Cerebral Cortex, February 2016;26: 783–796

doi: 10.1093/cercor/bhv041
Advance Access Publication Date: 12 March 2015
Original Article

ORIGINAL ARTICLE

Attention Stabilizes Representations in the Human Hippocampus

Mariam Aly[1] and Nicholas B. Turk-Browne[1,2]

[1]Princeton Neuroscience Institute and [2]Department of Psychology, Princeton University, Princeton, NJ 08544, USA

Address correspondence to Mariam Aly, Peretsman-Scully Hall 324, Department of Psychology, Princeton University, Princeton, NJ 08544, USA.
Email: aly@princeton.edu

Abstract

Attention and memory are intricately linked, but how attention modulates brain areas that subserve memory, such as the hippocampus, is unknown. We hypothesized that attention may stabilize patterns of activity in human hippocampus, resulting in distinct but reliable activity patterns for different attentional states. To test this prediction, we utilized high-resolution functional magnetic resonance imaging and a novel "art gallery" task. On each trial, participants viewed a room containing a painting, and searched a stream of rooms for a painting from the same artist (art state) or a room with the same layout (room state). Bottom-up stimulation was the same in both tasks, enabling the isolation of neural effects related to top-down attention. Multivariate analyses revealed greater pattern similarity in all hippocampal subfields for trials from the same, compared with different, attentional state. This stability was greater for the room than art state, was unrelated to univariate activity, and, in CA2/CA3/DG, was correlated with behavior. Attention therefore induces representational stability in the human hippocampus, resulting in distinct activity patterns for different attentional states. Modulation of hippocampal representational stability highlights the far-reaching influence of attention outside of sensory systems.

Key words: attentional modulation, high-resolution fMRI, hippocampal subfields, medial temporal lobe, task representations

Introduction

Attention is critical for the encoding and retrieval of long-term memory (Chun and Turk-Browne 2007; Hardt and Nadel 2009), and long-term memory can serve to guide attention (Summerfield et al. 2006; Stokes et al. 2012). Such memory depends on the integrity of the medial temporal lobe (MTL), namely the parahippocampal cortex (PHc), perirhinal cortex (PRc), entorhinal cortex (ERc), and hippocampus (Cohen and Eichenbaum 1993; Brown and Aggleton 2001). Given the tight connection between attention and memory, and between memory and the MTL, surprisingly little is known about how attention modulates the MTL.

Much of what is known about attentional modulation of brain activity comes from studies of the visual system. The central finding from this literature is that brain areas that respond selectively to a stimulus show enhanced neural activity when that stimulus is attended versus unattended (Kastner and Ungerleider 2000). This response enhancement has been found in various species, for different kinds of attention, and in multiple visual areas (Maunsell and Treue 2006; Gilbert and Li 2013). This approach has also been used within the MTL, revealing enhanced activity in PHc when attention is allocated to scenes (O'Craven et al. 1999; Dudukovic et al. 2010). There is also some evidence that attention can enhance activity in PRc and ERc (Dudukovic et al. 2010).

In contrast to the extensive work on attentional modulation of the visual system, and the handful of studies on attentional modulation of MTL cortex, little is known about how attention modulates the human hippocampus. There is some evidence that hippocampal long-term memory encoding and retrieval are affected by attention (e.g., Fernandes et al. 2005; Adcock et al. 2006; Dudukovic and Wagner 2007; Uncapher and Rugg 2009; Duncan et al. 2012; Hashimoto et al. 2012; Carr et al. 2013; Wolosin et al. 2013), and that the hippocampus supports the selection of relevant long-term memories during navigation (Brown et al. 2010; Brown and Stern 2014). When there are no long-term memory demands, however, the evidence for

attentional modulation of the hippocampus is equivocal (for null findings, see Yamaguchi et al. 2004; Dudukovic et al. 2010; cf. Newmark et al. 2013). Here, we argue that selective attention does indeed modulate the hippocampus, focusing on 2 reasons why this has yet to be established. First, and most importantly, the neural signature of attention in the hippocampus might be different than in cortical regions. Secondly, just as attention modulates visual areas selective for task-relevant features, attention may modulate the hippocampus most strongly when the relational "features" that it represents are task-relevant. Below we expand on both of these points.

The most robust neural signature of attentional modulation in cortex is enhanced activity. We hypothesized that attentional modulation of the hippocampus might manifest most strongly as modulation of *representational stability*. More concretely, distinct patterns of activity may be established for different attentional states, resulting in activity patterns that are similar to each other (or stable) across multiple instances of the same attentional state. We use the term "representation" to indicate the coding of information specific to that state, whether the attentional goal itself or the stimulus features that are goal-relevant. In the latter case, "representation" does not apply to the particular features that are relevant in any given instance (since they can vary), but rather to the type of features that tend to be relevant across distinct instances.

Evidence in support of our hypothesis that attention stabilizes hippocampal representations comes from studies of place cells in freely navigating rodents (Muzzio, Kentros, et al. 2009). Place cells tend to fire when an animal is in a particular location in the environment (i.e., the cell's place field), but their firing is surprisingly variable across passes through the firing field (e.g., Fenton and Muller 1998). Importantly, this variability is affected by manipulations of the presumed "attentional state" of mice: Tasks that increase the goal relevance of spatial cues in the environment also enhance place field stability, and this stability is positively correlated with behavioral performance (Kentros et al. 2004). Later studies demonstrated that stability is selective: Place fields stabilize when visuospatial cues are relevant and olfactory representations stabilize when odor cues are relevant (Muzzio, Levita, et al. 2009). Even within a single modality, different ensembles of hippocampal cells consistently co-activate when different spatial reference frames are relevant (Jackson and Redish 2007; Kelemen and Fenton 2010; see also Fenton et al. 2010). This work suggests that attention-like modulation of the hippocampus involves dynamic, network-level switching between cell assemblies that represent different attentional states.

These findings show the existence of distinct and reliable neural representations in the rodent hippocampus when different stimulus dimensions are task-relevant. Thus, manipulations of top-down attention that alter which information is task-relevant may also modulate hippocampal activity patterns in humans. Importantly, studies of the visual system show that modulation is strongest in a region when the task-relevant information is represented in that region. Thus, we sought to manipulate top-down attention to information preferentially represented in the hippocampus.

Decades of studies show that the basic currency of the hippocampus is relations, including spatial, temporal, and featural varieties (Cohen and Eichenbaum 1993; Brown and Aggleton 2001). These relational representations are integral to episodic memory (Cohen and Eichenbaum 1993; Brown and Aggleton 2001; Davachi 2006), working memory (Hartley et al. 2007; Hannula and Ranganath 2008), and perception (Lee et al. 2012; Aly et al. 2013). By analogy to visual cortex, we hypothesized that attending to relational information would modulate the hippocampus. Indeed, previous studies that did not observe modulation of the hippocampus employed tasks in which attention was directed to items rather than relations (Yamaguchi et al. 2004; Dudukovic et al. 2010).

To investigate these hypotheses, we used functional magnetic resonance imaging (fMRI) and examined the stability of hippocampal activity patterns as a function of attention to different kinds of relations in a novel "art gallery" task. The stimuli were indoor spaces rendered in 3D, each with a painting, a unique room layout, and several pieces of furniture (Fig. 1A). Participants were cued to attend to the painting (art state) or to the layout (room state). Participants were then presented with a "base image," followed by a series of 4 "search images." For the art state, participants examined the search images for a painting created by the same artist as the painting in the base image. Matching paintings were similar in style (e.g., use of color and brushstrokes) but not necessarily content. For the room state, participants examined the search images for a room with the same layout as the base image. Matching rooms had the same spatial layout and furniture configuration, but a changed appearance (i.e., different wall color, painting, and furniture pieces) and perspective (30° rotation). At the end of the trial, participants were probed about whether there had been a matching painting or room. The probe was valid (i.e., the same as the cue) on 80% of trials and invalid on 20% of trials. The comparison of detection performance for valid versus invalid probes provided a behavioral measure of attention.

This approach is analogous to studies of top-down attention in the visual system, which manipulate participants' internal goals to control how a given external stimulus is processed. This research generally concerns attention to low-level features such as an orientation, color, or motion direction. Here, we manipulated attention to high-level dimensions—artistic style for the art task and geometrical layout for the room task—that are abstracted away from the low-level features of any given painting or room. Such attention to high-level dimensions is common in the real world. For example, when you go to an art museum, you might want to visit a newly acquired painting from your favorite Impressionist artist. To help find this work, you could search for an "Impressionism" feature within the artistic style dimension. This feature has many properties, including brush stroke, color distribution, realism, etc., which collectively make up your attentional state. Likewise, when looking for a place to live, you might want to find a house with an open concept design. To help find such a house, you could search for an "open concept" feature within the geometrical layout dimension. This feature has many properties, including large space, long aspect ratio, few walls, many windows, etc., which collectively make up your attentional state.

In our design, the cue on every trial specified the high-level dimension to be attended, and the base image provided the target feature within that dimension for which to search. Orienting attention to the correct dimension was necessary for successfully detecting an image with that feature and for ignoring images with a feature that matched the base image on the unattended dimension. Importantly, across trials, the same base and search images were encountered in both the art and room tasks. In this way, any neural differences across tasks are a reflection of top-down attentional states rather than the availability of different kinds of information in the input.

