Cerebral Cortex Advance Access published August 6, 2013

Cerebral Cortex
doi:10.1093/cercor/bht194

Cortical Reinstatement Mediates the Relationship Between Content-Specific Encoding Activity and Subsequent Recollection Decisions

Alan M. Gordon[1], Jesse Rissman[1,4], Roozbeh Kiani[2,5] and Anthony D. Wagner[1,3]

[1]Department of Psychology, [2]Department of Neurobiology, [3]Neurosciences Program, Stanford University, Stanford, CA 94305, USA, [4]Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA and [5]Center for Neural Science, New York University, New York, NY, 10003, USA

Address correspondence to Alan M. Gordon, Stanford Memory Laboratory, Jordan Hall, Building 420, Stanford, CA 94305-2130, USA.
Email: gordonam@stanford.edu

Episodic recollection entails the conscious remembrance of event details associated with previously encountered stimuli. Recollection depends on both the establishment of cortical representations of event features during stimulus encoding and the cortical reinstatement of these representations at retrieval. Here, we used multivoxel pattern analyses of functional magnetic resonance imaging data to examine how cortical and hippocampal activity at encoding and retrieval drive recollective memory decisions. During encoding, words were associated with face or scene source contexts. At retrieval, subjects were cued to recollect the source associate of each presented word. Neurally derived estimates of encoding strength and pattern reinstatement in occipitotemporal cortex were computed for each encoding and retrieval trial, respectively. Analyses demonstrated that (1) cortical encoding strength predicted subsequent memory accuracy and reaction time, (2) encoding strength predicted encoding-phase hippocampal activity, and (3) encoding strength and retrieval-phase hippocampal activity predicted the magnitude of cortical reinstatement. Path analyses further indicated that cortical reinstatement partially mediated both the effect of cortical encoding strength and the effect of retrieval-phase hippocampal activity on subsequent source memory performance. Taken together, these results indicate that memory-guided decisions are driven in part by a pathway leading from hippocampally linked cortical encoding of event attributes to hippocampally linked cortical reinstatement at retrieval.

Keywords: fMRI, hippocampus, path analysis, pattern classification, subsequent memory

Introduction

When confronted with environmental stimuli, we often make decisions that are guided by the recollection of contextual and other associated details surrounding previous encounters with the stimuli. Much neuroimaging research is concerned with the neural mechanisms that support such recollective decisions (Davachi 2006; Diana et al. 2007; Mitchell and Johnson 2009; Danker and Anderson 2010; Rissman and Wagner 2012). In particular, 3 sets of findings concerning the functional neurobiology of recollection are of relevance to the present study. One examines how stronger content-specific cortical representations at encoding are correlated with more accurate decisions at retrieval (e.g., Kirchhoff et al. 2000; Davachi et al. 2003; Uncapher et al. 2006; Otten et al. 2007; Prince et al. 2009; Kuhl, Rissman, et al. 2012), suggesting that recollective decisions are partially dependent on the initial neural representation of associated details or event attributes. Another examines how content-specific cortical representations at encoding are reinstated at retrieval (e.g., Nyberg et al. 2000; Wheeler et al. 2000; Kahn et al. 2004; Polyn et al. 2005; Woodruff et al. 2005; Wheeler et al. 2006), suggesting that recollective decisions are also dependent on the reactivation of cortical patterns that initially represented associated details. A third examines how activity in the hippocampus supports both the encoding (e.g., Davachi et al. 2003; Kirwan and Stark 2004; Ranganath et al. 2004; Uncapher 2005) and retrieval (e.g., Eldridge et al. 2000; Kirwan and Stark 2004; Yonelinas et al. 2005; Montaldi et al. 2006) of associations. Although these 3 phenomena have been independently demonstrated, limited work has examined the relationship between the strength of neural representations at encoding, reinstatement of neural representations at retrieval, and memory performance. In the present study, we used functional magnetic resonance imaging (fMRI) and multivoxel pattern analyses (MVPA) (Haxby et al. 2001; for reviews, see Norman et al. 2006; Rissman and Wagner 2012; Tong and Pratte 2012) to test the hypothesis that cortical reinstatement at retrieval mediates the relationship between the establishment of cortical patterns at encoding and the later recollection of associated source (or contextual) information (operationalized as accurate responding on a 2-alternative forced-choice source memory task).

Numerous fMRI studies have used univariate analytic techniques to examine "subsequent memory effects" (SMEs), wherein the magnitude of encoding-phase blood oxygen level-dependent (BOLD) activity predicts later remembering or forgetting (e.g., Brewer et al. 1998; Wagner et al. 1998; Fernández et al. 1999; Henson et al. 1999; for reviews, see Paller and Wagner 2002; Blumenfeld and Ranganath 2007; Uncapher and Wagner 2009; Kim 2011). Of particular significance to the present experiment are studies that demonstrated content- or source-specific SMEs, wherein activation in regions thought to support the representation of specific classes of stimuli (e.g., Kirchhoff et al. 2000; Otten et al. 2007; Prince et al. 2009) or source features (e.g., Davachi et al. 2003; Uncapher et al. 2006; Staresina et al. 2011) was predictive of content- or source-specific subsequent memory. These SME studies suggest that stronger activity in cortical regions that selectively represent a specific stimulus or source category at encoding can lead to better memory performance at retrieval.

More recently, several neuroimaging studies have employed multivariate analysis strategies to examine how distributed patterns of trial-specific neural activity at encoding relate to later memory outcomes (Jenkins and Ranganath 2010; Xue et al. 2010; Watanabe et al. 2011; Kuhl, Rissman, et al. 2012; Zeithamova et al. 2012; LaRocque et al. 2013; Ritchey et al. forthcoming). Notably, Kuhl, Rissman, et al. (2012) found that,

during the viewing of word-face or word-house pairs, classifier-derived estimates of the strength of face- or house-specific patterns of activity in temporal and prefrontal cortices were higher for pairs for which subjects could subsequently recollect that a face or house was previously associated with the word.

Univariate and multivariate methods have also been used to demonstrate the phenomenon of "cortical reinstatement" (reviewed in Mitchell and Johnson 2009; Danker and Anderson 2010; Schacter et al. 2012), whereby content-specific cortical activity at encoding is reinstated at retrieval (e.g., Nyberg et al. 2000; Wheeler et al. 2000; Kahn et al. 2004; Polyn et al. 2005; Woodruff et al. 2005; Wheeler et al. 2006; Johnson et al. 2009; Kuhl et al. 2010, 2011). Moreover, memory decisions accompanied by subjective reports of conscious recollective experience have been shown to be associated with stronger content-specific cortical activity at retrieval (e.g., Wheeler and Buckner 2004; Johnson and Rugg 2007), and classifier-derived estimates of the fidelity of cortical reinstatement have been observed to relate to retrieval performance (Johnson et al. 2009; McDuff et al. 2009; Kuhl et al. 2011; Kuhl, Rissman, et al. 2012; Staresina et al. 2012; Ritchey et al. in press). Collectively, these studies document a link between the strength or fidelity of cortical reinstatement and behavioral expressions of event recollection.

In addition to a relationship between cortical encoding and retrieval patterns and memory behavior, extensive data implicate the hippocampus as critical for event memory (e.g., Squire 1992; Cohen and Eichenbaum 1993; Squire et al. 2004). Several fMRI studies indicate that hippocampal activity during episodic encoding is greater for trials in which subjects encode associations in which specific event details or associations can be subsequently retrieved (e.g., Davachi et al. 2003; Kirwan and Stark 2004; Ranganath et al. 2004; Staresina et al. 2011). Additionally, hippocampal–cortical connectivity during encoding has been shown to be greater for subsequently retrieved stimuli (Ranganath et al. 2005; Gagnepain et al. 2011; Schott et al. 2011). During retrieval, hippocampal activity is greater both in contrasts of correct versus incorrect associative retrieval (e.g., Eldridge et al. 2000; Cabeza et al. 2001; Dobbins et al. 2003; Chen et al. 2011) and contrasts of recognition with versus without recollection (e.g., Eldridge et al. 2000; Yonelinas et al. 2005; Montaldi et al. 2006; c.f., Wais et al. 2010; Smith et al. 2011). These studies suggest that, in conjunction with cortical representations of event content, the hippocampus is involved in event encoding and retrieval such that subjects can recollect past event or source details when presented a partial cue.

While prior studies have established that (1) content-specific cortical activity at encoding can predict later memory behavior, (2) cortical reinstatement is a neural component of acts of recollection, and (3) hippocampal activity at encoding and retrieval predicts memory expression, no study has tested the critical synthesizing hypothesis that the relationship between content-specific cortical encoding activity and later retrieval of that event content is mediated through the fidelity of cortical reinstatement at retrieval. To this end, the present study sought to bridge subsequent memory and cortical reinstatement effects, by employing MVPA to investigate (1) how content-specific patterns of activity at encoding are related to content-specific patterns at retrieval, and (2) how together these quantities predict behavioral variables related to retrieval decisions. While undergoing fMRI during incidental encoding and subsequent source retrieval, subjects performed a mental imagery-based source memory task. Specifically, during encoding, subjects encountered individual words and imagined either a corresponding face or scene for each word. Following encoding, subjects reencountered the studied words individually and engaged in a source retrieval task, wherein they were required to indicate whether they had performed face or scene imagery during encoding of the word. Because correct performance on this 2-alternative forced-choice task is dependent on successful recollection, correct versus incorrect source judgments were taken to index retrieval with versus without recollection (while acknowledging that some trials may be correct due to guessing). MVPA analyses of data from both the encoding and retrieval phases provided continuous, trial-specific measures of encoding strength and cortical reinstatement.

Using this design and analytic approach, we provide new insights into the neurocognitive processes subserving episodic memory. First, we demonstrate that a neurally derived measure of content-specific cortical encoding strength varies with encoding-phase hippocampal activity and with later source retrieval accuracy and decision time. Secondly, we demonstrate that this neural measure of encoding strength scales positively with cortical reinstatement at retrieval, suggesting that the magnitude of cortical reinstatement at retrieval is dependent in part on the strength of established cortical patterns at encoding. Thirdly, we demonstrate that retrieval-phase hippocampal activity scales with this cortical reinstatement measure. Finally, and critically, and we demonstrate that the relationship of both cortical encoding strength and retrieval-phase hippocampal activity to retrieval accuracy is partially mediated by the effect of cortical reinstatement. Taken together, these results indicate that recollection-dependent memory decisions are driven by a pathway leading from content-specific representational strength at encoding through hippocampally linked content-specific cortical reinstatement.

Materials and Methods

Subjects

Thirty-three right-handed subjects were recruited from the Stanford University community. Written informed consent was obtained in accordance with procedures approved by the institutional review board at Stanford University; subjects received monetary compensation ($20/h). Because many of the correct trials were likely to be guesses for subjects demonstrating poor retrieval performance, the data from 5 subjects were excluded due to poor performance (<65% correct source retrieval; performance range of excluded subjects was 49–62%). Additionally, because this study depends on good estimates of the strength of source representations at encoding, data from one subject were excluded due to poor encoding task performance (mental imagery was subjectively reported as successful on <75% of trials). While arbitrary, these exclusionary criteria were determined before analyses of the questions of interest were conducted. Data from 27 subjects were submitted to all analyses (15 females; mean age = 20.1 years, range = 18–24 years).

Behavioral Procedures

During an incidental encoding phase, subjects were cued to engage in 1 of 2 mental imagery generation tasks for each of a set of 180 adjectives (Fig. 1). At the onset of each 14-s encoding trial, a source/task cue—either "Scene" or "Person"—was centrally presented for 500 ms, followed by 500 ms of a blank screen. Subsequently, an adjective (e.g., "loud") was centrally presented for 750 ms and subjects were given an additional 3250 ms to imagine either a scene or a famous face that was compatible with the adjective, depending on the source task. After imagery, subjects were given 2 s to indicate whether or not they were

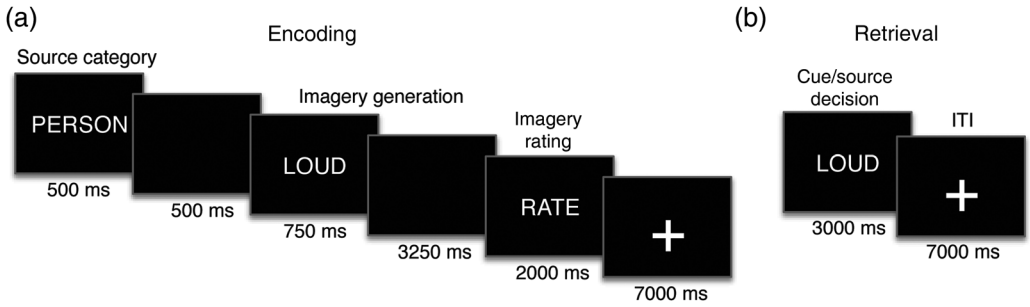


Figure 1. Schematic of behavioral tasks. (*a*) During encoding, subjects were presented with a source/task cue ("Person" or "Scene") followed by an adjective, and instructed to imagine either the face of a famous person or a scene described by the adjective. (*b*) At retrieval, subjects viewed adjectives previously presented at the encoding phase and made a response indicating whether they remembered associating each item with a face or scene.

able to successfully generate any details related to an appropriate scene or face. Subjects made this rating by pressing a button with either their left or right pointer finger. This rating was used to exclude encoding trials on which the subject could not properly perform the task (a mean of 10.3% of face and 5.8% of scene trials were excluded from the analysis). Adjectives were presented over 6 consecutive 7.2-min scanning runs; during each run, 15 adjectives were encoded via face imagery and 15 via scene imagery. To permit single-trial measurement of the observed hemodynamic response, a slow event-related design was implemented by including a 7-s fixation period following offset of the rating period. Condition order and button-press mapping were randomized; source assignment of words was counterbalanced across subjects.

During the retrieval phase, subjects performed a 2-alternative forced-choice source recognition task on all of the words presented during the encoding phase. On each 10-s trial, a studied word was presented for 3 s, and subjects made a button with either their right or left index finger to indicate which imagery task (face or scene) they remembered having previously performed on that word. A 7-s fixation period followed word offset. Across six 4.9-min retrieval scans, subjects made source decisions for 168 adjectives; during each scan, 14 words had been encoded via face imagery and 14 via scene imagery. The remaining 12 studied adjectives, which corresponded to the first and last item of each encoding run, were used in a practice version of the retrieval task that immediately preceded the retrieval scans. Condition order was randomized, and button-press mapping was counterbalanced across subjects. The average lag between a word's study and test presentations was 49.7 min (range = 44.9–58.1 min).

fMRI Acquisition

Functional data were acquired on a 3-T Signa MRI system (GE Medical Systems) using a T_2*-weighted, gradient-echo spiral in–out sequence (repetition time TR = 2 s, echo time = 30 ms, 30 axial slices, 3.3 × 3.3 × 4 mm spatial resolution). Six initial volumes from each run were discarded to allow for T_1 equilibration. High-resolution, T_1-weighted spoiled gradient recalled echo structural images were collected for anatomical visualization. Visual stimuli were projected onto a screen and viewed through a mirror; responses were collected through a magnet-compatible button box.

fMRI Analysis

Data were preprocessed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm, last accessed August 7, 2009). Functional images were corrected for differences in slice acquisition timing, followed by motion correction using sinc interpolation, and coregistration of structural images to each subject's mean functional image. The structural images were then segmented, and gray matter images were spatially normalized to a gray matter template image in Montreal Neurological Institute (MNI) stereotactic space. The resulting nonlinear transformation parameters were applied to all structural and functional images. Functional images were resampled into 3-mm isotropic voxels and smoothed with a 6-mm full-width at half maximum (FWHM) kernel.

Univariate Conjunction Analysis of Encoding/Retrieval Overlap

Separate statistical models were created for the encoding (imagery) and retrieval (source decision) phases, under the assumptions of the general linear model (GLM). The encoding model included event regressors corresponding to face and scene imagery trials for which subjects reported successfully performing the imagery task; imagery failure and no-response trials were coded as a condition of noninterest. The retrieval model included 4 event regressors of interest, corresponding to correct and incorrect face and scene source memory trials for which subjects had reported successful imagery during the encoding phase; retrieval trials for which imagery was unsuccessful at encoding or for which there was no response at encoding or retrieval were coded as a condition of noninterest. Session effects and individual volumes labeled as artifactual by the criteria of having TR-to-TR motion of >0.5 mm or global signal intensity values of >4 SD from the mean were modeled as nuisance covariates. A median 0.9% of encoding-phase volumes and 1.4% of retrieval-phase volumes met these criteria.

One-sample *t*-tests were used to identify voxels for which linear contrasts of GLM parameter estimates reliably differed from zero across subjects. To identify voxels at the group level that differed both during face versus scene encoding/imagery and during correct face versus scene source retrieval decisions, we performed a random-effects conjunction analysis across the encoding and retrieval data ($P < 0.005$, 5-voxel extent for each contrast). To correct for multiple comparisons, Monte Carlo simulations across 10 000 randomly generated data sets were run using AlphaSim (http://afni.nimh.nih.gov, last accessed Aug 13, 2010), using our full-brain anatomical mask to define the search space and assuming the empirical smoothness observed in our data (11.2 × 11.5 × 10.9 mm FWHM). This analysis revealed that a probability of false detection $P < 0.05$ corresponds to a voxel height threshold of $P < 0.005$ coupled with a spatial extent threshold of $k \geq 40$ voxels; these are the statistical and spatial thresholds at which data are reported. For visualization purposes, the conjunction maps are displayed on a mean structural image created from the normalized T_1 images from each of the 27 subjects.

MVPA Approach

Data preprocessing was performed using the Princeton MVPA Toolbox (http://www.pni.princeton.edu/mvpa/, last accessed Nov 1, 2012) and custom Matlab (Mathworks, Natick, MA, USA) scripts. Following realignment and spatial normalization, session-specific functional time series from each voxel were linearly detrended. Since we were interested in using MVPA to index the degree to which neural representations of stimulus features at encoding were reinstated at retrieval (i.e., the "fidelity of reinstatement"; Kuhl et al. 2011), data from occipital and temporal cortices, exclusive of the hippocampi, were used to train and test the classifier. The occipitotemporal mask (17 394 voxels) was created by drawing a mask over occipital and temporal cortices of the group-averaged anatomical brain and excluding the hippocampal mask of the Automated Anatomical Labeling toolbox (Tzourio-Mazoyer et al. 2002).

Trial-specific encoding and retrieval patterns in bilateral occipitotemporal cortex were computed by averaging the TRs corresponding

to 6–10 s post-task cue onset for encoding trials, and 4–8 s postprobe onset for retrieval trials. These were the TRs where the amplitudes of the mean hemodynamic response function of regions that demonstrated source-specific activity were greatest (Supplementary Fig. 1). Analysis of variance-based feature selection was employed to select the 1000 voxels within the occipitotemporal cortex for which activity maximally differentiated the classes within the training set of each classification. Prior to classification, activity of each voxel across patterns was z-scored. Classification was performed using a logistic regression classifier with L2 regression regularization (penalization parameter = 0.01 for all classifications), as instantiated in the LIBLINEAR classification library (http://www.csie.ntu.edu.tw/~cjlin/liblinear/, last accessed Dec 17, 2012). Prior to each iteration of classifier training, the data were subsampled to ensure equal numbers of face and scene trials.

