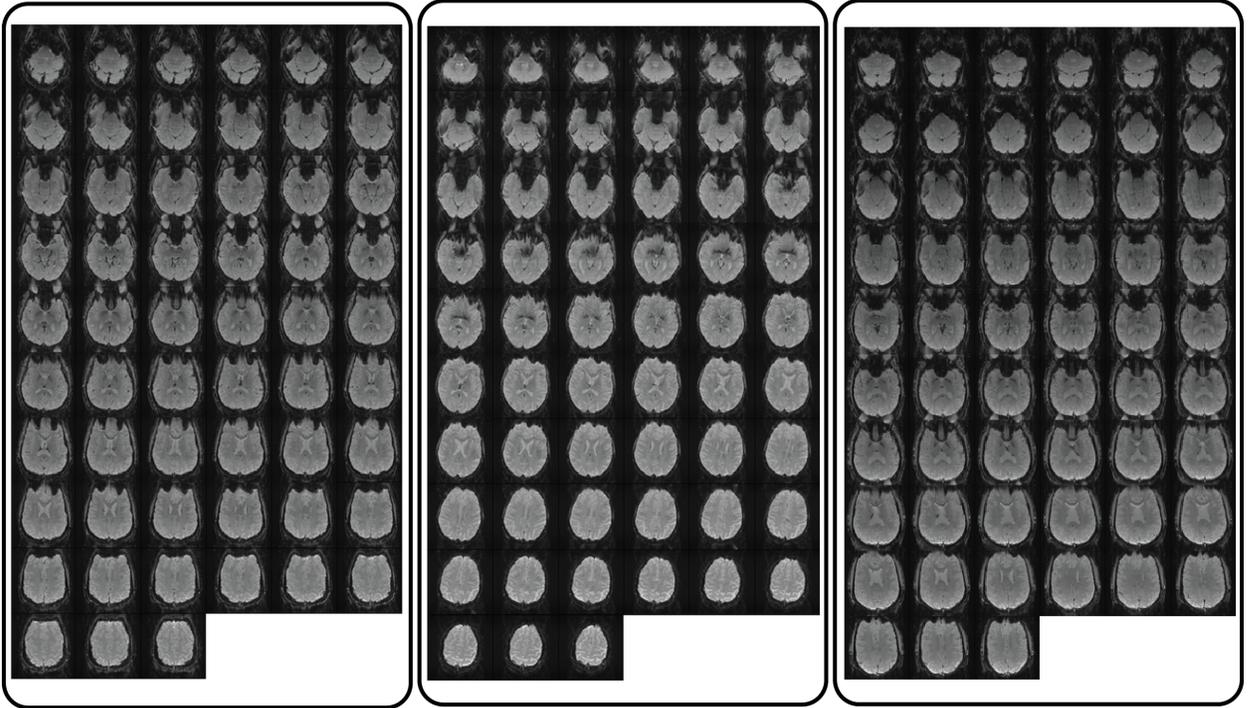
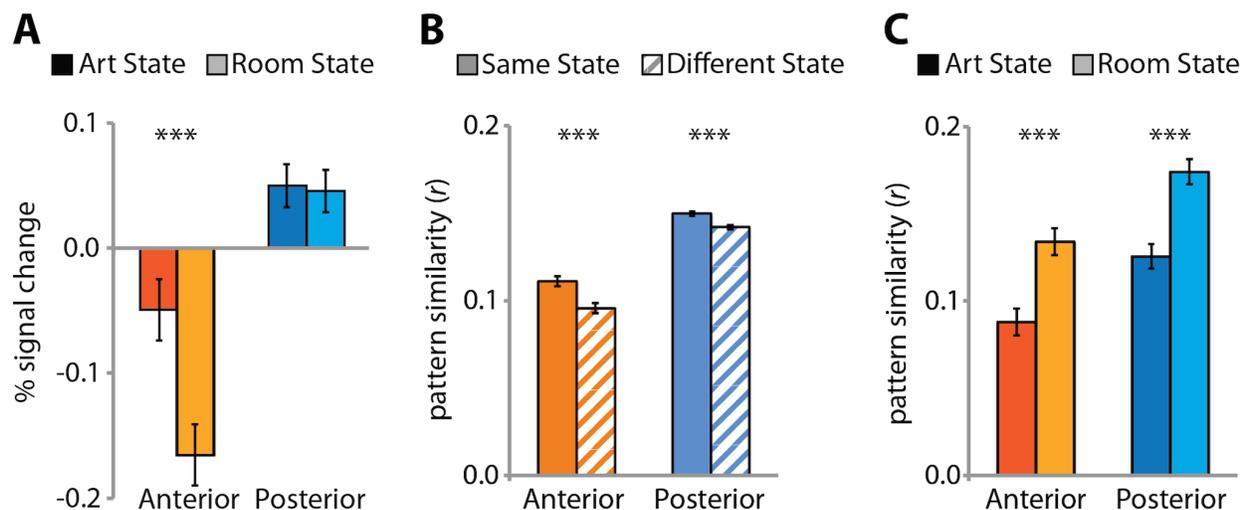