Because our hypotheses concerned the MTL, we acquired high-resolution structural MRIs and manually traced several regions of interest (ROIs) in the hippocampus and surrounding cortex (Fig. 2). Within each ROI, we measured the blood-oxygen-level

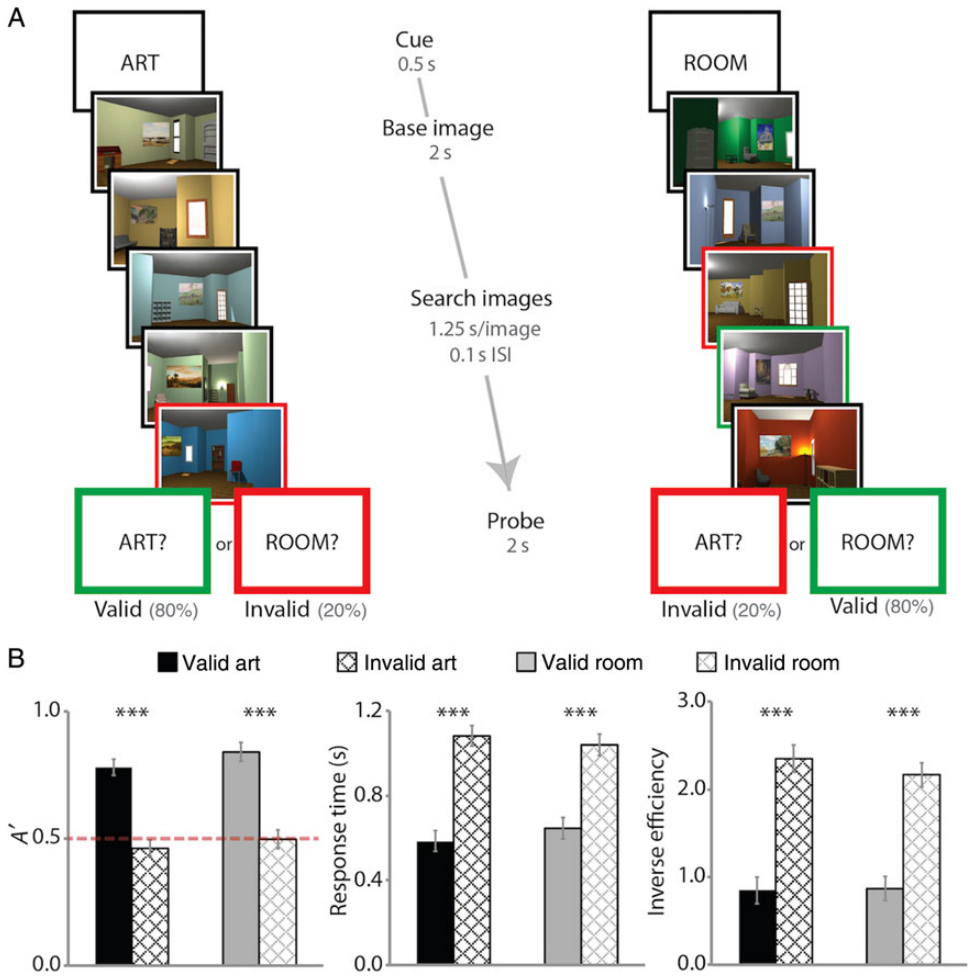

Figure 1. Behavioral task. Two sample trials are depicted (see text for details about task instructions). For visualization, cued matches are outlined in green and uncued matches in red. (*A*) Example of an art-state trial, with cued match absent and uncued match present, and a room-state trial, with cued match present and uncued match present. (*B*) Sensitivity, RT, and inverse efficiency in making present/absent judgment as a function of attentional state and probe type. Error bars depict ±1 SEM of the within-subject valid versus invalid difference. Dashed line indicates chance performance. ***$P < 0.001$.

dependent (BOLD) activity elicited by art- and room-state trials. We first verified that our task was effective by testing whether attention improved behavioral performance (faster and more accurate responses to valid vs. invalid probes) and enhanced univariate activity in the MTL cortex—in PHc and ERc for the room state due to their roles in spatial processing (e.g., Epstein and Kanwisher 1998; Jacobs et al. 2013) and in PRc for the art state due to its role in object processing (e.g., Brown and Aggleton 2001; Davachi 2006). We then tested 3 novel predictions about the human hippocampus: (1) That attention would stabilize activity patterns, with increased pattern similarity between trials from the same versus different attentional states; (2) that this stability would be greater for the room state, given the importance of the hippocampus to spatial processing (e.g., Muzzio, Kentros, et al. 2009; Lee et al. 2012; Aly et al. 2013); and (3) that this stability would be behaviorally relevant, with increased stability linked to better performance.

Materials and Methods

Participants

Twenty-four individuals (15 males) from the Princeton University community participated for monetary compensation (age: M = 22.5 years, SD = 4.0; education: M = 15.0 years, SD = 2.5). Written informed consent was obtained from all participants, and the study was approved by the Institutional Review Board at Princeton University. Five participants performed at or near chance on one or both tasks (>2.8 SDs below the mean of an independent pilot sample, and >2.65 SDs below the mean of the included scanned participants), and data for these participants were excluded from the reported analyses. Their inclusion did not change any of the patterns of results, and behavioral performance remained above chance overall.

Behavioral Task

Stimuli

The rooms were rendered with Sweet Home 3D (sweethome3d.com). Each room contained multiple pieces of furniture and had a unique shape and layout. Twenty rooms were used to create the experimental stimuli, and an additional 6 were used for practice. For each room, a second version was created with a 30° viewpoint rotation (half clockwise and half counterclockwise). This second version was also altered so that the content was different but the spatial layout was the same: Wall colors were changed, and furniture was changed to different exemplars of the same category (e.g., a chair was replaced by a different

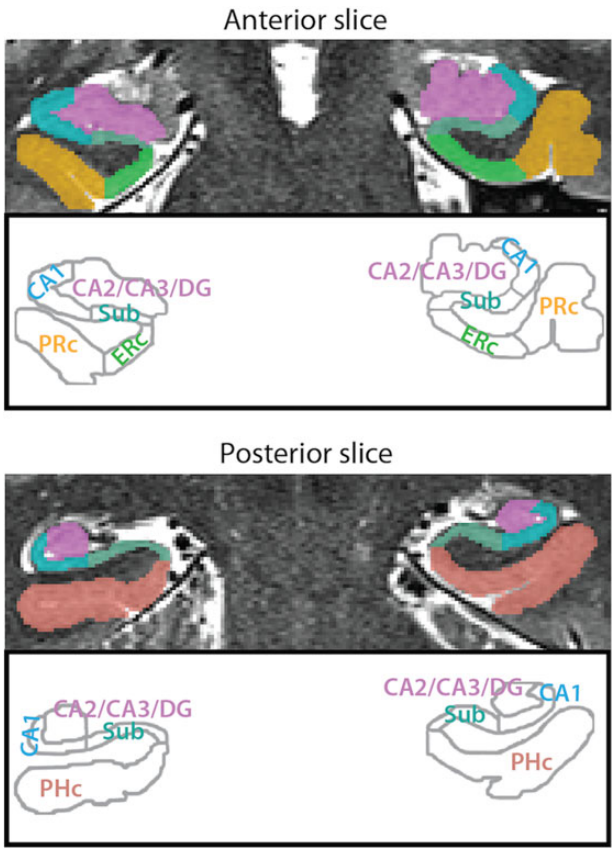

Figure 2. MTL ROIs. Example segmentation from one participant is depicted for an anterior and a posterior slice. ROIs consisted of 3 hippocampal subfields (subiculum [Sub], CA1, and CA2/CA3/DG) and 3 MTL cortical regions (PRc, ERc, and PHc). For segmentation guide, see Supplementary Methods.

chair). For the paintings, 2 works from the same artist, similar in style but not necessarily content, were chosen from the Google Art Project (20 artists/40 paintings for experimental stimuli and 6 artists/12 paintings for practice stimuli). None of the practice stimuli were used in the scanned experiment.

Design and Procedure

The stimulus set of 120 images was generated such that each of the 40 rooms (20 in 2 perspectives each) was paired with 3 paintings from different artists, and each of the 40 paintings (20 artists with 2 paintings each) was paired with 3 different rooms. For every participant, 20 of these images (unique art and room combinations) were chosen as "base images," and 20 trial "templates" were created by choosing (for each template) one of these base images, a room match (one of the remaining 100 images with the same layout but a different artist), an art match (a different remaining image with a painting from the same artist but a different room layout), and distractors (4 remaining images with a different layout and artist). The art and room matches for one template could serve as distractors for other templates. Base images were not used as distractors or matches for other templates; however, base images for each participant served as distractors, art matches, or room matches for other participants. Image selection was counterbalanced such that all of the 120 images were equally likely to appear. As a result, every painting and room appeared an equal number of times.

Each of the 20 templates was used to generate 10 trials. The cues for these trials were split between tasks (5 art and 5 room). Within each task, there was a 50% probability of the task-relevant match being present or absent (e.g., a room match on room-task trials), and independently, a 50% probability of the task-irrelevant match being present or absent (e.g., an art match on room-task trials). When they appeared, the art and room matches for each template were the same for all trials generated from that template. Distractors were selected to fill out the remaining slots in the search set (i.e., 2 distractors were selected if both the task-relevant and task-irrelevant match were present, 3 if only one match was present, and 4 if neither match was present). The probes for these trials matched the cue with 80% probability (valid trials); the remaining 20% of trials were invalid. The 200 total trials were divided evenly into 8 runs. We restricted correlations of BOLD activity patterns to adjacent runs because of generic time-dependent factors that may reduce pattern similarity (e.g., fatigue and motion), and thus included all trials generated from a given template in back-to-back runs. Trial order within run was randomized with the constraint that trials from the same template could not occur twice in a row, and the 10 trials from each template were equally divided between the 2 runs.

This design has several important features: The stimuli were identical across the 2 tasks; the presence of one match (e.g., an art match) was uninformative about the presence of the other (e.g., a room match); and art and room matches were equally likely to occur on trials of either task. Successful task performance therefore necessitated orienting attention based on the cue at the beginning of each trial, because nothing about the stimuli themselves differed between the 2 tasks. These design features also ensured that differences in brain activity between the art and room tasks would be related to participants' top-down attentional states rather than bottom-up stimulation.

