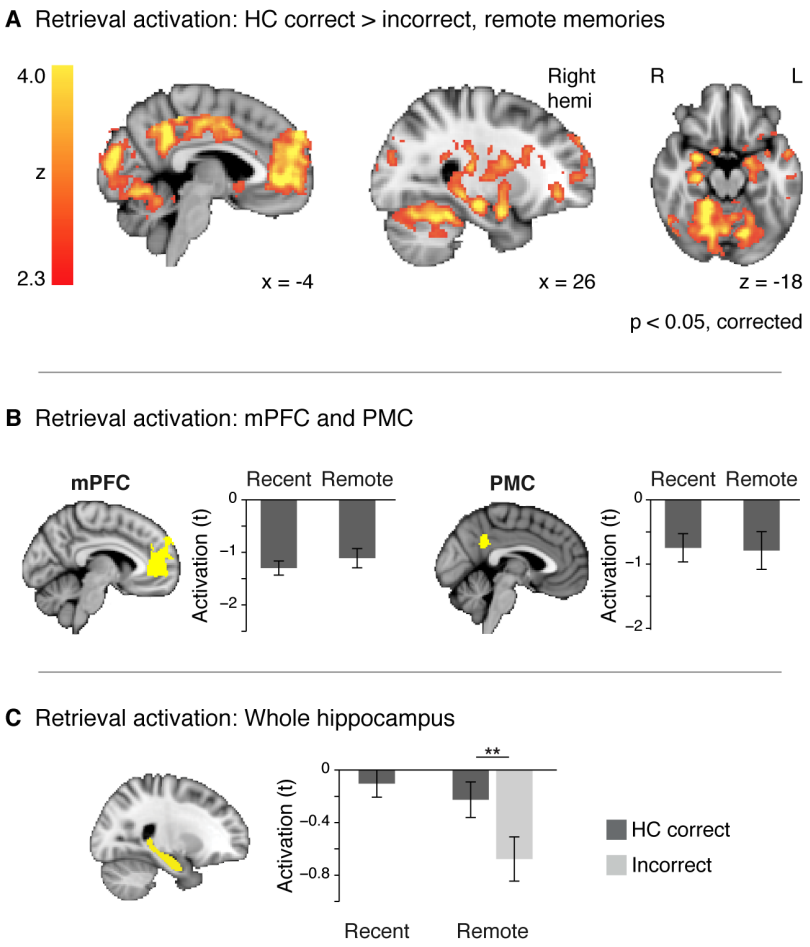
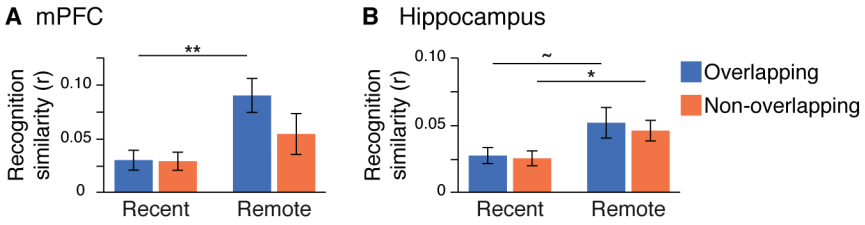
Figure S1, related to Figure 3