Encoding task classification was accomplished by randomly generated 10-fold cross-validation of occipitotemporal data from the encoding phase and, as described below, the value of probabilistic classifier output was further treated as a measure of trial-specific "encoding strength" (i.e., the degree to which the encoding pattern for a trial resembled the pattern associated with face or with scene imagery). Subsequently, reinstatement classification was accomplished by training a new classifier on all encoding-phase data and then testing on all retrieval-phase data; the value of probabilistic classifier output was further treated as a measure of the "fidelity of reinstatement" (i.e., the degree to which the retrieval pattern in occipitotemporal cortex resembled the encoding pattern associated with face or with scene imagery).

Trial-wise classifier accuracy was computed based on the value of probabilistic classifier output (ranging from 0, corresponding to a perfectly certain face classification, to 1, corresponding to a perfectly certain scene classification) relative to the chance value of 0.5. Classifier performance within each subject was calculated independently for faces and scenes; reported performance is the mean across these 2 values. The significance of classifier performance was calculated using permutation testing. First, within each subject, a null distribution of accuracy was generated, by shuffling the labels of the guesses over 10 000 iterations, and calculating the resulting accuracy for each iteration. A difference score was calculated by subtracting the median permutated classifier accuracy score from the observed classifier accuracy score. A t-statistic across subjects was computed by performing a 1-sample t-test of these difference scores across all subjects, against a null hypothesis of zero.

Encoding strength and fidelity of reinstatement were computed by taking the logit (log odds) of the trial-wise probabilistic classifier output for the encoding task and reinstatement classifications, respectively. Encoding strength and reinstatement are signed in the direction of the correct source for a given trial, such that, regardless of whether the trial was a face or scene trial, the evidence is positive when the classifier guessed correctly, and negative when the classifier guessed incorrectly. These variables are thus neutral with respect to which source class (face or scene) was retrieved.

To examine how the continuous measures of encoding strength and reinstatement scaled with decision-relevant behavioral variables—source retrieval accuracy and reaction time (RT)—we conducted mixed linear and logistic regression analyses involving these trial-wise neural measures of mnemonic evidence. Analyses were performed using mixed-effect GLMs as implemented by the lme4 statistical package (http://cran.r-project.org/web/packages/lme4/, last accessed Mar 2, 2012), in the R statistical environment (http://www.r-project.org/, last accessed Mar 2, 2012). A random intercept term modeling the mean subject-specific outcome value, along with a random slope term modeling the subject-specific effect of the independent variable of interest on the outcome, was included in each model. The approach of specifying random slopes and intercepts within mixed models allows for analysis of trial-wise variables across subjects, while also accounting for (1) subject-specific baselines in the dependent measure, and (2) subject-specific linear relationships between the independent variable of interest and the dependent variable.

Statistical models included relevant fixed-effect independent variables of noninterest. To control for the effect of retrieval accuracy status (correct or incorrect), which was moderately correlated with both encoding (mean $r = 0.08$) and reinstatement strength (mean $r = 0.18$), retrieval accuracy was included as a regressor of noninterest in all models, with the exception of models in which (1) only correct or incorrect trials were analyzed, or (2) retrieval accuracy was itself the dependent variable. We thus ensured that effects of interest could not only be attributed just to a trial being correct or incorrect, but rather could be attributed to the continuous estimates of encoding and reinstatement strength after controlling for retrieval accuracy. To control for the effect of stimulus content (face or scene) on various analyses of interest, stimulus category was included in all models as a regressor of noninterest. To test whether the effects of interest differed across face and scene categories, all reported mixed-effects analyses were repeated, testing for the significance of the relationship between the dependent variable and an independent variable modeling the interaction between the variable of interest and stimulus category. Regression equations are explicitly given in Supplementary Materials. The significance of effects within mixed-model regressions were obtained using log-likelihood ratio tests, resulting in χ^2 values and corresponding P-values.

In all reported statistical tests, classifier-derived variables were treated as continuous. For the purpose of visualization of data across subjects, however, when classifier-derived measures of encoding and reinstatement strength were used as independent variables, data within each subject were binned into quintiles based on the value of the independent variable of interest. The values of the dependent variable within each quintile were averaged across subjects. Error bars on these plots indicate ±within-subject error (Loftus and Masson 1994). Additionally, for each linear and logistic regression plot, subject-specific scatter and logistic histogram plots are included in the Supplementary Results (Supplementary Figs 3–10).

Mean Classifier Beta Map

To visualize which voxels in the occipitotemporal cortex most consistently drive a classifier to "face" or "scene" outcomes, a mean classifier beta map was computed. After the classifier was trained on the full set of encoding data, classifier beta values were assigned to each of the 1000 voxels used in the classifier. The mean beta maps across all subjects were then computed. For visualization purposes, these maps were arbitrarily thresholded at $|\beta| > 2$. Mean beta values that drove classification toward face outcomes were greater in magnitude in the positive direction; values that drove classification toward scene outcomes were greater in magnitude in the negative direction. Mean beta values that were small in magnitude were assigned to voxels that were less informative to classification, and voxels that were more frequently excluded from classification due to feature selection. For display, maps were projected onto inflated fiducial brains with the use of multifiducial mapping in Caret (http://brainvis.wustl.edu/, last accessed Jan 5, 2011); the extent of activation corresponds to the mean extent observed within template subjects from the PALS atlas.

Hippocampal Signal Analysis

To perform analyses linking hippocampal activity at encoding and retrieval to trial-wise encoding strength and cortical reinstatement measures, we extracted trial-wise hippocampal activity for each encoding and retrieval trial. Left and right hippocampal regions of interest (ROIs) were anatomically defined with the Anatomical Automatic Labeling toolbox for SPM (Tzourio-Mazoyer et al. 2002). Left and right hippocampal signal intensity values for each encoding and retrieval trial were extracted from each trial-wise pattern used in the MVPA approach described above. To remove effects correlated with the global signal, a linear regression of raw hippocampal signal against mean signal in the entire occipitotemporal lobe mask was run within each ROI, within each subject. The residuals from this analysis, representing the hippocampal activity controlling for the global mean signal, were used in all hippocampal analyses (note: analyses of hippocampal signal performed without this global signal residualization yielded results similar to that with the residualized hippocampal signal; see Supplementary Results).

Because left and right hippocampi were not hypothesized to have divergent functions, and because activity in these ROIs was highly

correlated (mean $r = 0.78$ at encoding and 0.77 at retrieval), analyses combined activity across the hippocampal ROIs. This bilateral hippocampal signal was computed by first z-scoring each individual ROI within each subject, and then computing the mean across ROIs.

Multiple Comparisons Correction for Regression-Derived Effects of Interest

Regressions linking neurally derived variables were run among 6 variables of interest: Cortical encoding strength, encoding-phase hippocampal activity, cortical reinstatement, retrieval-phase hippocampal activity, RT, and accuracy. A priori hypotheses specified relationships among these variables (see Path Analysis). Nevertheless, because these variables are densely interconnected, any pair of these variables might have been related to each other, either directly or indirectly (through intermediate variables or through variables that relate independently to both members of the pair), so a significant relationship among any pair of variables would be of interest. Because of this, our hypothesis space spanned comparisons over all pair-wise relationships between 2 neurally derived variables, or between a neurally derived variable and a behavioral variable. Since there are 14 such pair-wise relationships, we performed multiple comparisons correction accounting for 14 comparisons within this family of variables (Supplementary Table 1), using false discovery rate (Benjamini and Hochberg 1995).

Path Analysis

To examine how our neural and behavioral variables at encoding and retrieval interrelated in a single statistical framework, we performed a path analysis. A generalization of regression methods to account for sequential effects, path analysis can be used to determine the strength of both "direct effects" between variables and "indirect effects" in which the relationship between variables is mediated by their relationship with intermediate variables.

Separate path structures were created for the dependent variable of memory accuracy (for which all trials were included; reported in Results section) and for that of retrieval RT (for which only correct trials were included; reported in Supplementary Results). The path structure was created by specifying a unidirectional path between each pair of variables shown to significantly covary with each other. Note that the directions of these path arrows are not meant to convey a strong causal claim about the relationship among variables. Rather, as with any regression analysis, path analysis requires one variable to be the dependent variable, and another set of variables to be the independent variables. Influenced by the sequence of events through time and by previous theoretical and empirical research, the direction of path arrows mirrored the directionality of all presented regression analyses, which were chosen based on the following principles: (1) encoding-phase variables preceded retrieval-phase variables; (2) neural variables at retrieval preceded behavioral output variables, given the hypothesis that hippocampal and cortical activity lead to response implementation; (3) cortical encoding strength preceded encoding-phase hippocampal activity, a choice made to fit within the theoretical framework in which, during perception, information propagates from cortex to hippocampus (Squire and Zola-Morgan 1991; Rolls 1996; Mishkin et al. 1997, 1998; Eichenbaum 2000); and (4) retrieval-phase hippocampal activity preceded cortical reinstatement, a choice made to fit with the theoretical framework that pattern completion in the medial temporal lobes leads to reactivation of associate representations in the cortex (Marr 1971; McClelland et al. 1995; McClelland and Goddard 1996; Rolls and Treves 1998; Wallenstein et al. 1998).

Prior to path estimation, all variables were z-scored, except for memory accuracy, which was treated as a binary categorical variable. Path coefficients are thus standardized and estimate the standard deviation change in the outcome variable per standard deviation change in the predictor variable, with the exception of path coefficients leading to memory accuracy, which reflect the change in the logit of memory accuracy per standard deviation change in the given predictor variable.

The goodness of fit of the path structure was computed using a directional-separation test (Pearl and Verma 1987; Shipley 2000) generalized for mixed-effects models (Shipley 2009). For each of k pairs of nondirectly connected variables X_i and X_j, the probability that X_i and X_j are independent, conditional on the set of variables with paths directly leading to either member of the pair, was calculated using a mixed-effects regression. These k probabilities were then combined with Fisher's combined probability test and compared with a χ^2 distribution with $2k$ degrees of freedom. A resulting probability value greater than an alpha of 0.1 signifies that we may retain the model. Optimality of the model was assessed by testing whether the deletion of any path present in the model resulted in a significantly worse model fit, and whether the addition of any path not present in the model resulted in a significantly better model fit. Comparison of fits between each set of 2 nested models was performed by assessing the significance of the difference in the nested model fits using a χ^2 distribution.

Path coefficients were estimated by decomposing the path structure into a set of regressions in which each variable was predicted by the set of all variables with a path leading toward it. As with the mixed-model regressions discussed earlier, regressions used in path analysis included random slopes and random intercepts of interest, and included stimulus category and, where appropriate, memory accuracy, as covariates of no interest.

The coefficient of the indirect path across direct paths A and B was computed as $a \times b$, where a and b represent the direct effects for each corresponding path. The significance of this indirect effect was calculated with bootstrapping methods. A null distribution of the indirect effect $a \times b$ was calculated across 10 000 iterations of data sampled with replacement. Reported P-values constitute the proportion of this distribution greater than the null value of 0. The "index of mediation" $[a \times b \times (\sigma_X/\sigma_Y)]$ is also reported (Preacher and Kelley 2011), providing a measure of effect size for the indirect effects.

Results

Behavioral Results

Subjects correctly remembered the source on 79.3% (SD = 8.77%) of retrieval trials. Mean retrieval decision RTs were faster on correct (mean = 1659 ms; SD = 174 ms) than on incorrect (mean = 1943 ms; SD = 234 ms) trials [$t_{(26)} = 7.22$, $P < 10^{-4}$]. Accuracy did not differ across person versus scene trials [$t_{(26)} = 0.14$, $P > 0.5$], but correct-trial RT was significantly faster for person (mean = 1622 ms, SD = 181 ms) versus scene (mean = 1712 ms, SD = 188 ms) imagery trials [$t_{(26)} = 3.01$, $P < 0.01$].

fMRI-Derived Trial-Wise Quantification of Encoding Strength

We initially sought to quantify the strength of source encoding patterns from BOLD data in the bilateral occipitotemporal cortex on a trial-by-trial basis. To do so, we first conducted a 10-fold cross-validated multivariate pattern analysis on the fMRI data from the occipitotemporal cortex during encoding trials for which subjects reported correctly performing the source imagery task. Mean classifier accuracy in decoding the encoding task from occipitotemporal cortex (81.1%) was significantly above chance [$t_{(26)} = 18.4$, $P < 10^{-4}$]. Subsequently, we computed a continuous metric of cortical "source encoding strength" for each encoding trial, taking the log odds of the classifier's probabilistic estimate of the encoding source associated with each classification attempt. This procedure provided an estimate of the extent to which each encoding pattern resembled encoding patterns of one versus the other class, as such it can be construed as a measure of source-specific activity during encoding. Logistic regression revealed that a greater magnitude of encoding strength predicted a higher likelihood of classifier accuracy (Eq. 1, $\chi^2(1) = 65.3$, $P < 10^{-4}$, Supplementary

Fig. 2), providing validation that our continuous measure of encoding strength is tightly coupled with the extent to which a classifier can use the neural information to correctly determine the imagery task being performed at encoding. Across subjects, the occipitotemporal cortical features that significantly drove classification (1) toward faces included voxels in fusiform gyrus, and (2) toward scenes included voxels in parahippocampal and superior occipital gyri (Fig. 2*a*).

The multivariate approach employed in the present data set is an extension of univariate approaches in which mean activity is extracted across ROIs found to code for one versus other sources. To ascertain the extent to which our measure of encoding strength was driven by such univariate amplitude differences (vs. patterns), a univariate analysis was run on the encoding-phase voxels submitted to classification (Supplementary Methods). For each classifier iteration, sets of face- and scene-responsive voxels were defined from the training data, and trial-wise mean signal amplitude within these regions were extracted from the testing data. Analyses revealed that the measure of encoding strength scaled positively with univariate amplitude in voxels responsive to the correct class (Eq. S3, $\chi^2(1) = 57.8$, $P < 10^{-4}$) and negatively with univariate amplitude in voxels responsive to the incorrect class (Eq. S3, $\chi^2(1) = 56.3$, $P < 10^{-4}$). Taken together, univariate amplitude effects accounted for approximately 63% of the variance in our measure of encoding strength. Similar effects were found for the measure of cortical reinstatement, for which univariate amplitude effects accounted for 47% of the variance (Supplementary Results). These results indicate that the multivariate measures of encoding strength and cortical reinstatement are highly related to coarse-scale univariate amplitude effects. However, the remaining variance in the multivariate measures that is not explained by the univariate measures may be due to contributions to this multivariate signal from more fine-scale patterns. Finally, an exploratory analysis was conducted to determine whether source-specific activity patterns were also present in the hippocampus at encoding, using trial-wise data from the hippocampus. Classification of source within the hippocampus was significantly above chance [mean accuracy = 62.3%; $t_{(26)} = 7.61$, $P < 10^{-4}$, Supplementary Results].

Encoding Strength Predicts Subsequent Recollection Decisions

Motivated by recent data indicating that encoding patterns elicited by the perception of faces versus scenes predict later cued recall (Kuhl, Bainbridge, et al. 2012), we next sought to determine whether the strength of encoding patterns in the bilateral occipitotemporal cortex during face versus scene imagery was predictive of later source recollection decisions and decision latencies. Consistent with this possibility, logistic regression revealed that stronger occipitotemporal encoding patterns, as indexed by the classifier, predicted a higher likelihood of accurate retrieval (Eq. 2, $\chi^2(1) = 20.3$, $P < 10^{-4}$, Fig. 2*b*).

For encoding trials for which subjects subsequently recollected the correct face/scene source associate, greater encoding strength also predicted faster retrieval RTs (Eq. 3, $\chi^2(1) = 5.23$, $P < 0.05$, Fig. 2*c*). In contrast, for subsequently incorrect decisions, encoding strength did not predict retrieval RTs (Eq. 3, $\chi^2(1) = 1.60$, $P = 0.21$). An encoding strength by retrieval accuracy interaction confirmed that the predictive relationship between encoding strength and retrieval RT differed across subsequently correct versus incorrect recollection decisions (Eq. 4, $\chi^2(1) = 10.8$, $P < 0.005$). Collectively, these results indicate that stronger, source-specific encoding patterns in the bilateral occipitotemporal cortex during visual imagery predict more accurate and faster source memory decisions at retrieval.

Encoding Strength Predicts Hippocampal Activity at Encoding

Motivated by prior observations that greater hippocampal encoding activity (i.e., BOLD amplitude) predicts later recollection accuracy (e.g., Davachi et al. 2003; Ranganath et al. 2004), we investigated whether encoding-phase hippocampal activity predicted subsequent memory accuracy. In contrast to these prior studies, which used recognition tests that discriminated between recollected versus familiar versus forgotten trials, hippocampal activity at encoding did not significantly predict subsequent memory performance on our 2-alternative forced-choice retrieval task ($P > 0.1$), a finding that may reflect the presence of guesses

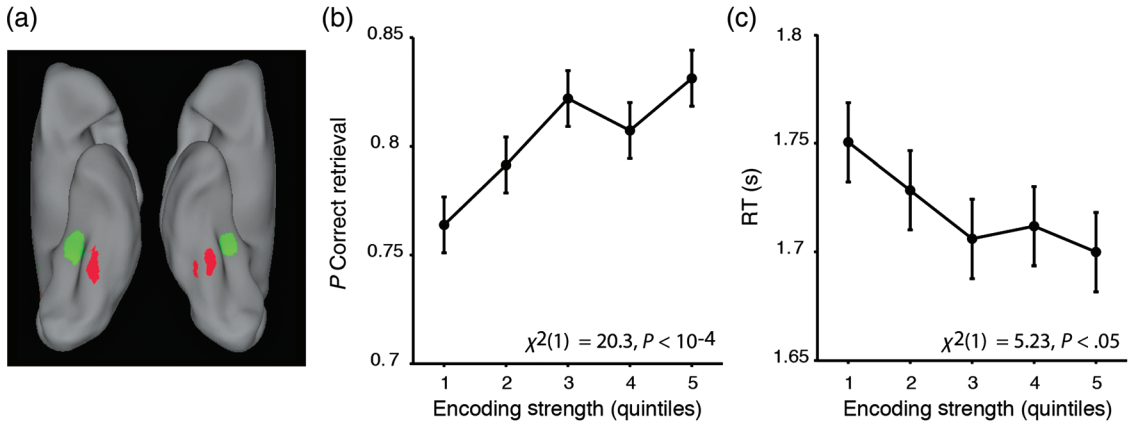


Figure 2. Patterns of BOLD response in the occipitotemporal cortex during face and scene imagery can be used to generate a metric of encoding strength that is related to later source retrieval performance. (*a*) Mean classification beta map, thresholded at $|\beta| > 2$. Red voxels more consistently drive classifier output toward "face" classification, and green voxels more consistently drive classifier output towards "scene" classification. (*b*) Plot of the probability of correct source retrieval as a function of encoding strength. (*c*) Plot of encoding strength against RT for correct retrieval trials. Error bars indicate ±within-subject SEM.

in the source correct condition of our study. We next examined whether encoding-phase hippocampal activity scaled with our trial-wise measure of cortical encoding strength. Indeed, cortical encoding strength scaled positively with mean hippocampal encoding activity (Eq. 5, $\chi^2(1) = 17.5$, $P < 10^{-4}$, Fig. 3). This result indicates that greater strength of cortical representation at encoding scales with greater concomitant hippocampal activity, suggesting a role for hippocampus in the representation and/or binding of source features.

