OXFORD

Cerebral Cortex Communications, 2021, **2**, 1–15

doi: 10.1093/texcom/tgab045
Original Article

ORIGINAL ARTICLE

Detailed Episodic Memory Depends on Concurrent Reactivation of Basic Visual Features within the Posterior Hippocampus and Early Visual Cortex

Michael B. Bone[1,2] and Bradley R. Buchsbaum[1,2]

[1]Rotman Research Institute at Baycrest, Toronto, Ontario, M6A 2E1, Canada and [2]Department of Psychology; University of Toronto, Toronto, Ontario, M5S 1A1, Canada

*Address correspondence to Michael B. Bone. Email: michael.bone@mail.utoronto.ca

Abstract

The hippocampus is a key brain region for the storage and retrieval of episodic memories, but how it performs this function is unresolved. Leading theories posit that the hippocampus stores a sparse representation, or "index," of the pattern of neocortical activity that occurred during perception. During retrieval, reactivation of the index by a partial cue facilitates the reactivation of the associated neocortical pattern. Therefore, episodic retrieval requires joint reactivation of the hippocampal index and the associated neocortical networks. To test this theory, we examine the relation between performance on a recognition memory task requiring retrieval of image-specific visual details and feature-specific reactivation within the hippocampus and neocortex. We show that trial-by-trial recognition accuracy correlates with neural reactivation of low-level features (e.g., luminosity and edges) within the posterior hippocampus and early visual cortex for participants with high recognition lure accuracy. As predicted, the two regions interact, such that recognition accuracy correlates with hippocampal reactivation only when reactivation co-occurs within the early visual cortex (and vice versa). In addition to supporting leading theories of hippocampal function, our findings show large individual differences in the features underlying visual memory and suggest that the anterior and posterior hippocampus represents gist-like and detailed features, respectively.

Key words: neocortex, MVPA, memory, hippocampus, fMRI

Introduction

The ability to mentally re-experience vivid imagery from a past event is a defining feature of episodic memory. A large body of evidence indicates that vivid episodic recollection is implemented by the reactivation of neural activity that occurred during the recalled episode—particularly within modality-specific neocortical regions, e.g., the visual cortex for visual memories (Buchsbaum et al. 2012; Kuhl et al. 2012; Johnson and Johnson 2014; St-Laurent et al. 2014; Naselaris et al. 2015; Wing et al. 2015; Bone et al. 2019; Bone et al. 2020). Leading theories of hippocampal function posit that it mediates neocortical reactivation by storing a compressed representation, or "index," of the neocortical activity that occurred during perception, thereby facilitating the encoding of arbitrary associations between an event's constituent features that can later be retrieved by the reactivation of a subset of the encoded features (Teyler and DiScenna 1986; McClelland et al. 1995; Nadel et al. 2000; Kumaran et al. 2016; Sekeres et al. 2018; Barron et al. 2020). Despite being a key element of current theories of episodic memory and hippocampal function,

Received: 22 June 2021; **Revised:** 22 June 2021; **Accepted:** 25 June 2021

direct evidence for the belief that episodic memories arise from the interaction of hippocampal and neocortical representations *of the same information* remains limited in humans.

Due to the anatomical position of the hippocampus near the end of the ventral visual pathway, it is generally assumed that the constituent features directly indexed by the hippocampus are limited to "high-level" representations, such as object category and the spatiotemporal relations between objects (Barron et al. 2020). However, humans have the capacity to vividly and accurately reconstruct a past visual experience from the original perspective—which is severely compromised with damage to the hippocampus (Tulving 1993; Rubin et al. 2003). High-level representations capture statistical regularities shared across category members, so they generally lack the event-specific information that would be required for vivid and accurate visual recall, such as object pose and lighting. Low-level features (e.g., edges, luminosity) are not abstracted to the same extent as high-level features, so indexing neocortical activity representing a sparse set of low-level features, in addition to high-level features, would provide the information necessary to constrain memory reactivation to be specific to the past event. Therefore, if the hippocampus supports detailed episodic memory, then retrieval should be facilitated by reactivation of a hippocampal network that indexes neocortical activity representing low-level visual features—particularly within the early visual cortex.

The patterns that can be indexed and the features that they represent are determined by the physical connections between the hippocampus and neocortex. This connectivity varies along the longitudinal axis, with the posterior hippocampus (pHC) reciprocally linked to sensory regions of the posterior neocortex, and the anterior hippocampus (aHC) connected to anterior neocortical structures implicated in the representation of schemas (Kier et al. 2004; Catenoix et al. 2011; Poppenk and Moscovitch 2011; Poppenk et al. 2013). Based on differences in connectivity and an approximately linear increase in receptive field size (Kjelstrup et al. 2008) along the hippocampal longitudinal axis, researchers (Poppenk et al. 2013; McCormick et al. 2015) postulate that the pHC indexes the fine-grained perceptual features of an event, which constitute vivid, perceptually rich memories, whereas the aHC indexes coarse-grained features, which support gist-like memories. Although the bulk of experimental evidence supports this hypothesis (Evensmoen et al. 2015; Schlichting et al. 2015; Brunec et al. 2018; Sekeres et al. 2018; Grady 2020), at least one recent finding implicates the aHC and pHC in the representation of detailed and gist-like memories, respectively (Dandolo and Schwabe 2018).

The current experiment addresses three questions: First, we ask whether detailed and accurate visual memory is associated with hippocampal reactivation of neural activity representing visual features—particularly low-level features. Second, is detailed visual memory more strongly associated with reactivation within the pHC than the aHC? And third, does the association between memory performance and hippocampal reactivation of low-level visual features depend upon concurrent reactivation of the same features within the visual cortex, as predicted by the hippocampus's role in indexing neocortical activity?

To this end, we combined functional magnetic resonance imaging (fMRI) and measures of neural reactivation applied to a challenging recall and recognition task. We defined reactivation in two ways: namely, image-specific and feature-specific reactivation. Image-specific reactivation refers to multivoxel reactivation of neural activity that occurred during perception of specific *images* (the most common approach within the memory reactivation literature), whereas feature-specific reactivation refers to reactivation of neural activity that occurred during perception of specific *features* shared across images. Cross-validation is used to exclude the encoding trials of the target image from the feature-specific training set, so the training sets for the two measures are different. Features were extracted from layer node activations of the VGG16 deep neural net (DNN) (Simonyan and Zisserman 2014). Activations from the convolutional layers (1–13), and the fully connected layers (14–16) were used, corresponding to low-visual (edges and luminosity; 1–4), middle-visual (simple object parts and patterns; 5–9), high-visual (complex object parts, e.g., faces; 10–13) and semantic (object category; 14–16) features, respectively. Because feature-specific reactivation is trained across images (excluding the target/cued image), the method requires neural representations of visual features to be at least partially shared between images. This would not be possible if memories with shared features are represented by non-overlapping neural patterns (Yassa and Stark 2011). Findings indicate that similar events are represented by partially overlapping neural patterns within the hippocampus, and that the degree of representational overlap varies between hippocampal subfields (Bakker et al. 2008; Yassa and Stark 2011; Rolls 2013; Berron et al. 2016), so feature-specific reactivation should be detectable within the hippocampus.

The experiment (see Bone et al. 2020, which addressed neocortical reactivation using the same experimental data) had two video viewing runs, used to train the feature-specific encoding models, and three sets of alternating encoding and retrieval runs (Fig. 1). During encoding runs participants memorized a set of thirty color images (per run) while performing a 1-back task. In the following retrieval runs, participants' recall and recognition memory of the images was assessed. Neural reactivation was measured while participants visualized a cued image within a light-gray rectangle, followed by a memory vividness rating. An image was then presented that was either identical to the cued image or a similar lure, and the participants judged whether they had seen the image during encoding and provided a confidence rating. Critically, the lure images carried the semantic and visual gist of the cued images but had different fine-grained details (see Supplementary Fig. 1 for example image pairs). Consequently, accuracy on the recognition task served as a measure of detailed, rather than gist-like, memory.