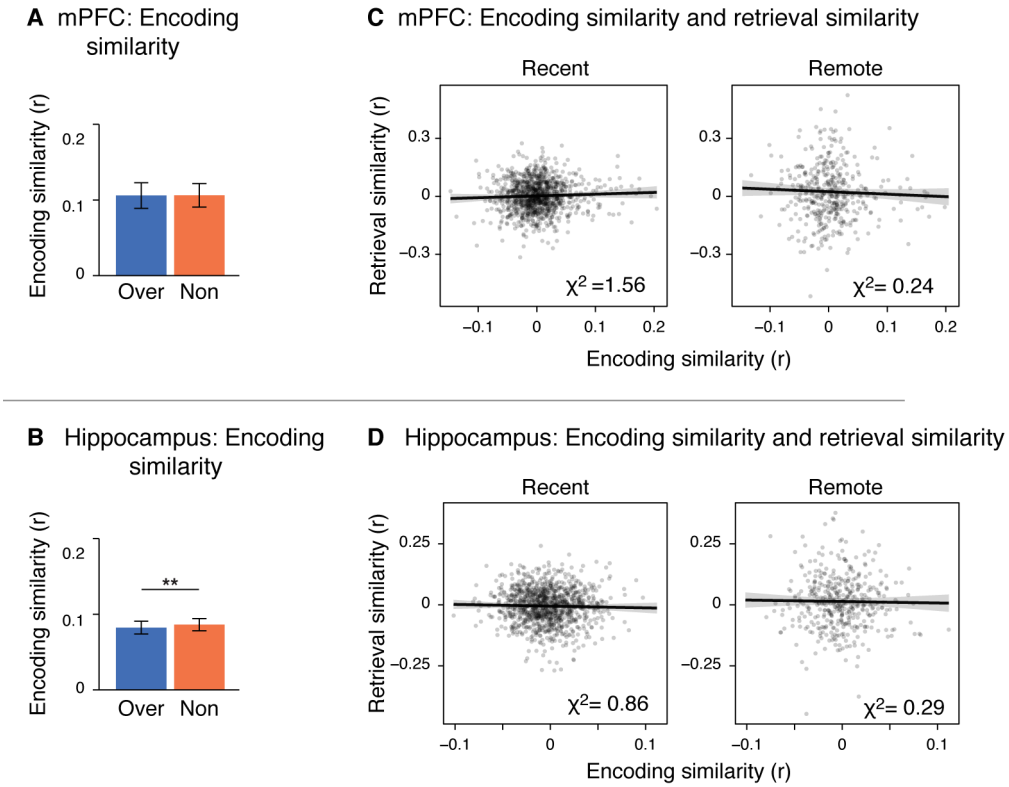
Univariate activation during retrieval. **A**. The retrieval similarity analyses were designed to measure representational changes in overlapping and non-overlapping memories over time. We focused on regions recruited during successful memory retrieval. We assessed how activation was influenced by the retrieval of remote memories, and then used these findings to define ROIs for use in the main analyses. A contrast of activation for HC correct vs. incorrect trials revealed clusters including bilateral hippocampus, medial prefrontal cortex, posterior medial cortex, and bilateral angular gyrus. These clusters survived correction for multiple comparisons using a cluster-mass threshold ($p < 0.05$, cluster-forming threshold $z = 2.3$). From this contrast, we extracted clusters in mPFC and PMC for use in similarity analyses. **B**. Average activation evoked by HC correct trials was extracted from mPFC and PMC separately for remote and recent retrieval. There was no reliable difference in activation for HC correct trials over time (mPFC: $t_{(18)} = 1.34$, $p = 0.20$, PMC: $t_{(18)} = -0.20$, $p = 0.84$). There were too few incorrect trials during recent retrieval to examine differences in activation by accuracy. Error bars signify SEM. **C**. The hippocampus exhibited increased activation for HC correct trials relative to incorrect trials at the remote retrieval session ($t_{(18)} = 5.14$, $p < 0.001$), and no difference in activation for HC correct trials over time ($t_{(18)} = -0.99$, $p = 0.34$). Error bars signify SEM. ** indicates $p < 0.01$.

