voxel across time = 0). Classifier analyses were based on penalized logistic regression using L2-norm regularization. All classification analyses involved training a classifier on a sample of fMRI data and testing the classifier on a distinct sample of data. To assess reactivation during retrieval, classification analyses were conducted for which the training set consisted of the encoding data and the testing set consisted of the retrieval data. For these cross-phase analyses, a penalty parameter of 100 was used. Additionally, classification analysis of the encoding data was performed, for which all but one of the encoding rounds constituted the training set, and the left-out set constituted the testing set; classification was then repeated such that every encoding round contributed to both the training and testing sets. For these intraphase analyses, a penalty parameter of 10 was used. Penalty parameters were based on preliminary analyses and not the result of optimization procedures.

For all classification analyses, voxel inclusion was restricted using an anatomically defined mask of the VOTC. The mask was generated using the Anatomical Automatic Labeling atlas (http://www.cyceron.fr/web/aal_anatomical_automatic_labeling.html) and consisted of the union of the masks labeled as left fusiform, right fusiform, left parahippocampal, and right parahippocampal. The mask consisted of 2,553 total voxels (Fig. S2). No additional feature selection was performed. Our use of a VOTC mask—as opposed to using a whole-brain classifier—was motivated by three factors: (i) prior work has shown the VOTC to be highly sensitive to differences between face and scene perception; (ii) we were interested in characterizing mnemonic reactivation specifically within higher-level visual areas, and (iii) reducing the dimensionality of the classification space in a principled way has been shown to improve classification performance by eliminating uninformative voxels (12).

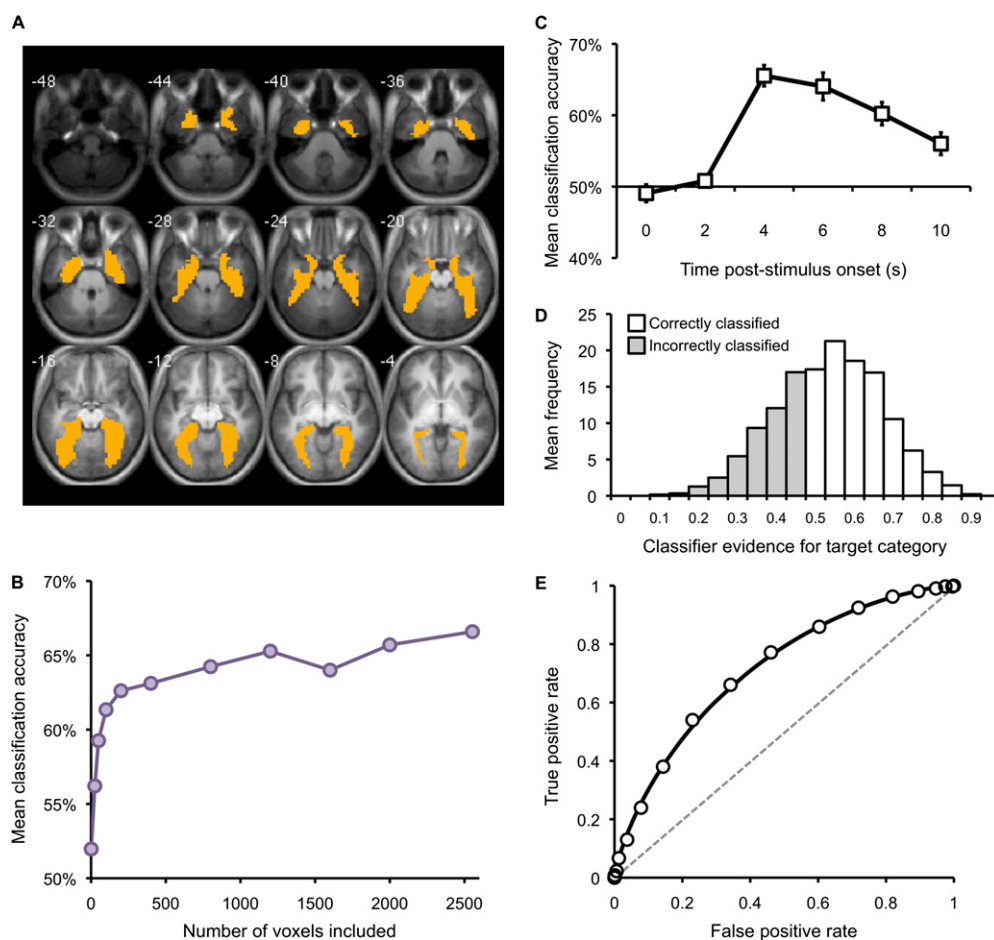
All of the classifications considered in the present study represented classification between face vs. scene categories. For each trial in the testing set, the logistic regression classifier generated