Cortical Reinstatement—Univariate Analyses

Having established that source encoding patterns predict subsequent memory accuracy and decision RTs, we next tested whether source-specific activity at encoding was reinstated at retrieval, using univariate GLM analyses of the fMRI data at encoding and retrieval. These analyses confirmed 2 effects: (1) The face and scene imagery tasks were associated with differential cortical activity at encoding, and (2) source-specific regions active at encoding were reactivated during correct source memory decisions at retrieval. Specifically, a conjunction analysis of face versus scene activity during encoding and retrieval ($P < 0.005$ for each contrast; corrected $P < 0.05$) revealed that initial imagery and subsequent retrieval of faces differentially activated multiple regions, including the ventromedial prefrontal cortex, whereas imagery and retrieval of scenes differentially activated posterior parahippocampal cortex, among other regions (Fig. 4*a*). These results add to a rich literature establishing that recollection is associated with reinstatement of

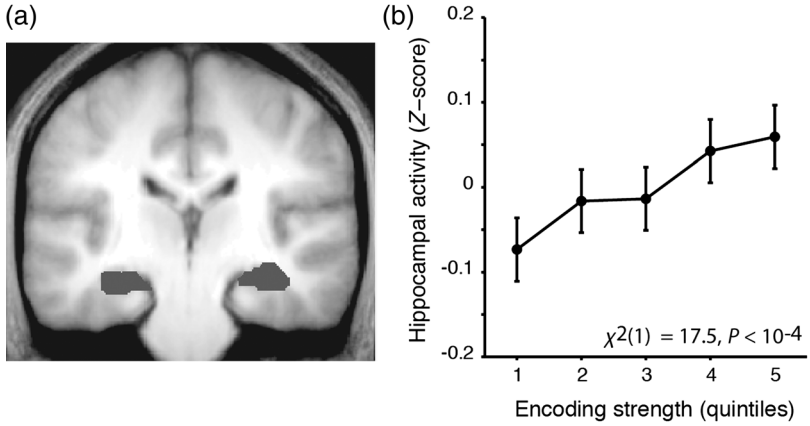


Figure 3. Encoding strength predicts encoding-phase hippocampal activity. (*a*) Location of anatomically defined left and right hippocampal ROIs, overlayed on a mean normalized group anatomical image. (*b*) Plot of hippocampal activity against encoding strength for correct retrieval trials. Error bars indicate ±within-subject SEM.

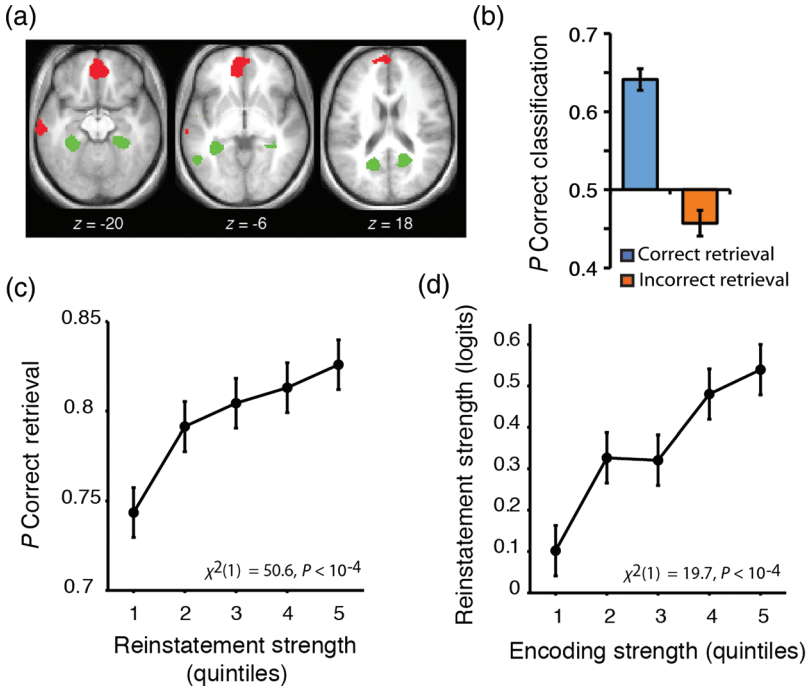


Figure 4. Cortical reinstatement tracks encoding strength and subjective memory decision. (*a*) Conjunction map of source-specific activity at encoding and retrieval, thresholded at $P < 0.005$ for both encoding and retrieval (conjoint $P < 0.05$, corrected for multiple comparisons). Red: face > scene; green: scene > face. Activations are displayed on the mean ($N = 27$ subjects) normalized anatomical image. (*b*) Classification performance for correct retrieval trials (blue) and incorrect retrieval trials (orange). Error bars indicate standard error of the mean. (*c*) Plot of the probability of correct source retrieval as a function of cortical reinstatement. (*d*) Plot of cortical reinstatement by encoding strength. Error bars indicate within-subject SEM.

activation in regions that demonstrated content-specific responses during initial learning (e.g., Nyberg et al. 2000; Wheeler et al. 2000; Kahn et al. 2004; Johnson et al. 2009; for a review, see Danker and Anderson 2010).

Cortical Reinstatement Predicts Retrieval Decision Variables—Multivariate Analyses

Multivariate methods can extend univariate reinstatement results by providing a quantitative estimate of cortical reinstatement on a trial-wise basis (Polyn et al. 2005; Johnson et al. 2009; McDuff et al. 2009; Kuhl et al. 2011; Staresina et al. 2012; Ritchey et al. forthcoming). One question of interest is how multivariate estimates of encoding pattern reinstatement relate to memory decisions. Accordingly, to generate trial-specific estimates of cortical reinstatement on each retrieval trial, a classifier was trained to dissociate face- and scene encoding trials and tested with the independent data from the retrieval phase. This procedure provides an estimate of the extent to which retrieval patterns resemble encoding patterns of one versus the other class, as such it can be construed as a trial-wise measure of the reinstatement of source-specific activity. We then asked whether the classifier's probabilistic prediction of the retrieved source, based on the obtained retrieval pattern, tracked the subject's encoding experience or the subject's source memory decision at retrieval.

Our cortical reinstatement measure strongly correlated with retrieval accuracy (Eq. 6, $\chi^2(1) = 50.6$, $P < 10^{-4}$, Fig. 4*c*), such that greater cortical reinstatement predicted a greater likelihood of correct retrieval. For trials on which subjects made a correct source decision, the classifier predicted the face/scene encoding source with a mean accuracy of 64.1%, which is significantly above chance [$t_{(26)} = 10.1$, $P < 10^{-4}$, Fig. 4*b*]. In contrast, for trials on which subjects made an incorrect source decision, the classifier performed modestly, but significantly below chance [mean accuracy = 45.7%, $t_{(26)} = -2.47$, $P < 0.05$, Fig. 4*b*], suggesting that retrieval errors may be driven in part by the activation of occipitotemporal cortical patterns tied to the erroneous source context. We also found that greater cortical reinstatement correlated with faster decision RTs, but with marginal significance ($P < 0.1$) and with differential contributions from face and scene trials (Supplementary Results).

Finally, an exploratory analysis was conducted to investigate whether reinstated source-specific patterns of activity were also present in the hippocampus. The classification rate of reinstated face/scene information in the hippocampus was modestly but significantly higher than chance [mean accuracy = 54.3%; $t_{(26)} = 2.96$, $P < 0.01$, Supplementary Results]. Taken together, these data demonstrate that reinstatement tracked subjective memory decisions, even when the decisions were incorrect (e.g., Kahn et al. 2004).

The Strength of Encoding Patterns Predicts the Subsequent Fidelity of Cortical Reinstatement at Retrieval

To the extent that the fidelity of a retrieved representation is dependent on the strength of the initially encoded representation, then the fidelity of cortical reinstatement should be predicted by the strength of source-specific patterns at encoding. Consistent with this hypothesis, a mixed-effects linear regression of encoding strength on cortical reinstatement demonstrated that events that gave rise to greater encoding strength were associated with a higher fidelity of cortical reinstatement at retrieval (Eq. 7, $\chi^2(1) = 19.7$, $P < 10^{-4}$, Fig. 4*d*). Critically, these data constitute novel evidence that future states of cortical brain patterns (1) are partially dependent on the strength of the cortical patterns established in the past and (2) have consequences for memory expression.

Hippocampal Activity at Retrieval Scales with Retrieval Accuracy and RT

Previous work has demonstrated that, relative to incorrect source decisions and correct rejections, correct recollection decisions are associated with greater hippocampal activity, implicating hippocampal computations in the recollection of past event details (e.g., Eldridge et al. 2000; Cabeza et al. 2001; Dobbins et al. 2003). Extending these prior categorical observations, mixed-effects linear regression analyses (Eqs 8 and 9) demonstrated that greater hippocampal retrieval activity was associated with (1) a higher likelihood of correct source recollection (Fig. 5*a*; $\chi^2(1) = 14.4$, $P < 0.0005$), and (2) faster retrieval decision RTs for correct trials (Fig. 5*b*; $\chi^2(1) = 13.7$, $P < 0.0005$). Taken together, these results demonstrate that

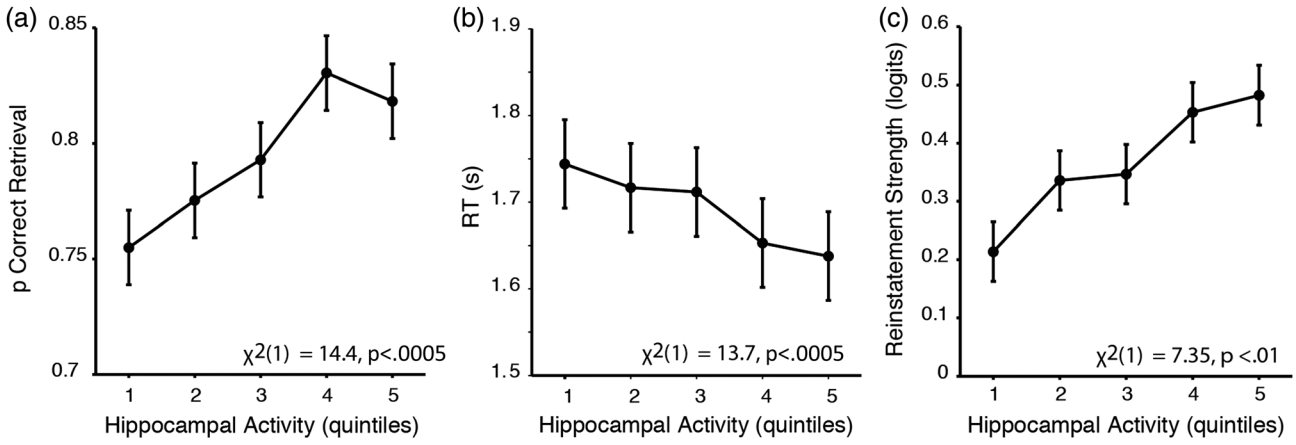


Figure 5. Hippocampal activity at retrieval tracks performance and strength of cortical reinstatement. (*a*) Plot of the probability of correct source retrieval as a function of hippocampal activity. (*b*) Plot of RT for correct trials as a function of hippocampal activity. (*c*) Plot of cortical reinstatement strength as a function of hippocampal activity. Error bars indicate within-subject SEM.

greater hippocampal activity at retrieval is associated with faster and more accurate source retrieval decisions.

The preceding analyses indicate that hippocampal retrieval activity varies with recollection accuracy and decision RT. Given that (1) hippocampal activity during encoding is thought to scale with greater binding of event details, (2) hippocampal activity at retrieval is thought to scale with greater recollection of the event details necessary for correct source decisions, and (3) stronger binding of details at encoding is thought to lead to stronger recollection of details at retrieval, we hypothesized that greater encoding-phase hippocampal activity would be associated with greater retrieval-phase hippocampal activity. Consistent with this hypothesis, a mixed-effects linear regression revealed that events associated with greater hippocampal activity at encoding were also associated with greater hippocampal activity at retrieval (Eq. 10, $\chi^2(1) = 4.94$, $P < 0.05$).

Hippocampal Activity at Retrieval Scales with Cortical Reinstatement

Recollection is thought to partially depend on hippocampal retrieval computations (pattern completion) that serve to reinstate source-specific cortical patterns (Marr 1971; McClelland et al. 1995; McClelland and Goddard 1996; Rolls and Treves 1998; Wallenstein et al. 1998). Supporting this hypothesis, we found that trial-wise retrieval activity in the hippocampus (Eq. 11, $\chi^2(1) = 7.35$, $P < 0.01$, Fig. 5*c*) predicted the classifier-measured strength of cortical reinstatement. Overall, these results provide evidence that retrieval events accompanied by greater hippocampal activity are associated with greater cortical reinstatement of content-specific representations observed at encoding.

Effects of Stimulus Content

To determine whether each reported effect of interest was differentially driven by the face or scene stimulus category, we tested for an interaction between the effect of interest and stimulus category within each analysis. Specifically, within each mixed linear model (Eqs 1–11; Supplementary Table 1), we tested the significance of the relationship between the dependent variable and an independent variable modeling the interaction of the effect of interest and stimulus category. Of these analyses, one stimulus category interaction proved significant: The relationship between activity in the hippocampus and reinstatement strength was significantly greater for face versus scene trials ($\chi^2(1) = 3.87$, $P < 0.05$). Besides this result, no other predictor–class interaction was found to be significant ($P > 0.1$, for all tested interactions) for any effect of interest. These analyses thus demonstrate that, for 8 of 9 reported regressions of interest involving measures of hippocampal activity and distributed cortical source representations, there is no evidence for a difference in these relationships across face and scene trials.

Multiple Comparisons Correction Across Regressions of Interest

To account for multiple comparisons among the regressions of interest, false discovery rate analysis (Benjamini and Hochberg 1995) was performed. An $N = 14$ was used, encompassing both the 8 regressions of interest reported above and the remaining 6 pair-wise comparisons (reported in Supplementary Table 1) that met the criteria of being a pair-wise comparison between variables of interest, and containing at least one neural variable. All above reported effects of interest (Eqs 1–11) have a q-value of <0.05, indicating that the expected rate of false discovery, or incorrectly rejecting the null hypothesis, across this family of reported effects is 5% (Supplementary Table 1). All nonreported effects of interest have a q-value of >0.05. Thus, all regression effects reported above are significant when accounting for multiple comparisons, whereas all other pair-wise comparisons among these variables do not survive correction.

Synthesizing Regressions of Interest into a Path Analysis

As a final analysis approach, we conducted path analyses that combined all analyses of interest into a single sequential statistical model, which allowed us to test for indirect effects (i.e., examining whether an independent variable predicts a dependent variable through its action on an intermediate variable). In a first path model, the variables of "cortical encoding strength," "cortical reinstatement," "hippocampal encoding activity," "hippocampal retrieval activity," and "memory accuracy" were included. The path structure was created (Fig. 6*a*)

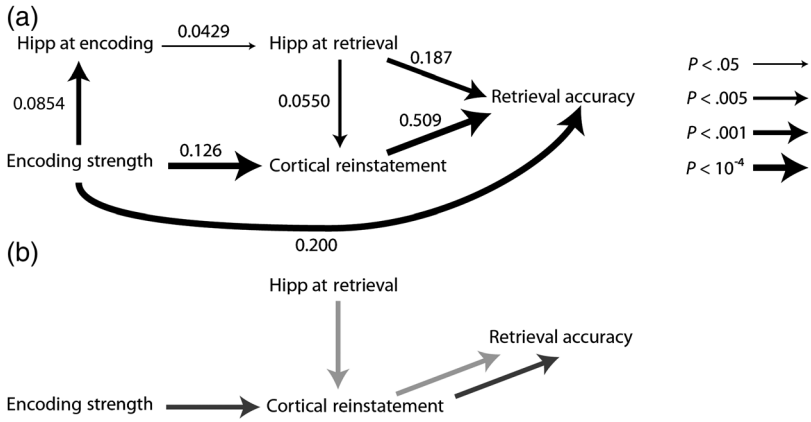


Figure 6. Path analysis reveals how pathways of neural activity at encoding and retrieval lead to source retrieval. (*a*) Path analysis relating neurally derived mnemonic variables and retrieval accuracy. Numeric labels indicate standardized path coefficients. Path thickness indicates the statistical significance of each given direct effect. (*b*) The indirect path from encoding strength to cortical reinstatement to retrieval accuracy (dark gray) and the indirect path from retrieval-phase hippocampal activity to cortical reinstatement to retrieval accuracy (light gray).

by specifying paths between each pair of variables for which a significantly predictive relationship was found in the preceding analyses (see above). A directional-separation test revealed that we could retain this model as a reasonable fit to the data (Eqs 12a–c, $\chi^2(6) = 8.26$, $P = 0.22$). Further analysis revealed that this model had a locally optimal fit; adding any path that was not already present in the structure did not significantly improve the model fit ($\chi^2(2) < 3.09$, $P > 0.13$ for each added path), while subtracting any link that was present in the structure significantly decreased the model fit ($\chi^2(2) > 8.34$, $P < 0.05$). All direct relationships between variables remained significant in the path model (Eqs 12a–d, $P < 0.05$ for each path; Fig. 6), indicating that each reported effect of interest was significant even when controlling for other reported effects of interest. A second path model encompassing only correct retrieval trials was created to model all variables from the first model with the exception of memory accuracy and the addition of RT (Supplementary Results).

Cortical Reinstatement Partially Mediates the Effects of Encoding Strength on Retrieval Accuracy

Importantly, path analysis allowed us to examine indirect (mediated) pathways of interest. First, we examined whether the effect of encoding-phase variables on retrieval behavior was mediated in part by retrieval-phase behavioral variables. Our path analyses contained 2 sets of indirect paths that fit this profile: (1) A path from encoding strength to retrieval strength to accuracy, and (2) a path from encoding strength to cortical reinstatement to RT. Indirect effect 1 was significant [$P < 10^{-4}$, $Z = 3.72$; index of mediation = 0.036, 95% confidence interval (CI) = 0.026–0.044, Fig. 6*b*], demonstrating that distributed activity at encoding predicted subsequent memory accuracy in part through its effect on distributed activity at retrieval. Indirect path 2 did not reveal a significant indirect pathway from encoding strength to cortical reinstatement to RT ($P = 0.19$, $Z = 1.32$; Supplementary Results). Taken together, these analyses demonstrate that the effect of encoding strength on retrieval accuracy is partially mediated by the variable of cortical reinstatement.

Cortical Reinstatement Partially Mediates the Effects of Retrieval-Phase Hippocampal Activity on Retrieval

A second set of path analyses sought to examine how hippocampally supported pattern completion may support memory retrieval in part through the intermediate effects on cortical representations. Specifically, we tested whether the effect of retrieval-phase hippocampal activity on (1) retrieval accuracy and (2) RT was mediated by cortical reinstatement. We found that the indirect path predicting accuracy was significant ($P < 10^{-4}$, $Z = 3.72$; index of mediation = 0.025, 95% CI = 0.010–0.043, Fig. 6*b*). The indirect path predicting RT was not significant ($P = 0.20$, $Z = 1.28$; Supplementary Results). These analyses demonstrate that retrieval-phase hippocampal activity predicted retrieval accuracy in part through its effects on cortical reinstatement.

