Expanded Method

Participants. For Experiment 1, 69 participants from the University of California, Davis, were recruited to participate in two sessions for partial course credit. One participate did not return for the second session, two participants were excluded for poor performance on the retrieval/restudy test (< 50% accuracy) and two were excluded for exceptionally poor performance on the final recall test (2.0 and 2.4 SD below group mean), resulting in a final sample size of 64 participants (32 per group; 50 females, M_{age} = 20.6).

For Experiment 2, 33 individuals from the University of California, Davis, and the surrounding community were recruited to participate in two sessions in exchange for $70. Five participants were excluded from final analysis; one did not complete the first session because she fell asleep, one opted to come out of the scanner during session 1, two exhibited excessive motion, and one was run on the wrong experimental program during Day 2. The final sample size included 28 participants (13 females, M_{age} = 22.8).

Materials. 120 objects were selected from Konkle, Brady, Alvarez, and Oliva (2010a). Additionally, 60 scene images were selected from Konkle, Brady, Alvarez, and Oliva (2010b).

Procedure. An institutional review board at the University of California, Davis, reviewed the experimental procedure and provided approval. Participants were provided with informed consent prior to the start of the experiments. On Day 1, participants completed 8 (Experiment 1) or 10 (Experiment 2) blocks involving encoding and practice. During encoding, participants were shown 12 scene-object pairs along with their verbal labels (see Fig. 1); two back-to-back trials shared the same scene and participants were to imagine each object in that scene to encourage integration. Scene-object pairs were presented for 4 s each with a 2-s interstimulus interval; at the 2-s mark, participants were prompted to indicate with a key press whether they could imagine the

www.pnas.org/cgi/doi/10.1073/pnas.1800006115

object in the scene, to encourage engagement. All pairs were shown a second time in the same order to ensure strong encoding.

Following encoding, participants performed retrieval or restudy on one of the two objects paired with each of the scenes. This practice was done 1 or 3 times in Experiment 1, and 3 times in Experiment 2. For all practice trials, the scene image was shown. For retrieval trials, participants also saw a one-letter word stem (Fig. 1); after 2 s, they were to indicate how well they remembered the object using a response scale from 1 ("Can picture it") to 4 ("Don't remember"). Feedback was not provided because performance was expected to be very high, given that participants were only learning 12 items in each block. For restudy trials, participants saw the full word; after 2 s, they were told which key to press (randomly selected on each trial), to control for the motor response and to encourage engagement.

After participants had complete all practice trials, they performed a final task in which they were prompted for each object to indicate the second letter of the word. This final task was performed both for retrieval and restudy trials, and it served as a manipulation check to ensure that their retrieval responses were accurate. For this test, participants saw 3 letter options (keys 1-3) and a "?" (key 4), which they should use if they could not remember the word. These test trials were not included in the fMRI analyses (performance was at ceiling on the letter identification test; Experiment 1: .92 and .95 for retrieval, .98 and .98 for restudy; Experiment 2: .97 for retrieval, .98 for restudy). In Experiment 1, participants were given feedback on proportion accurate after each block to encourage correct responding and task engagement. They were also asked to indicate how well they had integrated the objects into the scene, to ensure compliance. After completing a block, participants in both experiments were given a self-paced break before moving on to the next block.

Participants returned the next day to complete a cued-recall test. During the test, participants were shown a scene image along with a one-letter word stem to prompt recall of the correct word. The test was self-paced with a maximum of 20 s per trial. Retrieval and restudy targets were tested after all non-practiced objects to reduce any impact of output interference.

fMRI.

Image acquisition. MRI data were acquired using a 3T Siemens Trio scanner with a 32-channel headcoil. High-resolution T1-weighted structural images were acquired using a magnetization prepared rapid acquisition gradient echo (MPRAGE) pulse sequence (FOV = 256 mm; image matrix = 256 x 256; axial slices = 208; thickness = 1 mm). Functional images were acquired using a multi-band gradient echo planar imaging (EPI) sequence (TR = 1220 ms; TE = 24 ms; FOV = 192 mm; image matrix = 64 x 64; flip angle = 67; mult-band factor = 2; axial slices = 38; voxel size = 3 x 3 x 3 mm).