Stimuli were presented using the Psychophysics Toolbox for Matlab (psychtoolbox.org). Each trial began with a fixation dot 500 ms before cue onset (Fig. 1A). The "ART" or "ROOM" cue was presented for 500 ms and subtended 1.9° or 3.1° horizontally, respectively. Following the cue, the base image was presented for 2 s; this and all subsequent images subtended 16.6° × 12.6°. Then, the 4 images in the search set were sequentially presented for 1.25 s each, separated by a 100-ms interval. After the last image, the "ART" or "ROOM" probe was presented for a maximum of 2 s. Participants responded "yes" or "no" with a button box by using the index and middle fingers, respectively. The probe disappeared once a response was made. After the response window and an 8-s interval with a blank screen, the next trial began. At the end of each run, the percentage of correct responses made on that run was displayed along with feedback (e.g., "You are doing pretty well!" for 75–90% accuracy).

Participants came in for instructions and a practice session the day before the scan. They viewed a sample base image and its art and room matches, and then completed 10 practice trials (2 trials generated from each of 5 templates, none of which was used in the scanned experiment). The procedure was identical to the scanning task except that feedback ("You are correct!" or "You are incorrect.") was given after every trial, as well as overall accuracy at the end of the practice task. Participants repeated the practice until they reached at least 70% accuracy.

Eye Tracking

Eye position was monitored during the fMRI scan with a SensoMotoric Instruments iView X MRI-LR eye-tracking system sampling at 60 Hz, and the resulting data were analyzed using BeGaze software. We employed a free-viewing paradigm in which participants were allowed to move their eyes during the

trials. Because different types of information were relevant for the art and room tasks, restricting fixation may have disadvantaged one task over the other. Nevertheless, we collected eye-tracking data to know how participants moved their eyes and to enable follow-up analyses that considered how these eye movements related to the fMRI results (see Supplementary Methods).

MRI Acquisition

MRI data were collected on a 3-T Siemens Skyra scanner with a 20-channel head coil. Functional images were obtained with a multiband echo-planar imaging (EPI) sequence (repetition time = 2 s, echo time = 40 ms, flip angle = 71°, acceleration factor = 3, shift = 2, voxel size = 1.5 mm iso), with 57 oblique axial slices (16° transverse to coronal) acquired in an interleaved order. Whole-brain high-resolution (1.0 mm iso) T_1-weighted structural images were acquired with a magnetization-prepared rapid acquisition gradient echo sequence. Two T_2-weighted turbo spin-echo images were acquired (and averaged) for manual segmentation of hippocampal subfields and MTL cortex (see Supplementary Methods), consisting of 54 slices perpendicular to the long axis of the hippocampus (0.44 × 0.44 mm in-plane, 1.5 mm thick). Field maps were collected for registration, consisting of 40 oblique axial slices (3 mm iso).

Because multiband imaging is still an active area of development, we conducted extensive pilot testing of several multiband sequences in consultation with our MR physicist to maximize signal-to-fluctuation noise ratios and minimize distortions and ghosting. We settled on parameters that produced images of similar quality to our other parallel acquisition EPI sequences (e.g., iPAT). Representative mean functional scans from 3 participants are shown in Supplementary Figure 1.

fMRI Analysis

Software

Preprocessing and analyses were conducted using FEAT, FLIRT, and command-line functions (e.g., randomise, fslmaths) in FSL. ROI analyses (e.g., pattern similarity and percent signal change) were performed with custom Matlab scripts. Searchlight analyses were performed using Simitar (www.princeton.edu/~fpereira/simitar) and custom Matlab scripts.

Preprocessing

The first 3 volumes of each run were discarded to allow for T_1 equilibration. Preprocessing steps included: brain extraction, slice-timing correction, motion correction, high-pass filtering (maximum period = 128 s), and spatial smoothing (a 3-mm FWHM Gaussian kernel). Field map preprocessing was based on recommendations specified in the FUGUE user guide (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FUGUE/Guide) and carried out with a custom script. First, the 2 field map magnitude images were averaged together and skull stripped. The field map phase image was then converted to rad/s and smoothed with a 2-mm Gaussian kernel. The resulting preprocessed phase and magnitude images were included in the preprocessing step of FEAT analyses to unwarp the functional images and aid registration to anatomical space. Following registration, the distortion-corrected functional images were compared to the originals to ensure that unwarping was effective. In all cases, using the field maps reduced distortion in anterior temporal and frontal regions.

Univariate Analyses

The main GLM for univariate analyses contained 4 regressors of interest: Valid and invalid trials for the art and room states. These were modeled as 8-s epochs from cue onset to the offset of the last image. Additionally, there was a regressor for trials in which the participant did not respond (modeled the same way), and a regressor for the probe/response period, which was modeled as a 2-s epoch from probe onset. All regressors were convolved with a double-gamma hemodynamic response function and their temporal derivatives were also entered. Finally, the 6 directions of head motion were included as nuisance regressors. Autocorrelations in the time series were corrected with FILM prewhitening. Each run was modeled separately in first-level analyses, resulting in 8 different models per participant. Only valid trials (i.e., trials in which the text cue at the beginning of the trial matched the text probe at the end) were analyzed further.

First-level parameter estimates from each run were converted to percent signal change and registered to the participant's T_2 image (up-sampling to the T_2 resolution). These values were then extracted from each anatomical ROI and averaged across voxels and runs (for anterior/posterior hippocampal ROIs, see Supplementary Fig. 2). Group analyses consisted of random-effects paired *t*-tests across participants. Whole-brain analyses are reported in Supplementary Methods and shown in Supplementary Figures 3 and 4 for completeness.

Multivariate Analyses

Pattern similarity analyses were performed on parameter estimate images from a separate single-trial GLM; these images were registered to each individual participant's T_2 image and up-sampled to the T_2 resolution. This approach allowed us to bin trials in various ways to answer a series of questions. Each trial was modeled as an 8-s epoch from cue onset to the offset of the last image. These 25 regressors (one for every trial in a run) replaced the 5 trial regressors (valid/ invalid × art/room + missed responses) in the previous GLM, but everything was identical otherwise. Only valid trials, on which the cue matched the probe, were analyzed, and we included trials with both correct and incorrect responses to balance the number of trials per participant. As in the previous GLM, each run was modeled separately, resulting in 8 different models per participant. Pattern similarity was computed within pairs of runs (i.e., runs 1–2, 3–4, 5–6, and 7–8) because correlating patterns that are separated far in time might result in increased noise from factors such as gradual motion and fatigue. Whole-brain searchlight analyses are reported in Supplementary Methods and shown in Supplementary Figure 5.

Parameter estimates across voxels within each ROI for a given trial were reshaped into a vector, and the correlations between all pairs of vectors within adjacent runs were calculated (Kriegeskorte et al. 2008). Correlations were then averaged on the basis of the analysis of interest: First, correlations for trials of the same attentional state (i.e., art/art and room/room) were compared with those for trials of different states (i.e., art/room). Secondly, same-state correlations were compared between states (i.e., art/art vs. room/room). Thirdly, correlations for trials from the same versus different templates were compared (Supplementary Fig. 6A). Finally, correlations for trials in which a task-relevant match was present were compared with those in which a task-relevant match was absent (Supplementary Fig. 6B). Correlations were averaged across (pairs of) runs within participant, Fisher-transformed to ensure normality, and compared at the group level with random-effects paired *t*-tests across participants.

Brain/Behavior Correlations

For both the art and room tasks, we correlated pattern similarity in each ROI with behavioral performance across individuals. If

there was a reliable correlation within task for an ROI (i.e., room-state pattern similarity and room-state behavior), then pattern similarity in that task and ROI was correlated with behavioral performance in the other task to assess the specificity of the relationship (i.e., room-state pattern similarity and art-state behavior). These analyses were repeated for univariate activity and in a voxelwise manner over the whole brain using a searchlight approach (where patterns were defined over 27-voxel cubes centered on every voxel).

Multivariate–Univariate Dependence Analysis

Pattern similarity is often assumed to reflect the presence of a reliable pattern of activity that is not adequately captured in terms of a single mean response. But such similarity can also be observed when voxels within a region of interest consistently activate or deactivate in a univariate fashion (for discussion, see Coutanche 2013, Davis and Poldrack 2013, Davis et al. 2014). To determine whether this was the case in our data—that is, whether attentional modulation of univariate activity can account for pattern similarity—we examined whether the same voxels were contributing to both effects. Specifically, we developed the multivariate-univariate dependence (MUD) analysis to test whether the average amount of (positive or negative) univariate activity in a voxel was related to how much the voxel contributed to pattern similarity across trials.

The MUD analysis involved several steps: (1) Activity in each ROI for each trial was first normalized by subtracting the mean and dividing by the root sum-of-squares. (2) These normalized values for each voxel were then multiplied for each pair of trials of the same attentional state. This product was a measure of the extent to which a voxel contributed to pattern similarity: Voxels with 2 positive values or 2 negative values (i.e., positive products) contributed to a positive correlation, and the greater the magnitude of the product, the greater the contribution; voxels with one positive and one negative value (i.e., negative products) contributed to a negative correlation, again in proportion to the magnitude of the product. In fact, the sum of all of these normalized products is the Pearson correlation over voxels (see Worsley et al. 2005 for an application of the same technique over time for estimating functional connectivity). (3) These products were then averaged for each voxel across pairs of trials of the same attentional state, resulting in a measure of how much that voxel contributed to same-state pattern similarity. (4) For each voxel, we also determined the average univariate activity (percent signal change vs. baseline) across trials of the same attentional state. (5) Finally, for each ROI, we correlated the measure of pattern similarity "influence" with univariate activity across voxels. Insofar as univariate activity explains pattern similarity, then this correlation should be reliably positive (in the case of activation) or negative (in the case of deactivation) across participants in a random-effects one-sample *t*-test. See Supplementary Methods, Supplementary Table 1, and Supplementary Figure 7 for simulations that verify the utility of this approach.