Discussion

The present study demonstrated 5 key findings. First, a neurally derived measure of trial-wise content-specific "encoding strength" in the occipitotemporal cortex predicted (1) subsequent accuracy in remembering whether a face or scene had been imagined at encoding, and (2) subsequent retrieval decision latency. Secondly, cortical encoding strength scaled with encoding-phase hippocampal activity. Thirdly, cortical encoding strength and retrieval-phase hippocampal activity predicted the strength of cortical reinstatement at retrieval. Fourthly, the magnitude of cortical reinstatement predicted retrieval accuracy and RT. Finally, the effects of cortical encoding strength and retrieval-phase hippocampal activity on retrieval accuracy were partially mediated by the effect of cortical reinstatement. Collectively, these results document how the recollection of event details depends on cortical–hippocampal interactions during event encoding and subsequent retrieval.

Content-Specific Cortical Activity at Encoding

Regions of the human occipitotemporal cortex, such as fusiform gyrus and posterior parahippocampal cortex, demonstrate increased BOLD activity during the perception of specific visual categories such as faces and houses (e.g., Kanwisher et al. 1997; Epstein and Kanwisher 1998) and the generation of mental images specific to such categories (e.g., O'Craven and Kanwisher 2000; Davachi et al. 2003). Studies pairing categorical perception with a subsequent memory task demonstrate that content-specific activity during perception is greater for category members that are subsequently remembered relative to those subsequently forgotten (e.g., Kirchhoff et al. 2000; Davachi et al. 2003). In particular, greater activation in face-responsive regions of fusiform gyrus during face encoding is predictive of greater subsequent memory for faces (Sergerie et al. 2005; Nichols et al. 2006; Prince et al. 2009), whereas greater activation in scene-responsive regions of the parahippocampal cortex during scene encoding is predictive of subsequent memory for scenes (Brewer et al. 1998; Kirchhoff et al. 2000; Turk-Browne et al. 2006; Hayes et al. 2007; Awipi and Davachi 2008; Prince et al. 2009; Preston et al. 2010).

While nearly all prior studies of content-specific SMEs utilized univariate measures of encoding activity, the present study exploited multivariate techniques to quantify the "strength" of content-specific occipitotemporal cortical patterns at encoding and to relate this measure of cortical encoding strength to subsequent associative recollection (i.e., source memory). Through this approach, we demonstrated that when mental imagery at encoding is accompanied by stronger content-specific patterns across the occipitotemporal cortex, subjects are more likely to be able to subsequently remember having imagined an exemplar from the respective category (see also Kuhl, Rissman, et al. 2012 for similar findings with a perceptual-based encoding paradigm). Additionally, we observed that greater cortical encoding strength is predictive of faster correct retrieval decisions. As both correct decisions and faster RT have been linked to judgments made with greater decision evidence, these results support the view that greater encoding strength leads to greater source decision evidence. Future studies can further document the decision process by incorporating formal computational decision models into the analysis of the relationship of neural measures of reinstatement to behavioral retrieval variables.

A number of mechanisms may account for why cortical encoding strength predicts subsequent recollection. First, although analyses were restricted to encoding trials for which

subjects reported successful mental imagery, variability in cortical encoding strength may nevertheless reflect differences in prior semantic knowledge that differentially enabled mental imagery given the conceptual cue (i.e., the word). Encoding trials for which a subject had greater semantic knowledge may have resulted in generation of a richer or more vivid mental image (Kozhevnikov et al. 2005) and may have been associated with "deeper" (i.e., more semantically rich) encoding of the word/image association. Both factors would facilitate subsequent retrieval. Secondly, variability in cortical encoding strength may also reflect variability in task-focused attention. Episodic encoding has been shown to decrease in the presence of a competing task (e.g., Craik et al. 1996; Foerde 2006). Trials in which subjects deployed more focused attention on the mental imagery task versus other endogenous goals or thoughts may be associated with a more richly detailed mental image and a greater likelihood of subsequent accurate retrieval. As such, the relationship between cortical encoding strength and subsequent memory may partially reflect the role of attentional processes during encoding (e.g., Craik and Lockhart 1972; Rock and Gutman 1981; Chun and Johnson 2011; Uncapher et al. 2011).

Cortical–Hippocampal Interactions at Encoding

Much research suggests that event encoding depends on hippocampal–cortical interactions. Anatomically, the outputs of neocortical regions, including occipitotemporal cortical areas, project to the perirhinal and parahippocampal cortex in the medial temporal lobe, which project, via entorhinal cortex, to the hippocampus (Suzuki and Amaral 1994; Lavenex and Amaral 2000). The hippocampus is thus a convergence zone for a broad range of cortically represented event features (Squire and Zola-Morgan 1991; Rolls 1996; Mishkin et al. 1997, 1998; Eichenbaum 2000). Computationally, associative encoding of events (episodes) is thought to require the cortical representation of event features, which converge on and are bound in the hippocampus (Marr 1971; Squire 1992; Cohen and Eichenbaum 1993; McClelland et al. 1995; Norman and O'Reilly 2003).

A number of studies indicate that greater encoding-phase functional connectivity between hippocampus and cortical regions involved in perception predicts a higher likelihood of subsequent memory. For example, fMRI data suggest that greater hippocampal–medial occipital cortical connectivity at encoding predicts subsequent recognition memory for previously viewed objects (Ranganath et al. 2005) and subsequent free recall for viewed words (Schott et al. 2011). During the presentation of auditory stimuli, subsequent recollection is associated with greater BOLD signal connectivity between hippocampus and superior temporal gyrus (Gagnepain et al. 2011) and greater theta-phase intracranial electroencephalography (EEG) synchronicity between hippocampal and occipitotemporal regions (Babiloni et al. 2009). Intracranial EEG data also indicate that encoding-period gamma-phase synchronization and theta coherence between hippocampus and perirhinal and/or entorhinal cortex predict subsequent free recall of visually presented words (Fell et al. 2001, 2003).

We observed a positive correlation between trial-to-trial fluctuations in the magnitude of hippocampal univariate activity and our multivariate measure of content-specific cortical encoding strength. Two mechanisms may account for this finding. First, stronger cortical representation of the content-specific information may trigger greater hippocampally supported encoding operations. Specifically, the propagation of stronger perceptual information to the hippocampus may have a higher likelihood of eliciting hippocampal neural responses and ultimately mnemonic binding (i.e., hippocampal encoding). Secondly, greater hippocampal activity during the encoding phase may signify that subjects are engaging in hippocampally supported retrieval processes while performing the mental imagery task. Greater hippocampally supported retrieval may contribute to stronger mental imagery, which, in turn, would promote greater cortical encoding strength. These 2 purported mechanisms are not mutually exclusive; it is possible that stronger content-specific cortical representations are associated with both greater hippocampal encoding and retrieval operations during mental imagery.

Cortical Reinstatement

Cortical reinstatement, the reactivation of content-specific encoding patterns at retrieval, has been observed in a broad array of paradigms (Nyberg et al. 2000; Wheeler et al. 2000; Kahn et al. 2004; Polyn et al. 2005; Woodruff et al. 2005; Wheeler et al. 2006; Johnson et al. 2009; Kuhl et al. 2011; Kuhl, Bainbridge, et al. 2012). The present data revealed that the strength of cortical reinstatement is predictive of retrieval decision times and decision accuracy. As such, our results indicate that the fidelity of cortical reinstatement can affect multiple decision variables at retrieval, which suggests that the reinstated cortical patterns serve as a source of evidence driving the mnemonic decision. Previous research indicates that cortical reinstatement scales with many conditions that are thought to be associated with greater strength of recollection, including contrasts of (1) source recollection versus item recognition (e.g., Kahn et al. 2004), (2) subjective reports of recollection versus familiarity (Wheeler and Buckner 2004; Johnson and Rugg 2007), (3) conditions indexing graded memory strength (Johnson et al. 2009), (4) greater amounts of associated event details (Khader et al. 2005), (5) stronger subjective ratings of vividness of retrieved details (Daselaar et al. 2008), and (6) more specificity of retrieved details (e.g., Kuhl et al. 2011).

Memory-guided decisions are thought to depend on retrieved mnemonic evidence that is evaluated in relation to decision criteria (Ratcliff 1988; Dunn 2004; Wixted and Stretch 2004). The present data revealed that the strength of cortical reinstatement is predictive of retrieval decision times and decision accuracy. As such, our results indicate that the fidelity of cortical reinstatement can affect multiple decision variables at retrieval, which suggests that the reinstated cortical patterns serve as a source of evidence driving the mnemonic decision.

Our data also indicate that source-related cortical activity patterns during retrieval are more closely tied to subjective retrieval experiences (or decisions) than to the objective experiential history associated with a stimulus (Kahn et al. 2004; Slotnick and Schacter 2004; Rissman et al. 2010 for related findings concerning item memory). That is, rather than tracking the true mnemonic history of an item, the pattern classifier's prediction about the subject's memory state tracked the subject's reported decision, even when the decision was incorrect. This observation suggests that retrieval errors are driven in part by the activation of distributed cortical patterns that represent features of the incorrect source or content. This finding complements prior results derived from univariate fMRI analyses, which

demonstrate that erroneous memory for novel items (i.e., false recollection) is sometimes accompanied by reactivation of cortical regions selective for the content of the false memory (e.g., Kahn et al. 2004; Slotnick and Schacter 2004).

While a growing body of evidence links cortical reinstatement to retrieval decision variables, it is important to note that cortical reinstatement effects can arise for reasons not directly tied to the retrieval decision process. For instance, in the present data, it is possible that each retrieval trial is accompanied by subjects engaging in imagery of face and scene information related to the retrieval cue. Subjects who experience a greater familiarity for any imagined face versus scene information may be more prone to respond that a given memory cue was originally associated with a face at retrieval. In this way, it would be possible, in principle, to correctly identify a source associate without engaging in recollection. Given the relatively high level of source retrieval accuracy in this study, we believe that, if present, such a generate-recognize process was likely used in a minority of trials. Nevertheless, future work is needed to dissociate how different forms of task cognition may contribute to "cortical reinstatement" signals.

Hippocampal–Cortical Interactions at Retrieval

Consistent with the notion that hippocampal retrieval processes subserve the recollection of event details, we observed that greater hippocampal activity during retrieval correlated with source memory accuracy and correct retrieval decision time. This accords with an extensive literature demonstrating that hippocampal activity is greater for correct retrieval events (e.g., Eldridge et al. 2000; Cabeza et al. 2001; Dobbins et al. 2003), and extends this relationship to retrieval decision latency.

We also found that retrieval-phase hippocampal activity correlated with the strength of cortical reinstatement. These data complement findings in humans, demonstrating that hippocampal activity is correlated with measures of cortical reinstatement during retrieval (e.g., Kuhl et al. 2010; Staresina et al. 2012; Wimmer and Shohamy 2012; Zeithamova et al. 2012; Ritchey et al. in press). Taken together, our findings indicate that hippocampal BOLD signal amplitude during source recollection covaries with a multivariate measure of cortical reinstatement (a neural measure of retrieval strength), as well as with retrieval accuracy and decision latency (behavioral measures of retrieval strength).

Furthermore, we synthesized the results linking hippocampal activity to reinstatement and to memory performance by creating a path model to relate all neural and behavioral variables for which significant relationships were documented into a single statistical model (Fig. 6). This model is consistent with the perspective that a cascading series of neural responses, including cortical encoding strength, hippocampal univariate encoding activity, cortical reinstatement, and hippocampal univariate retrieval activity, drive memory behavior.

In particular, this model revealed that the effect of retrieval-phase hippocampal activity on subsequent retrieval was partially mediated by the magnitude of cortical reinstatement. This result provides evidence for theories positing that retrieval is supported by hippocampal processes that drive the reinstatement (or replay) of cortical patterns established at event encoding. Specifically, cue-related information represented in the cortex is thought to propagate to the hippocampus, where neurons linked to associates of the cue become active, a process known as pattern completion (Marr 1971; McClelland et al. 1995; McClelland and Goddard 1996; Rolls and Treves 1998; Wallenstein et al. 1998). Signals resulting from hippocampal pattern completion are thought to ultimately project back to cortical regions that code for the associated event details, reinstating cortical patterns that were established at encoding, and enabling retrieval task performance. The results of the present path analysis indicate that cortical reinstatement partially mediates the effect of retrieval-phase hippocampus activity on retrieval accuracy, thus providing a compelling account for how interactions between the hippocampus and cortex can lead to successful memory-dependent behavior.

From Cortical Encoding Strength to Cortical Reinstatement to Retrieval Accuracy

Critically, the path model also revealed the existence of a significant indirect path from cortical encoding strength to cortical reinstatement to subsequent memory performance. This finding documents one mechanistic pathway for content-specific SMEs: Content-specific activity at encoding predicts subsequent memory in part through the intermediate variable of cortical reinstatement strength. More generally, this result suggests that the state of the cortical pattern at encoding affects memory through its affect on the state of the cortical pattern elicited at retrieval. This finding complements those of Ritchey et al. (in press), who observed that correct item recognition is predicted by the similarity of the pattern of occipital cortical activity elicited by an item at encoding and retrieval, and that this similarity is related to retrieval-phase hippocampal activity.

It is worth noting that, while an indirect path through the variable of cortical reinstatement partly accounted for the relationship between cortical encoding strength and subsequent memory, the direct effect of encoding strength on subsequent memory remained significant. We note that complete mediation of the effect of encoding strength on retrieval behavior is unlikely to be observed because of the presence of multiple potential mediator variables (Baron and Kenny 1986), including variables that are not measured by our cortical encoding and cortical reinstatement assays. In this study, for instance, pattern typicality effects likely contributed to our measures of encoding strength and cortical reinstatement, but were not modeled out of our analyses. Thus, for a pattern in which the reinstated activity was strong but was more atypical of encoding trials of the same class, the "true" strength of reinstatement will be underestimated by our measurement. In a related point, item-specific representations at encoding and retrieval likely played important roles in establishing memory accuracy, but were not included in our analyses, because our approach measured only category-general representational strength. Future studies examining how item-specific neural patterns at encoding and retrieval affect memory performance (e.g., Staresina et al. 2012; Ritchey et al. in press) may further clarify how neural activity at encoding predicts retrieval behavior through an intermediate effect on that at retrieval.

Conclusion

This study demonstrates that retrieval of event details can be predicted by the strength of cortical representations at both encoding and retrieval. Furthermore, the strength of these representations scales with hippocampal activity at both encoding

and retrieval. Finally, path analysis indicates that the state of occipitotemporal cortex at encoding influences retrieval performance in part through its effects on the state of occipitotemporal cortex at retrieval. The approach of relating classifier-derived measures of cortical representational strength and hippocampal activity at encoding and retrieval promises to yield future insights into how cortical–hippocampal interactions support memory-based decisions.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by a grant from the National Institute of Mental Heath (R01–MH080309) and a National Science Foundation Graduate Research Fellowship to A.M.G.

Notes

Conflict of Interest: None declared.

References

Awipi T, Davachi L. 2008. Content-specific source encoding in the human medial temporal lobe. J Exp Psychol Learn Mem Cogn. 34:769–779.

Babiloni C, Vecchio F, Mirabella G, Buttiglione M, Sebastiano F, Picardi A, Di Gennaro G, Quarato PP, Grammaldo LG, Buffo P et al. 2009. Hippocampal, amygdala, and neocortical synchronization of theta rhythms is related to an immediate recall during grey auditory verbal learning test. Hum Brain Mapp. 30:2077–2089.

Baron RM, Kenny DA. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 51:1173–1182.

Benjamini Y, Hochberg Y. 1995. Controlling the fales discovery rate: a practical and powerful appraoch to multiple testing. J R Stat Soc Ser B (Methodological). 57:289–300.

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

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JD. 1998. Making memories: brain activity that predicts how well visual experience will be remembered. Science. 281:1185–1187.

Cabeza R, Rao SR, Wagner AD, Mayer AR, Schachter DL. 2001. Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. Proc Natl Acad Sci. 98:4805–4810.

Chen J, Olsen RK, Preston AR, Glover GH, Wagner AD. 2011. Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. Learn Mem. 18:523–528.

Chun MM, Johnson MK. 2011. Memory: enduring traces of perceptual and reflective attention. Neuron. 72:520–535.

Cohen NJ, Eichenbaum H. 1993. Memory, amnesia, and the hippocampal system. Cambridge (MA): The MIT Press.

Craik FIM, Govoni R, Naveh-Benjamin M, Anderson ND. 1996. The effects of divided attention on encoding and retrieval processes in human memory. J Exp Psychol Gen. 125:159–180.

Craik FIM, Lockhart RS. 1972. Levels of processing: a framework for memory research. J Verb Learn Verb Behav. 11:671–684.

Danker JF, Anderson JR. 2010. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol Bull. 136:87–102.

Daselaar SM, Rice HJ, Greenberg DL, Cabeza R, LaBar KS, Rubin DC. 2008. The spatiotemporal dynamics of autobiographical memory: neural correlates of recall, emotional intensity, and reliving. Cereb Cortex. 18:217–229.

Davachi L. 2006. Item, context and relational episodic encoding in humans. Curr Opin Neurobiol. 16:693–700.

Davachi L, Mitchell JP, Wagner AD. 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proc Natl Acad Sci USA. 100:2157–2162.

Diana RA, Yonelinas AP, Ranganath C. 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn Sci. 11:379–386.

Dobbins IG, Rice HJ, Wagner AD, Schacter DL. 2003. Memory orientation and success: separable neurocognitive components underlying episodic recognition. Neuropsychologia. 41:318–333.

Dunn JC. 2004. Remember-know: a matter of confidence. Psychol Rev. 111:524–542.

Eichenbaum H. 2000. A cortical-hippocampal system for declarative memory. Nat Rev Neurosci. 1:41–50.

Eldridge L, Knowlton B, Furmanski C, Bookheimer S, Engel S. 2000. Remembering episodes: a selective role for the hippocampus during retrieval. Nat Neurosci. 3:1149–1152.

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

Fell J, Klaver P, Elfadil H, Schaller C, Elger CE, Fernandez G. 2003. Rhinal-hippocampal theta coherence during declarative memory formation: interaction with gamma synchronization? Eur J Neurosci. 17:1082–1088.

Fell J, Klaver P, Lehnertz K, Grunwald T, Schaller C, Elger CE, Fernández G. 2001. Human memory formation is accompanied by rhinal–hippocampal coupling and decoupling. Nat Neurosci. 5:1259–1264.

Fernández G, Brewer JB, Zhao Z, Glover GH, Gabrieli JD. 1999. Level of sustained entorhinal activity at study correlates with subsequent cued-recall performance: a functional magnetic resonance imaging study with high acquisition rate. Hippocampus. 9:35–44.

Foerde K. 2006. Modulation of competing memory systems by distraction. Proc Natl Acad Sci. 103:11778–11783.

Gagnepain P, Henson R, Chételat G, Desgranges B, Lebreton K, Eustache F. 2011. Is neocortical–hippocampal connectivity a better predictor of subsequent recollection than local increases in hippocampal activity? New insights on the role of priming. J Cogn Neurosci. 23:391–403.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.

Hayes SM, Nadel L, Ryan L. 2007. The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. Hippocampus. 17:873–889.

Henson RN, Rugg MD, Shallice T, Josephs O, Dolan RJ. 1999. Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. J Neurosci. 19:3962–3972.

Jenkins LJ, Ranganath C. 2010. Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. J Neurosci. 30:15558–15565.