Figure S2, related to Figure 3



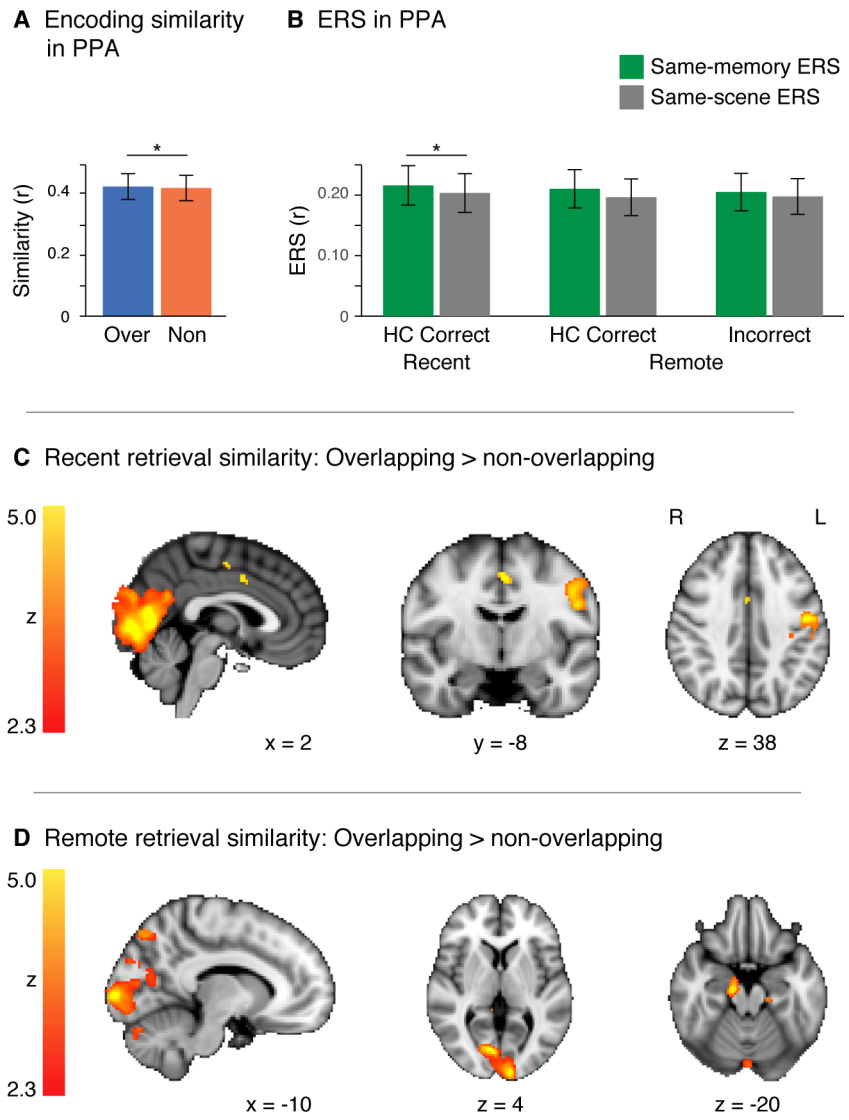
Recognition similarity. Recognition similarity was computed across all trials for which the target object was recognized with high confidence, using a 2 (Overlap: overlapping, non-overlapping) x 2 (Time: recent, remote) ANOVA. **A.** Recognition similarity in mPFC. There was no significant interaction between overlap and day ($F_{(1,18)}$ = 1.26, p = 0.28), nor was there a main effect of overlap ($F_{(1,18)}$ = 1.68, p = 0.21), in contrast to retrieval similarity. Instead, there was a significant main effect of time ($F_{(1,18)}$ = 24.19, p < 0.001). This main effect was driven by greater similarity for remote memories relative to recent memories, in particular for overlapping trials ($t_{(18)}$ = 3.33, p = 0.004) but not for non-overlapping trials ($t_{(18)}$ =1.45, p = 0.16). There was no difference in recognition similarity for overlapping versus non-overlapping remote memories ($t_{(18)}$ =1.21, p = 0.24). **B.** Recognition similarity in the whole hippocampus. In the hippocampus, there was a trend for a main effect of time ($F_{(1,18)}$ = 4.37, p = 0.05), but no main effect of overlap ($F_{(1,18)}$ = 2.69, p = 0.12) and no interaction ($F_{(1,18)}$ = 0.36, p = 0.56). Over time, recognition similarity increased amongst non-overlapping trials $t_{(18)}$ = 2.25, p = 0.04) and less so amongst overlapping trials ($t_{(18)}$ = 1.88, p = 0.08). There was no difference in similarity by overlap for either recognition period (both p's > 0.28). ** indicates p < 0.01. * indicates p < 0.05. ~ indicates p < 0.10. Error bars signify SEM.