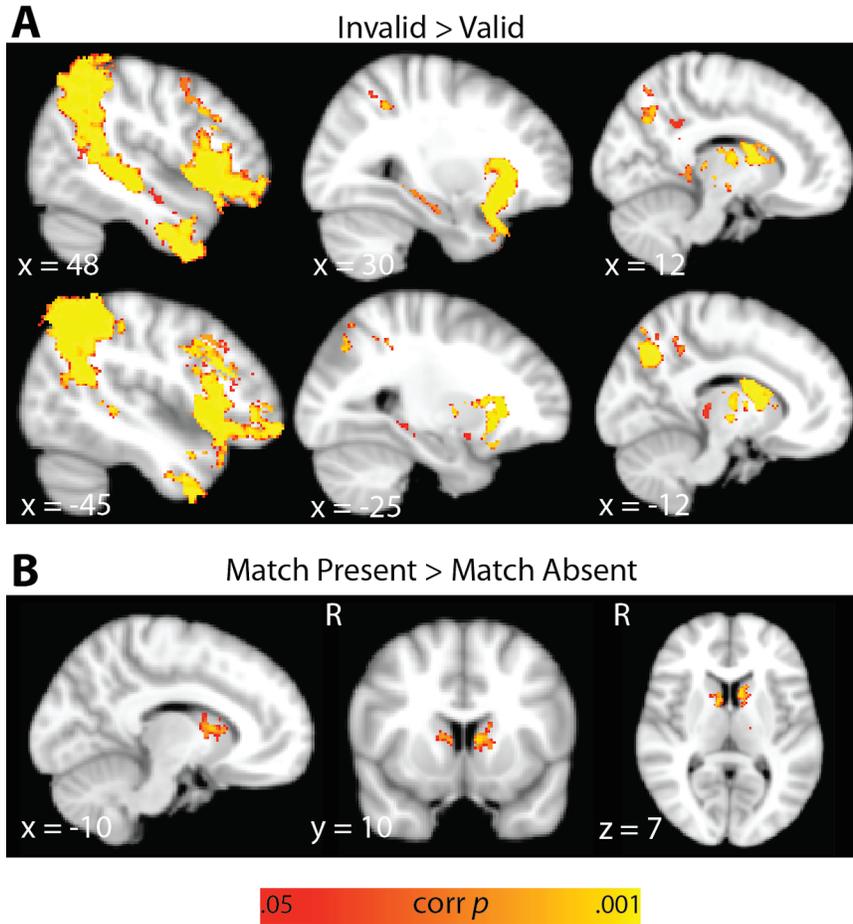
## Supplementary Figures and Tables



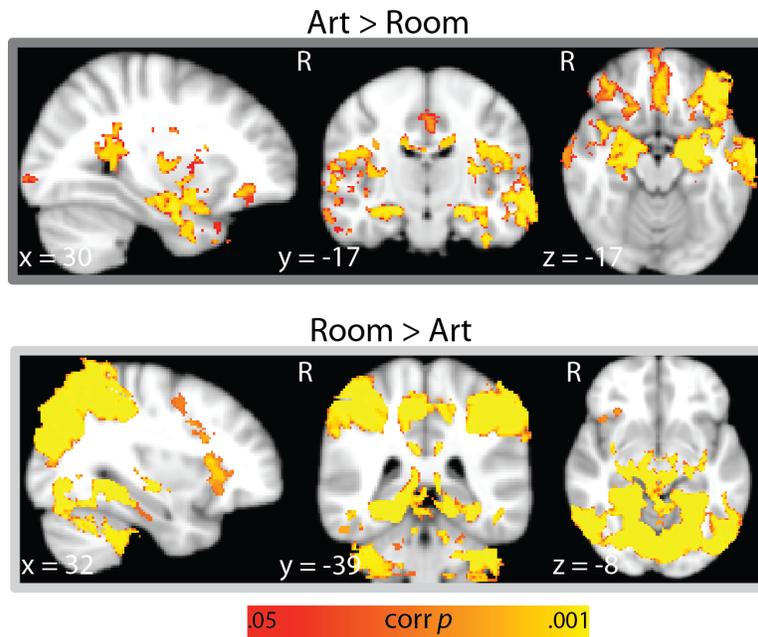
**Supplementary Figure 1.** Representative mean functional scans for three participants, collected using a multiband EPI sequence. Parameters for this sequence are given in Methods.