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

Johnson JD, Rugg MD. 2007. Recollection and the reinstatement of encoding-related cortical activity. Cereb Cortex. 17:2507–2515.

Kahn I, Davachi L, Wagner AD. 2004. Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. J Neurosci. 24:4172–4180.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 17:4302–4311.

Khader P, Burke M, Bien S, Ranganath C, Rösler F. 2005. Content-specific activation during associative long-term memory retrieval. Neuroimage. 27:805–816.

Kim H. 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. Neuroimage. 54:2446–2461.

Kirchhoff BA, Wagner AD, Maril A, Stern CE. 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. J Neurosci. 20:6173–6180.
Kirwan CB, Stark CEL. 2004. Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. Hippocampus. 14:919–930.
Kozhevnikov M, Kosslyn S, Shephard J. 2005. Spatial versus object visualizers: a new characterization of visual cognitive style. Mem Cogn. 33:710–726.
Kuhl BA, Bainbridge WA, Chun MM. 2012. Neural reactivation reveals mechanisms for updating memory. J Neurosci. 32:3453–3461.
Kuhl BA, Rissman J, Chun MM, Wagner AD. 2011. Fidelity of neural reactivation reveals competition between memories. Proc Natl Acad Sci USA. 108:5903–5908.
Kuhl BA, Rissman J, Wagner AD. 2012. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia. 50:458–469.
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.
LaRocque KF, Smith ME, Carr VA, Witthoft N, Grill-Spector K, Wagner AD. 2013. Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. J Neurosci. 33:5466–5474.
Lavenex P, Amaral DG. 2000. Hippocampal-neocortical interaction: a hierarchy of associativity. Hippocampus. 10:420–430.
Loftus GR, Masson MEJ. 1994. Using confidence intervals in within-subject designs. Psychon Bull Rev. 1:476–490.
Marr D. 1971. Simple memory: a theory for archicortex. Philos Trans R Soc Lond Ser B Biol Sci. 262:23–81.
McClelland JL, Goddard NH. 1996. Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. Hippocampus. 6:654–665.
McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev. 102:419–457.
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.
Mishkin M, Suzuki WA, Gadian DG, Vargha-Khadem F. 1997. Hierarchical organization of cognitive memory. Philos Trans R Soc B Biol Sci. 352:1461–1467.
Mishkin M, Vargha-Khadem F, Gadian DG. 1998. Amnesia and the organization of the hippocampal system. Hippocampus. 8:212–216.
Mitchell KJ, Johnson MK. 2009. Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? Psychol Bull. 135:638–677.
Montaldi D, Spencer TJ, Roberts N, Mayes AR. 2006. The neural system that mediates familiarity memory. Hippocampus. 16:504–520.
Nichols EA, Kao Y-C, Verfaellie M, Gabrieli JDE. 2006. Working memory and long-term memory for faces: evidence from fMRI and global amnesia for involvement of the medial temporal lobes. Hippocampus. 16:604–616.
Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev. 110:611–646.
Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci (Regul Ed). 10:424–430.
Nyberg L, Habib R, McIntosh AR, Tulving E. 2000. Reactivation of encoding-related brain activity during memory retrieval. Proc Natl Acad Sci USA. 97:11120–11124.
O'Craven KM, Kanwisher N. 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J Cogn Neurosci. 12:1013–1023.
Otten LJ, Sveen J, Quayle AH. 2007. Distinct patterns of neural activity during memory formation of nonwords versus words. J Cogn Neurosci. 19:1776–1789.
Paller KA, Wagner AD. 2002. Observing the transformation of experience into memory. Trends Cogn Sci (Regul Ed). 6:93–102.
Pearl J, Verma T. 1987. The logic of representing dependencies by directed graphs. In: AAAI'87 Proceedings of the sixth national conference on artificial intelligence. Cambridge (MA): MIT Press.
Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. Science. 310:1963–1966.
Preacher KJ, Kelley K. 2011. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. Psychol Methods. 16:93–115.
Preston AR, Bornstein AM, Hutchinson JB, Gaare ME, Glover GH, Wagner AD. 2010. High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. J Cogn Neurosci. 22:156–173.
Prince SE, Dennis NA, Cabeza R. 2009. Encoding and retrieving faces and places: distinguishing process- and stimulus-specific differences in brain activity. Neuropsychologia. 47:2282–2289.
Ranganath C, Heller A, Cohen MX, Brozinsky CJ, Rissman J. 2005. Functional connectivity with the hippocampus during successful memory formation. Hippocampus. 15:997–1005.
Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M. 2004. Dissociable correlates of recollection and familiarity within the medial temporal lobes. Neuropsychologia. 42:2–13.
Ratcliff R. 1988. Continuous versus discrete information processing modeling accumulation of partial information. Psychol Rev. 95:238–255.
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.
Rissman J, Wagner AD. 2012. Distributed representations in memory: insights from functional brain imaging. Annu Rev Psychol. 63:101–128.
Ritchey M, Wing EA, LaBar KS, Cabeza R. in press. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cerebral Cortex.
Rock I, Gutman D. 1981. The effect of inattention on form perception. J Exp Psychol Hum Percept Perform. 7:275–285.
Rolls E. 1996. A theory of hippocampal function in memory. Hippocampus. 6:601. –620.
Rolls E, Treves A. 1998. Neural networks and brain function. Oxford: Oxford University Press.
Schacter DL, Chamberlain J, Gaesser B, Gerlach K. 2012. Neuroimaging of true, false, and imaginary memories. In: Nadel L, Sinnot-Armstrong WP, editors. Memory and law. New York: Oxford University Press. p. 233–262.
Schott BH, Wüstenberg T, Wimber M, Fenker DB, Zierhut KC, Seidenbecher CI, Heinze H-J, Walter H, Düzel E, Richardson-Klavehn A. 2011. The relationship between level of processing and hippocampal-cortical functional connectivity during episodic memory formation in humans. Hum Brain Mapp. 407–424
Sergerie K, Lepage M, Armony JL. 2005. A face to remember: emotional expression modulates prefrontal activity during memory formation. Neuroimage. 24:580–585.
Shipley B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology. 90:363–368.
Shipley B. 2000. A new inferential test for path models based on directed acyclic graphs. Struct Equation Model. 7:206–218.
Slotnick SD, Schacter DL. 2004. A sensory signature that distinguishes true from false memories. Nat Neurosci. 7:664–672.
Smith CN, Wixted JT, Squire LR. 2011. The hippocampus supports both recollection and familiarity when memories are strong. J Neurosci. 31:15693–15702.
Squire L, Zola-Morgan S. 1991. The medial temporal lobe memory system. Science. 253:1380–1386.
Squire LR. 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev. 99: 195–231.
Squire LR, Stark CEL, Clark RE. 2004. The medial temporal lobe. Ann Rev Neurosci. 27:279–306.

Staresina BP, Duncan KD, Davachi L. 2011. Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. J Neurosci. 31:8739–8747.
Staresina BP, Henson RNA, Kriegeskorte N, Alink A. 2012. Episodic reinstatement in the medial temporal lobe. J Neurosci. 32:18150–18156.
Suzuki WA, Amaral DG. 1994. Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. J Neurosci. 14:1856–1877.
Tong F, Pratte MS. 2012. Decoding patterns of human brain activity. Ann Rev Psychol. 63:483–509.
Turk-Browne NB, Yi D-J, Chun MM. 2006. Linking implicit and explicit memory: common encoding factors and shared representations. Neuron. 49:917–927.
Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273–289.
Uncapher MR. 2005. Encoding and the durability of episodic memory: a functional magnetic resonance imaging study. J Neurosci. 25:7260–7267.
Uncapher MR, Hutchinson JB, Wagner AD. 2011. Dissociable effects of top-down and bottom-up attention during episodic encoding. J Neurosci. 31:12613–12628.
Uncapher MR, Otten LJ, Rugg MD. 2006. Episodic encoding is more than the sum of its parts: an fMRI investigation of multifeatural contextual encoding. Neuron. 52:547–556.
Uncapher MR, Wagner AD. 2009. Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. Neurobiol Learn Mem. 91:139–154.
Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 281:1188–1191.
Wais PE, Squire LR, Wixted JT. 2010. In search of recollection and familiarity signals in the hippocampus. J Cogn Neurosci. 22:109–123.
Wallenstein GV, Hasselmo ME, Eichenbaum H. 1998. The hippocampus as an associator of discontiguous events. Trends Neurosci. 21:317–323.
Watanabe T, Hirose S, Wada H, Katsura M, Chikazoe J, Jimura K, Imai Y, Machida T, Shirouzu I, Miyashita Y et al. 2011. Prediction of subsequent recognition performance using brain activity in the medial temporal lobe. Neuroimage. 54:3085–3092.
Wheeler ME, Buckner RL. 2004. Functional-anatomic correlates of remembering and knowing. Neuroimage. 21:1337–1349.
Wheeler ME, Petersen SE, Buckner RL. 2000. Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci USA. 97:11125–11129.
Wheeler ME, Shulman GL, Buckner RL, Miezin FM, Velanova K, Petersen SE. 2006. Evidence for separate perceptual reactivation and search processes during remembering. Cereb Cortex. 16:949–959.
Wimmer GE, Shohamy D. 2012. Preference by association: how memory mechanisms in the hippocampus bias decisions. Science. 338:270–273.
Wixted JT, Stretch V. 2004. In defense of the signal detection interpretation of remember/know judgments. Psychon Bull Rev. 11:616–641.
Woodruff CC, Johnson JD, Uncapher MR, Rugg MD. 2005. Content-specificity of the neural correlates of recollection. Neuropsychologia. 43:1022–1032.
Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. 2010. Greater neural pattern similarity across repetitions is associated with better memory. Science. 330:97–101.
Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. 2005. Separating the brain regions involved in recollection and familiarity in recognition memory. J Neurosci. 25:3002–3008.
Zeithamova D, Dominick AL, Preston AR. 2012. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron. 75:168–179.

Supplementary Materials

Supplementary Methods

A Comparison of Univariate and Multivariate Methods: To investigate how classifier-derived neural variables related to measures derived from univariate amplitude, we performed a univariate amplitude analysis using the same data structures utilized by our pattern classifier. As in the classification analysis, patterns were parsed into 10 leave-one-out cross validation sets (for the encoding analysis), and an encoding training set and a retrieval testing set (for the reinstatement analysis). For each training set, a set of face- and scene-responsive voxels was obtained by using an ANOVA of activity across classes to determine the 500 voxels with greatest activity for faces vs. scenes and the 500 voxels with greatest activity for scenes vs. faces. Within each trial of each testing set, mean activity across the face- and scene-responsive voxels was computed. Mixed-effects linear regressions (Eq. S3, S4) were used to relate the classifier-derived measure of encoding strength and cortical reinstatement to mean univariate amplitude in voxels responsive to the correct and incorrect source. R^2 values for these regressions were calculated as the correlation between observed values and fitted values of the dependent variable.

Supplementary Results

Analyses of Hippocampal Signal Without Global Signal Residualization: To reduce the effects of global noise on the hippocampal signal, global mean signal was used to residualize the hippocampal signal in analyses reported in the main text. To

compare results with and without global mean signal residualization, all the hippocampal analyses were recomputed, using unresidualized hippocampal activity. Without global signal residualization, hippocampal activity scaled with other relevant variables in much the same way as with global signal residualization. Specifically, encoding-phase hippocampal activity scaled with cortical encoding strength ($p < .05$), and retrieval-phase hippocampal activity scaled with cortical reinstatement ($p < .05$), retrieval accuracy ($p < .005$), and RT ($p < .0001$). The one reported relationship that was no longer significant when unresidualized hippocampal activity was used was that between hippocampal activity at encoding and retrieval ($p = .22$)

Source Classification Within the Hippocampus: To assess whether the imagined and reinstated source categories were dissociable within patterns of hippocampal BOLD data, classification was performed on encoding and retrieval data from the hippocampus. Classification of the encoding data (62.3%, $t(26) = 7.71$, $p < 10^{-4}$) was significantly above chance, though at a considerably lower rate than classification using cortical patterns (81.1%). Similarly, a classifier trained on encoding-phase hippocampal data and tested on retrieval-phase hippocampal data discriminated the correctly retrieved source at a rate (54.3%, $t(26) = 2.91$, $p < .01$) greater than chance, but considerably lower than the rate based on cortical patterns (64.1%). Neither trialwise hippocampal encoding strength or trialwise hippocampal reinstatement magnitude scaled with retrieval accuracy or RT ($p > .1$ for all

comparisons), a null effect that may be related to the poor trialwise estimates of neural encoding and reinstatement strength obtained for the hippocampus.

Cortical Reinstatement Marginally Predicts Correct Retrieval Reaction Time (RT): To test whether cortical reinstatement predicted RT, we performed a linear regression analysis. We found that the classifier-derived measure of strength of cortical reinstatement marginally predicted retrieval RT for correct trials, with greater strength of reinstatement being associated with faster retrieval RTs (Eq. S1, $\chi^2(1) = 3.53$, $p < .1$). These results suggest that stronger, or more representative, cortical reinstatement may also predict shorter latencies for correct memory decisions, though future experiments are needed to further assess this possibility.

Relationship of Univariate Amplitude Effects to Multivariate Measures: The classifier-derived variables of encoding strength and cortical reinstatement may index patterns of activity distributed across cortex, but they may also track mean univariate amplitude differences in sets of voxels that are responsive to one vs. the other class. To investigate the relationship between univariate amplitude effects and classifier-derived measures, we conducted an analysis in which sets of face- and scene-specific voxels were independently defined, and then mean activity across each set was extracted from each pattern of encoding and retrieval data. Submitting these data to analysis, we found that this univariate measure of mean amplitude in voxels coding for the correct source was highly positively related to our measures of encoding strength (Eq. S3 $\chi^2(1) = 57.76$, $p < 10^{-4}$) and cortical reinstatement (Eq. S4,

$\chi^2(1) = 36.7, p < 10^{-4}$). Additionally, univariate measures of mean amplitude in voxels coding for the incorrect source were also highly negatively related to our measures of encoding strength (Eq. S3 , $\chi^2(1) = 56.3, p < 10^{-4}$) and cortical reinstatement(Eq. S4 , $\chi^2(1) = 45.7, p < 10^{-4}$). Statistical models including correct and incorrect univariate amplitude explained much, but not all, of the variance in encoding strength ($R^2 = .63$) cortical reinstatement ($R^2 = .47$). Thus, the multivariate measures employed here are highly related to univariate amplitude measures, but may also be driven in part by non-amplitude quantities, such as the effects of distributed patterns.

Path Analysis Predicting Correct-Trial RT: To test for the indirect effects related to RT, we conducted a path analysis including the variables of encoding strength, cortical reinstatement, hippocampal activity at encoding and retrieval, and RT (Supplementary Figure 11). Data were taken from correct trials only. The path structure was created by specifying paths between each pair of variables for which at least a marginally significant predictive relationship was found. A directional-separation test ($\chi^2(6) = 5.55, p = .53$) revealed that we could retain this model as a reasonable fit to the data.

Controlling for other reported effects of interest, all direct paths in the structure remained significant, except for the effect of cortical reinstatement on RT. Additionally, we tested for the existence of an indirect path from encoding strength to RT via cortical reinstatement, and an indirect path from retrieval-phase

hippocampal activity to RT via cortical reinstatement. Both indirect pathways proved non-significant ($p > .1$)

Effects of Cortical Reinstatement on RT are Differentially Driven by Scenes vs. Faces: To determine whether the effect of cortical reinstatement on correct trial RT was differentially driven by the face or scene stimulus class, we tested for an interaction between face/scene class and cortical reinstatement in predicting RT (Eq. S5). We found a significant interaction of cortical reinstatement and class on decision RT, which demonstrates that cortical reinstatement scales with RT more for scene than face trials ($\chi^2(1) = 7.26$, $p < .01$).

Supplementary Figure 1. Timecourse of TR-by-TR classification. (Left) Plot of probability of correct source classification of encoding trials. Shaded TRs, corresponding to 6-10 s after source category presentation were included in the analysis used to generate a metric of encoding strength. (Right) Plot of probability of correct source classification of retrieval trials, when the classifier was trained on encoding-phase data. Shaded TRs, corresponding to 4-8 s period after retrieval cue presentation, were used in the analysis to generate a metric of cortical reinstatement. Error bars indicate ± SEM.

Supplementary Figure 2. Plot of the probability of correct encoding task classification against the magnitude of encoding strength (i.e., certainty of the classifier). Error bars indicate ± within-subject SEM.

Supplementary Figure 3. Subjectwise plot of probability of correct source retrieval as a function of encoding strength. Each of the 27 plots represents an individual subject. Histograms at the top and bottom of each plot indicate probability distributions of encoding strength for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 2b.

Supplementary Figure 4. Subjectwise plot of encoding strength against reaction time for correct retrieval trials. For group data, see Figure 2c.

Supplementary Figure 5. Subjectwise plot of hippocampal activity against encoding strength for correct retrieval trials. For group data, see Figure 3b.

Supplementary Figure 6. Subjectwise plot of probability of correct source retrieval as a function of cortical reinstatement. Histograms at the top and bottom of each plot indicate probability distributions of cortical reinstatement for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 4c.

Supplementary Figure 7. Subjectwise plot of cortical reinstatement by encoding strength. For group data, see Figure 4d.

Supplementary Figure 8. Subjectwise plot of probability of correct source retrieval as a function of retrieval-phase hippocampal activity. Histograms at the top and bottom of each plot indicate probability distributions of retrieval-phase hippocampal activity for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 5a.

Supplementary Figure 9. Subjectwise plot of reaction time for correct trials as a function of hippocampal activity. For group data, see Figure 5b.

Supplementary Figure 10. Subjectwise plot of cortical reinstatement strength as a function of hippocampal activity. For group data, see Figure 5c.

Supplementary Figure 11. Path analysis relating neurally-derived mnemonic variables and RT for correct retrieval trials. Numeric labels indicate standardized path coefficients. Path thickness indicates the significance of each given effect. Neither the indirect path from encoding strength to cortical reinstatement to RT nor the indirect path from hippocampal activity at retrieval to cortical reinstatement to RT are significant ($p > .1$).

Supplementary Table 1

IV	DV	Interaction with stimulus class (p)	Uncorrected Significance (p)	FDR-corrected Significance (q)
Encoding Strength	Hipp Enc	0.557	**$2.86 * 10^{-5}$**	**$8.58 * 10^{-5}$**
Encoding Strength	Hipp Ret	0.532	0.239	0.256
Encoding Strength	Cortical Reinstatement	0.764	**$9.13 * 10^{-6}$**	**$3.42 * 10^{-5}$**
Encoding Strength	RT	0.108	**$2.22 * 10^{-2}$**	**$3.70 * 10^{-2}$**
Encoding Strength	Retrieval Accuracy	0.979	**$6.6 * 10^{-6}$**	**$3.30 * 10^{-5}$**
Hipp Enc	Hipp Ret	0.349	**$2.62 * 10^{-2}$**	**$3.94 * 10^{-2}$**
Hipp Enc	Cortical Reinstatement	0.544	0.145	0.181
Hipp Enc	RT	0.791	0.223	0.256
Hipp Enc	Retrieval Accuracy	0.751	0.960	0.960
Hipp Ret	Cortical Reinstatement	**0.048**	**$6.69 * 10^{-3}$**	**$1.25 * 10^{-2}$**
Hipp Ret	RT	0.557	**$2.13 * 10^{-4}$**	**$4.56 * 10^{-4}$**
Hipp Ret	Retrieval Accuracy	0.181	**$1.50 * 10^{-4}$**	**$3.75 * 10^{-4}$**
Cortical Reinstatement	RT	**0.007**	*$6.01 * 10^{-2}$*	*$8.20 * 10^{-2}$*
Cortical Reinstatement	Retrieval Accuracy	0.545	**$1.12 * 10^{-12}$**	**$1.68 * 10^{-11}$**

Supplementary Table 1. Summary of all regression analyses between two neurally-derived variables, or between a neurally-derived variable and a behavioral variable. Bold entries correspond to significant values ($p < .05$), italicized entries correspond to marginally significant values ($p < .1$). Column 1: the independent variable. Column 2: the dependent variable. Column 3: the interaction of each independent variable with the effect of face/scene stimulus class in predicting the dependent variable. Column 4: the uncorrected p-value of the effect of the independent variable on the dependent variable. Column 5: the q-value of the given regression, obtained through false discovery rate analysis. The q-value corresponds to the maximal FDR at which the regression could be considered significant.