ROI selection. Anatomical ROIs were identified individually for each participant in native space. Medial temporal lobe ROIs (HC, PHC, PRC) were manually traced using the guidelines supplied by Frankó et al. [3], and cortical ROIs were parcellated using Freesurfer [4,5] (Destrieux atlas). PM ROIs included PCC, precuneus, angular gyrus, RSC, and PHC. AT ROIs included OFC, PRC, and temporal pole. Manual tracing also included segmenting the HC into head, body, and tail.

Image pre-processing. SPM12 was used to pre-process images. Functional images were realigned, corrected for motion, and re-sliced. MPRAGE images and the ROI tracings were co-registered to the 7th volume of the first EPI scan (Artifact Detection Tools (NITRIC.org) was used to detect bad volume; one participant's data were co-registered to the 8th volume instead). Single trial estimates were extracted using LS-S method [6].

RSA analysis. All RSA analyses were run on unsmoothed native-space images. The RSA Toolbox [7] was used to generate dissimilarity matrices for each ROI for each participant using Pearson's *r*. These dissimilarity matrices were inverted and the similarity values of interest were extracted and averaged for each condition. For each of the networks, we assessed reactivation in each ROI independently, and then averaged across all ROIs to ensure that all ROIs were equally weighted in the analysis. Pattern similarity was assessed for all trials irrespective of accuracy during retrieval practice or the final test; this was done to ensure equal treatment of restudy and retrieval practice conditions.

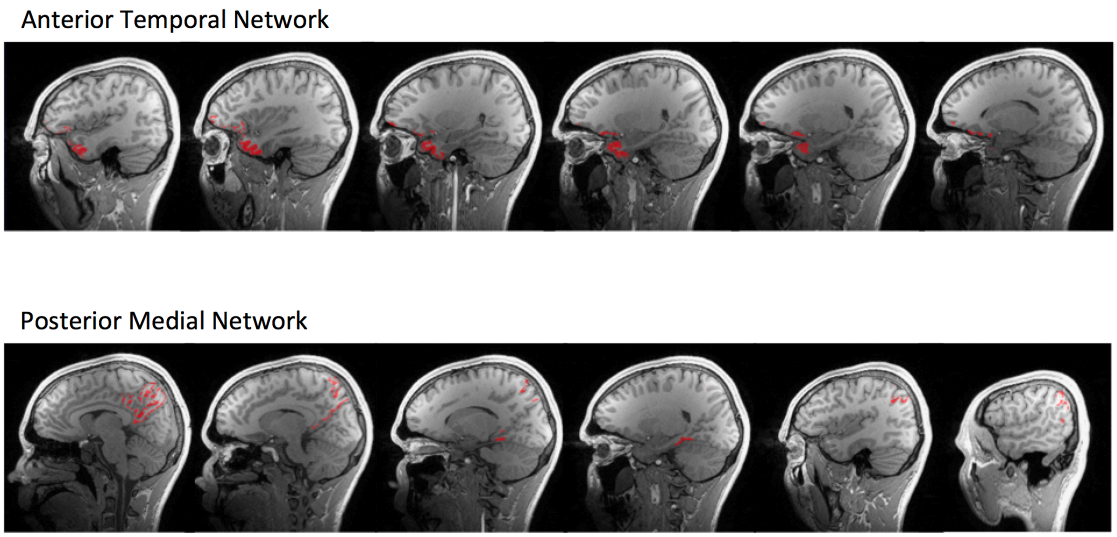


Figure S1 Examples of the AT and PM networks from one participant.