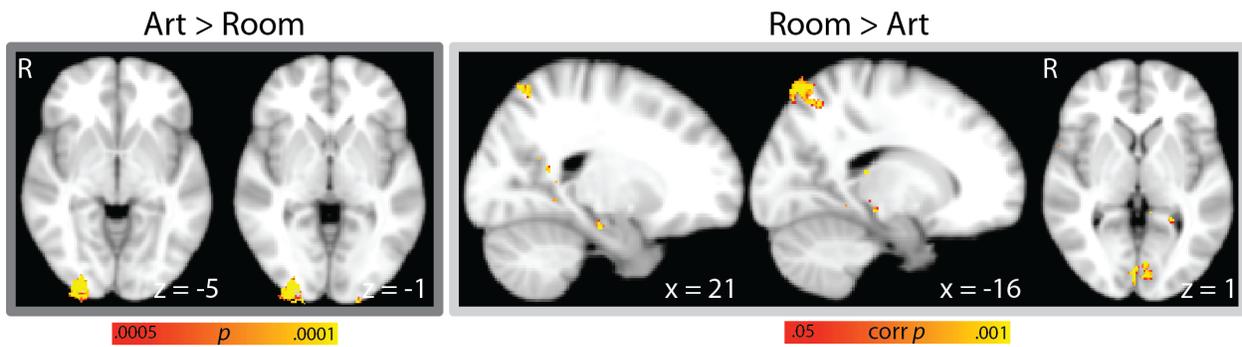
**Supplementary Figure 2.** Univariate and multivariate results for anterior and posterior hippocampus. **(A)** Univariate activity. **(B)** Pattern similarity for same vs. different states. **(C)** Pattern similarity for art vs. room states. Attentional modulation of univariate activity was present in anterior [ $t(18) = 4.73, p = .0002$ ] but not posterior hippocampus [ $t(18) = .25, p = .80$ ], whereas modulation of pattern similarity was observed in both anterior [same vs. different state:  $t(18) = 5.57, p < .0001$ ; art vs. room state:  $t(18) = 5.95, p < .0001$ ] and posterior hippocampus [same vs. different state:  $t(18) = 5.81, p < .0001$ ; art vs. room state:  $t(18) = 6.63, p < .0001$ ]. Error bars depict  $\pm 1$  SEM of the within-subject difference. Results for (B) and (C) are shown as Pearson correlations, but statistical tests were performed only after applying the Fisher transformation. \*\*\* =  $p < .001$ .



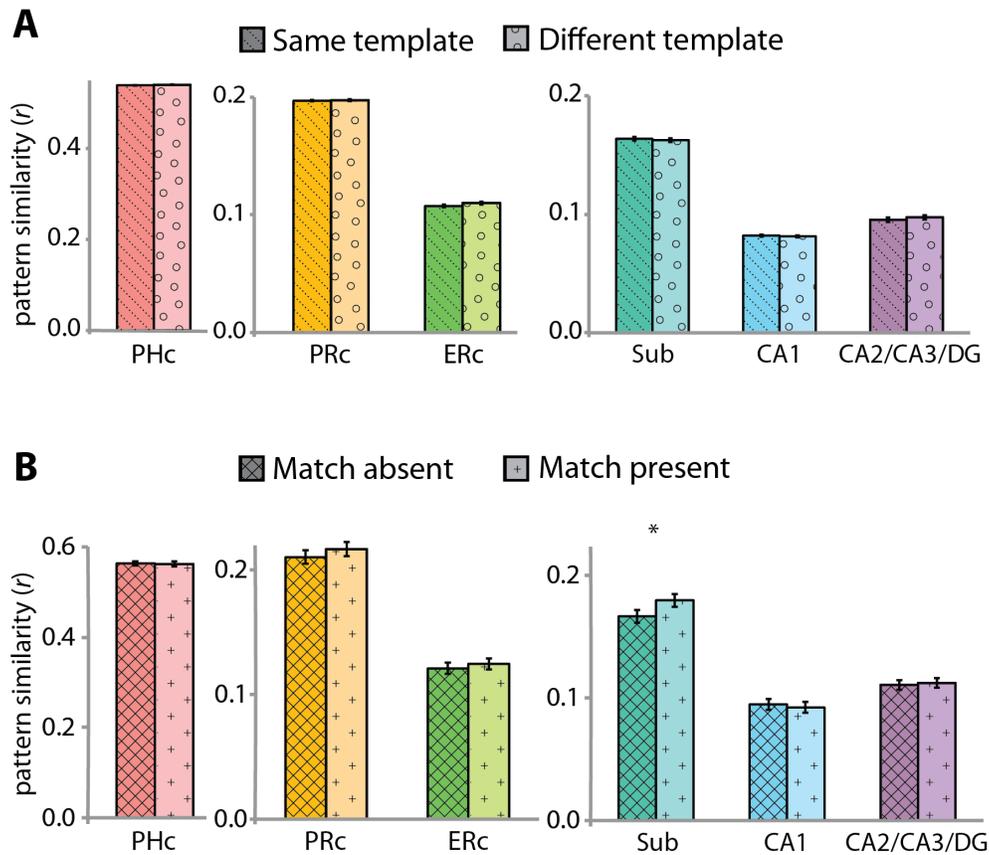
**Supplementary Figure 3.** Whole-brain univariate analyses. **(A)** Several regions, including the key nodes of the reorienting network (Corbetta and Shulman 2002), showed increased activity for invalid compared to valid probes ( $p < .05$  TFCE corrected): bilateral inferior frontal gyrus, orbitofrontal cortex, inferior parietal lobule, superior parietal lobule, intraparietal sulcus, precuneus, superior temporal sulcus, middle and inferior temporal gyri, temporal pole, hippocampus, thalamus, and caudate nucleus. Not shown are activations in the bilateral putamen, globus pallidus, and cingulate cortex. **(B)** Presence vs. absence of the task-relevant match on validly-cued trials. The caudate nucleus, a key region of the striatal reward circuit (Delgado 2007), showed greater activity for match present vs. match absent trials ( $p < .05$  TFCE corrected). There were also small clusters (not shown) in the left thalamus and left insula.