a scalar probability estimate that the trial corresponded to a face vs. scene (by construction, these probability estimates summed to unity). On each trial, the classifier's guess represented the category with the higher probability and was coded as correct or incorrect based on whether the guess corresponded to the target category for that trial. Critically, for retrieval trials, the target category represented the category of the image that subjects were intended to retrieve, as opposed to the category of what they reported retrieving.

Classification data were considered in three ways. First, we computed classification accuracy—the percentage of trials that the classifier correctly categorized. Second, we computed mean classifier evidence—the average probability that the classifier assigned to the target category of each trial. Capitalizing on the fact that the classifier's predictions were probabilistic rather than binary, this measure of classification performance potentially provided a more sensitive index of category discriminability than classification accuracy. Third, receiver operating characteristic (ROC) curves were generated for some of the core analyses. ROC curves were constructed by ranking the trial-by-trial classifier evidence scores according to how strongly each favored faces vs. scenes, and then charting the relationship between the true positive rate [arbitrarily defined here as the probability of classifying a face as a face, or $P(\text{face}|\text{face})$] and the false positive rate [$P(\text{face}|\text{scene})$] across a range of potential classifier decision thresholds. AUC was used as an index of classifier performance derived from the ROC curves, reflecting the mean accuracy with which a randomly chosen pair of trials from each class would be correctly classified.

Given that each trial in the encoding and retrieval phases corresponded to several volumes of fMRI data (six 2-s volumes per trial), trial-level classifier data were obtained by averaging several temporally contiguous volumes to generate a single brain activity pattern for each trial. For data from the encoding phase, TRs 3–4 (corresponding to 4–8 s poststimulus onset) were averaged; for the retrieval phase, TRs 3–6 (corresponding to 4–12 s poststimulus onset) were averaged. A wider window was used for retrieval trials because of variability in subjects' reaction times. Though all statistical analyses were based on trial-level classifier data, in some cases we display TR-by-TR classification performance for retrieval trials. In such cases, classification was separately applied to each of the six volumes corresponding to a single retrieval trial, producing a time course of classification performance. TR-by-TR classification is reported for illustrative purposes and to confirm that temporal evolution of classifier performance generally conformed to the shape of the canonical hemodynamic response function.