Figure S3, related to Figure 3



Encoding similarity. **A and B.** To investigate whether the hippocampus and mPFC represented overlap over encoding, we computed overlapping and non-overlapping similarity measures across encoding trials. There was no significant difference in encoding similarity for overlapping versus non-overlapping trials in mPFC ($t_{(18)}$ = -0.10, p = 0.92), consistent with pattern similarity measured during recent retrieval. Interestingly, in bilateral hippocampus, we found that overlapping similarity was significantly lower than non-overlapping similarity during encoding ($t_{(18)}$ = -3.47, p = 0.003). ** indicates p < 0.01. Error bars signify SEM. **C and D.** Trial-level relationship between encoding similarity and retrieval similarity We indexed how much each retrieval trial carried information about overlap by computing the difference between overlapping and non-overlapping similarity for each trial. We entered this difference score into a mixed-effects linear regression with time (recent, remote) as a predictor. When we included encoding similarity as an additional predictor of retrieval similarity, the fit of the model did not improve in mPFC (χ^2 = 0.07, p = 0.79) or bilateral hippocampus (χ^2 = 0.30, p = 0.38). This suggests that the variability in representational structure during learning did not significantly contribute to the restructuring of memories seen at remote retrieval. Gray points represent all trials included in the analysis. Black lines represent the best fit line representing the relationship between retrieval similarity and encoding similarity. Gray ribbons signify 95% confidence intervals.

Figure S4, related to Figure 3



Pattern similarity in visual regions. **A.** Encoding similarity in PPA was significantly greater for overlapping trials relative to non-overlapping trials ($t_{(18)}$ = 2.50, p = 0.02). Retrieval similarity was not modulated by time or overlap (all p's > 0.21; not pictured). **B.** PPA exhibited memory-specific reinstatement, as indexed by a 2 (Time) x 2 (ERS) ANOVA applied to HC correct trials. This revealed an effect of ERS ($F_{(1,18)}$ = 7.29, p = 0.01), but no reliable effect of time or interaction (both p's > 0.65). There was greater same-memory ERS relative to same-scene ERS during recent ($t_{(18)}$= 3.02, p = 0.007) but not remote retrieval ($t_{(18)}$= 1.27, p = 0.22). A 2 (Accuracy) x 2 (ERS) ANOVA applied to remote trials revealed a marginal effect of ERS ($F_{(1,18)}$ = 4.30, p = 0.05) and no effect of accuracy or interaction (both p's > 0.64). * indicates p < 0.05. Error bars indicate SEM. **C.** Clusters where overlapping similarity was greater than non-overlapping similarity during recent retrieval. **D.** Clusters where overlapping similarity was greater than non-overlapping similarity during remote retrieval. Clusters were corrected for family-wise error using FSL's TFCE (p < 0.05). See Table S4 for a complete list.