We found that the within-subject correlation between recognition accuracy and image-specific reactivation during recall was significantly greater within the pHC relative to the aHC for individuals with high lure accuracy. Feature-specific reactivation during recall produced a similar result, with the correlations limited to visual features (i.e., excluding semantic features). Moreover, low-level visual reactivation within the pHC positively interacted with low-level visual reactivation within the calcarine sulcus, indicating that the correlation between recognition accuracy and hippocampal reactivation depended upon reactivation of the same information co-occurring within the early visual cortex (and vice-versa). Overall, our results show that hippocampal representations extend to low-level visual features, and we confirm that the precision of these representations (including low-, mid-, and high-level visual features) varies along the long-axis, such that gist-like and detailed representations occur within the aHC and pHC, respectively. Our findings also expand upon previous work (Bone et al. 2020) by showing that individual differences in the features underlying visual memory extend to the hippocampus. Finally,

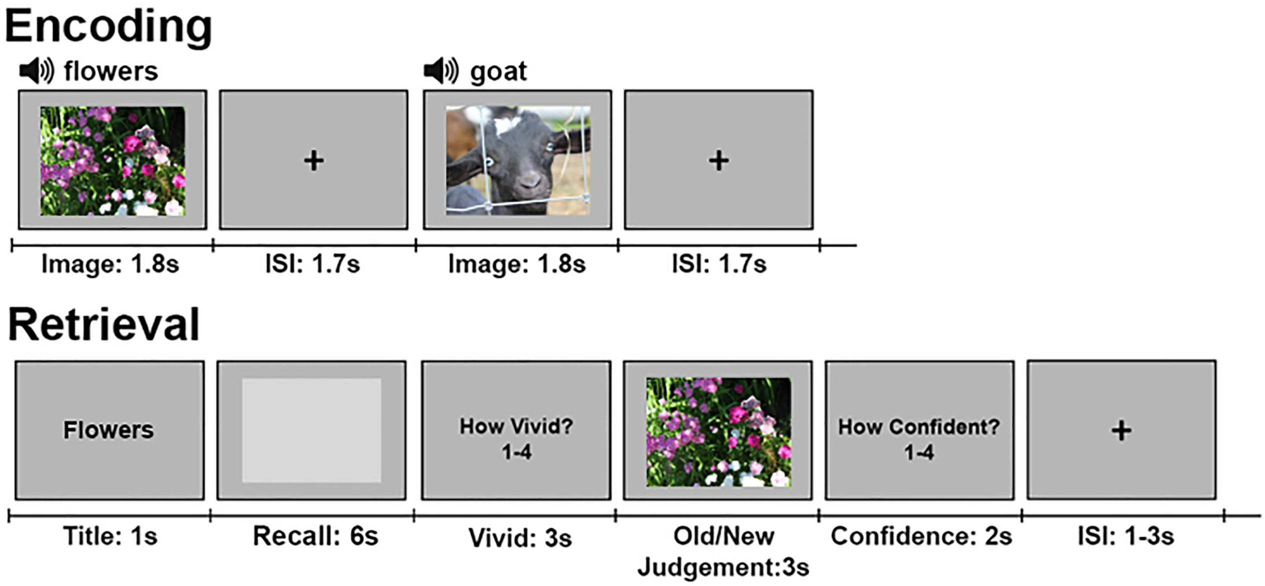

Figure 1. Procedure. Alternating image encoding and retrieval tasks. During encoding, participants performed a 1-back task while viewing a sequence of color photographs accompanied by matching auditory labels. During retrieval, participants 1) were cued with a visually-presented label, 2) retrieved and maintained a mental image of the associated photograph over a 6-s delay, 3) indicated the vividness of their mental image using 1–4 scale, 4) decided whether a probe image matched the cued image, and 5) entered their confidence rating with respect to the old/new judgment. Due to copyright concerns, images used in the study could not be included in the figure. The images depicted in the figure are for explanatory purposes only.

our results support leading theories that claim the hippocampus implements episodic recall by storing and reactivating a sparse index of neocortical activity.

Materials and Methods

Participants

Thirty-seven right-handed young adults with normal or corrected-to-normal vision and no history of neurological or psychiatric disease were recruited through the Baycrest subject pool, tested, and paid for their participation. Informed consent was obtained, and the experimental protocol was approved by the Rotman Research Institute's Ethics Board. Subjects were either native or fluent English speakers and had no contraindications for MRI. Data from 12 of these participants was excluded from the final analyses for the following reasons: excessive head motion (5; removed if > 5 mm within run maximum displacement in head motion), fell asleep (2), did not complete experiment (3), trial labeling error (1), second video run was cut short due to technical difficulties (1). Thus, 25 participants were included in the final analysis (13 males and 12 females, 20–32 years old).

Stimuli

In total, 111 colored photographs (800 by 600) were gathered from online sources. For each image, an image pair was acquired using Google's similar image search function, for a total of 111 image pairs (222 images). Twenty-one image pairs were used for practice, and the remaining 90 were used during the in-scan encoding and retrieval tasks (see Supplementary Fig. 1 for example image pairs). Each image was paired with a short descriptive audio title in a synthesized female voice (https://neospeech.com; voice: Kate) during encoding runs; this title served as a visually presented retrieval cue during the in-scan retrieval task. Two videos used for model training (720 by 480 pixels; 30 fps; 10 m 25 s and 10 m 35 s in length) comprised a series of short (~4 s) clips drawn from YouTube and Vimeo, containing a wide variety of themes (e.g., still photos of bugs, people performing manual tasks, animated text, etc.). One additional video cut from "Indiana Jones: Raiders of the Lost Ark" (1024 by 435 pixels; 10 m 6 s in length) was displayed while in the scanner, but the associated data were not used in this experiment because the aspect ratio (widescreen) did not match the images.

Procedure

Before undergoing MRI, participants were trained on a practice version of the task incorporating 21 practice image pairs. Inside the MRI scanner, participants completed three video viewing runs and three encoding-retrieval sets. The order of the runs was as follows: first video viewing run (short clips 1), second video viewing run (short clips 2), third video viewing run (Indiana Jones clip), first encoding-retrieval set, second encoding-retrieval set, third encoding-retrieval set. A high-resolution structural scan was acquired between the second and third encoding-retrieval sets, providing a break.

Video viewing runs were 10 m 57 s long. For each run, participants were instructed to pay attention while the video (with audio) played within the center of the screen. The order of the videos was the same for all participants.

Encoding-retrieval sets were composed of one encoding run followed by one retrieval run. Each set required the participants to first memorize and then recall 30 images drawn from 30 image pairs. The image pairs within each set were selected randomly, with the constraint that no image pair could be used in more than one set. The image selected from each image pair to be presented during encoding was counterbalanced across subjects. This experimental procedure was designed to limit the concurrent memory load to 30 images for each of three consecutive pairs of encoding-retrieval runs.

Encoding runs were 6 m 24 s long. Each run started with 10s during which instructions were displayed on-screen. Trials began with the appearance of an image in the center of the screen (1.8 s), accompanied by a simultaneous descriptive audio cue (e.g., a picture depicting toddlers would be coupled with the spoken word "toddlers"). Images occupied 800 by 600 pixels of a 1024 by 768 pixel screen. Between trials, a crosshair appeared centrally (font size = 50) for 1.7 s. Participants were instructed to pay attention to each image and to encode as many details as possible so that they could visualize the images as precisely as possible during the imagery task. The participants also performed a 1-back task requiring the participants to press "1" if the displayed image was the same as the preceding image, and "2" otherwise. Within each run, stimuli for the 1-back task were randomly sampled with the following constraints: 1) each image was repeated exactly four times in the run (120 trials per run; 360 for the entire session), 2) there was only one immediate repetition per image, and 3) the other two repetitions were at least 4 items apart in the 1-back sequence.