Results

Behavior

As measures of behavioral performance, we examined sensitivity (A') and response times (RTs) when detecting matches for valid and invalid probes in both attentional states (Fig. 1*B*). On valid trials, sensitivity was above chance [0.5; art: $t_{(18)} = 16.45$, $P < 0.0001$; room: $t_{(18)} = 17.15$, $P < 0.0001$] and higher than on invalid trials [art: $t_{(18)} = 9.90$, $P < 0.0001$; room: $t_{(18)} = 9.24$, $P < 0.0001$]. On invalid trials, sensitivity was not different from chance [art: $t_{(18)} = 1.08$, $P = 0.29$; room: $t_{(18)} = 0.10$, $P = 0.92$], suggesting that attention was effectively and selectively engaged by the cue. RTs were also faster on valid than invalid trials [art: $t_{(18)} = 10.41$, $P < 0.0001$; room: $t_{(18)} = 7.77$, $P < 0.0001$], inconsistent with a tradeoff between speed and accuracy for valid versus invalid trials. (For further evidence from whole-brain fMRI analyses that our attentional manipulation was effective, see Supplementary Fig. 3.)

Comparing the attentional states revealed that sensitivity was higher for the room versus art state on valid trials [$t_{(18)} = 3.05$, $P = 0.007$], but not on invalid trials [$t_{(18)} = 0.63$, $P = 0.53$]; the attentional modulation effect (difference between valid and invalid trials) also did not differ [$t_{(18)} = 0.48$, $P = 0.64$]. Importantly, higher sensitivity for the room state was accompanied by slower RTs relative to the art state [$t_{(18)} = 3.85$, $P = 0.001$], suggesting that the difference between states reflects a speed/accuracy tradeoff. To verify this, we calculated inverse efficiency scores (i.e., RT/accuracy; Townsend and Ashby 1978), which did not differ between art and room states [valid: $t_{(18)} = 0.46$, $P = 0.65$; invalid: $t_{(18)} = 1.24$, $P = 0.23$].

Thus, the 2 tasks were comparable in difficulty. This is not to say that the tasks were identical: The features that were attended differed, as did the need for abstraction and object versus spatial processing. These differences were reflected in the whole-brain distribution of activity for the art versus room states (see Supplementary Fig. 4), and we consider such differences in processing characteristics to be essential components of an attentional manipulation.

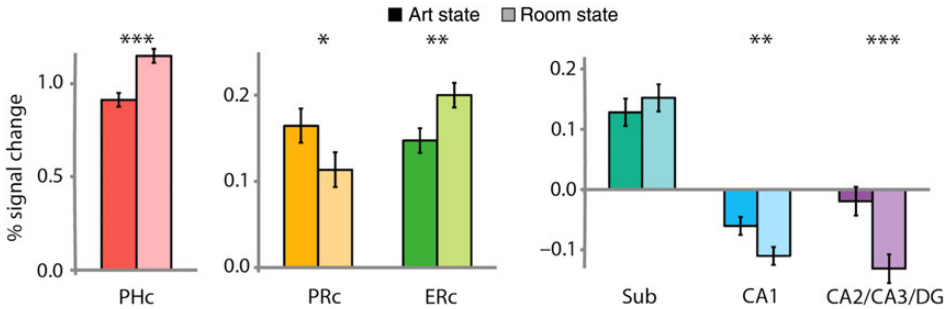

Figure 3. Attentional modulation of univariate activity. BOLD activity evoked in the art and room states was extracted from all voxels in each ROI and averaged. Baseline corresponds to unmodeled periods of passive viewing of a blank screen. In MTL cortex, PHc and ERc were more active in the room state and PRc was more active in the art state. In the hippocampus, CA1 and CA2/CA3/DG were also more active in the art state (or deactivated in the room state). Error bars depict ±1 SEM of the within- subject art versus room state difference. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

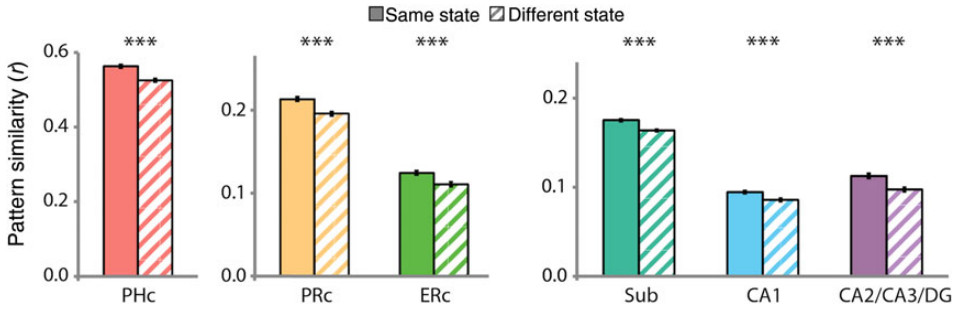

Figure 4. State-dependent pattern similarity. BOLD activity evoked in the art and room states was extracted from all voxels in each ROI and correlated across trials of the same versus different states. In MTL cortex, all regions showed greater pattern similarity for same versus different states. In the hippocampus, all subfields showed greater pattern similarity for same versus different states. Results are shown as Pearson correlations, but statistical tests were performed only after applying the Fisher transformation. Error bars depict ±1 SEM of the within-subject same versus different state difference. ***$P < 0.001$.

Univariate Activity

In MTL cortex, we expected attentional states to differentially affect the overall level of activity. Specifically, we expected enhanced activity for the room state in PHc and ERc, and enhanced activity for the art state in PRc. To test this prediction, we extracted the percent signal change in BOLD activity for art and room states from the PHc, PRc, and ERc ROIs, and averaged over the voxels within each ROI. For this and subsequent analyses, we collapsed across left and right hemispheres because we had no a priori predictions about hemispheric differences; in all cases, the pattern of results was identical for both hemispheres.

This analysis yielded the expected region × state double dissociation (Fig. 3): PHc and ERc were more active for room compared with art states [$t_{(18)} = 6.47$, $P < 0.0001$ and $t_{(18)} = 3.67$, $P = 0.002$, respectively], while PRc was more active for art compared with room states [$t_{(18)} = 2.55$, $P = 0.02$]. Thus, we verified that attention modulates MTL cortex in a manner similar to how it modulates visual cortex, by enhancing neural activity. Furthermore, in PRc, this provides the first evidence of modulation by selective attention, extending beyond prior findings of nonspecific modulation (Dudukovic et al. 2010). Nevertheless, we are cautious in interpreting this effect: Although predicted a priori, it does not survive correction for multiple comparisons (Bonferroni threshold across regions, $P < 0.008$).

Although we did not expect similar modulation of the hippocampus based on prior studies, we also examined univariate activity in each hippocampal subfield ROI for the sake of completeness (for anterior/posterior hippocampal ROIs, see Supplementary Fig. 2A). Surprisingly, overall activity in the cornu ammonis (CA) fields and dentate gyrus (DG) was modulated by attention, with more activity for art versus room states [CA1: $t_{(18)} = 3.33$, $P = 0.004$; CA2/CA3/DG: $t_{(18)} = 4.63$, $P = 0.0002$]; the subiculum showed no difference [$t_{(18)} = 1.07$, $P = 0.30$]. One possible explanation for this effect—in contrast to prior null results (e.g., Yamaguchi et al. 2004; Dudukovic et al. 2010)—is that by requiring abstraction of object and spatial information over surface details, our task may have placed greater demands on the flexible, relational representations that are the hallmark of hippocampal processing (e.g., Cohen and Eichenbaum 1993).

Pattern Similarity

Same Versus Different States

In the hippocampus, we predicted that attention would modulate multivoxel patterns of activity. Specifically, we hypothesized that attention would induce state-dependent activity patterns—that is, patterns that would be stable across repeated occurrences of the same attentional state. We tested for this stability by extracting a vector of BOLD activity across voxels in a given hippocampal ROI for each trial, and then correlating these vectors as a function of whether they were obtained from trials of the same versus different states (Fig. 4; for anterior/posterior hippocampal ROIs, see Supplementary Fig. 2B). We predicted that pattern similarity would be greater for trials from the same (i.e., art/art and room/room) compared with different (i.e., art/room) attentional states. This prediction was borne out in the data [subiculum: $t_{(18)} = 8.09$, $P < 0.0001$; CA1: $t_{(18)} = 5.08$, $P < 0.0001$; CA2/CA3/DG: $t_{(18)} = 5.67$, $P < 0.0001$]. These results provide clear support for our representational stability hypothesis regarding attentional modulation of the hippocampus.

Although we developed this hypothesis for the hippocampus (in light of the lack of other forms of modulation in that region, and based on animal models), we also tested for stability in MTL cortex. Indeed, the MTL cortical ROIs showed the same effect as the hippocampus, with greater pattern similarity for same versus different states [PHc: $t_{(18)} = 7.04$, $P < 0.0001$; ERc: $t_{(18)} = 5.00$, $P < 0.0001$; PRc: $t_{(18)} = 6.86$, $P < 0.0001$]. This modulation of activity patterns is novel with respect to prior studies of attentional modulation in MTL cortex, which focused exclusively on univariate activity (e.g., Dudukovic et al. 2010). Moreover, although the effects were similar in MTL cortex and hippocampus, later analyses suggest that they are dissociable.

Art Versus Room States

To examine whether stability differed between the art and room states, we examined pattern similarity for the 2 states separately. Specifically, we focused only on the correlation between vectors from trials of the same state, and compared the average correlation for art versus room trials (Fig. 5; for anterior/posterior hippocampal ROIs, see Supplementary Fig. 2C). We hypothesized that pattern similarity would be greater for the room state in the hippocampus, given its importance for spatial processing. Indeed, this prediction was confirmed in all subfields [subiculum: $t_{(18)} = 7.98$, $P < 0.0001$; CA1: $t_{(18)} = 3.22$, $P = 0.005$; CA2/CA3/DG: $t_{(18)} = 6.09$, $P < 0.0001$]. In addition, greater pattern similarity for the room state was found in PHc and ERc [$t_{(18)} = 5.78$, $P < 0.0001$ and $t_{(18)} = 2.86$, $P = 0.01$, respectively], which are also involved in spatial processing; there was no difference in PRc [$t_{(18)} = 1.19$, $P = 0.25$].