Table S1, related to Figure 3

	β	**SE**	***t***
mPFC			
Intercept	-0.036	0.032	-1.114
Time	0.211	0.091	**2.328***
Activation	-0.038	0.036	-1.463
Whole hippocampus			
Intercept	-0.051	0.033	-1.575
Time	0.225	0.080	**2.815****
Activation	-0.018	0.026	-0.703
Anterior hippocampus			
Intercept	-0.057	0.035	-1.646
Time	0.224	0.077	**2.924****
Activation	-0.015	0.026	-0.580
Posterior hippocampus			
Intercept	-0.066	0.031	-2.150
Time	0.250	0.066	**3.820****
Activation	-0.002	-0.027	0.060

Influence of univariate activation on retrieval similarity. Mixed-effects linear regressions were conducted to investigate the relationship between trial-level fluctuations in univariate activation and trial-level estimates of retrieval similarity. Time (recent, remote) and activation were included in models predicting retrieval similarity (overlapping – non-overlapping) across HC correct trials, separately for each ROI. Additionally, comparisons to models without activation as a predictor confirmed that the inclusion of this predictor to each model did not improve its fit (whole hippocampus: $\chi^2 = 0.49$, $p = 0.48$, mPFC: $\chi^2 = 2.12$, $p = 0.15$). * indicates $p < 0.05$, ** indicates $p < 0.01$.

Table S2, related to Figure 3

Procedure	Contrast	dF	*t*	*p*
mPFC				
Permutation test	Recent	18	0.662	0.516
	Remote	18	2.130	**0.047** *
	Remote > Recent	18	2.184	**0.042** *
Subsampling	Recent	18	0.562	0.581
	Remote	18	2.006	0.060 ~
	Remote > Recent	18	1.972	0.064 ~
Whole hippocampus				
Permutation test	Recent	18	-1.747	0.098 ~
	Remote	18	1.842	0.082 ~
	Remote > Recent	18	2.814	**0.012** *
Subsampling	Recent	18	-1.853	0.080 ~
	Remote	18	1.819	0.086 ~
	Remote > Recent	18	2.976	**0.008** **

Non-parametric tests of retrieval similarity. The effects of time and overlap on retrieval similarity were confirmed with two non-parametric tests: (1) a permutation test where the number of remembered trials and their respective scenes were kept constant for each participant with a shuffling procedure, and (2) a sub-sampling procedure where the number of non-overlapping trials was randomly reduced to match to the number of overlapping trials that were used to compute retrieval similarity for each trial (see STAR Methods). These tests were conducted for the following statistics reported in the main text: overlapping versus non-overlapping similarity for recent trials, overlapping versus non-overlapping similarity for remote trials, and overlapping versus non-overlapping similarity for remote versus recent trials (i.e. time x overlap interaction). The results of both procedures were consistent with the findings reported in the main text. ~ indicates $p < 0.10$, * indicates $p < 0.05$, ** indicates $p < 0.01$.

Table S3, related to Figure 3 and Figure S2

Effect	DFn	DFd	F	*p*
mPFC				
Time	1	18	6.944	0.017 *
Overlap	1	18	3.800	0.067 ~
Test	1	18	5.860	**0.026** *
Time x Overlap	1	18	3.147	0.093 ~
Time x Test	1	18	6.018	**0.025** *
Overlap x Test	1	18	0.056	0.815
Time x Overlap x Test	1	18	0.074	0.788
Whole hippocampus				
Time	1	18	3.362	0.083 ~
Overlap	1	18	2.670	0.120
Test	1	18	1.956	0.179
Time x Overlap	1	18	4.469	**0.048** *
Time x Test	1	18	0.930	0.348
Overlap x Test	1	18	0.043	0.837
Time x Overlap x Test	1	18	2.268	0.149

Influence of memory test on pattern similarity in mPFC and whole bilateral hippocampus. Results of separate repeated-measures ANOVAs, with pattern similarity in whole, bilateral hippocampus and mPFC as dependent variables. Time (recent, remote), overlap (overlapping, non-overlapping) and memory test (retrieval, recognition) were included as independent variables in each model. Critically, there was no reliable interaction between memory test, time and overlap in either region. * indicates $p < 0.05$, ~ indicates $p < 0.10$.