Retrieval runs were 9 m 32 s long. Each run started with 10s during which instructions were displayed on-screen. Thirty images were then cued once each (the order was randomized), for a total of 30 trials per run (90 for the entire scan). Trials began with an image title appeared in the center of the screen for 1 s (font = Courier New, font size = 30). After 1 s, the title was replaced by an empty rectangular box shown in the center of the screen (6 s), and whose edges corresponded to the edges of the stimulus images (800 by 600 pixels). Participants were instructed to visualize the image that corresponded to the title as accurately as they could within the confines of the box. Once the box disappeared, participants were prompted to rate the subjective vividness (defined as the relative number of recalled visual details specific to the cued image presented during encoding) of their mental image on a 1–4 scale (1 = a very small number of visual details were recalled, 4 = a very large number of visual details were recalled) (3 s) using a four-button fiber optic response box (right hand; 1 = right index finger; 4 = right little finger). This was followed by the appearance of a probe image (800 by 600 pixels) in the center of the screen (3 s), which was either the same as or similar to the trial's cued image (i.e., either the image shown during encoding or its pair). While the image remained on the screen, the participants were instructed to respond with "1" if they thought that the image was the one seen during encoding (old), or "2" if the image was new (responses made using the response box). Following the disappearance of the image, participants were prompted to rate their confidence in their old/new response on a 1–4 scale (2 s) using the response box. Between each trial, a crosshair (font size = 50) appeared in the center of the screen for either 1, 2 or 3 s.

Randomization sequences were generated such that both images within each image pair (image A and B) were presented equally often during the encoding runs across subjects. During retrieval runs each image appeared equally often as a matching (encode A—> probe A) or mismatching (encode A—> probe B) image across subjects. Due to the need to remove several subjects from the analyses, stimulus versions were approximately balanced over subjects.

Setup and Data Acquisition

Participants were scanned with a 3.0-T Siemens MAGNETOM Trio MRI scanner using a 32-channel head coil system. Functional images were acquired using a multiband EPI sequence sensitive to BOLD contrast (22 × 22 cm field of view with a 110 × 110 matrix size, resulting in an in-plane resolution of 2 × 2 mm for each of 63 2-mm axial slices; repetition time = 1.77 s; echo time = 30 ms; flip angle = 62°). A high-resolution whole-brain magnetization prepared rapid gradient echo (MP-RAGE) 3-D T1 weighted scan (160 slices of 1 mm thickness, 19.2 × 25.6 cm field of view) was also acquired for anatomical localization.

The experiment was programmed with the E-Prime 2.0.10.353 software (Psychology Software Tools, Pittsburgh, PA). Visual stimuli were projected onto a screen behind the scanner made visible to the participant through a mirror mounted on the head coil.

fMRI Preprocessing

Functional images were converted into NIFTI-1 format, motion-corrected and realigned to the average image of the first run with AFNI's (Cox 1996) 3dvolreg program. The maximum displacement for each EPI image relative to the reference image was recorded and assessed for head motion. The average EPI image was then co-registered to the high-resolution T1-weighted MP-RAGE structural using the AFNI program align_epi_anat.py (Saad et al. 2009).

The functional data for each experimental task (video viewing, 1-back encoding task, retrieval task) was then projected to a subject-specific cortical surface generated by Freesurfer 5.3 (Dale et al. 1999). The target surface was a spherically normalized mesh with 32 000 vertices that was standardized using the resampling procedure implemented in the AFNI program MapIcosahedron (Argall et al. 2006). To project volumetric imaging data to the cortical surface we used the AFNI program 3dVol2Surf with the "average" mapping algorithm, which approximates the value at each surface vertex as the average value among the set of voxels that intersect a line along the surface normal connecting the white matter and pial surfaces.

The three video scans (experimental runs 1–3), because they involved a continuous stimulation paradigm, were directly mapped to the surface without any pre-processing to the cortical surface. The three retrieval scans (runs 5, 7, 9) were first divided into a sequence of experimental trials with each trial beginning ($t = -2$) 2 s prior to the onset of the retrieval cue (verbal label) and ending 32 s later in 2-s increments. These trials were then concatenated in time to form a series of 90 trial-specific time-series, each of which consisted of 16 samples. The resulting trial-wise data blocks were then projected onto the cortical surface. To facilitate separate analyses of the "recall" and "old/new judgment" retrieval data, a regression approach was implemented. For each trial, the expected hemodynamic response associated with each task was generated by convolving a series of instantaneous impulses (i.e., a delta function) over the task period (10 per second; imagery: 61; old/new: 31) with the SPM canonical hemodynamic response. Estimates of beta coefficients for each trial and task were computed via a separate linear regression per trial (each with 16 samples: one per time point), with vertex activity as the dependent variable, and the expected hemodynamic response values for the "recall" and "old/new judgment" tasks as independent variables. The "recall" beta coefficients were used in all subsequent neural analyses. Data from the three encoding scans (runs 4, 6, 8) were first processed in volumetric space using a trial-wise regression approach, where the onset of each image stimulus was modeled with a separate regressor formed from a convolution of the instantaneous impulse with the SPM canonical hemodynamic response. Estimates of trial-wise beta coefficients were then computed using the "least squares sum" (Mumford et al. 2012) regularized regression approach as implemented in the AFNI program 3dLSS. The 360 (30 unique images per run, 4 repetitions per run, 3 total runs) estimated beta coefficients were then projected onto the cortical surface with 3dVol2Surf.

Hippocampal ROI Definition

To define anterior and posterior hippocampal ROIs, we used the Freesurfer's (version 5.3) automated parcellation of the left and right hippocampi on the T1-weighted image of each participant.

Based on the finding that receptive field size varies linearly along the longitudinal axis in rodents (Kjelstrup et al. 2008), we expected that the level of detail represented would gradually increase from the anterior to posterior potions of the hippocampus in humans. Consequently, the precise boundaries of the "aHC" and "pHC" were not considered to be critical, so the left and right hippocampal ROIs were equally divided into five sections along the antero-posterior axis, yielding five ROIs per hemisphere. These ROIs were then used as masks to extract time-series from the pre-processed and co-registered fMRI data.

We used five equal longitudinal sections because that would allow us to exclude the middle (to reduce signal bleed between the regions and increase the expected difference between them), while maintaining a reasonable number of voxels for decoding when the two posterior and anterior regions are grouped together.

Deep Neural Network Image Features

We used the pretrained TensorFlow implementation of the VGG16 deep neural network (DNN) model (Simonyan and Zisserman 2014; see http://www.cs.toronto.edu/&#x007E;frossard/post/vgg16 for the implementation used). Like AlexNet (the network used in previous studies, e.g., Güçlü and van Gerven 2015), VGG16 uses Fukushima's (1980) original visual-cortex inspired architecture, but with greatly improved top-5 (out of 1000) classification accuracy (AlexNet: 83%, VGG16: 93%). The network's accuracy was particularly important for this study because we did not hand-select stimuli (images and video frames) that were correctly classified by the net. The VGG16 model consists of a total of thirteen convolutional layers and three fully connected layers. 90 image pairs from the memory task and 3775 video frames (3 frames per second; taken from the two short-clip videos; video 1: 1875 frames; video 2: 1900 frames; extracted using "Free Video to JPG Converter" https://www.dvdvideosoft.com/products/dvd/Free-Video-to-JPG-Converter.htm) were resized to 224 × 224 pixels to compute outputs of the VGG16 model for each image/frame. The outputs from the units in all layers were treated as vectors corresponding to low-level visual features (layers 1–4), mid-level visual features (layers 5–9), high-level visual features (layers 10–13) and semantic features (layers 14–16).

Convolutional layers (layers 1–13) were selected to represent visual features because they are modeled after the structure of the visual cortex (Fukushima 1980), and previous work showed that the features contained within the convolutional layers of AlexNet (which has a similar architecture to VGG16) corresponded to the features represented throughout the visual cortex (Güçlü and van Gerven 2015). The layer activations were visually inspected to confirm whether they represent the appropriate features. The low-level layers were required to have similar outputs to edge filters. Layers 1–4 best fit that description. The high-level layer was required to have features that selectively respond to complex objects (e.g., faces). Layers 10–13 contained such features. There were no a priori demands on the type of features represented by the middle layer, so layers 5–9 were selected. We used the fully connected layers (14–16) to approximate semantic features because, unlike the convolutional layers, they are not modeled upon the visual cortex. Instead, the fully connected layers are designed to learn features (derived from high-level visual features in layer 13) that directly contribute to the semantic classification of images.