We additionally examined the standard deviation of pattern similarity in the art versus room states, but found no reliable effects in any ROI at the corrected statistical threshold [PRc: $t_{(18)} =$

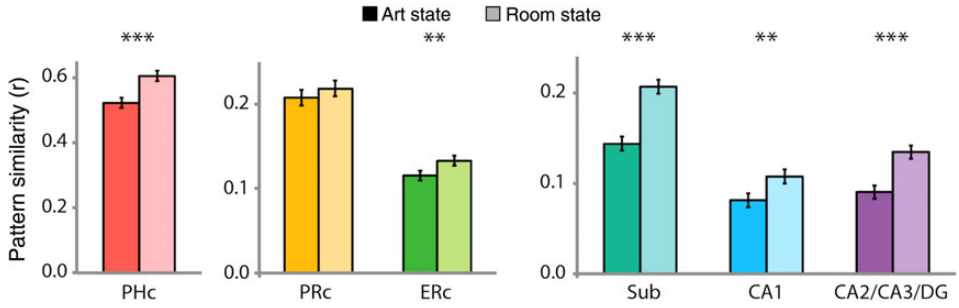

Figure 5. Comparison of pattern similarity between states. BOLD activity was extracted from all voxels in each ROI and separately correlated across trials of the art and room states, respectively. In MTL cortex, PHc and ERc showed greater pattern similarity for room versus art states, and PRc showed no difference. In the hippocampus, all subfields showed greater pattern similarity for room versus art states. Results are shown as Pearson correlations, but statistical tests were performed only after applying the Fisher transformation. Error bars depict ±1 SEM of the within-subject art- versus room-state difference. $^{**}P < 0.01$, $^{***}P < 0.001$.

2.19, $P = 0.04$; all other Ps > 0.27]. Thus, our attentional manipulation affected the mean but not variability of pattern similarity.

In the univariate analyses above, we found lower activity in the CA fields and DG for the room state compared with the art state. This might have reflected a reduction in information content in the room state, which would have resulted in less pattern similarity compared with the art state. Instead, we observed greater pattern similarity for the room state, consistent with prior findings that regions with attenuated univariate activity can contain more multivariate information (e.g., Kok et al. 2012). Note that this relationship was not uniform across ROIs, with PHc and ERc showing greater activity and greater pattern similarity for the room state, PRc showing lower activity for the room state and no difference in pattern similarity, and subiculum showing no difference in activity but greater pattern similarity for the room state. This suggests that these 2 manifestations of attentional modulation are not interchangeable and may provide distinct signatures of attention.

Relationship Between Univariate Activity and Pattern Similarity

To more directly test the relationship between univariate activity and pattern similarity, we performed additional analyses on ROIs that showed differences between the art and room states in both measures (PHc, ERc, CA1, and CA2/CA3/DG; not PRc or subiculum). In PHc and ERc, this modulation was in the same direction for both measures: more activity and pattern similarity for the room state. Conversely, in the CA fields and DG, the modulation was in opposite directions: Less activity but more pattern similarity for the room state. Given the role of all of these regions in spatial processing, one possibility is that changes in univariate activity on room-state trials are driving changes in multivariate pattern similarity—that is, that the 2 measures are different manifestations of the same underlying effect. If so, then in PHc and ERc, the voxels with greater room-state activity should make a larger contribution to room-state pattern similarity. In contrast, in the CA fields and DG, the voxels with lower room-state activity should make a larger contribution to room-state pattern similarity.

To quantify the relationship between activity and pattern similarity in the room state, we developed a new multivariate-univariate dependence (MUD) analysis in which the contribution of each voxel to pattern similarity was estimated and then correlated with its level of activity (Fig. 6). For a given pair of trials, the contribution to pattern similarity was estimated by normalizing the vector of activity for each trial and then computing the pairwise product of these normalized values at each voxel within an ROI. Voxels with positive products increase pattern similarity and voxels with negative products decrease pattern similarity—in both cases proportional to the magnitude of the product. For each voxel, the products for all pairs of trials were averaged, resulting in one contribution score per voxel. These scores were then correlated across voxels with the average activity level in those voxels to produce an index of the dependence between activity and pattern similarity within each ROI (for more detail, see Materials and Methods).

In PHc and ERc, the MUD analysis revealed a positive relationship, suggesting that pattern similarity could in part be related to changes in overall activity [PHc: mean $r = 0.58$, $t_{(18)} = 22.29$, $P < 0.0001$; ERc: mean $r = 0.23$, $t_{(18)} = 4.95$, $P < 0.0001$]. However, in the CA fields and DG, there was no relationship between activity and pattern similarity [CA1: mean $r = 0.08$, $t_{(18)} = 1.39$, $P = 0.18$; CA2/CA3/DG: mean $r = 0.01$, $t_{(18)} = 0.21$, $P = 0.83$]. Thus, the parallel enhancement of activity and pattern similarity in PHc and ERc for the room state was driven in part by modulation of the same voxels. In the CA fields and DG, however, activity and pattern similarity went in opposite directions, and moreover, partly non-overlapping sets of voxels made the biggest contributions to these 2 effects. This suggests that, in the hippocampus, pattern similarity reflects the operation of a distinct attentional mechanism than what modulates overall activity.

In the preceding analyses, the hippocampal ROIs had both lower room-state pattern similarity (Fig. 5) and lower MUD (Fig. 6) than PHc. This raises a concern that the lack of a MUD relationship in hippocampal ROIs might be a floor effect from low pattern similarity values. However, this is not a general issue with MTL cortex versus the hippocampus: ERc and CA2/CA3/DG showed robust and identical pattern similarity in the room state [ERc: mean $r = 0.13$, $t_{(18)} = 10.01$, $P < 0.0001$; CA2/CA3/DG: mean $r = 0.13$, $t_{(18)} = 13.07$, $P < 0.0001$], whereas ERc [mean $r = 0.23$, $t_{(18)} = 4.95$, $P < 0.0001$] but not CA2/CA3/DG [mean $r = 0.01$, $t_{(18)} = 0.21$, $P = 0.83$] showed a reliable MUD effect.

To more closely examine whether MUD effects are constrained by the magnitude of pattern similarity, we performed 2 additional analyses: A median-split analysis and simulations. In the median-split analysis, we divided the participants into high and low pattern similarity groups for ERc and CA2/CA3/DG based on the rank of their room-state pattern similarity relative to the median participant (resulting in $n = 9$ per group). We then reversed the floor effect concern by focusing on the low ERc group [mean pattern similarity $r = 0.09$, $t_{(8)} = 13.36$, $P < 0.0001$] and the high CA2/CA3/DG group [$r = 0.17$, $t_{(8)} = 18.90$, $P < 0.0001$]. Nevertheless, we again found that ERc [mean $r = 0.13$, $t_{(8)} = 2.32$, $P < 0.05$], but not CA2/CA3/DG [mean $r = -0.03$, $t_{(8)} = 0.31$, $P = 0.77$], showed a reliable MUD effect. This

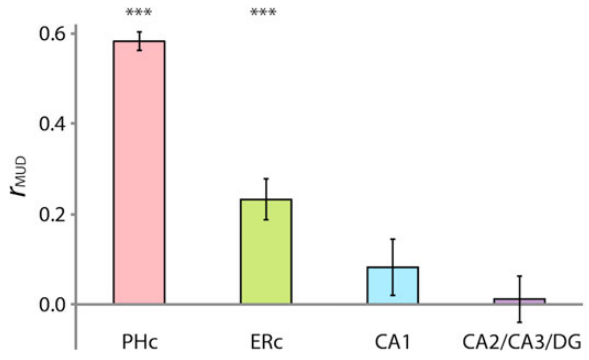

Figure 6. Multivariate-univariate dependence (MUD) analysis. The contribution of each voxel to pattern similarity was estimated by normalizing BOLD activity over voxels within an ROI for each trial and computing pairwise products across trials. Average products from room trials were then correlated with average activity in room trials over voxels to estimate MUD. In MTL cortex, PHc and ERc showed a positive relationship between activity and pattern similarity. In the hippocampus, CA1 and CA2/CA3/dentate gyrus (DG) showed no relationship. Error bars depict ±1 SEM across participants. Results are shown as Pearson correlations, but statistical tests were performed only after applying the Fisher transformation. ***$P < 0.001$.

analysis rules out the possibility that low pattern similarity necessarily mandates a non-significant MUD effect.

Moreover, we conducted simulations to explore the possible relationships between univariate activity, pattern similarity, and MUD (see Supplementary Methods and Supplementary Table 1). These simulations showed that a given level of univariate activity and pattern similarity can produce various MUD effects (see Supplementary Fig. 7). Namely, the magnitude and sign of MUD are controlled by the extent to which the activity levels of the voxels that are most stable across patterns are high or low relative to the other voxels in the ROI.

Finally, it is important to clarify what information the MUD analysis does and does not provide. The magnitude of the MUD effect indicates the extent to which voxels' univariate activity is predictive of their contribution to multivariate pattern similarity. The sign of the MUD effect indicates whether relatively high or low univariate activity is driving pattern similarity. Neither of these conclusions has any consequences for the "shape" of the activity pattern: that is, a zero MUD effect can reflect a truly distributed, high-dimensional activity pattern, but it can also be observed if there are spatially localized clusters of activation and deactivation within an ROI. Thus, the lack of a MUD effect cannot be used to infer the presence of a distributed, high-dimensional representation. We use it here specifically with the goal of asking whether pattern similarity reflects a signed overall shift in activity over a subset of the ROI.

Brain/Behavior Relationships

The fact that attentional modulation of activity and pattern similarity can be dissociated raises the question of which effect is more behaviorally relevant. To address this question, we correlated individual differences in behavioral performance (i.e., A') with overall activity and pattern similarity in the hippocampal and MTL ROIs. Due to our small sample size for estimating correlations, we used a robust correlation method in which outliers from the minimum covariance determinant are removed to prevent them from exerting disproportionate leverage (Pernet et al. 2013). Additionally, we corrected for multiple comparisons (Bonferroni threshold across regions, $P < 0.008$).