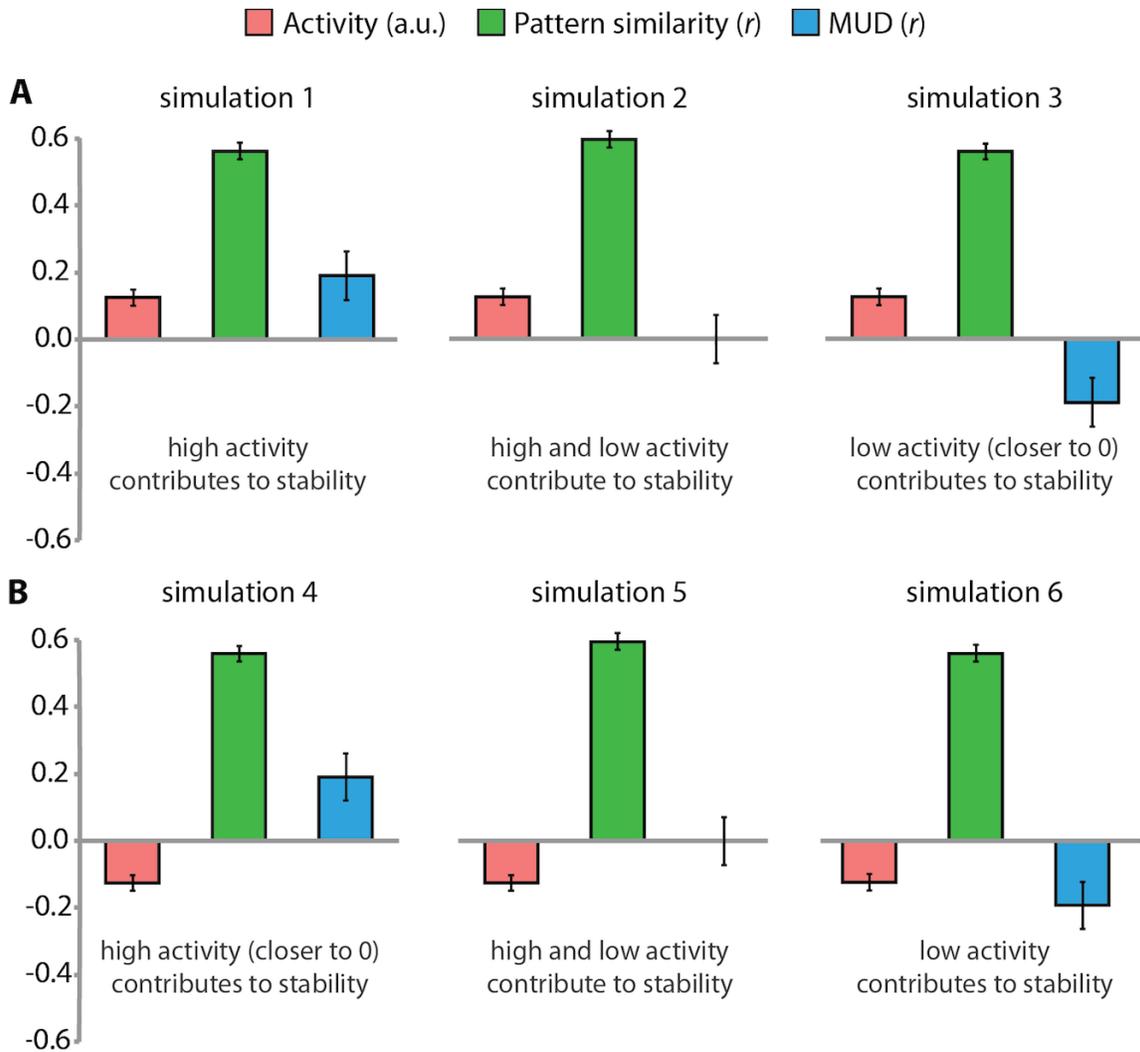
**Supplementary Figure 4.** Whole-brain univariate analysis of art vs. room states. Regions showing greater activity for the art compared to the room state (top) were distributed anteriorly in the brain ( $p < .05$  TFCE corrected): bilateral superior temporal sulcus, temporal pole, hippocampus, perirhinal cortex, amygdala, putamen, insula, orbitofrontal cortex, as well as right early visual cortex. Regions showing greater activity for the room compared to the art state (bottom) were distributed posteriorly: bilateral primary visual cortex, thalamus, lateral occipital cortex, lingual gyrus, parahippocampal gyrus, precuneus, posterior cingulate/retrosplenial cortex, intraparietal sulcus, inferior parietal lobule, and superior parietal lobule. The whole-brain distribution of univariate activity for the art vs. room states highlights potential differences in the nature of these tasks. For example, the art task may have required more abstraction than the room task, which is consistent with the greater involvement of the orbitofrontal cortex and temporal pole (e.g., [Peelen and Caramazza 2012](#); [Wilson et al. 2014](#)). Additionally, the room vs. art task may have involved more spatial processing, consistent with increased activity in parietal areas linked to spatial attention, the intraparietal sulcus and inferior parietal lobule ([Corbetta and Shulman 2002](#)).