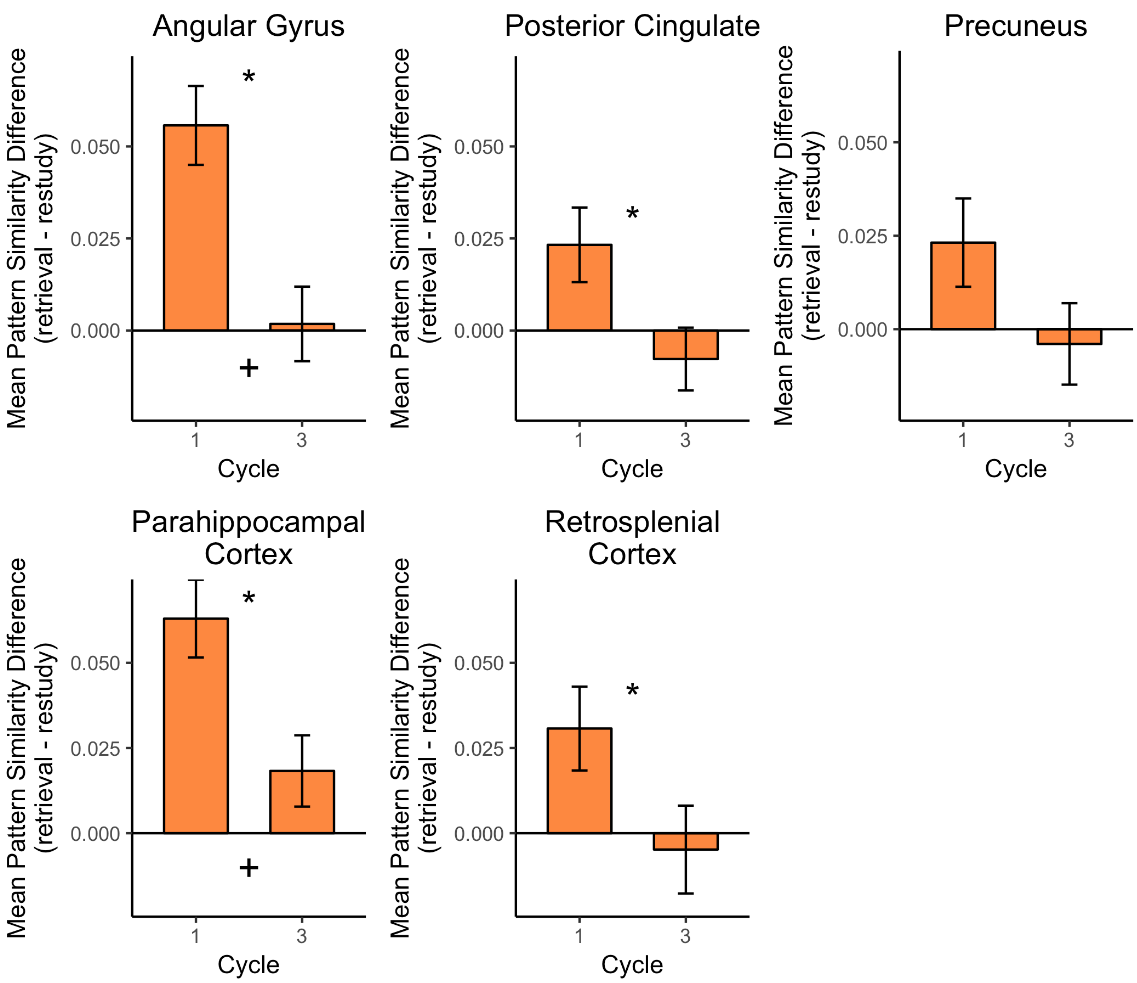


Figure S2 Pattern similarity for the testing effect drops off at later practice cycles. Asterisk represents a significant interaction between Practice Type (retrieval versus restudy) and Cycle; plus sign represents a significant main effect of Practice Type (retrieval > restudy).