There was a positive correlation between room-state pattern similarity in CA2/CA3/DG and behavior in the room task [$r_{(17)} = 0.78$, $P < 0.0001$], such that greater pattern similarity was related to better performance (Fig. 7). No other correlation involving these variables survived the corrected threshold: in none of the other MTL ROIs did room-state pattern similarity significantly correlate with behavior in the room task [PHc: $r_{(16)} = 0.06$, $P = 0.81$; PRc: $r_{(16)} = 0.31$, $P = 0.20$; ERc: $r_{(17)} = 0.52$, $P = 0.02$; subiculum: $r_{(17)} = 0.55$, $P = 0.01$; CA1: $r_{(16)} = 0.20$, $P = 0.41$]; nowhere else in the brain (from a searchlight analysis) did room-state pattern similarity correlate with room-state behavior; room-state activity in CA2/CA3/DG did not significantly correlate with room-state behavior [$r_{(17)} = -0.39$, $P = 0.10$; all other ROIs, $P > 0.10$]. Finally, the relationship between CA2/CA3/DG pattern similarity and behavior persisted after partialling out MUD [$r_{(17)} = 0.79$, $P < 0.0001$].

There are 2 forms of attention that could produce the brain/behavior correlation in CA2/CA3/DG for the room task: (1) A generic attention effect (e.g., arousal, alertness, and motivation) shared across both art and room tasks, and (2) a selective attention effect unique to the room task. If this correlation is partly attributable to a generic effect, then art-state pattern similarity (containing the shared but not unique component) should also predict room behavior. This relationship was in fact reliable [$r_{(17)} = 0.56$, $P = 0.01$], suggesting that a generic factor contributed (although this was not evident in the reverse correlation of room-state pattern similarity and art behavior [$r_{(16)} = 0.20$, $P = 0.40$]). However, the presence of a generic effect does not preclude an additional selective effect, which could be isolated by removing the shared component. We therefore re-ran the room-state pattern similarity and room behavior correlation after controlling for art-state pattern similarity, and the relationship persisted [partial $r_{(17)} = 0.66$, $P = 0.0005$]. Confirming that the correlation between art-state pattern similarity and room behavior only reflected a generic effect, it was eliminated by controlling for room-state pattern similarity [partial $r_{(17)} = -0.001$, $P = 0.996$]. Moreover, room behavior was more strongly correlated with room- versus art-state pattern similarity (dependent-correlation test, $P = 0.04$). Taken together, these results suggest that, above and beyond any generic attentional effects, representational stability in CA2/CA3/DG for the room task was modulated by selective attention in a behaviorally meaningful way.

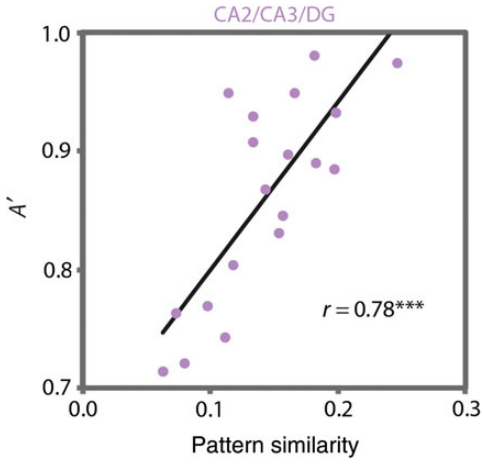

Figure 7. Brain–behavior relationships. Individual differences in room-state pattern similarity in CA2/CA3/DG were strongly correlated with individual differences in behavioral performance (A') on the room task. This effect was specific to this region, to the room task, and to the pattern similarity measure. ***$P < 0.001$.

Discussion

The consequences of attention for neural processing have been investigated extensively in sensory systems, but much less so in memory systems. At the same time, in terms of behavior, attention not only affects perception but also learning and memory. In the current study, we address this gap by asking how attention modulates regions in the human brain that are critical for long-term episodic memory, namely the hippocampus and surrounding MTL cortex. We found that attention modulates both the level of activity and the stability of activity patterns in these regions.

In MTL cortex, PRc showed enhanced activity when attention was directed to art, consistent with its role in processing object information (e.g., Brown and Aggleton 2001; Davachi 2006), whereas PHc and ERc showed enhanced activity when attention was directed to the layout of a room, consistent with their roles in processing spatial information (e.g., Epstein and Kanwisher 1998; Jacobs et al. 2013). Additionally, activity patterns in these regions were more similar when comparing trials from the same (vs. different) attentional states, and PHc and ERc showed more such stability for room than art states. Thus, in MTL cortex, attention modulated both the strength of overall activity and the stability of activity patterns, and furthermore, these 2 measures depended on each other.

The effect of attention on the hippocampus was distinct from its effects on MTL cortex. Attention modulated the level of activity and the stability of activity patterns in opposite directions in CA1 and CA2/CA3/DG, with these regions showing lower activity but greater pattern similarity for room versus art states. Moreover, there was a correlation between performance on the room task and room-state pattern similarity in CA2/CA3/DG. Finally, again in contrast to the MTL cortex, there was no dependence between activity and pattern similarity in the CA fields and DG. Thus, attentional modulation of hippocampal pattern similarity is distinct from modulation of overall activity, and only the former was behaviorally meaningful.

We interpret the similarity of hippocampal activity patterns within task as reflecting selective top-down attention to the same kind of information across trials of that task. Thus, the "representation" being stabilized may be related to the features that are attended or the abstract goal that defines what those features are. For example, in the room task, the attended features are walls and furniture, which share similarities across images and trials of the room task even if there are differences in their specific low-level properties. Thus, a reliable activity pattern for the room state might reflect similarities in the general types of features being attended, though not the trial-specific features. Alternatively, the pattern of activity might reflect the abstract goal of attending to geometric layout; in this case, the features per se are not represented in the hippocampus, but rather the attentional filter that specifies which types of features are to be selected. This filter is common to all trials of a given state, resulting in a shared activity pattern. Our data are consistent with both possibilities, and thus we emphasize that "representational stability" refers to representations of either abstract features or attentional goals. Indeed, whereas this distinction cleanly maps onto sensory (e.g., visual cortex) versus control systems (e.g., prefrontal cortex), which are both affected by goal-directed attention but are thought to represent features and goals, respectively, the content of hippocampal representations remains an active area of inquiry (e.g., Liang et al. 2013).

There is one additional possibility, namely that hippocampal activity patterns reflect the specific target geometric layout or artistic style on a given trial. We found no evidence for this, however, as there was no difference in hippocampal pattern similarity across trials generated from the same versus different templates, even when restricting the analysis to trials of the same attentional state (see Supplementary Fig. 6).

Relation to Other Studies of Hippocampal Pattern Similarity

Our findings complement an emerging body of work utilizing multivariate pattern analysis to examine information represented in the human hippocampus. For example, patterns of hippocampal activity reliably discriminate between recall of different episodic memories (Chadwick et al. 2010), different locations within a spatial environment (Hassabis et al. 2009), different spatial environments altogether (Stokes et al. 2014), different scenes retrieved from long-term memory (Bonnici et al. 2012), different facing directions and locations in highly familiar environments (Vass and Epstein 2013), and different reward contexts that incentivize long-term memory encoding (Wolosin et al. 2013). The current work adds to these findings by showing that hippocampal activity patterns contain information about individuals' current top-down attentional state, when stimuli are held constant and retrieval from long-term memory is not required.

At first blush, prior findings of distinct activity patterns in the hippocampus for different spatial locations and environments (e.g., Hassabis et al. 2009; Vass and Epstein 2013; Stokes et al. 2014) and the mnemonic relevance of distinct, rather than similar, hippocampal representations (LaRocque et al. 2013) may seem inconsistent with our report of greater pattern similarity in the room task, in which different layouts were attended across trials. However, these 2 effects are not mutually exclusive: Activity patterns in the hippocampus may have a shared component across different layouts and trials that reflects similarities in the general features that are attended (walls and furniture) or the abstract goal of attending to spatial information, as well as unique components that reflect specific stimulus details about each layout. Our task was designed to identify the shared component reflective of attentional states, and thus, unlike prior studies, we did not seek to identify image- or environment-specific activity patterns. Instead, we focused on the pattern of activity across many images within a trial, and intentionally re-used images across trials with different target layouts. A different study design that enables measurement of image-specific activity patterns would be necessary to examine whether the hippocampus also represents unique aspects of each room.

Relation to Place Field Stability in Rodents

Studies of freely navigating rodents have found that manipulations of spatial attention—operationalized by varying task demands—affect the stability of place cell firing (e.g., Kentros et al. 2004; Muzzio, Levita, et al. 2009; see Muzzio, Kentros, et al. 2009). Tasks that place demands on olfactory rather than spatial cues do not increase place field stability, but do affect the stability of odor representations in the hippocampus (Muzzio, Levita, et al. 2009). The latter finding suggests that selective attention per se modulates different kinds of representational stability, rather than overall arousal or motivation. Finally, the stability of place cell firing correlates with spatial task performance (Kentros et al. 2004), highlighting the behavioral relevance of stable representations.

The preceding work has focused on the firing stability of individual place cells, but the stability of networks of cells may also

be affected by attentional states. Global, network-level switches between cell assemblies occur when animals utilize different spatial reference frames, suggesting that these assemblies code for the animals' attentional state (e.g., Jackson and Redish 2007; Kelemen and Fenton 2010; see also Fenton et al. 2010). Taken together, these studies suggest that attention induces representational stability in the hippocampus, both in individual cells and across broader networks.

We also found evidence for representational stability, but in this case in the human hippocampus, at the scale of voxels in functional neuroimaging, and for selective attention between 2 states from the same modality. Consistent with prior studies at a different level of analysis (Kentros et al. 2004), this stability correlated with individual differences in performance on a spatial task. Additionally, studies with rodents have found that attention does not modulate overall firing rates in the hippocampus (e.g., Kentros et al. 2004; Muzzio, Levita, et al. 2009), but here we show that selective attention can have opposite effects on activity and representational stability. Importantly, we also show that these effects are dissociable, and may result from modulation of partly non-overlapping sets of voxels.