To account for the low retinotopic spatial resolution resulting from participants eye movements, the spatial resolution of the convolutional layers (the fully connected layers have no explicit spatial representation) was reduced to 3 by 3 (original resolution for layers 1–2: 224 by 224; layers 3–4: 112 by 112; layers 5–7: 56 by 56; layers 8–10: 28 by 28; layers 11–13: 14 by 14). Convolutional layer activations were log-transformed to improve prediction accuracy (Naselaris et al. 2015).

Encoding–Decoding Analysis for Feature-Specific Reactivation

Our feature-specific reactivation measure is based on a method (Naselaris et al. 2015) involving two steps, an encoding step and a decoding step. The goal of the encoding step is to predict activity for each voxel when viewing or recalling an image based upon the relationship (at perception) between DNN feature activations (from a given layer) and neural activity associated with the target image. Once trained, a single encoding model takes DNN features from one layer of the network as the input and outputs the predicted activity for one voxel. Multiple models are used to cover all voxels and layers.

The goal of the decoding step is to measure reactivation of feature-specific neural activity patterns for each retrieval trial and feature-level. This is accomplished by correlating (over voxels within a ROI) the actual neural activity pattern during recall with the encoding models' (layer-specific) predictions for each encoded image (i.e., the particular combination of features comprising the image) and then comparing how strong the correlation with the target/cued image prediction is relative to all the non-target image predictions using a rank measure. A final additional step of residualization is used to remove trial-by-trial variance of the reactivation measure (from the decoding step) that is shared by other feature levels and between the aHC and pHC. The three steps are described in more detail in the following three sections.

Feature-Specific Encoding Model

Separate encoding models were estimated for all combinations of subject, feature level and brain surface vertex (Naselaris et al. 2015). Let v_{it} be the signal from vertex i during trial t. The encoding model for this vertex for a given feature level, l, is:

$$v_{it} = \mathbf{h}^{\mathrm{T}}\mathbf{f}_{lt} + \epsilon \tag{1}$$

Here, $\mathbf{f}_{lt}$ is a 100 × 1 vector of 100 image features from the layer of VGG16 representing the target feature level, l, associated with the current trial/image, t (only the 100 features from layer l with the largest positive correlations with the vertex activity, v_i, were selected to make the computation tractable. Correlations were performed immediately before each non-negative lasso regression using data from the movie and encoding tasks), **h** is a 100×1 vector of model parameters that indicate the voxel/vertex's sensitivity to a particular feature (the superscript T indicates transposition) and ϵ is zero-mean Gaussian additive noise.

The model parameters **h** were fit using non-negative lasso regression (R package "nnlasso"; Mandal and Ma 2016) trained on data drawn from the encoding and movie viewing tasks (excluding the Indiana Jones video because its widescreen aspect ratio differed significantly from the encoded images) using 3-fold cross validation over the encoding data (cross validation was performed over images, so trials containing presentations of the to-be-predicted image were not included in the training set; all movie data was used in each fold). The non-negative constraint was included to reduce the possibility that a complex linear combination of low-level features may approximate one or more high-level features. The regularization parameter (lambda) was determined by testing 5 log-spaced values from approximately 1/10000 to 1 (using the nnlasso function's path feature). For each value of the regularization parameter, the model parameters **h** were estimated for each vertex and then prediction accuracy (sum of squared errors; SSE) was measured on the held-out encoding data. For each vertex, the regularization parameter (lambda) that produced the highest prediction accuracy was retained for image decoding during recall.

Image Decoding

For feature-specific reactivation, encoding models were used to predict neural activity during recall for each combination of subject, feature-level, ROI, and retrieval trial (148 cortical FreeSurfer ROIs and 10 hippocampal ROIs). The accuracy of this prediction was assessed as follows: 1) for each combination of subject, feature-level, and ROI, the predicted neural activation patterns for the 90 images viewed during the encoding task were generated using a model that was trained on the movie and encoding task data, excluding data from encoding trials wherein the predicted image was viewed using 3-fold cross validation; 2) for each retrieval trial, the predictions were correlated (Pearson correlation across vertices within the given ROI) with the observed neural activity during recall resulting in 90 correlation coefficients per trial. 3) For each retrieval trial, the 90 correlation coefficients were ranked in descending order, and the rank of the prediction associated with the recalled image was recorded (1 = highest accuracy, 90 = lowest accuracy). 4) This rank was then subtracted from the mean rank (45.5) so that 0 was chance, and a positive value indicated greater-than-chance accuracy for the given trial (44.5 = highest accuracy, −44.5 = lowest accuracy). 5) The ranks were placed into four groups by layer (1–4, 5–9, 10–13, 14–16) and averaged together within each group, reducing the feature-levels from 16 to 4.

For image-specific reactivation, a similar decoding method was used (steps 1–4 ignoring references to feature-levels), except the predicted neural activation patterns for the 90 images were the average activation patterns (over four trials) when the participant viewed each image during encoding.

The reactivation results were averaged over bilateral ROI pairs (for cortical and hippocampal ROIs) to produce reactivation values for 74 bilateral cortical ROIs, and 5 bilateral hippocampal ROIs. To acquire anterior and posterior hippocampal reactivation values, the two most anterior and the two most posterior hippocampal ROIs (the middle ROI was not included) were averaged together, respectively.

Removing Shared Variance Between ROIs and Feature Levels

To remove the shared variance between feature levels and the anterior and posterior hippocampus for the within-subject analyses, residuals extracted from linear models were used in the place of the reactivation measure. Linear models were run for all combinations of ROI and feature-level. For image-specific reactivation within the hippocampus, trial-by-trial reactivation within the aHC (pHC) was the DV and reactivation within the pHC (aHC) was the IV. For feature-specific reactivation within the neocortex, reactivation of the target feature level was the DV and reactivation of the three non-target feature levels were three IVs. For feature-specific reactivation within the hippocampus, reactivation of the target feature level within the aHC (pHC) was the DV, reactivation of the three non-target feature levels within the aHC (pHC) were three IVs, and reactivation of all feature levels within the pHC (aHC) were four IVs. The residuals from the models were used as measures of feature-specific reactivation in all within-subject analyses, replacing the reactivation measures used as the DVs.

Bootstrap Statistics

For the within-subject LME models, confidence intervals and *P*-values were calculated with bootstrap statistical analyses (1000 samples) using the BootMer function (Bates et al. 2015). For the between-subject linear models, confidence intervals and *P*-values were generated with bootstrap statistical analyses (1000 samples) with random sampling over subjects.

Results

Recognition Accuracy

Recognition accuracy, averaged across participants, was 83.3% (SD = 6.8%; chance = 50%). Accuracy on old and lure trials was 81.7% (SD = 9.8%) and 85.0% (SD = 10.7%), respectively, with no significant difference in accuracy between the two conditions ($t(24) = -1.07$, $P = 0.295$, paired samples, two-tailed t-test) (Supplementary Fig. 2). Participants failed to respond within the 3 s old/new response period on 1.0% (SD = 1.5%) of trials. Those trials were classified as incorrect.

Recognition Accuracy and Hippocampal Reactivation during Recall

To determine whether detailed low-level visual features are represented by the hippocampus—particularly the pHC—we first set out to examine the relationship between trial-by-trial recognition accuracy and feature-specific reactivation within the aHC and pHC. We relied upon the participants' ability to distinguish between their memory of the previously seen image and a similar lure to indicate whether they recalled a detailed and accurate representation of the encoded image. The lures had similar low-level features and almost identical semantic features to the old image, so, while low-level features would be the most discriminative, the participant would still require a detailed and accurate memory of those features to perform well, i.e., a "low-resolution"/gist-like representation of low-level features would be insufficient and potentially misleading.