- Jonides J, Smith EE, Marshuetz C, Koeppe RA, Reuter-Lorenz PA (1998) Inhibition in verbal working memory revealed by brain activation. *Proc Natl Acad Sci USA* 95: 8410–8413.
- Badre D, Wagner AD (2005) Frontal lobe mechanisms that resolve proactive interference. *Cereb Cortex* 15:2003–2012.
- Badre D, Wagner AD (2006) Computational and neurobiological mechanisms underlying cognitive flexibility. *Proc Natl Acad Sci USA* 103:7186–7191.
- Kuhl BA, Kahn I, Dudukovic NM, Wagner AD (2008) Overcoming suppression in order to remember: Contributions from anterior cingulate and ventrolateral prefrontal cortex. *Cogn Affect Behav Neurosci* 8:211–221.
- Wimber M, et al. (2008) Neural markers of inhibition in human memory retrieval. *J Neurosci* 28:13419–13427.
- Kuhl BA, Shah AT, DuBrow S, Wagner AD (2010) Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat Neurosci* 13: 501–506.
- Shohamy D, Wagner AD (2008) Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* 60:378–389.
- Anderson MC, McCulloch JC (1999) Integration as a general boundary condition on retrieval-induced forgetting. *J Exp Psychol Learn Mem Cogn* 25:608–629.
- Johnson MK, Hashtroudi S, Lindsay DS (1993) Source monitoring. *Psychol Bull* 114: 3–28.
- Kuhl BA, Dudukovic NM, Kahn I, Wagner AD (2007) Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat Neurosci* 10:908–914.
- Stark CEL, Squire LR (2000) Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *J Neurosci* 20:7776–7781.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.