**Supplementary Figure 5.** Whole-brain searchlight analysis of art- vs. room-state pattern similarity. At the corrected threshold, no regions showed significantly greater pattern similarity for the art compared to the room state; lowering the threshold ( $p < .0005$  uncorrected) revealed an effect in left and right early visual cortex. Greater pattern similarity for the room compared to the art state was found in several regions ( $p < .05$  FWE corrected): bilateral primary visual cortex, superior parietal lobule, posterior cingulate (not shown), and right hippocampus.



**Supplementary Figure 6.** Supplementary pattern similarity analyses. **(A)** Pattern similarity for trials generated from the same vs. different templates. BOLD activity was extracted from all voxels in each ROI and correlated across trials generated from the same template and from different templates (as a reminder, trials from the same template have the same “base” image, art and room matches, and sets of possible distractors). There were no differences between these trial types in any ROI. An additional analysis restricted the calculation of pattern similarity for same and different templates to trials from the same state. The resulting similarity scores were submitted to a 2 (template: same vs. different) by 2 (state: art vs. room) ANOVA. There was a main effect of template in PHc, but it did not survive correction for multiple comparisons [ $F(1, 18) = 4.61, p = .05$ ]; no other ROI showed a main effect of template [PRc:  $F(1, 18) = 2.09, p = .17$ ; ERc:  $F(1, 18) = 0.68, p = .42$ ; Sub:  $F(1, 18) = 3.98, p = .06$ ; CA1:  $F(1, 18) = 1.12, p = .30$ ; CA2/CA3/DG:  $F(1, 18) = 1.10, p = .31$ ], nor were there any template x state interactions [all  $ps > .30$ ]. These results suggest that the hippocampal pattern similarity effects are related to general attentional states that are shared across stimuli, rather than bottom-up sensory information. **(B)** Pattern similarity for trials in which the task-relevant match was absent vs. present. This analysis was restricted to trials of the same attentional state. Specifically, we examined correlations among valid art trials in which an art match was present and separately among trials in which an art match was absent and, likewise, correlations among valid room trials in which a room match was present and absent. Subiculum showed increased pattern similarity on match present vs. absent trials [ $t(18) = 2.32, p = .03$ ], but this effect did not survive correction for multiple comparisons. Error bars depict  $\pm 1$  SEM of the within-subject difference. Results are shown as Pearson correlations, but statistical tests were performed only after applying the Fisher transformation. \* =  $p < .05$  (uncorrected).



**Supplementary Figure 7.** Results of simulations designed to test the MUD analysis. **(A)** Simulations demonstrating that a given level of positive univariate activity and positive pattern similarity can be associated with positive (left), zero (middle), and negative (right) MUD correlations. **(B)** Simulations demonstrating that a given level of negative univariate activity and positive pattern similarity can be associated with positive (left), zero (middle), and negative (right) MUD correlations. Each bar depicts the mean ( $\pm 1$  SD) across 1000 iterations of each simulation. Activity is in arbitrary units; pattern similarity and MUD are Fisher-transformed correlations. Details of the purpose of each simulation are reported in Supplementary Experimental Procedures.

		MUD effect		
		+	0	-
Univariate activity	+	<p>simulation 1</p> <p>50% of voxels: stable across patterns <math>N(0.25,1)</math></p> <p>50% of voxels: random across patterns <math>N(0,1)</math></p>	<p>simulation 2</p> <p>25% of voxels: stable across patterns <math>N(-0.25,1)</math></p> <p>25% of voxels: stable across patterns <math>N(0.50,1)</math></p> <p>50% of voxels: random across patterns <math>N(0.125,1)</math></p>	<p>simulation 3</p> <p>50% of voxels: stable across patterns <math>N(0,1)</math></p> <p>50% of voxels: random across patterns <math>N(0.25,1)</math></p>
	-	<p>simulation 4</p> <p>50% of voxels: stable across patterns <math>N(0,1)</math></p> <p>50% of voxels: random across patterns <math>N(-0.25,1)</math></p>	<p>simulation 5</p> <p>25% of voxels: stable across patterns <math>N(-0.50,1)</math></p> <p>25% of voxels: stable across patterns <math>N(0.25,1)</math></p> <p>50% of voxels: random across patterns <math>N(-0.125,1)</math></p>	<p>simulation 6</p> <p>50% of voxels: stable across patterns <math>N(-0.25,1)</math></p> <p>50% of voxels: random across patterns <math>N(0,1)</math></p>

**Supplementary Table 1.** MUD simulation details. For a given level of pattern similarity, we varied whether univariate activity was positive or negative (rows) and the distribution of activity across voxels that were stable ( $r = 1$ ) vs. random ( $r = 0$ ) over patterns. The resulting MUD effect is listed in columns. Each cell describes the details of one simulation. The normal distribution from which activity levels were drawn is described in terms of mean and standard deviation; for example,  $N(0,1)$  refers to the standard normal distribution with mean 0 and standard deviation 1. For each simulation, half of the voxels had stable activity across patterns and the remaining half had variable activity across patterns. In simulations 1 and 4, voxels with higher relative activity contributed more to pattern similarity. In simulations 2 and 5, voxels with higher and lower activity contributed equally to pattern similarity. In simulations 3 and 6, voxels with lower relative activity contributed more to pattern similarity. For more details, see Supplementary Experimental Procedures and Supplementary Fig. 7.