We did not solely rely upon our reactivation measure to distinguish representations with small differences because the amount of visual detail that can be decoded from fMRI scans is limited—even in best-case scenarios such as scans of the neocortex during perception (Wen et al. 2018, Fig. 8). Fine distinctions between neural representations would be considerably more difficult to detect within the hippocampus due to its size, the variety of information stored there, and the compression that must therefore occur within the region. Instead, the reactivation measure was used to indicate whether participants were able to *at least* reinstate an accurate "low-resolution" representation on a given trial that was sufficient to distinguish the target image from the other images seen during encoding. More detailed reactivation sufficient to distinguish the target image from the similar lure was inferred from the relation between reactivation and recognition accuracy. A positive trial-by-trial correlation was expected only if accurate fine-grained features were reactivated within the ROI because reactivation of a "low-resolution" gist-like representation alone would not be sufficient to distinguish the old image from a lure. Conversely, a negative trial-by-trial correlation was expected if successful reactivation of a "low-resolution" representation was generally associated with reactivation of inaccurate and misleading fine-grained features within the ROI, as would be the case for individuals who tend to have a strong memory for the gist of a visual scene but a weak memory for visual details.

Aside from allowing one to distinguish gist-like and detailed reactivation, correlating reactivation with recognition accuracy also accounts for memory variability, both within- and between-subjects. Within-subjects, one must account for the fact that the participants did not successfully reactivate the relevant neural patterns and recall the target image on all trials. Incorrect trials can be associated with *negative* reactivation (if the reactivation pattern is closer to the non-target image patterns than the target image pattern), resulting in a positive reactivation mean on correct trials potentially being counteracted by negative reactivation on incorrect trials. This issue was mitigated by including trial-by-trial recognition accuracy within our model as the dependent variable (DV) and reactivation as independent variables (IV). Between-subjects, one must account for individual differences in detailed episodic recall. Accounting for individual differences in the accuracy of detailed visual memory was necessary because the trial-by-trial correlation between recognition accuracy and low-level visual reactivation was expected to be *positive* for individuals who tend to have a strong memory for visual details and *negative* for individuals who tend to have a strong memory for the gist of a visual scene but a weak memory for visual details. Individual differences in mean accuracy, particularly lure accuracy, were expected to relate to how successful the participant was at reactivating neural patterns representing the fine-grained visual features that could be used to distinguish the lure from the encoded image. We therefore accounted for individual differences in fine-grained visual memory by including an interaction between reactivation and each subject's mean lure accuracy.

A binomial mixed linear effects (MLE) model was constructed with trial-by-trial accuracy as the dependent variable (DV), feature-specific reactivation of four feature levels (low-, mid-, high-level visual and semantic) within the aHC and pHC as eight independent variables (IV), the interaction between participants' average lure accuracy and the four levels of feature-specific reactivation within the aHC and pHC as eight IVs, probe type (old or lure) as an IV (control), and subject and image pair as crossed random effects (random intercept only due to model complexity limitations). In addition to the feature-specific reactivation model, a similar binomial MLE model was constructed in which image-specific reactivation replaced the four levels of feature-specific reactivation. To focus on the unique contributions of the aHC, pHC, and the four feature levels, residuals extracted from linear models (with the variance shared between the aHC, pHC and the four feature levels removed) were used in the place of the reactivation measures in the above analysis (see "Accounting for Shared Variance Between ROIs and Feature Levels" in section Materials and Methods). Residuals were used in this way for all subsequent analysis. The coefficients from these models are depicted in Figures 2 and 3 (feature-specific: 2*a*-*b*, 3*a*-*c*; image-specific: 2*c*, 3*d*).

Consistent with previous findings of individual differences within the visual cortex (Bone et al. 2020), a significant positive interaction was found between participants' average lure accuracy and low-level visual reactivation within the pHC [$\beta = 0.21$, $P = 0.001$; one-tailed 1000 sample bootstrap—all bootstrap statistics used 1000 samples] (Fig. 2*a*). In contrast, the interaction with low-level visual reactivation within the aHC was negative [albeit not significantly so: $\beta = -0.10$, $P = 0.102$; two-tailed bootstrap] and significantly less than the pHC interaction [$P = 0.001$; paired-samples one-tailed bootstrap]. A qualitative examination of the interaction coefficients in Figure 2*a* indicates that this relationship between the pHC and aHC is consistent across visual feature levels. To assess this trend quantitatively, we averaged the coefficients across feature levels and found that the averaged pHC interaction coefficient was significantly greater than zero and the aHC coefficient [pHC: $\beta = 0.09$, $P = 0.045$; one-tailed bootstrap; aHC: $\beta = -0.08$, $P = 0.112$; two-tailed bootstrap; difference: $P = 0.010$; paired-samples one-tailed bootstrap] (Fig. 2*b*). Qualitatively similar albeit quantitatively weaker interaction coefficients were found using the image-specific reactivation model [pHC: $\beta = 0.07$, $P = 0.120$; one-tailed bootstrap; aHC: $\beta = -0.07$, $P = 0.204$; two-tailed bootstrap; difference: $P = 0.042$; paired-samples one-tailed bootstrap] (Fig. 2*c*). Overall, these findings reveal pronounced individual differences, which suggest that reactivation within the posterior hippocampus, particularly low-level visual reactivation, is more strongly associated with recognition performance for participants with high lure accuracy.

To account for these striking individual differences, the partial regression coefficients for the relation between recognition accuracy and neural reactivation were assessed for participant lure accuracy one standard deviation above average (95%; Fig. 3*a*) and one standard deviation below average (75%; Fig. 3*b*). We first examine the results for individuals with high lure accuracy (Fig. 3*a*). As hypothesized, the low-level visual reactivation coefficient within the pHC was significantly greater than zero and was also significantly greater than the corresponding aHC coefficient [pHC: $\beta = 0.20$, $P = 0.021$; one-tailed bootstrap; aHC: $\beta = -0.18$, $P = 0.076$; two-tailed bootstrap; pHC—aHC: $P = 0.006$; paired-samples one-tailed bootstrap]. Looking beyond low-level features, we see that the pHC coefficient was significantly greater than zero for high-level features [$\beta = 0.26$, $P = 0.018$; one-tailed bootstrap; FDR corrected over feature levels], and the pHC coefficients were significantly greater than the aHC coefficients for all feature levels [mid-level visual: $P = 0.022$; high-level visual: $P = 0.006$; semantic: $P = 0.045$; paired-samples one-tailed bootstrap; FDR corrected over feature levels]. Qualitative analysis of Figure 3*a* suggests that interaction coefficients are generally positive within the pHC and negative within the aHC. The coefficients averaged over feature levels (Fig. 3*c*) confirm this trend [pHC: $\beta = 0.15$, $P = 0.031$; one-tailed bootstrap; aHC: $\beta = -0.18$, $P = 0.024$; two-tailed bootstrap; pHC—aHC: $P = 0.004$; paired-samples one-tailed bootstrap]. As with the interaction coefficients, the image-specific reactivation model produced qualitatively similar yet quantitatively weaker results