In contrast to the rodent work, where the representations being stabilized were of spatial locations or olfactory stimuli (e.g., Kentros et al. 2004; Muzzio et al. 2009), in the current work it is unlikely that specific stimulus properties such as viewpoints, colors, or shapes were the basis of representational stability. This is because such cues were not diagnostic for accurate performance, differed greatly over trials, and were held constant across tasks. Instead, reliable within-task representations could be related to general features in the focus of attention (walls and furniture vs. art) or, at a higher level of abstraction, the goal state itself. Thus, our findings converge with—but also extend and complement—single-unit recordings in rodents.

When Does Attention Modulate the Hippocampus?

Previous studies of attentional modulation in the hippocampus in tasks without demands on long-term memory have failed to find effects (e.g., Yamaguchi et al. 2004; Dudukovic et al. 2010; cf. Newmark et al. 2013). However, several studies in the memory literature have shown goal-directed modulation of the hippocampus. At encoding, different ways of orienting attention can affect the magnitude of univariate subsequent memory effects (Uncapher and Rugg 2009; Carr et al. 2013; but see Schott et al. 2013). For example, attention to the location of objects at encoding results in greater hippocampal activity for subsequently remembered versus forgotten locations, but not colors, in a source memory test (Uncapher and Rugg 2009). Likewise, at retrieval, certain forms of divided attention reduce hippocampal activity (Fernandes et al. 2005; but see Iidaka et al. 2000) and attentional or goal states affect the magnitude and nature of univariate hippocampal signals (Dudukovic and Wagner 2007; Duncan et al. 2012; Hashimoto et al. 2012). More broadly, if attention is construed as reflecting how a person's internal state affects the selection of goal-relevant information, then an additional literature on motivated encoding may be relevant (Adcock et al. 2006; Wolosin et al. 2013). These studies show that reward cues that incentivize remembering certain items or associations alter the motivational state of a participant, with high- versus low-reward states linked to better memory, stronger subsequent memory effects in the hippocampus, and different hippocampal activity patterns. Indeed, the discriminability of activity patterns in CA2/CA3/DG for different motivational states correlates with subsequent associative memory (Wolosin et al. 2013), which nicely complements our finding of behavioral correlations with CA2/CA3/DG pattern similarity. An important difference, however, is that we investigated behavior in an online attention task as opposed to long-term memory encoding.

Prior evidence of goal-directed modulation of mnemonic processes in the hippocampus raises the possibility that such processes were engaged during our tasks. That is, although we did not encourage or require the use of long-term memory, our findings might be interpreted as reflecting enhanced incidental encoding of task-relevant information by the hippocampus. Long-term memory might be useful in the current task, because base images and their matches were repeated across trials. On the other hand, long-term memory might have hurt performance because of proactive interference from earlier trials with highly similar or identical stimuli (Chadwick et al. 2014).

If long-term memory was beneficial, it is unclear which task in the current study was associated with better memory, because of the opposite sign of univariate and multivariate hippocampal effects across tasks. Greater pattern similarity (observed in the room task) and greater univariate activity (observed in the art task) have both been linked to enhanced encoding (e.g., Carr et al. 2013; Wolosin et al. 2013). Importantly, beyond any relationship to long-term memory, we found that modulation of the hippocampus was predictive of online behavior in the attention task, as shown by a robust brain/ behavior correlation in CA2/CA3/DG. This suggests that the hippocampus can be involved in attentional processing without overt long-term memory demands. Future studies will be needed to directly relate online measures of attentional modulation in the hippocampus to subsequent episodic memory.

Why did we observe modulation of the hippocampus by the immediate focus of attention when other studies have not? One reason may be related to the attention tasks used in prior studies, which involved orienting attention to locations or objects. This is a common approach for studying attentional modulation of visual cortex (see Kastner and Ungerleider 2000; Maunsell and Treue 2006; Gilbert and Li 2013), but may not be well suited for studying modulation of the hippocampus. In particular, these tasks and stimuli may not place sufficient demands on the computational repertoire of the hippocampus (Shohamy and Turk-Browne 2013). The hippocampus is supramodal—not only affected by multiple sensory modalities but also abstract factors such as goal state, task, context, and prospective decisions (e.g., Johnson and Redish 2007)—and it is also fundamentally relational, configural, and contextual in nature (Cohen and Eichenbaum 1993; Brown and Aggleton 2001; Davachi 2006). Thus, in order to study attentional modulation of the hippocampus, it may be important to use tasks that tap into more flexible relational representations, rather than representations of particular stimuli.

We designed the current tasks to place demands on these kinds of representations. The art task required abstraction and generalization from a given painting to identify a stylistically similar one. The room task required abstraction and generalization from a given room to identify one with the same spatial layout from a different perspective. In both tasks, low-level visual information was not particularly useful—across scenes, the content of the paintings, the perspective, the wall color, and the furniture changed. Thus, both tasks were designed to recruit relational processing, albeit in different ways.

The finding of stronger hippocampal pattern similarity and behavioral correlations for the room task, however, is consistent with an interpretation that this task placed greater emphasis on relational processing than did the art task. Alternatively, it may

be the case that the processing of spatial relations is privileged in the hippocampus relative to other kinds of relations. Future studies will be needed to determine whether the need for relational processing or the type of relation is the critical determinant of when attention modulates the hippocampus.

The current findings show that studies of attention in the hippocampus may have to proceed differently from those targeted at cortical areas: Attentional modulation may be most apparent in representational stability, and the tasks used to manipulate attention may require more complexity and abstraction.

Future Directions

There are 2 potential ways that attention can influence the activity of a brain region: (1) by strengthening the output of an earlier area, resulting in stronger input to the region, or (2) by directly modulating computations within the region. It is beyond the scope of the current study to adjudicate between these possibilities, and by either account, our results provide clear evidence of attentional effects in the hippocampus. Nevertheless, some of our data are supportive of the second account—that attention directly modulates the hippocampus, distinct from its effects on areas of MTL cortex that provide hippocampal input. First, the main cortical input to the hippocampus comes from ERc, which showed increased activity for room versus art states, whereas the opposite was observed in the CA fields and DG. Secondly, correlations between room-state pattern similarity and behavior were strongest within the hippocampus (a weaker correlation, significant only at an uncorrected threshold, was observed in ERc).

One caveat regarding the opposite univariate effects in ERc and hippocampus is that the relationship between BOLD activity and neural activity may be different in cortex and hippocampus. Unlike cortical regions, the hippocampal BOLD signal is less consistently tied to local field potentials (Ekstrom 2010). Thus, different directions of BOLD modulation may be related to different mappings between BOLD and neural activity in different regions, rather than a difference in the nature of the effect itself. Future studies with more invasive methods (e.g., intracranial recordings in humans) will be informative in this regard.

If attention directly modulates the hippocampus, then by what mechanism does this occur? The hippocampus receives afferent projections from all of the main neuromodulatory systems implicated in various aspects of attention, including dopaminergic, cholinergic, and noradrenergic inputs (see Muzzio, Kentros, et al. 2009)—and manipulating these systems affects place field stability (e.g., Kentros et al. 2004). These neurotransmitters may have also played a modulatory role in the current tasks. For example, acetylcholine amplifies afferent signals into the hippocampus and suppresses excitatory recurrent connections in CA3 (Newman et al. 2012), potentially leading to the observed increases in stability and reductions in activity, respectively, in the room versus art tasks. This suggests the possibility that selective attention to spatial information is associated with enhanced cholinergic modulation in the hippocampus. This link will need to be investigated in future research, however, since strengthening of afferent signals could also conceivably amplify noise from the environment and reduce representational stability.

Another important question concerns the relationship between univariate activity and pattern similarity in the hippocampus. Unlike the MTL cortical regions in the current study, univariate activity and pattern similarity went in opposite directions in the CA fields and DG. This could be explained by a sharpening mechanism, whereby neurons tuned for the current task inhibit ones that are not, resulting in a sparser and more selective pattern of activity—thus, less activity and more stability, respectively (Kok et al. 2012; Hulme et al. 2014). This can be reconciled with the results of the MUD analysis—where the activity of CA1 and CA2/CA3/DG voxels was unrelated to their contribution to pattern similarity—if both excitation and inhibition contribute to stable patterns (see Supplementary Fig. 7 and Supplementary Table 1). If this is the case, then carrying out the MUD analysis with absolute, rather than signed, univariate activity should reveal a relationship between the activity of voxels in these hippocampal ROIs and their contribution to pattern similarity. Indeed, we found this to be the case [MUD in CA1: mean $r = 0.43$, $t_{(18)} = 8.33$, $P < 0.0001$; CA2/CA3/DG: mean $r = 0.42$, $t_{(18)} = 7.44$, $P < 0.0001$]. This suggests that the absence of a MUD effect with signed univariate activity indicates a balance of activation and deactivation that contribute to the stability of multivariate patterns in the hippocampus. Future studies using neural recordings will be necessary to elucidate the conditions under which activity and pattern similarity do and do not align with each other, and will complement the current fMRI results by tying the effects directly to neural activity.

Finally, the current tasks required that the first image be held in mind while searching the following set of images for an art or room match. The hippocampus is thought to be important for binding together items or events that are separated in time (Howard and Eichenbaum 2013) and for some aspects of working memory more generally (e.g., Hannula and Ranganath 2008; also see Yonelinas 2013). Thus, our findings could arguably be conceptualized in terms of working memory. However, we consider working memory to be an essential component of top-down attention—it is required for maintaining one's current goal(s), which guide attentional selection and behavior (see Chun et al. 2011). Moreover, even if the current task is viewed through the lens of working memory, our findings still provide novel evidence that different goal states are represented in distinct hippocampal activity patterns. Nevertheless, future studies could directly examine the contribution of working memory to attentional modulation of the hippocampus by parametrically manipulating working memory load.

Conclusions

In the current study, we demonstrated that attention modulates the MTL. In MTL cortex, attention increased both the strength of the response and the stability of activity patterns, and these 2 outcomes were related. In the hippocampus, attention again modulated overall activity and pattern similarity, but these outcomes were dissociable, and only pattern similarity was behaviorally meaningful. These findings show that there are multiple signatures of attention throughout the human brain, including in systems not traditionally linked to sensory processing, like the hippocampus.