## Supplementary Methods

### MTL segmentation guide

Manual segmentation of hippocampal subfields and MTL cortex were done using published criteria ([Insausti 1993](#); [Insausti et al. 1998](#); [Goncharova et al. 2001](#); [Pruessner et al. 2002](#); [Duvernoy 2005](#); [Mueller and Weiner 2009](#); [Yushkevich et al. 2010](#)). All ROIs were traced on coronal slices using FSLview. MTL cortical ROIs were the parahippocampal cortex (PHc), perirhinal cortex (PRc), and entorhinal cortex (ERc). Hippocampal subfield ROIs were subiculum, CA1, and the combination of CA2, CA3, and dentate gyrus (DG). These subfields were traced along the entire length of the hippocampus (also see [Schapiro et al. 2012](#)).

The first step was to identify the transition between the head and body of the hippocampus, because the boundaries of MTL cortical areas depended in part on this transition. The last slice containing the head of the hippocampus was defined as the last slice in which the uncal apex was visible ([Duvernoy 2005](#); [Poppenk et al. 2013](#)). All subfields in the head of the hippocampus comprised the anterior hippocampal ROI in the analyses shown in Supplementary Fig. 2. All subfields in the body and tail of the hippocampus comprised the posterior hippocampal ROI.

The anterior boundary of the PRc was determined by the appearance of the collateral sulcus, and the posterior boundary was the last slice containing the head of the hippocampus. In the most anterior PRc slice, the dorsal boundary was just past the gyrus of Schwalbe; if that gyrus was not present, the boundary was at the midpoint between the medial and lateral edges of the temporal cortex ([Insausti et al. 1998](#)). In subsequent slices without the ERc, the dorsal boundary was perpendicular to the medial edge of the temporal cortex. The ventral boundary of the PRc was perpendicular to the bend at the lateral bank of the collateral sulcus.

The anterior boundary of the ERc was the slice anterior to the limen insulae (frontotemporal junction; [Goncharova et al. 2001](#)). The posterior boundary was the last slice containing the head of the

hippocampus. The ventrolateral boundary of the ERc was perpendicular to the bend at the medial bank of the collateral sulcus; this was the border with PRc. The dorsomedial boundary was the projected continuation of the inferior boundary of the hippocampus, perpendicular to the dorsomedial edge of the temporal cortex. This was the border with subiculum.

The anterior boundary of the PHc was the first slice with the body of the hippocampus. The posterior boundary was the last slice in which the hippocampus was identifiable. The ventrolateral boundary was perpendicular to the bend at the lateral bank of the collateral sulcus, and the dorsomedial boundary was the border with subiculum (as with ERc in anterior portions of the hippocampus).

The anterior boundary of the subiculum was the first slice in which the hippocampal head could be discerned, and the posterior boundary was the last slice in which the hippocampus was visible. The medial border of the subiculum was ERc in slices containing the hippocampal head and PHc in slices containing the hippocampal body and tail (i.e. the projected continuation of the inferior boundary of the hippocampus, perpendicular to the dorsomedial edge of the temporal cortex). At its most anterior slice, the subiculum comprised the entire ventral aspect of the hippocampus ([Duvernoy 2005](#)); the lateral boundary (with CA1) gradually moved medially until, at the body of the hippocampus, the lateral boundary was at the medial edge of the hippocampus at the point where it pinches into a tear shape.

The anterior boundary of CA1 was the first slice of the hippocampal head and the posterior boundary was the last slice in which the hippocampus was visible. The ventral boundary with subiculum was defined above. CA1 curved around the lateral edge of the hippocampus and bordered CA2/3 at the dorsal aspect of the hippocampus. The transition point between CA1 and CA2/3 was determined by the thickness of CA1 at that slice (see [Mueller and Weiner 2009](#)). The boundary between the CA fields and DG was defined by the presence of a dark band, comprised of stratum moleculare, lacunosum, and radiatum. DG was not clearly identifiable in the most anterior slices of the

hippocampus (Duvernoy 2005), so DG typically began 2-3 slices from the anterior extent of the hippocampus, and the posterior boundary was the last slice in which the hippocampus was visible.

### **Whole-brain analyses**

*Univariate analyses.* The GLM from the main ROI analyses (including regressors for valid and invalid art- and room-state trials) was also used to examine whole-brain activity. First-level parameter estimates were entered into second-level fixed-effects analyses for each participant and registered to 1.5 mm MNI space. Registration was performed using the field maps and the brain-extracted MPAGE image. Second-level contrast images for valid art vs. valid room trials were then entered into third-level random-effects analyses across participants and corrected for multiple comparisons with threshold-free cluster enhancement (TFCE; Smith and Nichols 2009). The resulting corrected  $p$  maps were thresholded at  $p < .05$ .