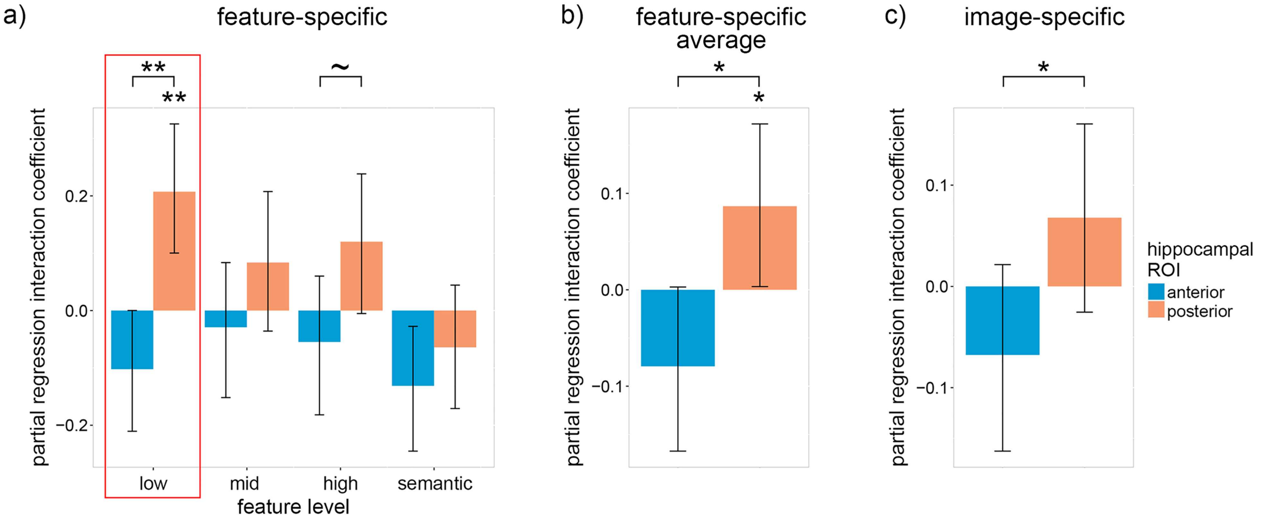

Figure 2. The trail-by-trial relation between recognition accuracy and neural reactivation within the anterior and posterior hippocampus differs according to individual differences in average lure accuracy. (*a*) Within-subject partial regression coefficients for the interaction between feature-specific neural reactivation and each subjects' average recognition lure accuracy with respect to trial-by-trial recognition accuracy. (*b*) The coefficients in (*a*) averaged over feature levels. (*c*) The same interaction as (*a*) except with item-specific (instead of feature-specific) reactivation. Error bars are 90% CIs; ∼ indicates $P < 0.10$, * indicates $P < 0.05$, **$P < 0.01$, one-tailed bootstrap; FDR corrected over visual feature levels except for low-level features because, in accordance with our hypotheses, low-level features were prioritized (indicated by the red boxes).

[pHC: $\beta = 0.15$, $P = 0.031$; one-tailed bootstrap; aHC: $\beta = -0.12$, $P = 0.148$; two-tailed bootstrap; pHC—aHC: $P = 0.005$; paired-samples one-tailed bootstrap] (Fig. 3*d*). Taken as a whole, our results support the claim that the pHC represents detailed low- and high-level visual features (which were expected to facilitate recognition accuracy), while the aHC represents more gist-like representations (which were expected to hinder recognition accuracy).

The above findings only held for individuals with high lure accuracy. For individuals with low lure accuracy, we found the opposite (Fig. 3*b*), i.e., the low-level visual coefficient within the pHC was significantly less than zero and marginally less than the aHC coefficient [pHC coefficient: $\beta = -0.22$, $P = 0.032$; two-tailed bootstrap; aHC coefficient: $\beta = 0.03$, $P = 0.380$; one-tailed bootstrap; pHC—aHC difference: $P = 0.060$; paired-samples two-tailed bootstrap]. No other coefficients were significant. Our results indicate that participants with low lure accuracy did not simply fail to reactivate low-level visual details, but instead reactivated inaccurate and misleading details that hurt their performance on the recognition task.

If individuals with low lure accuracy are unable to accurately recall detailed low-level features within the pHC, how did they attempt to compensate? Although we see no positive coefficients within Figure 3*b* that does not necessarily mean that the participants did not recall the associated features. An alternative explanation for the null findings is that the features were not sufficient for the difficult recognition task, as we hypothesized for sematic features and the gist-like features of the aHC. If individuals with low lure accuracy relied to a greater extent upon these suboptimal features, then we should see a negative relationship between the participants' average lure accuracy and feature-specific reactivation. To test this hypothesis, a between-subject linear model was constructed with participants' average lure accuracy as the dependent variable (DV) and feature-specific reactivation of the four feature levels within the aHC and pHC as eight independent variables (IV). Reactivation values from trials with incorrect old/new responses were excluded because we were interested in between-subject differences in the location (and features) of reactivation during successful recall.

As predicted, a significantly negative coefficient was found for low-level visual features within the aHC [$\beta = -0.10$, $P = 0.002$; two-tailed bootstrap], whereas the coefficient for low-level features within the pHC trended positive [$\beta = 0.04$, $P = 0.064$; one-tailed bootstrap; pHC—aHC: $P < 0.001$; paired-samples one-tailed bootstrap] (Fig. 4*a*). Positive coefficients were also found for mid-level, high-level and (marginally) semantic features within the pHC, indicating that individuals with high lure accuracy relied to a greater extent upon representations within the pHC [mid-level visual: $\beta = 0.05$, $P = 0.020$; high-level visual: $\beta = 0.06$, $P = 0.008$; semantic: $\beta = 0.02$, $P = 0.064$; one-tailed bootstrap; FDR corrected over feature levels]. The coefficients averaged over feature levels (Fig. 4*b*) support this interpretation [pHC: $\beta = 0.04$, $P = 0.003$; one-tailed bootstrap; aHC: $\beta = -0.04$, $P = 0.188$; two-tailed bootstrap; pHC—aHC: $P = 0.011$; paired-samples one-tailed bootstrap]. As with previous results, a variant of the above model using image-specific reactivation produced qualitatively similar yet quantitatively weaker results [pHC: $\beta = 0.02$, $P = 0.173$; one-tailed bootstrap; aHC: $\beta = -0.04$, $P = 0.092$; two-tailed bootstrap; pHC—aHC: $P = 0.029$; paired-samples one-tailed bootstrap] (Fig. 4*c*). Our findings support our hypothesis that individuals with low lure accuracy attempted to compensate for inaccurate low-level representations within the pHC by relying upon gist-like representations within the aHC, although the lack of a significantly positive low-level aHC coefficient within Figure 3*b* indicates that this strategy was unsuccessful for the current task.

Interaction Between the Hippocampus and Calcarine Sulcus

According to leading theories (Teyler and DiScenna 1986; McClelland et al. 1995; Nadel et al. 2000; Sekeres et al. 2018), episodic memories are encoded in distributed neural networks comprising hippocampal and neocortical neurons. Therefore, hippocampal reactivation should facilitate episodic recognition accuracy only when it co-occurs within relevant neocortical regions (and vice versa). To test this claim, a linear model was