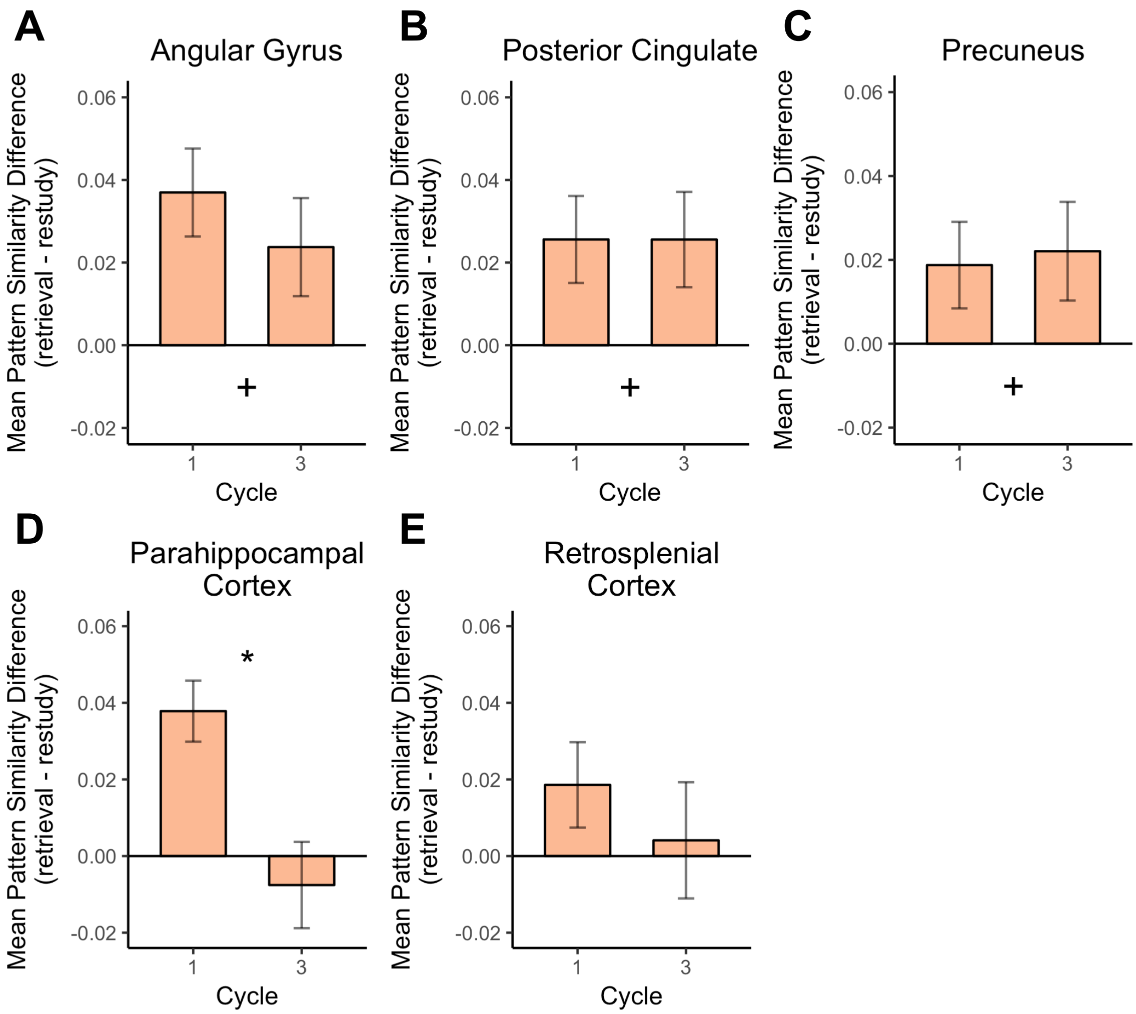


Figure S3 Pattern similarity for RIFA is sustained across cycle in the angular gyrus, PCC, and precuenus, whereas the PHC and RSC showed no sustained effect. Asterisk represents a significant interaction between Practice Type (retrieval versus restudy) and Cycle; plus sign represents a significant main effect of Practice Type (retrieval > restudy).