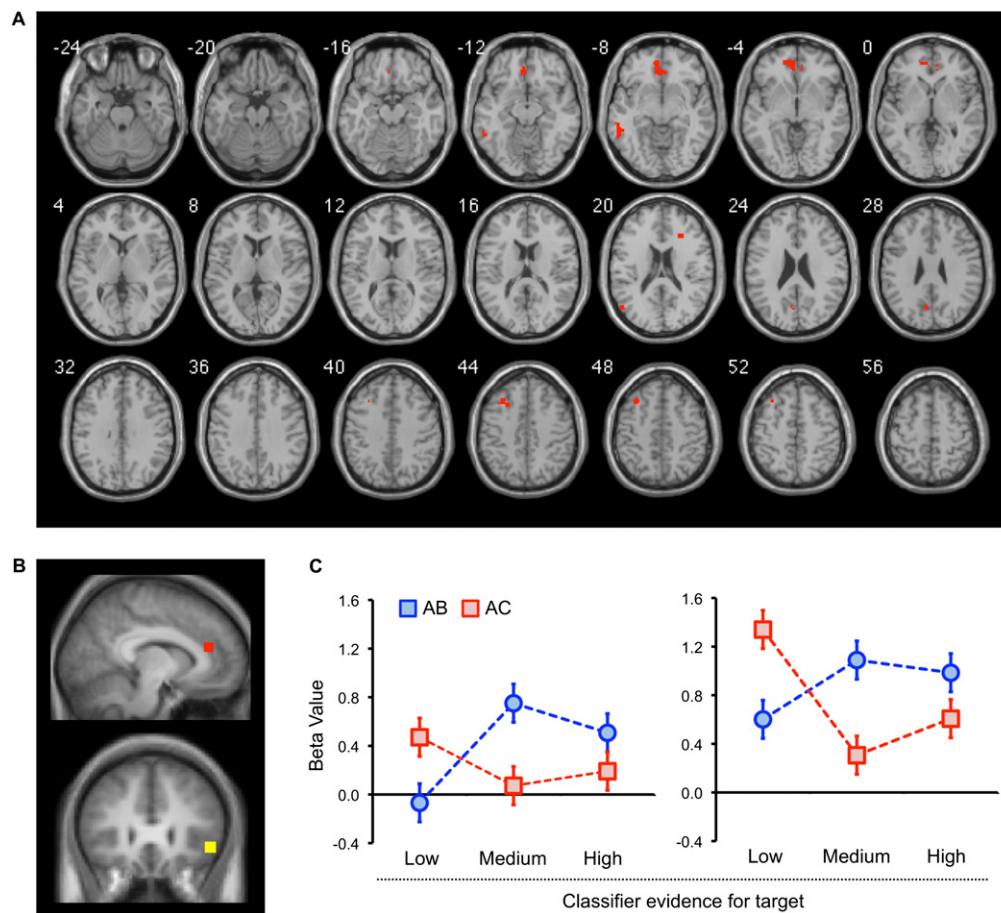


Fig. S3. (A) Results of conjunction analysis described in main test, revealing regions that displayed each of the following criteria: (i) a main effect of retrieval success for DE pairs (specific hit trials > general hit and don't know trials), $P < 0.005$, and (ii) an interaction between degree of classifier-based evidence for VOTC reactivation and level of competition [(high- > low-evidence AB trials) > (high- > low-evidence AC trials)], $P < 0.005$. (B) Anterior cingulate cortex and right inferior frontal gyrus ROIs generated from foci previously implicated in competitive retrieval (10). Anterior cingulate cortex ROI: 6-mm sphere, centered at MNI coordinates $x = -9$, $y = 36$, $z = 18$. Right inferior frontal gyrus ROI: 6-mm sphere centered at $x = 48$, $y = 27$, $z = -6$. (C) Beta values from each ROI revealed an interaction between classifier evidence and competition (anterior cingulate cortex: $F_{2,34} = 6.82$, $P < 0.005$; right inferior frontal gyrus: $F_{2,34} = 5.10$, $P < 0.05$). For each ROI, this interaction was characterized by greater responses for low-evidence AC trials than low-evidence AB trials (anterior cingulate cortex: $t_{17} = 2.45$, $P < 0.05$; right inferior frontal gyrus: $t_{17} = 2.81$, $P < 0.05$). Error bars indicate within-subject SEM.

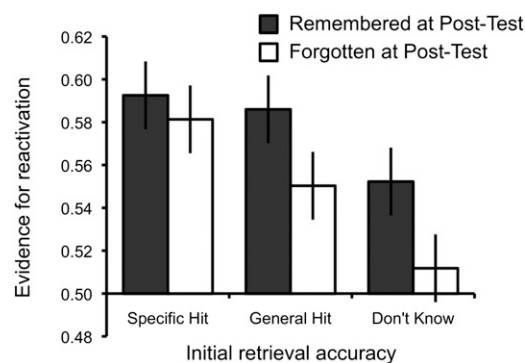
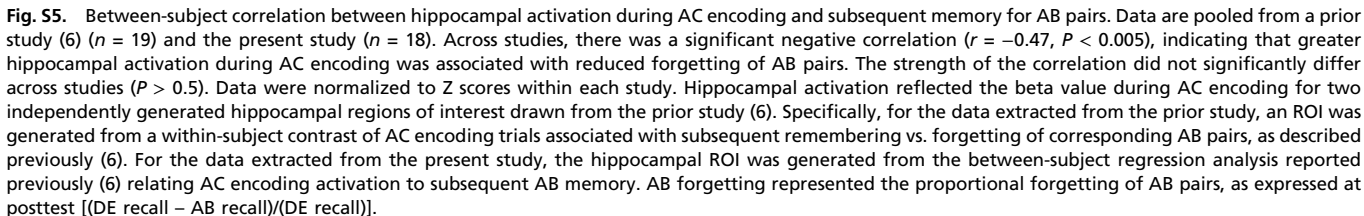


Fig. S4. Fidelity of reactivation for noncompetitive retrieval trials (AB and DE pairs) as a function of initial retrieval success and subsequent memory at posttest. Data are from 13 subjects (five subjects were excluded because there were zero trials in at least one of the relevant bins). A significant main effect of subsequent memory at posttest ($F_{1,12} = 12.73$, $P < 0.005$) indicated that higher-fidelity reactivation of AB/DE pairs during initial retrieval attempts was associated with better subsequent memory for these pairs at posttest. This effect did not interact with initial retrieval success ($F < 1$). Error bars indicate within-subject SEM.



