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