Table S4, related to Figure 3 and Figure S4

Cluster Index	Extent	*t*	X	Y	Z
Recent similarity					
Occipital pole/lingual gyrus	7509	6.49	-0.87	-78.6	5.12
L. Precentral gyrus	543	6.01	-49	-15	37.7
R. Parahippocampal gyrus	149	4.47	17.3	-39.4	-16
Juxtapositional lobule	47	5.56	-1.93	-8.81	47.7
White matter	20	3.74	-25.9	-15.2	45
Anterior cingulate	12	4.83	1.36	4.13	38.9
Remote similarity					
Occipital pole/lingual gyrus	2199	6.08	-1.71	-84.6	1.26
L. Superior lateral occipital cortex	417	4.43	-13.9	-76.4	47.2
Brainstem	305	5.7	11.1	-18	-27.7
Brainstem	86	4.34	-16.2	-22.9	-30
Occipital pole	32	3.49	12.3	-94.7	20.7

Whole-brain searchlights of retrieval similarity. Clusters identified where overlapping similarity was reliably greater than non-overlapping similarity, separately for recent and remote retrieval. Other searchlights conducted to assess differences in retrieval similarity over time (e.g. remote > recent retrieval of HC correct overlapping memories) yielded no significant clusters. R and L indicate right and left hemisphere. Extent is size of clusters in millimeters. X, Y, and Z coordinates indicate the center of gravity in MNI space (mm). *t* corresponds to the maximum t-statistic within each cluster. All clusters survive correction for family-wise error at $p = 0.05$ using threshold-free cluster enhancement (TFCE).

Table S5, related to Figure 4

	β	**SE**	***t***
Right hippocampus			
Intercept	-0.010	0.009	-1.094
Accuracy	0.031	0.012	**2.540 ***
Activation	0.008	0.005	1.478

Influence of univariate activation on ERS. A mixed-effects linear regression was computed to investigate the trial-level relationship between univariate activation, ERS and accuracy in right hippocampus. Accuracy (HC correct, incorrect) and activation were included in a model predicting memory-specific ERS (same-memory ERS – same-scene ERS) of remote memories in right hippocampus. Comparison to a model without Activation included confirmed that this variable does not reliably influence the relationship between accuracy and ERS ($\chi 2 = 2.09$ $p = 0.15$). * indicates $p < 0.05$.

Table S6, related to Figure 4

Included Trials	Contrast	dF	*t*	*p*
Remote	HC Correct > Incorrect	18	2.114	**0.048 ***
HC Correct	Remote > Recent	18	-0.125	0.905

Permutation test of ERS. The influence of memory accuracy on ERS during remote retrieval was confirmed with a non-parametric test. In this test, the number of remembered and forgotten trials were kept constant for each participant with a shuffling procedure (see STAR Methods). * indicates $p < 0.05$.

Table S7, related to Figure 5

Cluster Index	Extent	*t*	X	Y	Z
Hippocampus seed					
L. Middle frontal gyrus	1757	4.22	-40.9	28.4	24.9
L. Ventral temporal cortex, cerebellum	502	3.80	-46.2	-62.8	-11.9
L. Frontal operculum cortex	376	3.44	-44.1	9.46	1.33
L. Superior lateral occipital cortex	310	3.40	-34.5	-79.6	25.3
mPFC seed					
L. Supramarginal gyrus	484	3.56	-55.5	-37.9	34.8
Anterior cingulate gyrus	466	3.29	-1.21	13.3	32.4
L. Frontal pole	399	3.55	-36.5	41.9	24.5
L. Middle temporal cortex	309	3.44	-53.9	-61.2	1.65

Encoding-related changes in rest connectivity. R and L indicate right and left hemisphere. Extent is size of clusters in millimeters. X, Y, and Z coordinates indicate the center of gravity in MNI space (mm). *t* corresponds to the maximum t-statistic within each cluster. All clusters survive correction for family-wise error at $p = 0.05$ using cluster mass thresholding ($z = 2.3$).