Region	BA	Z	MNI coordinates		
			x	y	z
Hippocampus	–	4.86	–24	–12	–12
Hippocampus	–	4.73	27	–9	–15
Angular gyrus	39	4.58	–45	–66	33
Supramarginal gyrus	40	3.74	–48	–48	33
Hippocampus	–	4.23	36	–33	–9
Cerebellum	–	3.82	30	–69	–24
Cerebellum	–	3.28	42	–54	–27
Middle temporal gyrus	21	3.69	–57	–36	–9
Middle/superior temporal gyrus	21, 22	3.60	–57	–45	0
Middle temporal gyrus	21	3.63	–57	–12	–15
Angular gyrus	39	3.61	48	–72	36
Angular gyrus	39	3.58	39	–63	30
Cerebellum	–	3.60	–30	–36	–33
Postcentral gyrus	43	3.45	–60	–21	21
Postcentral gyrus	43	3.32	–51	–21	21
Cerebellum	–	3.40	0	–63	–12
Cerebellum	–	3.27	0	–72	–6
Middle temporal gyrus	21	3.30	60	–12	–12
Cerebellum	–	3.27	–21	–57	–21
Superior parietal lobule	7	3.27	33	–72	51

Kuhl et al. www.pnas.org/cgi/content/short/1016939108

Table S2. Regions more active for low-evidence AC trials vs. low-evidence AB trials

Region	BA	Z	MNI coordinates		
			<i>x</i>	<i>y</i>	<i>z</i>
Precuneus	31	4.34	−18	−60	24
Precuneus	7	4.30	−12	−63	30
Precuneus	7	3.97	12	−63	30
Posterior cingulate gyrus	31	3.53	9	−48	24
Precuneus	7, 31	3.53	18	−51	30
Posterior cingulate gyrus	31	3.43	−3	−39	27
Middle frontal gyrus	10, 46	4.10	−36	51	9
Middle frontal gyrus	46	3.58	−45	42	18
Insula	13	4.01	39	6	6
Anterior cingulate/medial PFC	32	3.69	6	33	−15
Anterior cingulate/medial PFC	32	3.27	0	39	−9
Precuneus	7	3.56	−6	−60	63
Thalamus	−	3.51	3	−12	0
Fusiform	19	3.46	36	−72	−18
Cingulate gyrus	23	3.38	9	−27	33
Inferior parietal lobule	40	3.35	−45	−54	51
Medial superior frontal gyrus	6, 8	3.32	3	36	42
Cerebellum	−	3.32	21	−78	−18

$P < 0.001$, five-voxel extent threshold. Local maxima within each cluster (>8 mm apart) are indicated by indentation. BA, Brodmann's area.

Table S3. Relationship between DLPFC foci in the present study and a DLPFC region previously implicated in perceptual decision-making

	Z	P_{svc}	MNI coordinates		
			x	y	z
Specific hits > (general hits/don't knows), DE pairs only	3.06	0.017*	-18	27	39
	3.00	0.020*	-30	21	39
(High- > low-evidence AB) > (high- > low- evidence AC)	2.81	0.036*	-30	27	39
	2.72	0.046*	-27	21	42
High evidence > low evidence, AB/DE pairs only	2.80	0.036*	-24	21	42
	2.62	0.054	-27	27	42

A DLPFC ROI was generated, centered at coordinates previously associated with tracking the strength of perceptual evidence (MNI coordinates: $x = -24$, $y = 24$, $z = 36$; 8-mm radius sphere) (1). This ROI was used for small volume correction (SVC) of three contrasts: (i) a contrast of retrieval trials associated with detailed episodic retrieval vs. less-detailed or failed retrieval, (ii) a test for an interaction between pair type (AB vs. AC) and strength of classifier evidence (high vs. low), and (iii) a contrast of noncompetitive retrieval trials associated with high- vs. low-classifier evidence. Each contrast revealed responses within the DLPFC ROI that were significant following SVC, confirming that the DLPFC foci reported here were anatomically consistent with the DLPFC region previously implicated in perceptual decision-making. For each contrast: $P < 0.005$, five-voxel extent threshold. Small volume correction was performed at the voxel level using a family-wise error rate correction. Asterisk (*) denotes foci that were significant following SVC.

1. Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for perceptual decision-making in the human brain. *Nature* 431:859–862.