Two additional GLMs were run as manipulation checks. First, to verify that participants used the cues, we compared brain activity for invalid vs. valid probes: insofar as the cues caused orienting, then an invalid probe would require reorienting more than a valid probe and engage the ventral frontoparietal attention network (Corbetta and Shulman 2002). This GLM included 4 regressors: invalid and valid probe/response periods (modeled as 2-s epochs from probe onset), missed responses (modeled in the same way), and the image presentation period of all trials (modeled as 8-s epochs from cue onset to offset of the last image). The contrast of interest was invalid > valid probes. Second, to complement the behavioral data and show that participants detected the match image, we compared brain activity for trials in which there was vs. was not a match to the target in the search set: insofar as these matches were detected, then they might elicit activity in reward systems such as the striatum (Delgado 2007). This GLM included 6 regressors: match present and absent, both when the probe was valid and invalid (modeled as 8-s epochs from cue onset to offset of the last image), probe/response periods, and missed responses. Match present/absent was defined with reference to

the probe (e.g., art match present and art probe would be match present). However, we were only interested in cases where the cue matched the probe (i.e., valid trials), and thus the contrast of interest was valid match present > valid match absent.

*Multivariate analyses.* Pattern similarity analyses were performed with a searchlight approach over the whole brain. The process was the same as for the ROIs, with correlations computed over first-level parameter estimates from the single-trial GLM, but this was repeated for all possible 27-voxel cubes (3 x 3 x 3) in the brain. The results for each cube were assigned to the center voxel. Group analyses consisted of random-effects non-parametric tests (using randomise in FSL), and corrected for family-wise error (FWE) across all voxels. We changed from a cluster-based to a (more conservative) voxel-based correction because of the spatial autocorrelation introduced by the searchlight procedure; although cluster-based methods attempt to incorporate smoothness in the null distribution, we found that the searchlight analysis resulted in unrealistically large clusters. The room > art contrast was thresholded at  $p < .05$  corrected. No voxels survived this correction for the art > room contrast; since this was an exploratory analysis, we relaxed the threshold to  $p < .0005$  uncorrected to detect whether there were any sub-threshold effects.

*Brain/behavior correlations.* For both the art and room tasks, we correlated each voxel's pattern similarity value (from the searchlight analysis) with behavioral performance in the corresponding task across individuals. Group analyses consisted of random-effects non-parametric tests (using randomise in FSL), threshold at  $p < .05$  corrected. This whole-brain correlation analysis did not yield any reliable results (unlike the correlation analysis for the CA2/CA3/DG ROI).

## **MUD simulations**

We ran simulations to verify that the MUD analysis was able to detect different types of relationships between univariate activity and pattern similarity. In these simulations, the region of interest was represented by a vector of 1000 "voxels", with several "patterns" constructed by randomly

assigning activity levels to each voxel with the constraints listed in Supplementary Table 1. Univariate activity was defined as the mean activity level averaged across voxels and patterns. Pattern similarity was defined as the mean pairwise Pearson correlation between the patterns. The MUD effect was defined as the mean pairwise Pearson correlation between the patterns. The MUD effect was calculated by correlating the univariate activity of each voxel averaged over patterns with the pairwise product of normalized activity averaged over pairs of patterns. The whole process was repeated 1000 times for each simulation.

We conducted six simulations, in which we manipulated the level of activity over voxels as a function of whether this activity was stable across patterns (Supplementary Fig. 7). For the first three simulations, univariate activity and pattern similarity were positive: in simulation #1, the highest levels of activity were most stable across patterns, and we obtained a positive MUD effect (similar to PHc and ERc); in simulation #2, both high and low activity were stable, and we obtained a zero MUD effect; and in simulation #3, the lowest levels of activity (closer to zero because the mean was positive) were most stable, and we obtained a negative MUD effect. For the final three simulations, univariate activity was negative and pattern similarity was positive: in simulation #4, the highest levels of activity (closer to zero because the mean was negative) were most stable across patterns, and we obtained a positive MUD effect; in simulation #5, both high and low activity were stable, and we obtained a zero MUD effect (similar to CA1 and CA2/CA3/DG); and in simulation #6, the lowest levels of activity were most stable, and we obtained a negative MUD effect.

These simulations show that the sign and magnitude of the MUD effect is orthogonal to the mean univariate activity and pattern similarity. That is, a given level of univariate activity and pattern similarity can be associated with positive, zero, and negative MUD effects, depending upon how activity and stability are distributed with respect to each other.

## Eye-tracking

We deliberately employed a free-viewing paradigm in which participants were allowed to move their eyes. However, we measured these eye movements so that we could examine how they differed across tasks and how they related to the fMRI results. These analyses were based on 15 (of 19) participants: due to calibration problems, no or incomplete eye-tracking data were available for the other participants.

We assessed several different eye-tracking measures during the trials. There were more saccades in the room vs. art states [room: mean = 18.40, SD = 3.12; art: mean = 13.97, SD = 3.31;  $t(14) = 6.94, p < .0001$ ], and correspondingly a greater number of fixations as well [room: mean = 18.79, SD = 3.28; art: mean = 14.49, SD = 3.24;  $t(14) = 7.20, p < .0001$ ]. However, there was no difference between states in the total time spent making saccades [room: mean = .77 s, SD = .48 s; art: mean = .70 s, SD = .42 s;  $t(14) = 1.63, p = .13$ ] or fixating [room: mean = 8.56 s, SD = 1.31 s; art: mean = 8.72 s, SD = 1.26 s;  $t(14) = 1.73, p = .11$ ].