Authors' Contributions

M.A. and N.B.T.-B. conceived and designed the experiment. M.A. performed the experiment and analyzed the data. M.A. and N.B.T.-B. wrote the paper.

Supplementary Material

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

Funding

This work was supported by the National Institutes of Health (R01-EY021755 to N.B.T.-B.).

Notes

We thank Ken Norman and Jordan Poppenk for feedback on a previous version of this manuscript. *Conflict of Interest*: None declared.

References

Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JDE. 2006. Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron. 50:507–517.

Aly M, Ranganath C, Yonelinas AP. 2013. Detecting changes in scenes: the hippocampus is critical for strength-based perception. Neuron. 78:1127–1137.

Bonnici HM, Kumaran D, Chadwick MJ, Weiskopf N, Hassabis D, Maguire EA. 2012. Decoding representations of scenes in the medial temporal lobes. Hippocampus. 22:1143–1153.

Brown MW, Aggleton JP. 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat Rev Neurosci. 2:51–61.

Brown TI, Ross RS, Keller JB, Hasselmo ME, Stern CE. 2010. Which way as I going? Contextual retrieval supports the disambiguation of well learned overlapping navigational routes. J Neurosci. 30:7414–7422.

Brown TI, Stern CE. 2014. Contributions of medial temporal lobe and striatal memory systems to learning and retrieving overlapping spatial memories. Cereb Cortex. 24:1906–1922.

Carr VA, Engel SA, Knowlton BJ. 2013. Top-down modulation of hippocampal encoding activity as measured by high-resolution functional MRI. Neuropsychologia. 51:1829–1837.

Chadwick MJ, Bonnici HM, Maguire EA. 2014. CA3 size predicts the precision of memory recall. Proc Natl Acad Sci. 111: 10720–10725.

Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA. 2010. Decoding individual episodic memory traces in the human hippocampus. Curr Biol. 20:544–547.

Chun MM, Golomb JD, Turk-Browne NB. 2011. A taxonomy of external and internal attention. Annu Rev Psychol. 62:73–101.

Chun MM, Turk-Browne NB. 2007. Interactions between attention and memory. Curr Opin Neurobiol. 17:177–184.

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

Coutanche MN. 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? Cogn Affect Behav Neurosci. 13:667–673.

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

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage. 97:271–283.

Davis T, Poldrack RA. 2013. Measuring neural representations with fMRI: practices and pitfalls. Ann NY Acad Sci. 1296:108–134.

Dudukovic NM, Preston AR, Archie JJ, Glover GH, Wagner AD. 2010. High-resolution fMRI reveals match enhancement and attentional modulation in the human medial temporal lobe. J Cogn Neurosci. 23:670–682.

Dudukovic NM, Wagner AD. 2007. Goal-dependent modulation of declarative memory: neural correlates of temporal recency decisions and novelty detection. Neuropsychologia. 45:2608–2620.

Duncan K, Ketz N, Inati SJ, Davachi L. 2012. Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus. 22:389–398.

Ekstrom A. 2010. How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. Brain Res Rev. 62:233–244.

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

Fenton AA, Lytton WW, Barry JM, Lenck-Santini PP, Zinyuk LE, Kubík S, Bureš J, Poucet B, Muller RU, Olypher AV. 2010. Attention-like modulation of hippocampus place cell discharge. J Neurosci. 30:4613–4625.

Fenton AA, Muller RU. 1998. Place cell discharge is extremely variable during individual passes of the rat through the firing field. Proc Natl Acad Sci USA. 95:3182–3187.

Fernandes MA, Moscovitch M, Ziegler M, Grady C. 2005. Brain regions associated with successful and unsuccessful retrieval of verbal episodic memory as revealed by divided attention. Neuropsychologia. 43:1115–1127.

Gilbert CD, Li W. 2013. Top-down influences on visual processing. Nat Rev Neurosci. 14:350–363.

Hannula DE, Ranganath C. 2008. Medial temporal lobe activity predicts successful relational memory binding. J Neurosci. 28:116–124.

Hardt O, Nadel L. 2009. Cognitive maps and attention. In: Srinivasan N, editor. Progress in brain research. Vol. 176. The Netherlands: Elsevier. p. 181–194.

Hartley T, Bird CM, Chan D, Cipolotti L, Husain M, Vargha-Khadem F, Burgess N. 2007. The hippocampus is required for short-term topographical memory in humans. Hippocampus. 17:34–48.

Hashimoto R, Abe N, Ueno A, Fujii T, Takahashi S, Mori E. 2012. Changing the criteria for old/new recognition judgments can modulate activity in the anterior hippocampus. Hippocampus. 23:141–148.

Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA. 2009. Decoding neuronal ensembles in the human hippocampus. Curr Biol. 19:546–554.

Howard MW, Eichenbaum HB. 2013. The hippocampus, time, and memory across scales. J Exp Psych Gen. 142:1211–1230.

Hulme OJ, Skov M, Chadwick MJ, Siebner HR, Ramsøy TZ. 2014. Sparse encoding of automatic visual association in hippocampal networks. NeuroImage. 102:458–464.

Iidaka T, Anderson ND, Kapur S, Cabeza R, Craik FIM. 2000. The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. J Cogn Neurosci. 12:267–280.

Jackson J, Redish AD. 2007. Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. Hippocampus. 17:1209–1229.

Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei X-X, Suthana N, Sperling MR, Sharan AD, Fried I, et al. 2013. Direct recordings of grid-like neuronal activity in human spatial navigation. Nat Neurol. 16:1188–1191.

Johnson A, Redish AD. 2007. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J Neurosci. 27:12176–12189.

Kastner S, Ungerleider LG. 2000. Mechanisms of visual attention in the human cortex. Annu Rev Neurosci. 23:315–341.

Kelemen E, Fenton AA. 2010. Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. PLoS Biol. 8:e1000403. 1–14.

Kentros CG, Agnihotri NT, Streater S, Hawkins RD, Kandel ER. 2004. Increased attention to spatial context increases both place field stability and spatial memory. Neuron. 42:283–295.

Kok P, Jehee JFM, de Lange FP. 2012. Less is more: expectation sharpens representations in the primary visual cortex. Neuron. 75:265–270.

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. Front Hum Neurosci. 2:1–28.

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.

Lee ACH, Yeung LK, Barense MD. 2012. The hippocampus and visual perception. Front Hum Neurosci. 6:91. , 1–17.

Liang JC, Wagner AD, Preston AR. 2013. Content representation in the human medial temporal lobe. Cereb Cortex. 23:80–96.

Maunsell JHR, Treue S. 2006. Feature-based attention in visual cortex. Trends Neurosci. 29:317–322.

Muzzio IA, Kentros C, Kandel E. 2009. What is remembered? Role of attention on the encoding and retrieval of hippocampal representations. J Physiol. 587:2837–2854.

Muzzio IA, Levita L, Kulkarni J, Monaco J, Kentros C, Stead M, Abbott LF, Kandel ER. 2009. Attention enhances the retrieval and stability of visuospatial and olfactory representations in the dorsal hippocampus. PLoS Biol. 7:e1000140. 1–20.

Newman EL, Gupta K, Climer JR, Monaghan CK, Hasselmo ME. 2012. Cholinergic modulation of cognitive processing: insights drawn from computational models. Front Behav Neurosci. 6:24. , 1–19.

Newmark RE, Schon K, Ross RS, Stern CE. 2013. Contributions of the hippocampal subfields and entorhinal cortex to disambiguation during working memory. Hippocampus. 23:467–475.

O'Craven KM, Downing PE, Kanwisher N. 1999. fMRI evidence for objects as the units of attentional selection. Nature. 401:584–587.

Pernet CR, Wilcox R, Rousselet GA. 2013. Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. Front Psych. 3:606.

Schott BH, Wustenberg T, Wimber M, Fenker DB, Zierhut KC, Seidenbecher CI, Heinze H-J, Walter H, Düzel E, Richardson-Klavehn A. 2013. The relationship between level of processing and hippocampal-cortical functional connectivity during episodic memory formation in humans. Hum Brain Mapp. 34:407–424.

Shohamy D, Turk-Browne NB. 2013. Mechanisms for widespread hippocampal involvement in cognition. J Exp Psych Gen. 142:1159–1170.

Stokes J, Kyle C, Ekstrom AD. 2014. Complementary roles of human hippocampal subfields in differentiation and integration of spatial context. J Cogn Neurosci. doi:10.1162/jocn_a_00736.

Stokes MG, Atherton K, Patai EZ, Nobre AC. 2012. Long-term memory prepares neural activity for perception. Proc Natl Acad Sci. 109:E360–E367.

Summerfield JJ, Lepsien J, Gitelman DR, Mesulam MM, Nobre AC. 2006. Orienting attention based on long-term memory experience. Neuron. 49:905–916.

Townsend JT, Ashby FG. 1978. Methods of modeling capacity in simple processing systems. In: Castellan J, Restle F, editors. Cognitive Theory. Vol. 3. Hillsdale (NJ): Erlbaum. p. 200–239.

Uncapher MR, Rugg MD. 2009. Selecting for memory? The influence of selective attention on the mnemonic binding of contextual information. J Neurosci. 29:8270–8279.

Vass LK, Epstein RA. 2013. Abstract representations of location and facing direction in the human brain. J Neurosci. 33:6133–6142.

Wolosin SM, Zeithamova D, Preston AR. 2013. Distributed hippocampal patterns that discriminate reward context are associated with enhanced associative binding. J Exp Psych Gen. 142:1264–1276.

Worsley KJ, Chen J-I, Lerch J, Evans AC. 2005. Comparing functional connectivity via thresholding correlations and singular value decomposition. Philos Trans R Soc B. 360:913–920.

Yamaguchi S, Hale LA, D'Esposito M, Knight RT. 2004. Rapid prefrontal-hippocampal habituation to novel events. J Neurosci. 24:5356–5363.

Yonelinas AP. 2013. The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behav Brain Res. 254:34–44.