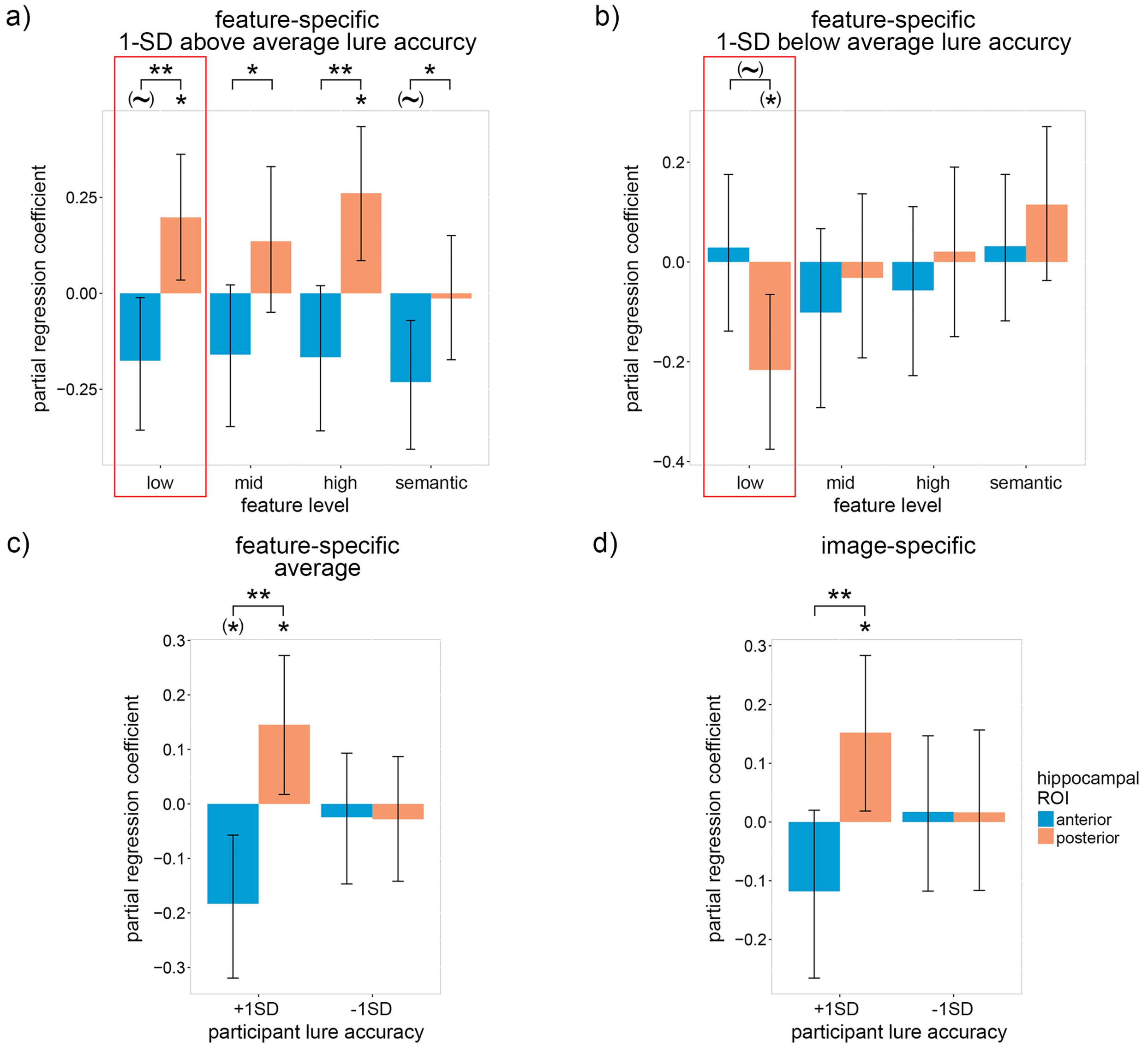

Figure 3. The trial-by-trial relation between recognition accuracy and neural reactivation within the anterior and posterior hippocampus for participants with low- and high-average lure accuracy. (*a*) Within-subject partial regression coefficients from the same model as Figure 2 for the relation between feature-specific neural reactivation and trial-by-trial recognition accuracy for participant lure accuracy 1 standard deviation above average (95%). (*b*) the same as (*a*) but for participant lure accuracy 1 standard deviation below average (75%). (*c*) The coefficients in (*a*) and (*b*) averaged over feature levels. (*d*) The same coefficients as (*a*) and (*b*) except with item-specific (instead of feature-specific) reactivation. Error bars are 90% CIs; * indicates $P < 0.05$, **$P < 0.01$, one-tailed bootstrap; (~) indicates $P < 0.10$, (*) indicates $P < 0.05$, two-tailed bootstrap; FDR corrected over visual feature levels except for low-level features because, in accordance with our hypotheses, low-level features were prioritized (indicated by the red boxes).

used to investigate whether a positive interaction exists between low-level visual reactivation within the pHC and calcarine sulcus (the neocortical region wherein V1 is concentrated; DeYoe et al. 1996), with respect to recognition accuracy (Fig. 5*a*). We focused on low-level features for this analysis because low-level reactivation correlated with recognition accuracy (indicating that the features are useful for the task), and it is known a priori where low-level visual features are represented within the neocortex. Given that participants with low lure accuracy showed no positive relationship between recognition accuracy and hippocampal reactivation (Fig. 3*b*), indicating that they were unable to reactivate an accurate index of early visual cortical activity, a positive interaction between the hippocampus and calcarine sulcus was expected only for participants with high lure accuracy.

The binomial MLE model for the interaction between the hippocampus and calcarine sulcus consisted of trial-by-trial accuracy as the dependent variable (DV), reactivation of all four feature levels within the anterior and posterior hippocampus as 8 IVs, low-visual reactivation within the calcarine sulcus as an IV, the interactions between low-visual reactivation within the calcarine sulcus and all 8 hippocampal reactivation measures as 8 IVs, the interactions between participants' average lure accuracy and all measures of reactivation (low-level within the calcarine sulcus and all levels within the aHC and pHC) as 9 IVs, the three-way interactions between average lure accuracy,

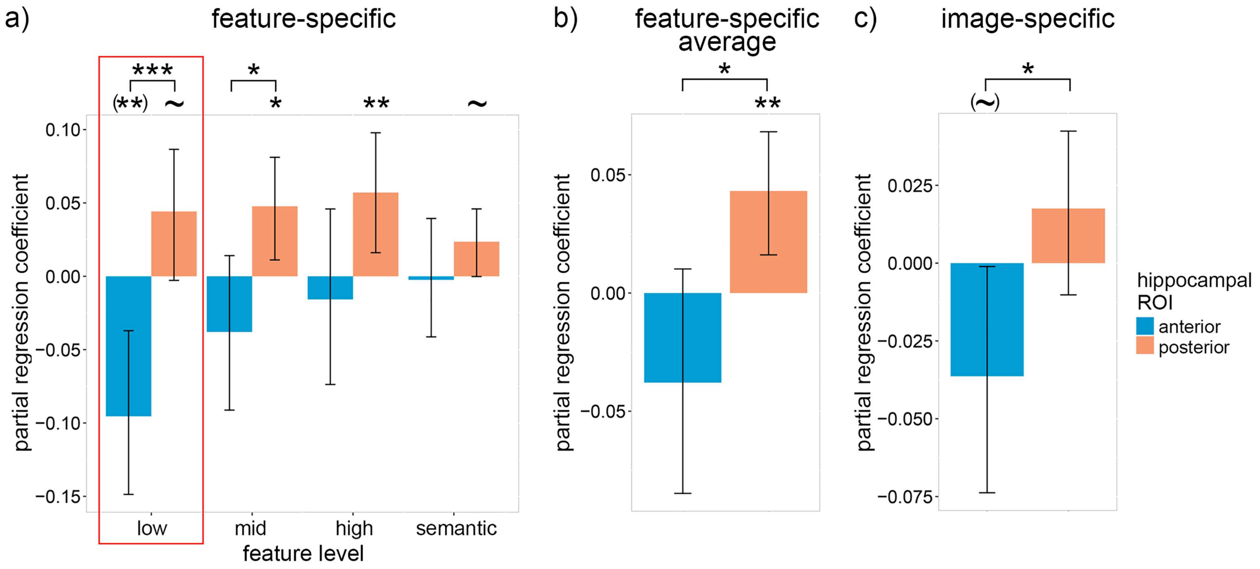

Figure 4. Relation between participants' average recognition lure accuracy and neural reactivation within the hippocampus on correct trials. (*a*) Between-subject partial regression coefficients for the relation between feature-specific neural reactivation within the anterior and posterior hippocampus and each subjects' average recognition lure accuracy. Only reactivation values on trials with correct old/new responses were included. (*b*) The coefficients in (*a*) averaged over feature levels. (*c*) The same coefficients as (*a*) except with item-specific (instead of feature-specific) reactivation. Error bars are 90% CIs; ~ indicates $P < 0.10$, *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-tailed bootstrap; (~) indicates $P < 0.10$, (**) $P < 0.01$, two-tailed bootstrap; FDR corrected over visual feature levels except for low-level features because, in accordance with our hypotheses, low-level features were prioritized (indicated by the red boxes).

low-visual reactivation within the calcarine sulcus and all 8 hippocampal reactivation measures as 8 IVs, probe type (old or lure) as an IV (control), and subject and image pair as crossed random effects (random intercept only due to model complexity limitations).