Appendix

Regression equations

To test whether magnitude of encoding strength predicted the true stimulus class (see Figure 2a), the following mixed-effects logistic regression was performed:

Eq. 1: $Pc_j = [\ 1 + e^{-(\beta_0 + \beta_1|g| + \beta_2 s + \beta_3 r + b_{0j} + b_{1j}|g|\ )}\ ]^{-1}$

where Pc is the probability of correct classification, g is encoding strength, s is stimulus class (face/scene), r is the retrieval performance status (correct/incorrect), β indicates a fixed-effect coefficient, b indicates a random-effect coefficient, and j indexes the subject.

To test whether encoding strength predicts subsequent memory performance (see Figure 3a), the following mixed-effects logistic regression was performed within each subject:

Eq. 2: $Pr_j = [\ 1 + e^{-(\beta_0 + \beta_1 g + \beta_2 s + b_{0j} + b_{1j} g\ )}\ ]^{-1}$

where Pr is the probability of correct memory retrieval.

To test whether encoding strength predicted retrieval reaction time in correct retrieval trials (see Figure 3b) and incorrect retrieval trials, the following mixed-effects linear regression was performed on the subset of correct and incorrect trials respectively:

Eq. 3: $RT_j = \beta_0 + \beta_1 g + \beta_2 s + b_{0j} + b_{1j} g$

where RT is retrieval reaction time.

To test whether encoding strength interacted with memory accuracy status in predicting reaction time, the following mixed-effects linear regression was performed:

Eq. 4: $RT_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + \beta_4 (g*r) + b_{0j} + b_{1j}(g*r)$

To test whether cortical encoding strength scaled positively with hippocampal activity (see Figure 4b), the following mixed-effects linear regression was performed:

Eq. 5: $v_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} g$

where v is encoding-phase hippocampal activity.

To test whether cortical reinstatement predicted retrieval accuracy (see Figure 5c), the following mixed-effects logistic regression was performed:

Eq. 6: $Pr_j = [\ 1 + e^{-(\beta_0 + \beta_1 I + \beta_2 s + b_{0j} + b_{1j} I\ )}\ ]^{-1}$

To test whether encoding strength predicted subsequent cortical reinstatement, (see Figure 5d) the following mixed-effects linear regression was performed:

Eq. 7: $I_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} g$

To test whether hippocampal activity at encoding predicts hippocampal activity at retrieval, the following mixed-effects linear regression was performed:

Eq. 8: $t_j = \beta_0 + \beta_1 v + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} v$

where t is retrieval-phase hippocampal activity.

To test whether hippocampal activity at retrieval predicts the likelihood of retrieval accuracy, (see Figure 6a) the following mixed-effects logistic regression was performed:

Eq. 9: $Pr_j = [\, 1 + e^{-(\beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t)} \,]^{-1}$

To test whether hippocampal activity at retrieval predicts RT for correct trials (see Figure 6b), the following mixed-effects linear regression was performed:

Eq. 10: $RT_j = \beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t$

To test whether hippocampal activity at retrieval predicts cortical reinstatement for correct trials (see Figure 6c), the following mixed-effects linear regression was performed:

Eq. 11: $I_j = \beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t$

A path analysis was created linking all effects of interest for which both correct and incorrect trials were included (see Figure 7.) The goodness-of-fit of this model was tested by determining the probability that the first listed predictor variable in equations 12a - 12c was independent from the outcome variable. These probabilities were the combined and tested against a chi-squared distribution with df = 6.

Eq. 12a: $Pr_j = \beta_0 + \beta_1 v + \beta_2 t + \beta_3 g + \beta_4 I + b_{0j} + b_{1j} v + b_{2j} t + b_{3j} g + b_{4j} I$

Eq. 12b: $t_j = \beta_0 + \beta_1 g + \beta_2 v + b_{0j} + b_{1j} g + b_{2j} v$

Eq. 12c: $I_j = \beta_0 + \beta_1 v + \beta_1 t + \beta_1 g + b_{0j} + b_{1j} v + b_{1j} t + b_{1j} g$

Path coefficients for this analysis were created by decomposing the path structure into the following regressions:

Eq. 13a: $Pr_j = \beta_0 + \beta_1 g + \beta_2 I + \beta_3 t + \beta_4 s + b_{0j} + b_{1j} g + b_{2j} I + b_{3j} t + b_{4j} s$

Eq. 13b: $I_j = \beta_0 + \beta_1 g + \beta_2 t + b_{0j} + b_{1j} g + b_{2j} t$

Eq. 13c: $t_j = \beta_0 + \beta_1 v + b_{0j} + b_{1j} v$

Eq. 13d: $v_j = \beta_0 + \beta_1 g + b_{0j} + b_{1j} g$

To test whether reinstatement strength predicted retrieval RT for correct trials, the following mixed-effects linear regression was performed:

Eq S1: $RT_j = \beta_0 + \beta_1 I + \beta_2 s + b_{0j} + b_{1j} I$

where I is the reinstatement strength and s is stimulus class.

To test whether reinstatement strength interacted with memory accuracy status in predicting retrieval RT, the following mixed-effects linear regression was performed:

Eq S2: $RT_j = \beta_0 + \beta_1 I + \beta_2 r + \beta_3 (I*r) + \beta_4 s + b_{0j} + b_{1j} (I*r)$

where r is accuracy status (correct/incorrect).

To test whether encoding strength is related to measures of univariate amplitude in source-specific voxels, the following mixed-effects linear regression was performed:

Eq S3: $g_j = \beta_0 + \beta_1 c + \beta_2 n + \beta_3 s + b_{0j} + b_{1j} c + b_{2j} r$

Where g is encoding strength, c is univariate amplitude in voxels responsive to the correct class and n is univariate amplitude in voxels responsive to the incorrect class.

To test whether encoding strength is related to measures of univariate amplitude in source-specific voxels, the following mixed-effects linear regression was performed:

Eq S4: $I_j = \beta_0 + \beta_1 c + \beta_2 n + \beta_3 s + b_{0j} + b_{1j} c + b_{2j} r$

Where I is reinstatement strength,

To test whether the effect of cortical reinstatement on RT differed for face vs. scene trials, the following mixed-effects linear regression was performed:

Eq S5: $RT_j = \beta_0 + \beta_1 I + \beta_2 s + \beta_3 (I*s) + b_{0j} + b_{1j} (I*s)$

Supplementary Materials

Supplementary Methods

A Comparison of Univariate and Multivariate Methods: To investigate how classifier-derived neural variables related to measures derived from univariate amplitude, we performed a univariate amplitude analysis using the same data structures utilized by our pattern classifier. As in the classification analysis, patterns were parsed into 10 leave-one-out cross validation sets (for the encoding analysis), and an encoding training set and a retrieval testing set (for the reinstatement analysis). For each training set, a set of face- and scene-responsive voxels was obtained by using an ANOVA of activity across classes to determine the 500 voxels with greatest activity for faces vs. scenes and the 500 voxels with greatest activity for scenes vs. faces. Within each trial of each testing set, mean activity across the face- and scene-responsive voxels was computed. Mixed-effects linear regressions (Eq. S3, S4) were used to relate the classifier-derived measure of encoding strength and cortical reinstatement to mean univariate amplitude in voxels responsive to the correct and incorrect source. R^2 values for these regressions were calculated as the correlation between observed values and fitted values of the dependent variable.

Supplementary Results

Analyses of Hippocampal Signal Without Global Signal Residualization: To reduce the effects of global noise on the hippocampal signal, global mean signal was used to residualize the hippocampal signal in analyses reported in the main text. To

compare results with and without global mean signal residualization, all the hippocampal analyses were recomputed, using unresidualized hippocampal activity. Without global signal residualization, hippocampal activity scaled with other relevant variables in much the same way as with global signal residualization. Specifically, encoding-phase hippocampal activity scaled with cortical encoding strength ($p < .05$), and retrieval-phase hippocampal activity scaled with cortical reinstatement ($p < .05$), retrieval accuracy ($p < .005$), and RT ($p < .0001$). The one reported relationship that was no longer significant when unresidualized hippocampal activity was used was that between hippocampal activity at encoding and retrieval ($p = .22$)

Source Classification Within the Hippocampus: To assess whether the imagined and reinstated source categories were dissociable within patterns of hippocampal BOLD data, classification was performed on encoding and retrieval data from the hippocampus. Classification of the encoding data (62.3%, $t(26) = 7.71$, $p < 10^{-4}$) was significantly above chance, though at a considerably lower rate than classification using cortical patterns (81.1%). Similarly, a classifier trained on encoding-phase hippocampal data and tested on retrieval-phase hippocampal data discriminated the correctly retrieved source at a rate (54.3%, $t(26) = 2.91$, $p < .01$) greater than chance, but considerably lower than the rate based on cortical patterns (64.1%). Neither trialwise hippocampal encoding strength or trialwise hippocampal reinstatement magnitude scaled with retrieval accuracy or RT ($p > .1$ for all

comparisons), a null effect that may be related to the poor trialwise estimates of neural encoding and reinstatement strength obtained for the hippocampus.

Cortical Reinstatement Marginally Predicts Correct Retrieval Reaction Time (RT): To test whether cortical reinstatement predicted RT, we performed a linear regression analysis. We found that the classifier-derived measure of strength of cortical reinstatement marginally predicted retrieval RT for correct trials, with greater strength of reinstatement being associated with faster retrieval RTs (Eq. S1, $\chi^2(1) = 3.53$, $p < .1$). These results suggest that stronger, or more representative, cortical reinstatement may also predict shorter latencies for correct memory decisions, though future experiments are needed to further assess this possibility.

Relationship of Univariate Amplitude Effects to Multivariate Measures: The classifier-derived variables of encoding strength and cortical reinstatement may index patterns of activity distributed across cortex, but they may also track mean univariate amplitude differences in sets of voxels that are responsive to one vs. the other class. To investigate the relationship between univariate amplitude effects and classifier-derived measures, we conducted an analysis in which sets of face- and scene-specific voxels were independently defined, and then mean activity across each set was extracted from each pattern of encoding and retrieval data. Submitting these data to analysis, we found that this univariate measure of mean amplitude in voxels coding for the correct source was highly positively related to our measures of encoding strength (Eq. S3 $\chi^2(1) = 57.76$, $p < 10^{-4}$) and cortical reinstatement (Eq. S4,

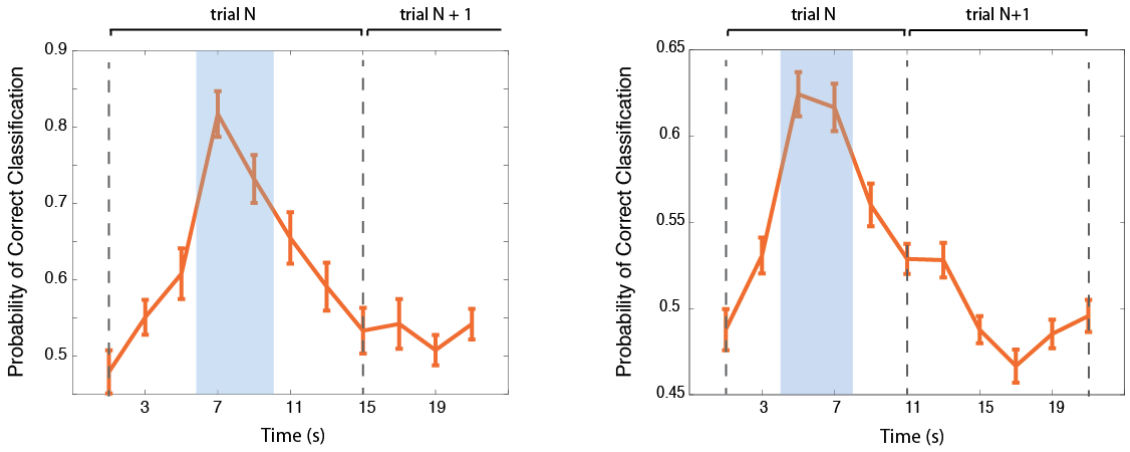
$\chi^2(1) = 36.7, p < 10^{-4}$). Additionally, univariate measures of mean amplitude in voxels coding for the incorrect source were also highly negatively related to our measures of encoding strength (Eq. S3 , $\chi^2(1) = 56.3, p < 10^{-4}$) and cortical reinstatement(Eq. S4 , $\chi^2(1) = 45.7, p < 10^{-4}$). Statistical models including correct and incorrect univariate amplitude explained much, but not all, of the variance in encoding strength ($R^2 = .63$) cortical reinstatement ($R^2 = .47$). Thus, the multivariate measures employed here are highly related to univariate amplitude measures, but may also be driven in part by non-amplitude quantities, such as the effects of distributed patterns.

Path Analysis Predicting Correct-Trial RT: To test for the indirect effects related to RT, we conducted a path analysis including the variables of encoding strength, cortical reinstatement, hippocampal activity at encoding and retrieval, and RT (Supplementary Figure 11). Data were taken from correct trials only. The path structure was created by specifying paths between each pair of variables for which at least a marginally significant predictive relationship was found. A directional-separation test ($\chi^2(6) = 5.55, p = .53$) revealed that we could retain this model as a reasonable fit to the data.

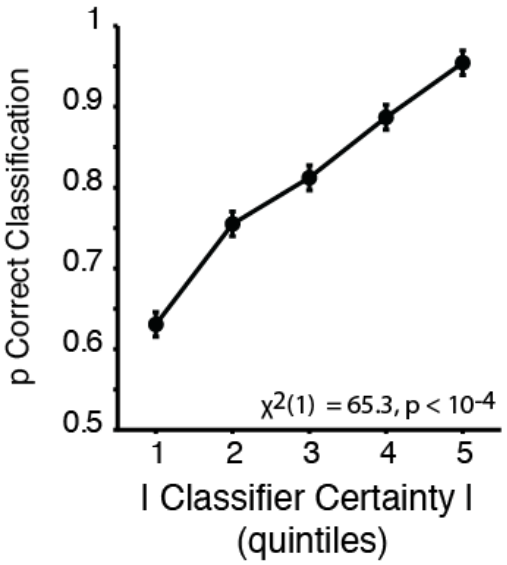
Controlling for other reported effects of interest, all direct paths in the structure remained significant, except for the effect of cortical reinstatement on RT. Additionally, we tested for the existence of an indirect path from encoding strength to RT via cortical reinstatement, and an indirect path from retrieval-phase

hippocampal activity to RT via cortical reinstatement. Both indirect pathways proved non-significant ($p > .1$)

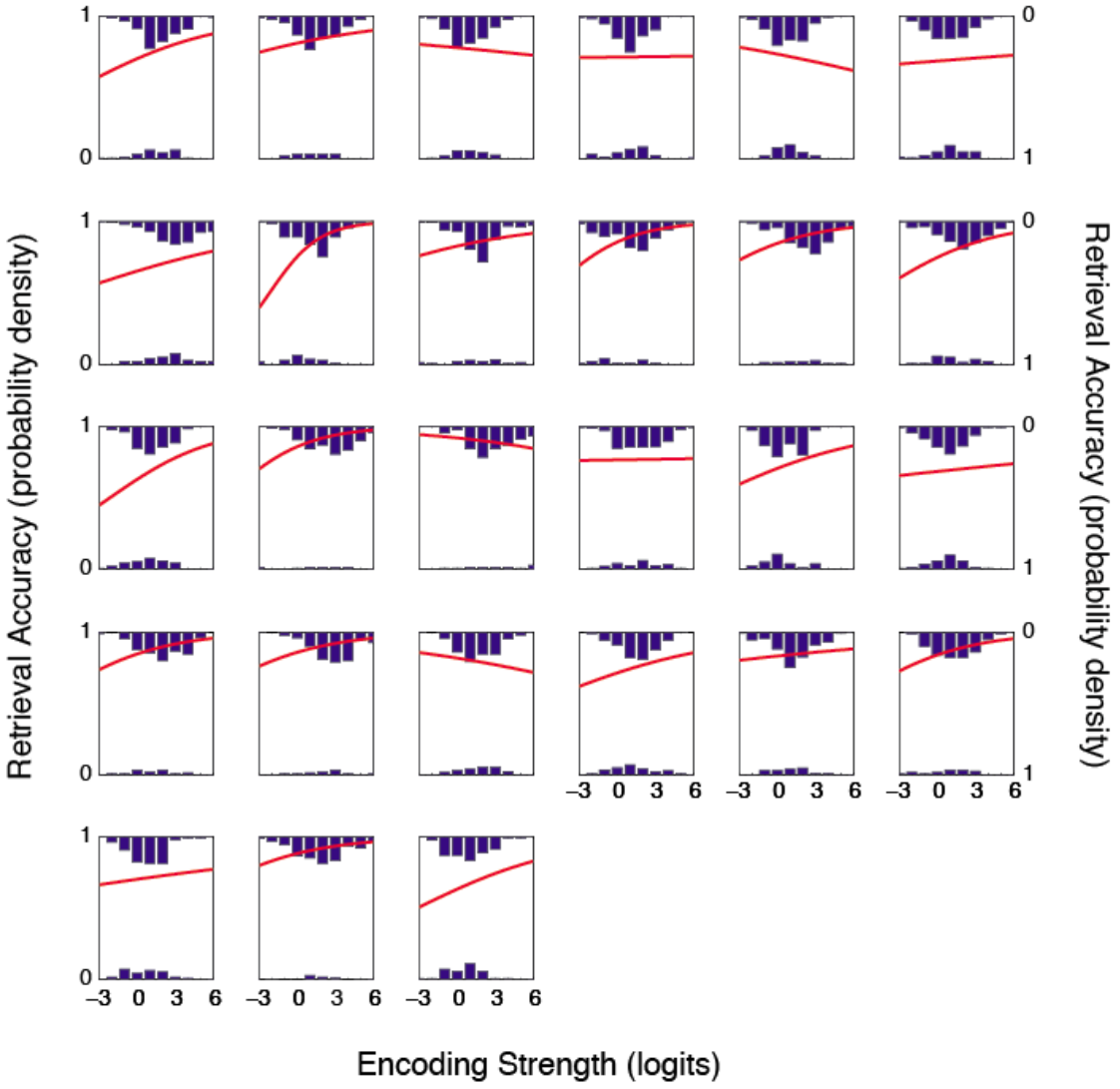
Effects of Cortical Reinstatement on RT are Differentially Driven by Scenes vs. Faces: To determine whether the effect of cortical reinstatement on correct trial RT was differentially driven by the face or scene stimulus class, we tested for an interaction between face/scene class and cortical reinstatement in predicting RT (Eq. S5). We found a significant interaction of cortical reinstatement and class on decision RT, which demonstrates that cortical reinstatement scales with RT more for scene than face trials ($\chi^2(1) = 7.26$, $p < .01$).