The differences in the number of saccades and fixations were not unexpected: different kinds of information were useful in the art vs. room states, so good performance required differential sampling of the images in order to find the relevant information for each task. This is why we allowed eye movements — we felt that they were an integral part of the attentional manipulation, and restricting them may have altered the nature and ecological validity of the attentional states we aimed to study. Nevertheless, it is important to verify that our fMRI results are not reflective of these eye movements *per se*. Below, we propose several reasons why the effects observed in univariate activity and pattern similarity cannot be attributed to differences in eye movements for the art vs. room states.

First, although there were differences in the number of saccades and fixations between states, this result was inconsistent and did not apply to all eye movement measures. Namely, as reported above, there was no difference in the total amount of time spent fixating or moving one's eyes during trials in the two states. These two results can be reconciled by the fact that saccades in the room state

were shorter in duration than in the art state. Regardless, the lack of a difference in overall fixation time means that the fMRI results do not reflect greater sensory input in one state or the other.

Second, saccade frequency tends to uniformly increase BOLD activity across brain regions (Kimmig et al. 2001), and thus if saccades were causing our effects, we would expect increased univariate activity for the room state, and more generally, that these effects would be uniform across regions. However, increases for room vs. art states were only found in a subset of regions (PHc and ERc), and the others showed no difference (subiculum) or reduced activity for room vs. art states (PRc, CA1, CA2/CA3/DG). Indeed, double dissociations like this cannot be explained by any single factor such as saccade frequency.

Third, if saccade frequency caused our effects, then the difference in the number of saccades for room vs. art states should be related to the difference in univariate activity and pattern similarity for room vs. art states. We examined this relationship across individuals for ROIs that showed state differences, but found no reliable correlations for univariate activity ( $ps > .47$ ) and only one reliable correlation for pattern similarity (ERc:  $r = 0.78, p < .001$ ; all other ROIs,  $ps > .26$ ). For all ROIs except ERc, this is inconsistent with an effect of saccades on our key data. We are hesitant to put much stock in the ERc correlation, as there were multiple comparisons performed and none of the four other ROIs with greater pattern similarity for room vs. art states showed hints of the same relationship.

Finally, for the ROI and state where pattern similarity was correlated with behavior (room state in CA2/CA3/DG), there was no correlation between pattern similarity and saccade frequency either in a bivariate analysis [ $r = .04, p = .89$ ] or when controlling for behavioral performance [partial  $r = .14, p = .63$ ]. Indeed, the correlation between room-state behavioral performance and room-state pattern similarity in CA2/CA3/DG remained significant when controlling for the number of saccades in the room state — even with the reduced sample size of the eye-tracking analysis [partial  $r = .67, p = .002$ ].

We are glad to have collected eye-tracking data to enable these analyses, but have convinced ourselves that differences in eye movements are not responsible for the observed fMRI results.

### Supplementary References

- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neuro.* 3: 201-215.
- Delgado MR. 2007. Reward-related responses in the human striatum. *Ann NY Acad Sci.* 1104: 70-88.
- Duvernoy HM. 2005. *The human hippocampus.* Third Ed. Springer, NY.
- Goncharova II, Dickerson BC, Stoub TR, deToledo-Morrell L. 2001. MRI of human entorhinal cortex: A reliable protocol for volumetric measurement. *Neurobiol Aging.* 22: 737-745.
- Insausti R. 1993. Comparative anatomy of the entorhinal cortex and hippocampus in mammals. *Hippocampus.* 3: 19-26.
- Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A. 1998. MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Am J Neuroradiol.* 19: 659-671.
- Kimmig H, Greenlee MW, Gondan M, Schira M, Kassubek J, Mergner T. 2001. Relationship between saccadic eye movements and cortical activity as measured by fMRI: Quantitative and qualitative aspects. *Exp Brain Res.* 141: 184-194.
- Mueller SG, Weiner MW. 2009. Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. *Hippocampus.* 19: 558-564.
- Peelen MV, Caramazza A. 2012. Conceptual object representations in human anterior temporal cortex. *J Neurosci.* 32: 15728-15736.
- Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. 2013. Long-axis specialization of the human hippocampus. *Trends Cog Sci.* 17: 230-240.
- Pruessner JC, Kohler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL, Evans AC. 2002. Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: Considering the variability of the collateral sulcus. *Cereb Cortex.* 12: 1342-1353.

- Schapiro AC, Kustner LV, Turk-Browne NB. 2012. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol.* 22: 1622-1627.
- Smith SM, Nichols TE. 2009. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localization in cluster inference. *NeuroImage.* 44: 83-93.
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. 2014. Orbitofrontal cortex as a cognitive map of task space. *Neuron.* 81: 267-279.
- Yushkevich PA, Wang H, Pluta J, Das SR, Craige C, Avants BB, Weiner MW, Mueller S. 2010. Nearly automatic segmentation of hippocampal subfields in *in vivo* focal T2-weighted MRI. *NeuroImage.* 53: 1208-1224.