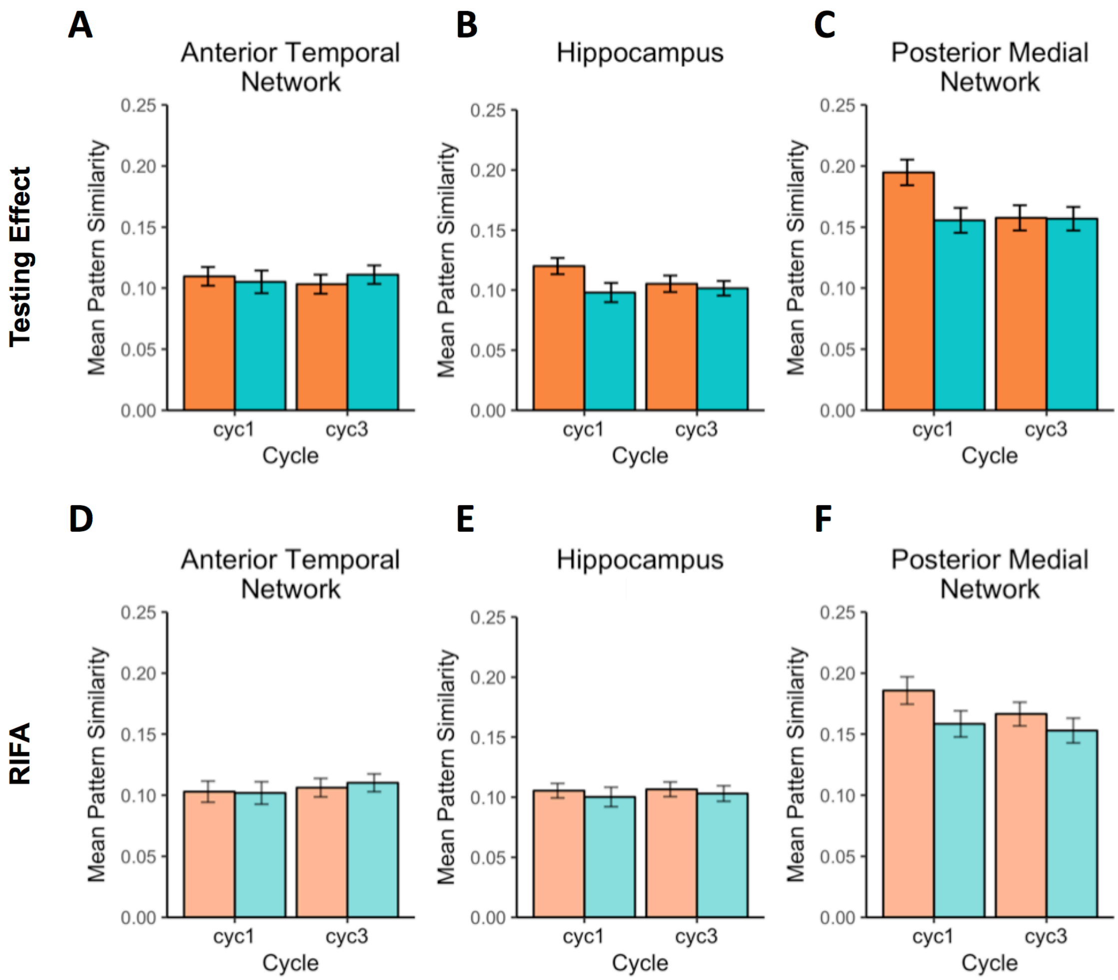
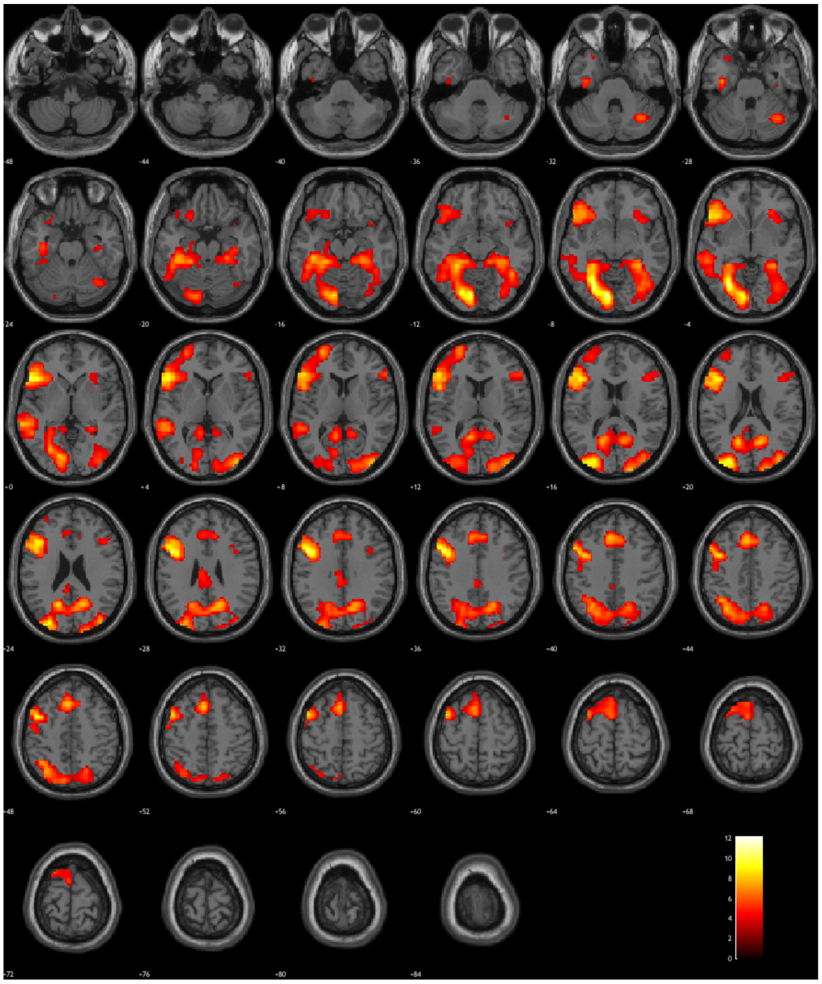


Figure S4 Pattern similarity by condition type (orange is retrieval, teal is restudy). This figure expands the results of Figure 3, which displays only difference scores. Panels A-C present results for the testing effect; Panels D-F present results for RIFA.