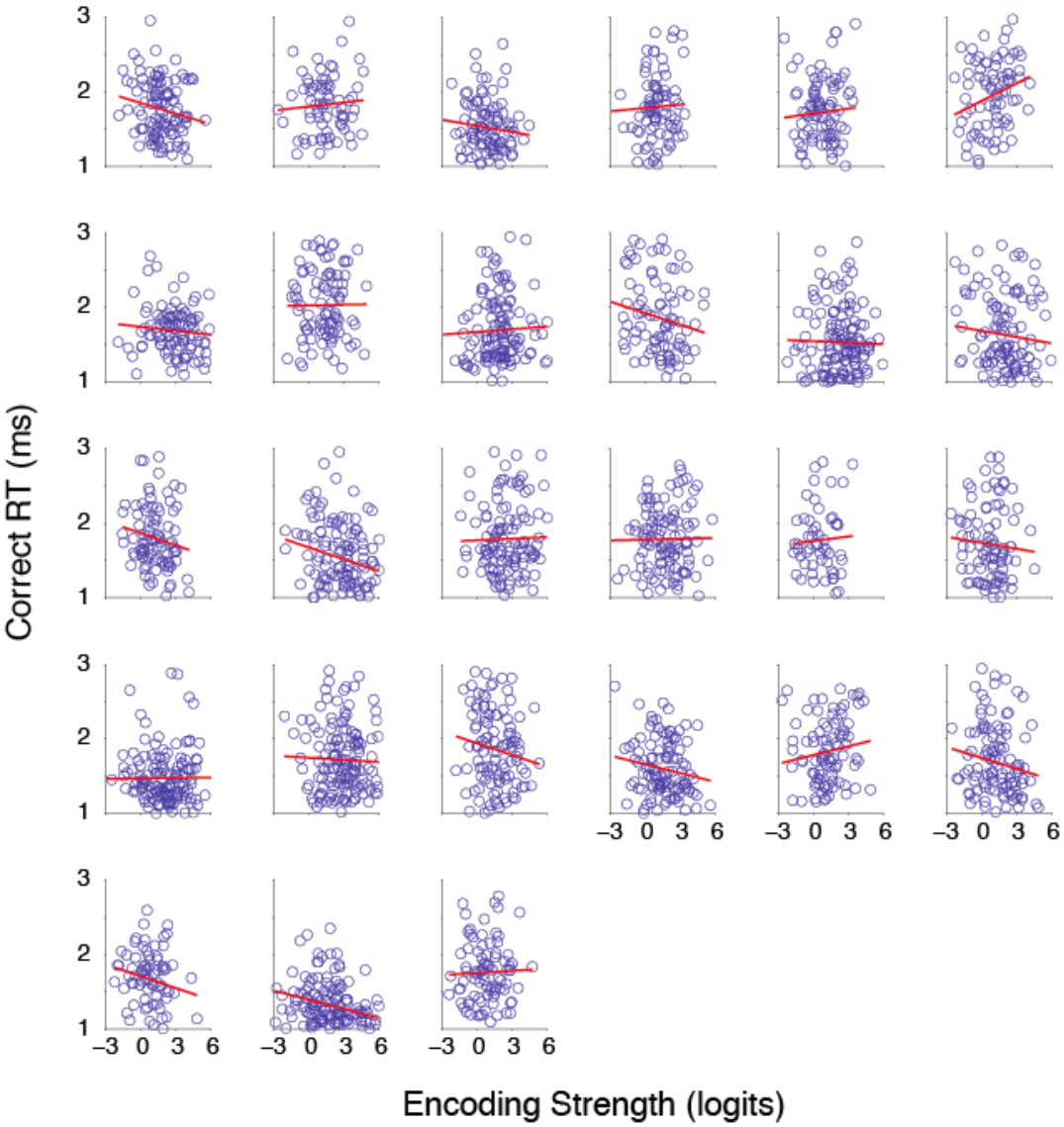
Supplementary Figure 1. Timecourse of TR-by-TR classification. (Left) Plot of probability of correct source classification of encoding trials. Shaded TRs, corresponding to 6-10 s after source category presentation were included in the analysis used to generate a metric of encoding strength. (Right) Plot of probability of correct source classification of retrieval trials, when the classifier was trained on encoding-phase data. Shaded TRs, corresponding to 4-8 s period after retrieval cue presentation, were used in the analysis to generate a metric of cortical reinstatement. Error bars indicate ± SEM.



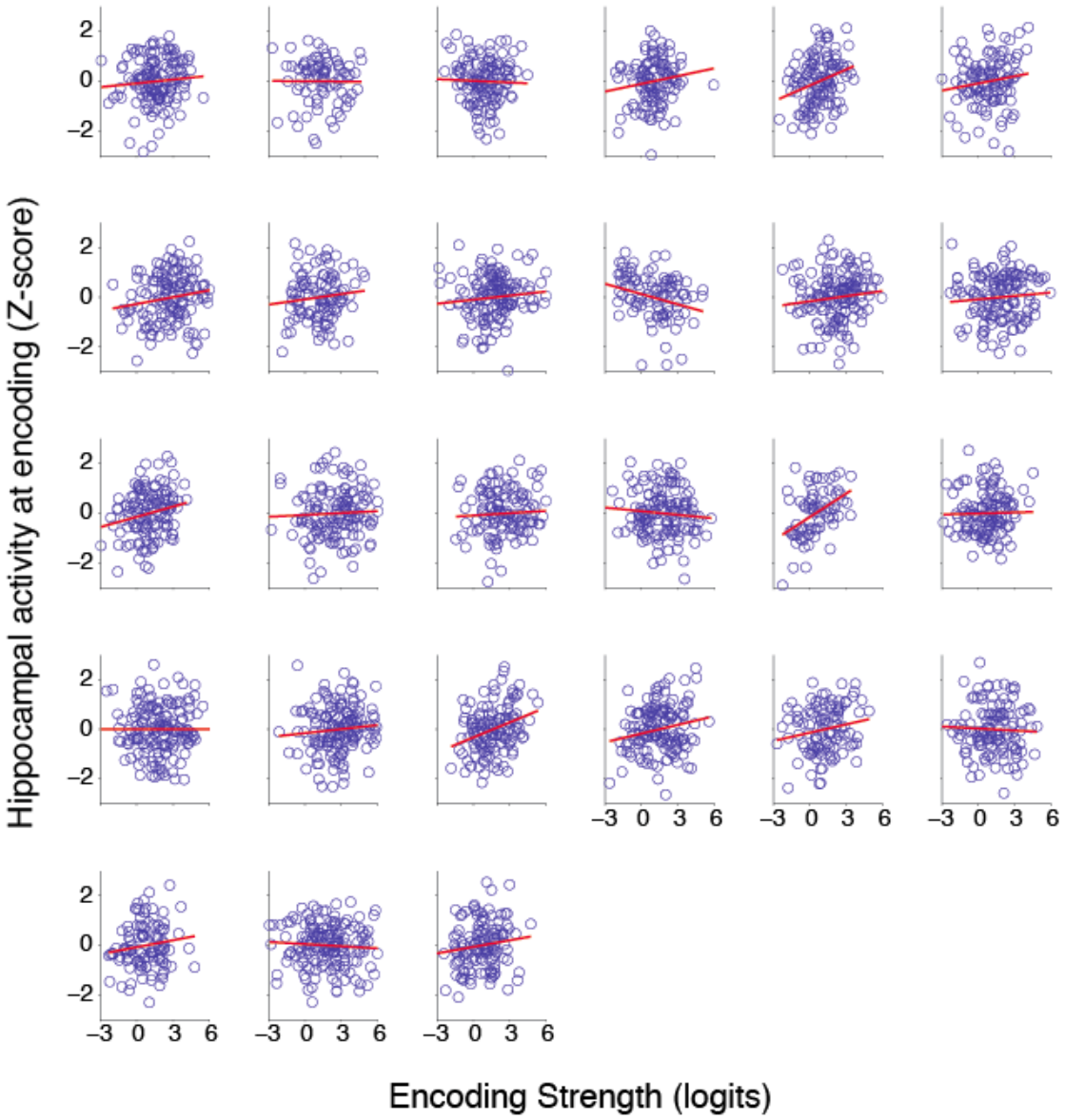
Supplementary Figure 2. Plot of the probability of correct encoding task classification against the magnitude of encoding strength (i.e., certainty of the classifier). Error bars indicate ± within-subject SEM.



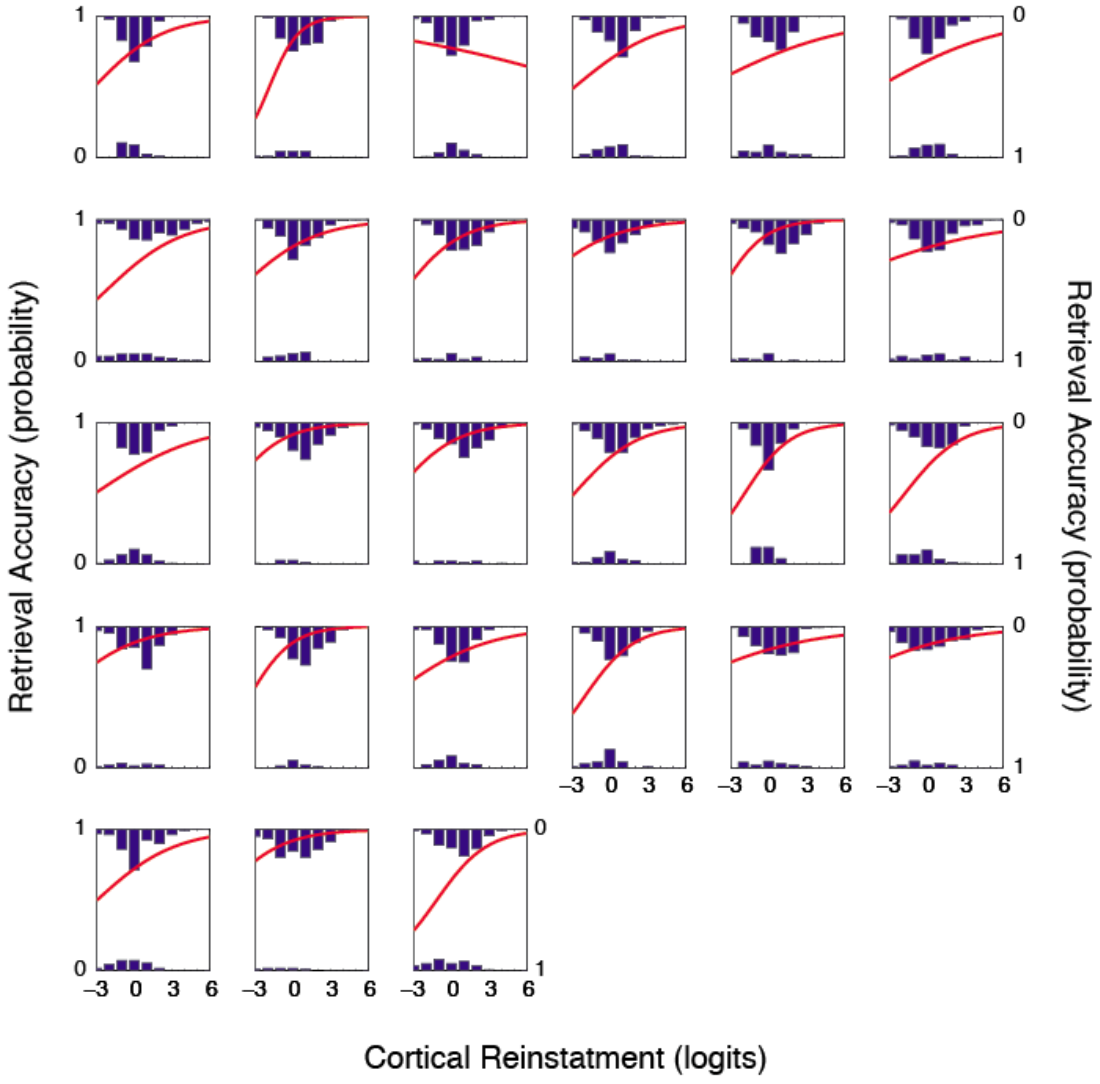
Supplementary Figure 3. Subjectwise plot of probability of correct source retrieval as a function of encoding strength. Each of the 27 plots represents an individual subject. Histograms at the top and bottom of each plot indicate probability distributions of encoding strength for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 2b.



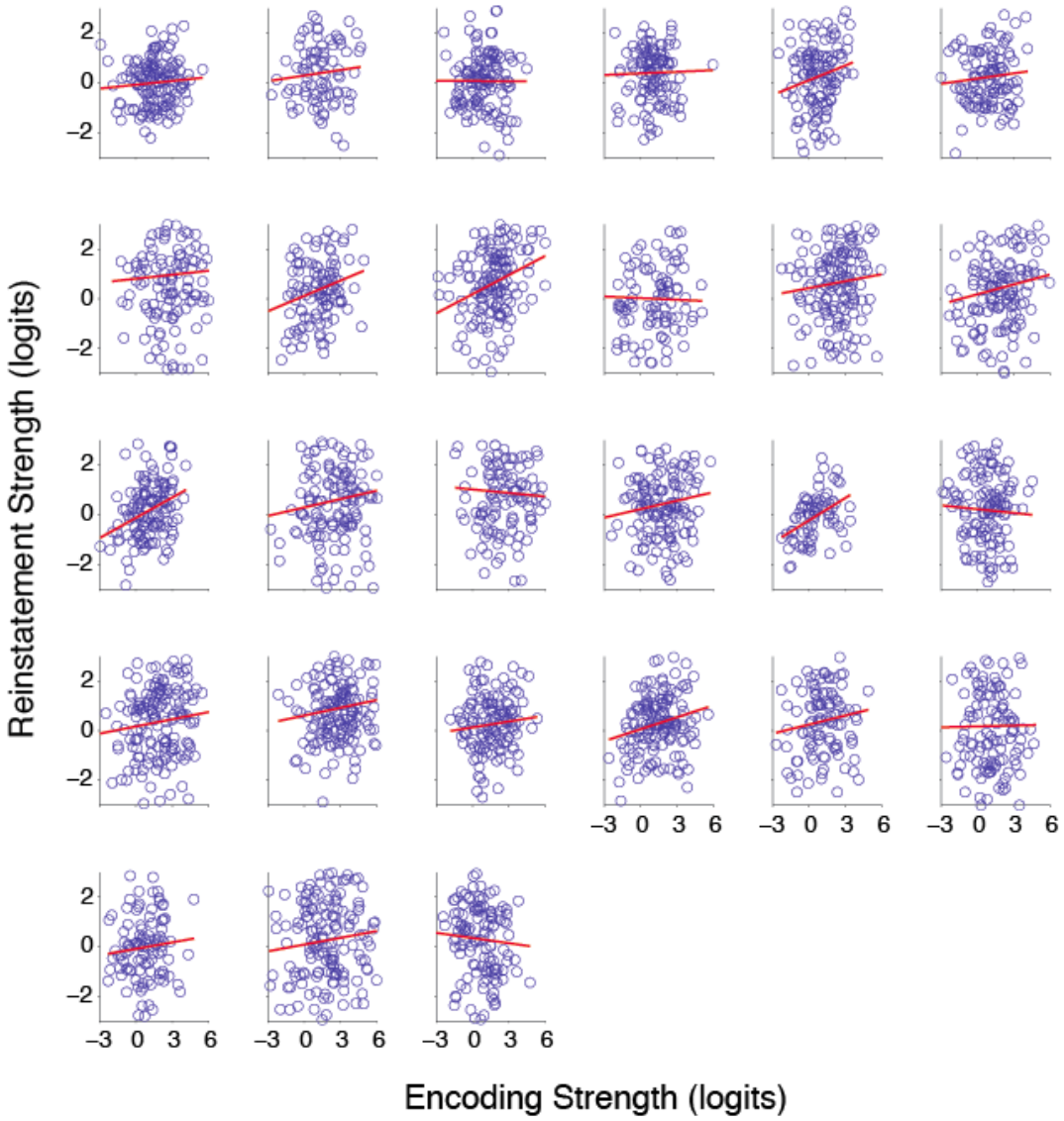
Supplementary Figure 4. Subjectwise plot of encoding strength against reaction time for correct retrieval trials. For group data, see Figure 2c.



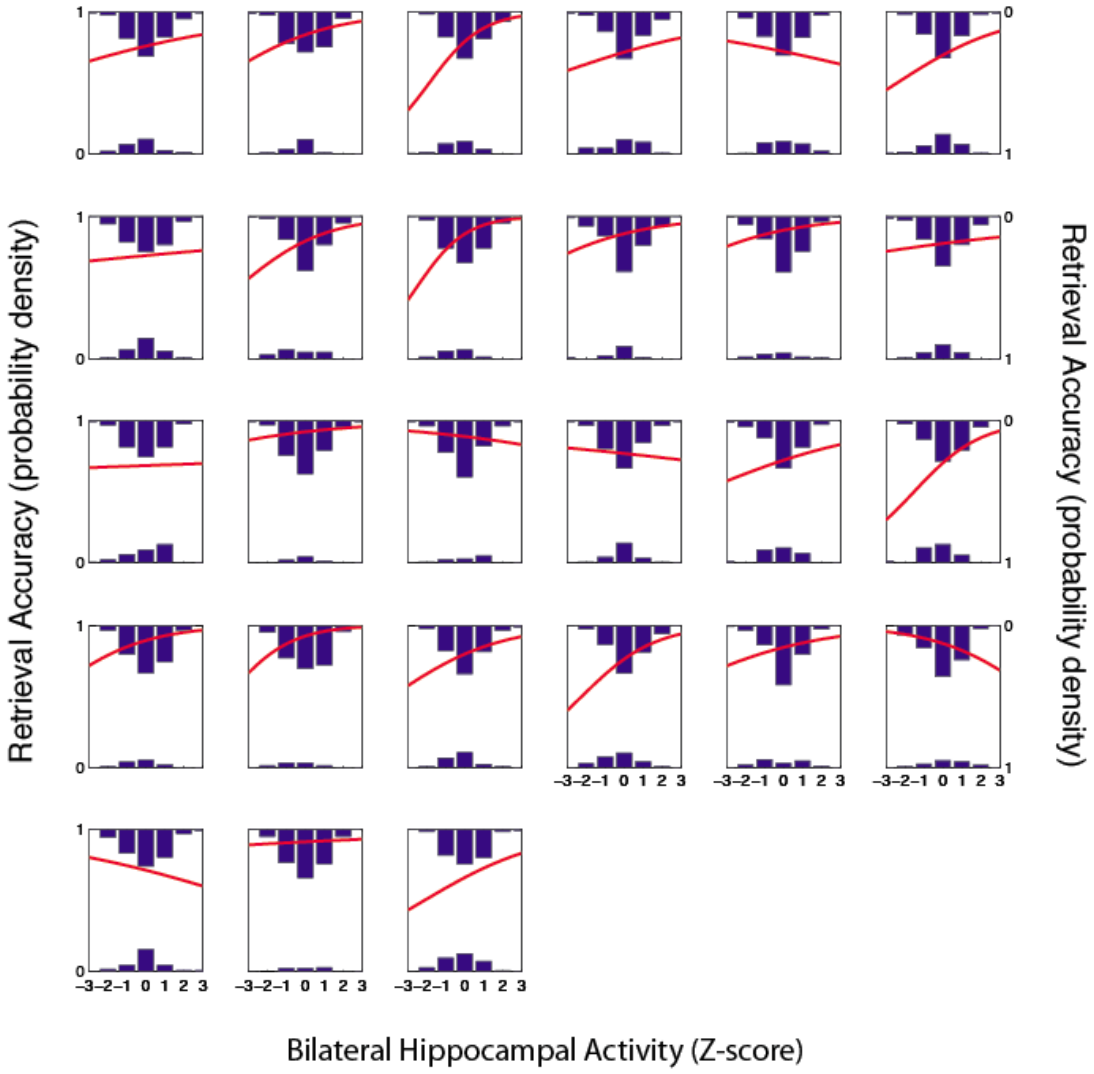
Supplementary Figure 5. Subjectwise plot of hippocampal activity against encoding strength for correct retrieval trials. For group data, see Figure 3b.