A **Retrieval > Restudy**



B **Restudy > Retrieval**

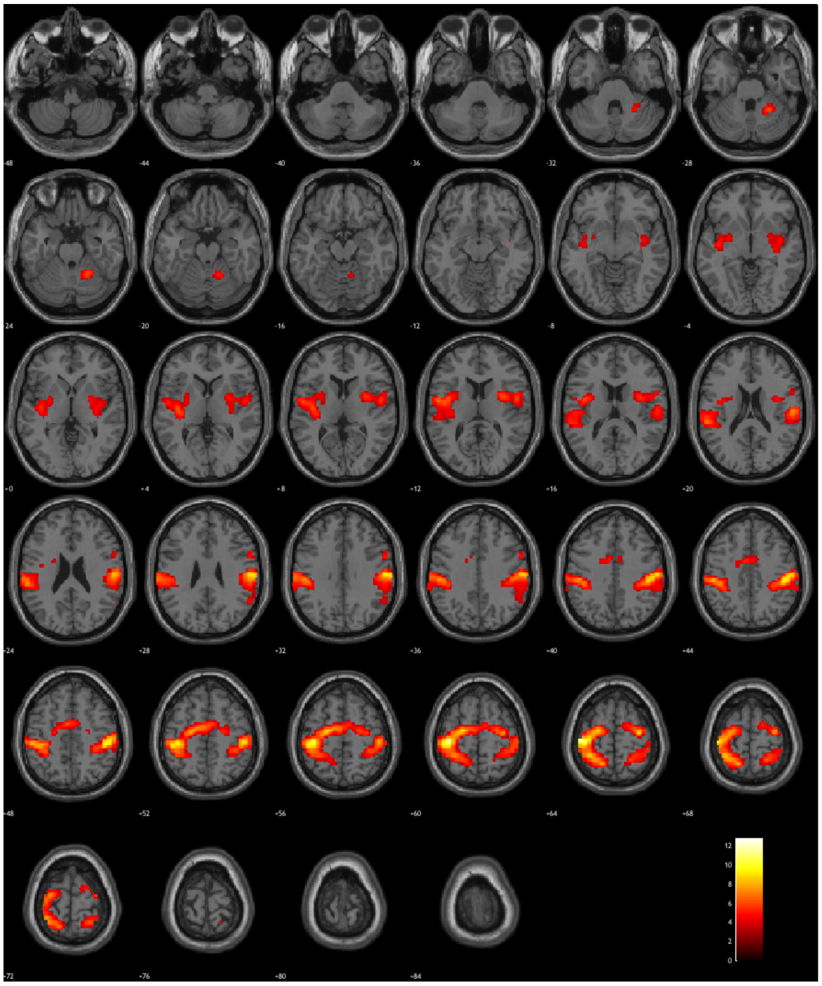


Figure S5 Univariate activation during the Practice phase across all cycles.

Table S1

Peak activation in significant clusters comparing retrieval to restudy during the practice phase.

	MNI coordinates				Cluster Size
Region	x	y	z	*t*	*k*
Retrieval > Restudy					
	-9	-82	-10	12.20	6054
	-54	23	2	11.51	2987
	57	29	8	5.53	159
	30	29	-4	5.20	90
Restudy > Retrieval					
	-39	-25	62	12.78	4687
	24	-52	-25	6.18	100