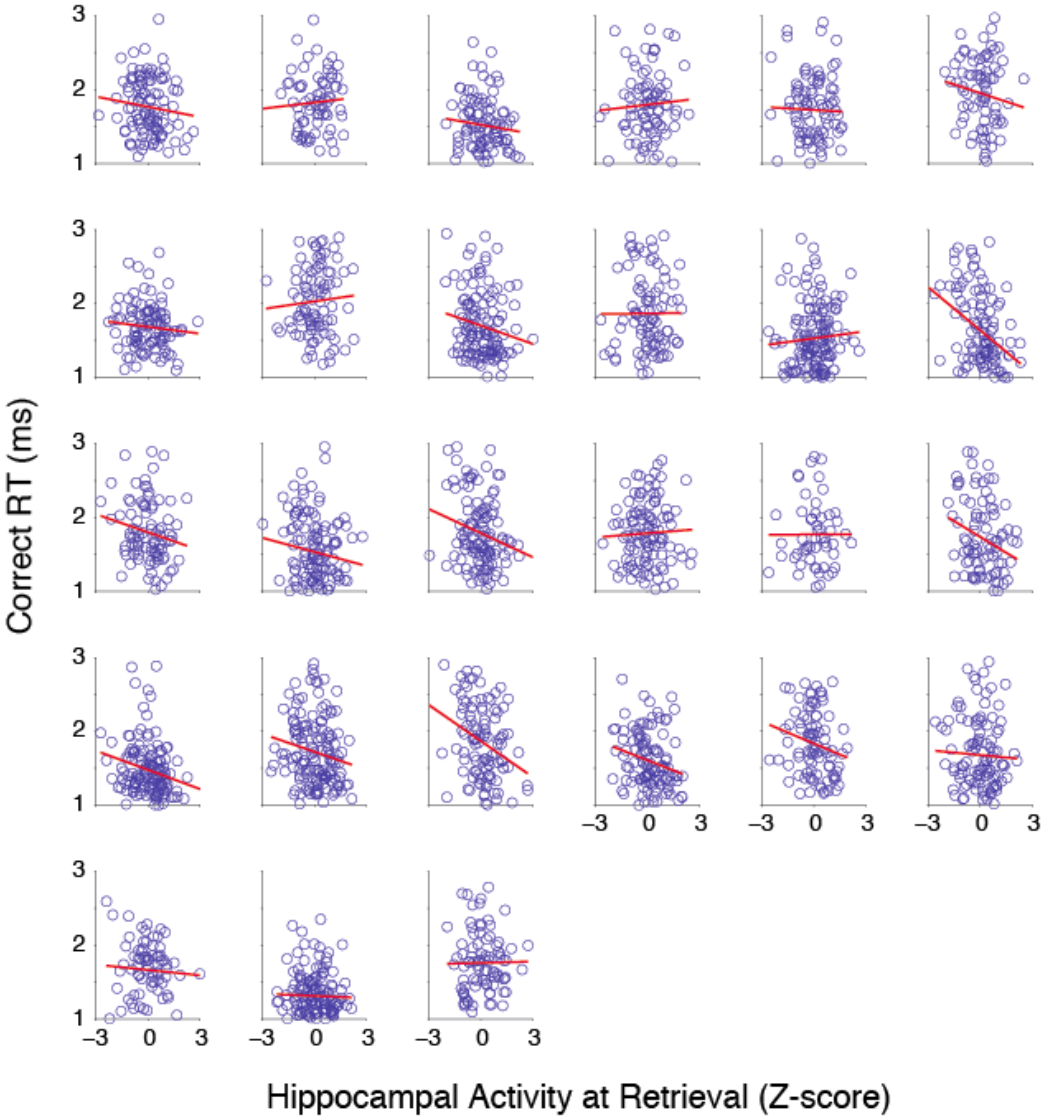
Supplementary Figure 6. Subjectwise plot of probability of correct source retrieval as a function of cortical reinstatement. Histograms at the top and bottom of each plot indicate probability distributions of cortical reinstatement for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 4c.



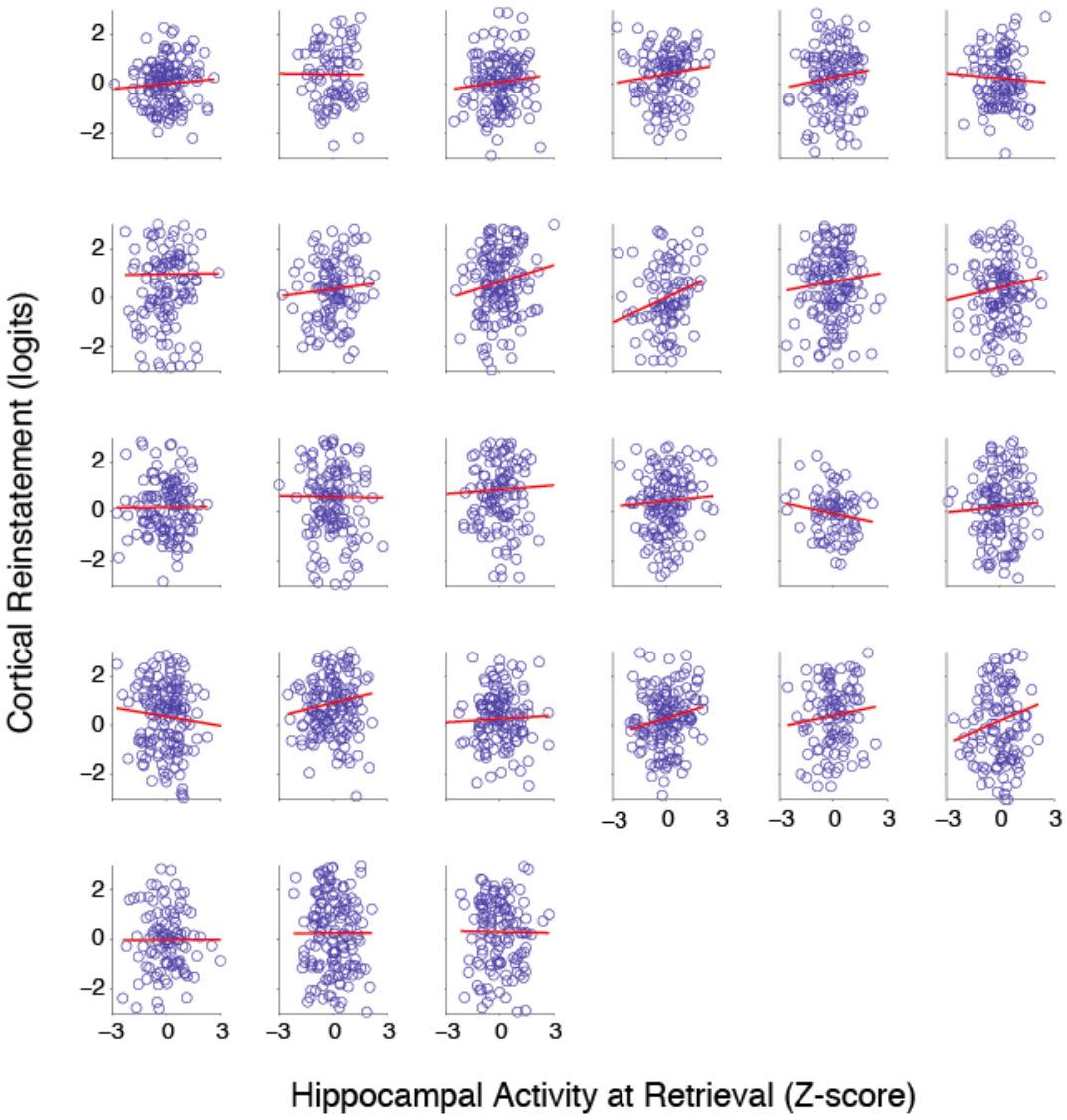
Supplementary Figure 7. Subjectwise plot of cortical reinstatement by encoding strength. For group data, see Figure 4d.



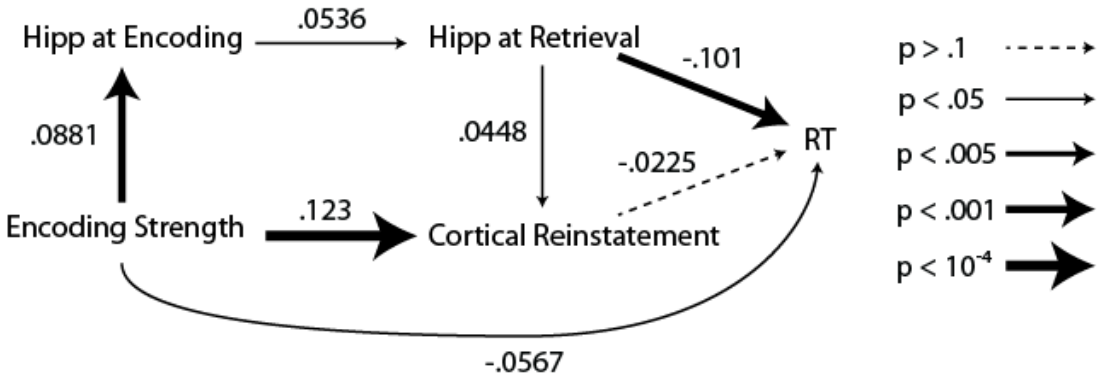
Supplementary Figure 8. Subjectwise plot of probability of correct source retrieval as a function of retrieval-phase hippocampal activity. Histograms at the top and bottom of each plot indicate probability distributions of retrieval-phase hippocampal activity for correct and incorrect retrieval trials, respectively. The red line indicates the logistic regression curve. For group data, see Figure 5a.



Supplementary Figure 9. Subjectwise plot of reaction time for correct trials as a function of hippocampal activity. For group data, see Figure 5b.



Supplementary Figure 10. Subjectwise plot of cortical reinstatement strength as a function of hippocampal activity. For group data, see Figure 5c.



Supplementary Figure 11. Path analysis relating neurally-derived mnemonic variables and RT for correct retrieval trials. Numeric labels indicate standardized path coefficients. Path thickness indicates the significance of each given effect. Neither the indirect path from encoding strength to cortical reinstatement to RT nor the indirect path from hippocampal activity at retrieval to cortical reinstatement to RT are significant ($p > .1$).

Supplementary Table 1

IV	DV	Interaction with stimulus class (p)	Uncorrected Significance (p)	FDR-corrected Significance (q)
Encoding Strength	Hipp Enc	5.57E-01	**2.86E-05**	**8.58E-05**
Encoding Strength	Hipp Ret	5.32E-01	2.39E-01	2.56E-01
Encoding Strength	Cortical Reinstatement	7.64E-01	**9.13E-06**	**3.42E-05**
Encoding Strength	RT	1.08E-01	**2.22E-02**	**3.70E-02**
Encoding Strength	Retrieval Accuracy	9.79E-01	**6.60E-06**	**3.30E-05**
Hipp Enc	Hipp Ret	3.49E-01	**2.62E-02**	**3.94E-02**
Hipp Enc	Cortical Reinstatement	5.44E-01	1.45E-01	1.81E-01
Hipp Enc	RT	7.91E-01	2.23E-01	2.56E-01
Hipp Enc	Retrieval Accuracy	7.51E-01	9.60E-01	9.60E-01
Hipp Ret	Cortical Reinstatement	4.83E-02	**6.69E-03**	**1.25E-02**
Hipp Ret	RT	5.57E-01	**2.13E-04**	**4.56E-04**
Hipp Ret	Retrieval Accuracy	1.81E-01	**1.50E-04**	**3.75E-04**
Cortical Reinstatement	RT	**7.04E-03**	*6.01E-02*	*8.20E-02*
Cortical Reinstatement	Retrieval Accuracy	5.45E-01	**1.12E-12**	**1.68E-11**

Appendix

Regression equations

To test whether magnitude of encoding strength predicted the true stimulus class (see Figure 2a), the following mixed-effects logistic regression was performed:

Eq. 1: $Pc_j = [\ 1 + e^{-(\beta_0 + \beta_1|g| + \beta_2 s + \beta_3 r + b_{0j} + b_{1j}|g|\ )}\ ]^{-1}$

where Pc is the probability of correct classification, g is encoding strength, s is stimulus class (face/scene), r is the retrieval performance status (correct/incorrect), β indicates a fixed-effect coefficient, b indicates a random-effect coefficient, and j indexes the subject.

To test whether encoding strength predicts subsequent memory performance (see Figure 3a), the following mixed-effects logistic regression was performed within each subject:

Eq. 2: $Pr_j = [\ 1 + e^{-(\beta_0 + \beta_1 g + \beta_2 s + b_{0j} + b_{1j} g\ )}\ ]^{-1}$

where Pr is the probability of correct memory retrieval.

To test whether encoding strength predicted retrieval reaction time in correct retrieval trials (see Figure 3b) and incorrect retrieval trials, the following mixed-effects linear regression was performed on the subset of correct and incorrect trials respectively:

Eq. 3: $RT_j = \beta_0 + \beta_1 g + \beta_2 s + b_{0j} + b_{1j} g$

where RT is retrieval reaction time.

To test whether encoding strength interacted with memory accuracy status in predicting reaction time, the following mixed-effects linear regression was performed:

Eq. 4: $RT_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + \beta_4 (g*r) + b_{0j} + b_{1j} (g*r)$

To test whether cortical encoding strength scaled positively with hippocampal activity (see Figure 4b), the following mixed-effects linear regression was performed:

Eq. 5: $v_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} g$

where v is encoding-phase hippocampal activity.

To test whether cortical reinstatement predicted retrieval accuracy (see Figure 5c), the following mixed-effects logistic regression was performed:

Eq. 6: $Pr_j = [\ 1 + e^{-(\beta_0 + \beta_1 I + \beta_2 s + b_{0j} + b_{1j} I\ )}\ ]^{-1}$

To test whether encoding strength predicted subsequent cortical reinstatement, (see Figure 5d) the following mixed-effects linear regression was performed:

Eq. 7: $I_j = \beta_0 + \beta_1 g + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} g$

To test whether hippocampal activity at encoding predicts hippocampal activity at retrieval, the following mixed-effects linear regression was performed:

Eq. 8: $t_j = \beta_0 + \beta_1 v + \beta_2 s + \beta_3 r + b_{0j} + b_{1j} v$

where t is retrieval-phase hippocampal activity.

To test whether hippocampal activity at retrieval predicts the likelihood of retrieval accuracy, (see Figure 6a) the following mixed-effects logistic regression was performed:

Eq. 9: $Pr_j = [\ 1 + e^{-(\beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t)}\ ]^{-1}$

To test whether hippocampal activity at retrieval predicts RT for correct trials (see Figure 6b), the following mixed-effects linear regression was performed:

Eq. 10: $RT_j = \beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t$

To test whether hippocampal activity at retrieval predicts cortical reinstatement for correct trials (see Figure 6c), the following mixed-effects linear regression was performed:

Eq. 11: $I_j = \beta_0 + \beta_1 t + \beta_2 s + b_{0j} + b_{1j} t$

A path analysis was created linking all effects of interest for which both correct and incorrect trials were included (see Figure 7.) The goodness-of-fit of this model was tested by determining the probability that the first listed predictor variable in equations 12a - 12c was independent from the outcome variable. These probabilities were the combined and tested against a chi-squared distribution with df = 6.

Eq. 12a: $Pr_j = \beta_0 + \beta_1 v + \beta_2 t + \beta_3 g + \beta_4 I + b_{0j} + b_{1j} v + b_{2j} t + b_{3j} g + b_{4j} I$

Eq. 12b: $t_j = \beta_0 + \beta_1 g + \beta_2 v + b_{0j} + b_{1j} g + b_{2j} v$

Eq. 12c: $I_j = \beta_0 + \beta_1 v + \beta_1 t + \beta_1 g + b_{0j} + b_{1j} v + b_{1j} t + b_{1j} g$

Path coefficients for this analysis were created by decomposing the path structure into the following regressions:

Eq. 13a: $Pr_j = \beta_0 + \beta_1 g + \beta_2 I + \beta_3 t + \beta_4 s + b_{0j} + b_{1j} g + b_{2j} I + b_{3j} t + b_{4j} s$

Eq. 13b: $I_j = \beta_0 + \beta_1 g + \beta_2 t + b_{0j} + b_{1j} g + b_{2j} t$

Eq. 13c: $t_j = \beta_0 + \beta_1 v + b_{0j} + b_{1j} v$

Eq. 13d: $v_j = \beta_0 + \beta_1 g + b_{0j} + b_{1j} g$

To test whether reinstatement strength predicted retrieval RT for correct trials, the following mixed-effects linear regression was performed:

Eq S1: $RT_j = \beta_0 + \beta_1 I + \beta_2 s + b_{0j} + b_{1j} I$

where I is the reinstatement strength and s is stimulus class.

To test whether reinstatement strength interacted with memory accuracy status in predicting retrieval RT, the following mixed-effects linear regression was performed:

Eq S2: $RT_j = \beta_0 + \beta_1 I + \beta_2 r + \beta_3 (I*r) + \beta_4 s + b_{0j} + b_{1j} (I*r)$

where r is accuracy status (correct/incorrect).

To test whether encoding strength is related to measures of univariate amplitude in source-specific voxels, the following mixed-effects linear regression was performed:

Eq S3: $g_j = \beta_0 + \beta_1 c + \beta_2 n + \beta_3 s + b_{0j} + b_{1j} c + b_{2j} r$

Where g is encoding strength, c is univariate amplitude in voxels responsive to the correct class and n is univariate amplitude in voxels responsive to the incorrect class.

To test whether encoding strength is related to measures of univariate amplitude in source-specific voxels, the following mixed-effects linear regression was performed:

Eq S4: $I_j = \beta_0 + \beta_1 c + \beta_2 n + \beta_3 s + b_{0j} + b_{1j} c + b_{2j} r$

Where I is reinstatement strength,

To test whether the effect of cortical reinstatement on RT differed for face vs. scene trials, the following mixed-effects linear regression was performed:

Eq S5: $RT_j = \beta_0 + \beta_1 I + \beta_2 s + \beta_3 (I*s) + b_{0j} + b_{1j} (I*s)$