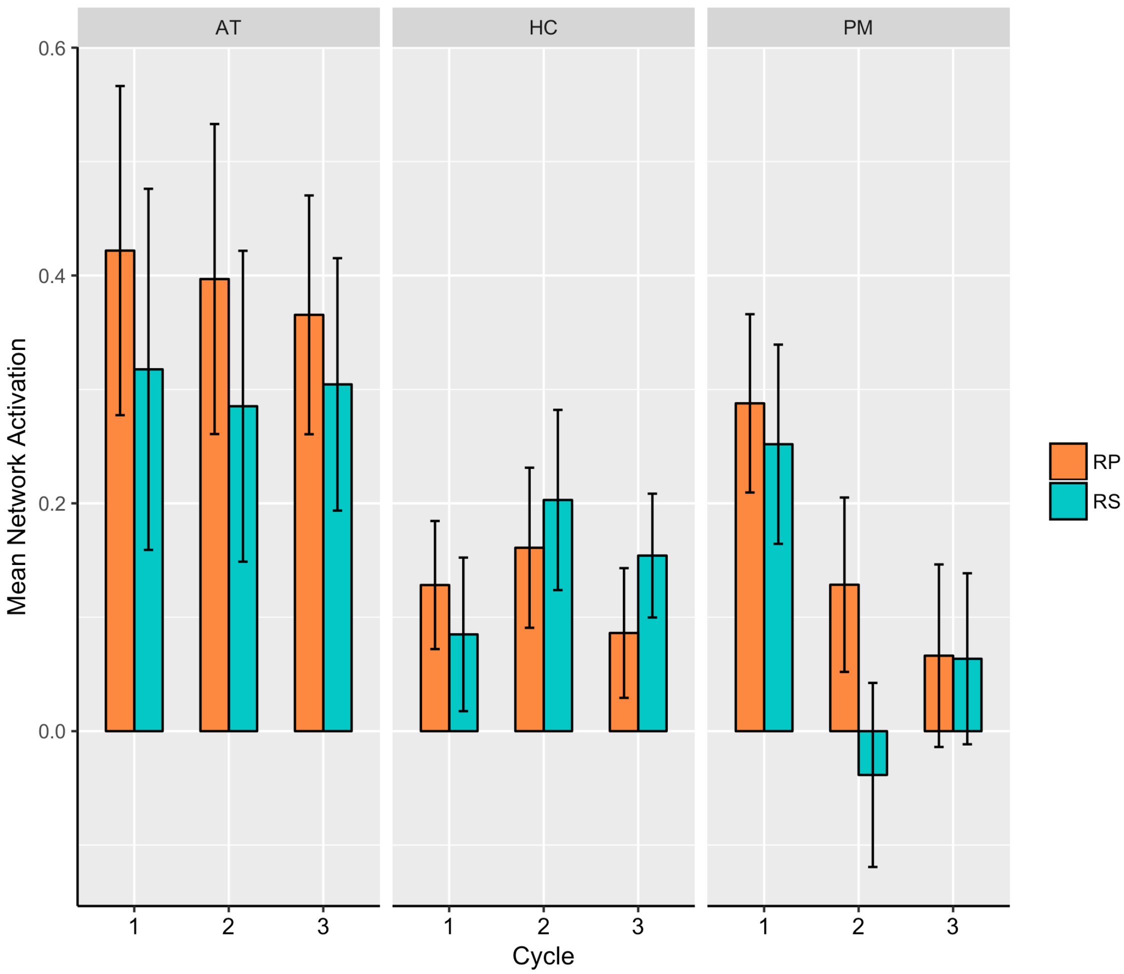


Figure S6 Mean univariate activation (condition > baseline) across voxels within each network by practice type and practice cycle. Average activation was extracted for each participant using his/her own network or hippocampal mask in standard space. A significant interaction between Cycle and Practice Type (retrieval vs. restudy) was found in the hippocampus, [$F(1.74,46.96) = 3.55$, $P = .04$]), and a significant main effect of Cycle was found in the posterior medial network, [$F(1.61,43.34) = 4.81$, $P = .02$]). All other main effects and interactions failed to achieve significance.

1. Konkle, T., Brady, T. F., Alvarez, G. A. & Oliva, A. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. Exp. Psychol. Gen.* **139,** 558–578 (2010).

2. Konkle, T., Brady, T. F., Alvarez, G. A. & Oliva, A. Scene Memory Is More Detailed Than You Think. *Psychol. Sci.* **21,** 1551–1556 (2010).

3. Frankó, E., Insausti, A. M., Artacho-Pérula, E., Insausti, R. & Chavoix, C. Identification of the human medial temporal lobe regions on magnetic resonance images. *Hum. Brain Mapp.* **35,** 248–256 (2014).

4. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31,** 968–980 (2006).

5. Fischl, B. *et al.* Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14,** 11–22 (2004).

6. Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* **59,** 2636–43 (2012).

7. Nili, H. *et al.* A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10,** e1003553 (2014).