Figure 5*b* depicts the interaction between low-level visual reactivation within the calcarine sulcus and the hippocampus (see Supplementary Fig. 3 for all interaction coefficients). As predicted by the hippocampus's role in indexing neocortical activity, the pHC interaction coefficient was significantly greater than zero for individuals with high lure accuracy [pHC: $\beta = 0.24$, $P = 0.017$; aHC: $\beta = 0.12$, $P = 0.105$; one-tailed bootstrap]. For individuals with low lure accuracy, the interaction coefficients were approximately zero [pHC: $\beta = 0.02$, $P = 0.466$; aHC: $\beta = 0.01$, $P = 0.476$; one-tailed bootstrap], with the pHC coefficient marginally lower than the corresponding coefficient for individuals with high lure accuracy [$P = 0.058$; paired-samples one-tailed bootstrap]. To elaborate upon the observed pHC interaction, Figure 5*c* depicts the pHC partial regression coefficients when reactivation within the calcarine sulcus is either high (1; Z-scored) or low (−1), and Figure 5*d* depicts the calcarine sulcus partial regression coefficients when reactivation within the pHC is either high (1) or low (−1). For individuals with high lure accuracy, low-level visual reactivation within both hippocampal and early visual ROIs was positively associated with recognition accuracy only when reactivation within the other ROI was high [high reactivation in the other ROI: pHC: $\beta = 0.46$, $P = 0.001$; calcarine sulcus: $\beta = 0.35$, $P = 0.007$; one-tailed bootstrap; low reactivation in the other ROI: pHC: $\beta = -0.02$, $P = 0.890$; calcarine sulcus: $\beta = -0.12$, $P = 0.352$; two-tailed bootstrap]. The results indicate that trial-by-trial recognition accuracy was only associated with low-level visual reactivation when it co-occurred within the pHC and calcarine sulcus, thereby supporting the claim that the hippocampus facilitates episodic memory by storing and retrieving a representation, or "index," of the neocortical activity that occurred during perception.

Discussion

We investigated the relationship between feature-specific reactivation within the hippocampus and neocortex and performance on a recognition task that required retrieval of visual details. We showed that image-specific and feature-specific reactivation within the pHC, and not the aHC, positively correlated with recognition accuracy, indicating that the pHC indexes more detailed features relative to the aHC. Moreover, striking individual differences were observed such that recognition accuracy was positively associated with low-level visual reactivation within the pHC for individuals with above-average recognition lure accuracy, whereas the opposite relationship was observed for individuals with below-average recognition lure accuracy. Our results show that representations within the hippocampus extend to low-level visual features, and suggest that individuals with below-average recognition performance reactivated inaccurate low-level details within the pHC (i.e., representations with small receptive fields) and relied—ineffectually—upon the reactivation of less detailed low-level visual features within the aHC (i.e., representations with large receptive fields) which are more likely to overlap with the features of the lure images. Lastly, the correlation between recognition accuracy and low-level reactivation within the hippocampus was found to depend upon low-level reactivation within the early visual cortex (calcarine sulcus), and vice versa. This mutual dependence between hippocampal and neocortical feature-specific reactivation supports the claim of leading theories of episodic memory that the hippocampus mediates neocortical reactivation by storing a sparse representa-

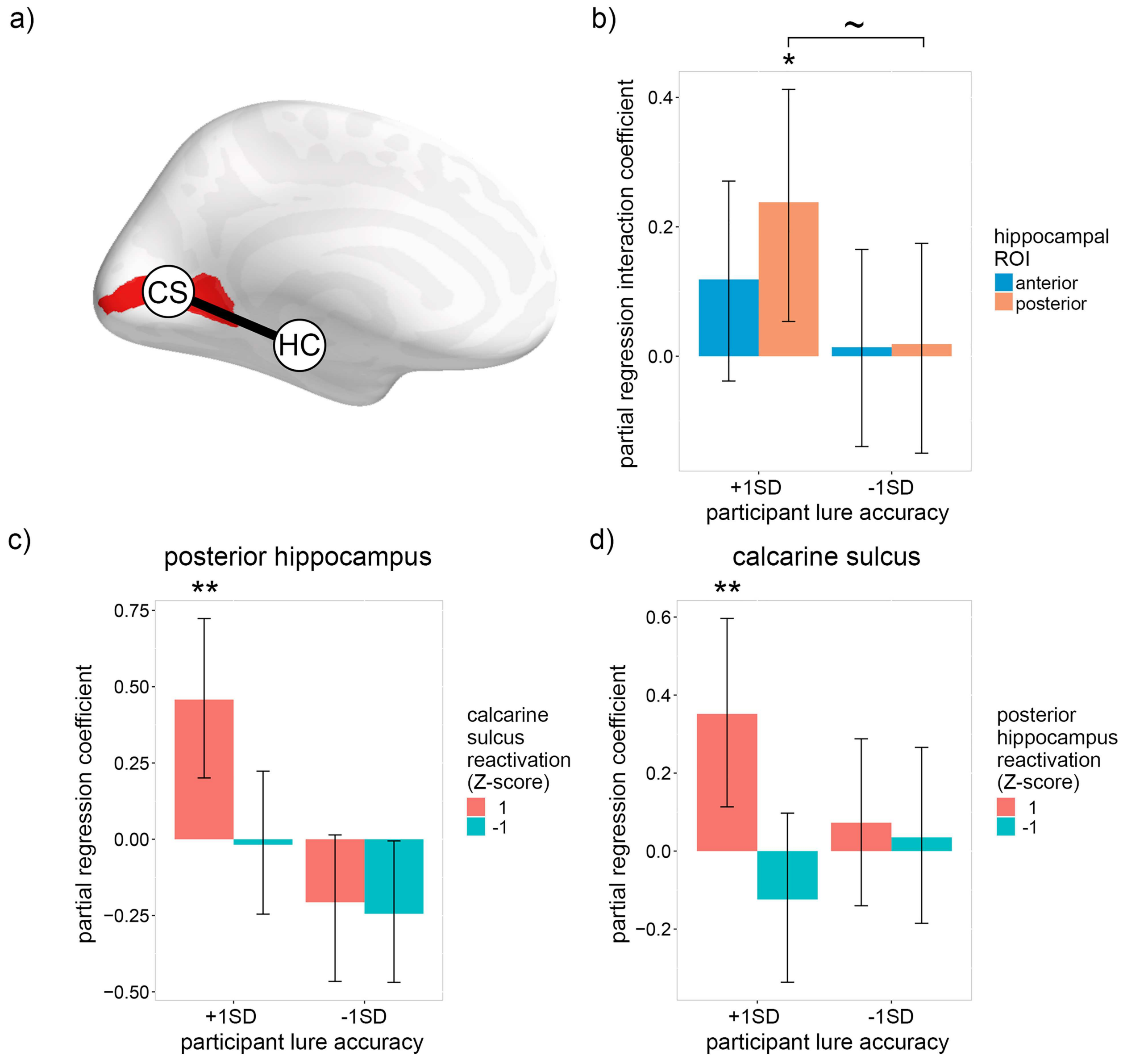

Figure 5. Relation between recognition accuracy and low-level visual neural reactivation within the hippocampus, calcarine sulcus and their interaction during recall. (*a*) Diagram of the regions of interest: CS = calcarine sulcus, HC = hippocampus. The bilateral calcarine sulcus ROI is indicated in red. The line between the two ROIs represents the interaction between the regions. (*b*) Within-subject partial regression coefficients for the interaction of low-level visual reactivation within the hippocampus and calcarine sulcus with respect to recognition accuracy. Interaction coefficients are displayed for participant lure accuracy 1 standard deviation above (95%) and below (75%) average. (*c*) Within-subject partial regression coefficients from the same model as (*b*) for the relation between low-level visual reactivation within the posterior hippocampus and recognition accuracy for participant lure accuracy 1 standard deviation above (95%) and below (75%) average and low-level visual reactivation within the calcarine sulcus 1 standard deviation above and below average. (*d*) Within-subject partial regression coefficients from the same model as (*b*) for the relation between low-level visual reactivation within the calcarine sulcus and trial-by-trial recognition accuracy for participant lure accuracy 1 standard deviation above (95%) and below (75%) average and low-level visual reactivation within the posterior hippocampus 1 standard deviation above and below average.

tion, or “index,” of the pattern of neocortical activity at encoding (Teyler and DiScenna 1986; McClelland et al. 1995; Nadel et al. 2000; Sekeres et al. 2018; Barron et al. 2020).

The fact that it was possible to detect reactivation within the hippocampus using a model trained to recognize feature-specific patterns shared across images strongly implies that hippocampal indexes of different events are not entirely pattern-separated (i.e., there must be some overlap between the indexes of memories that share similar features), but a couple of caveats must be considered. First, feature-specific reactivation could potentially be detected if pattern-separated indexes of related events (i.e., events with overlapping features, including, but not limited to, perception of the images in the current study) are consistently coactivated during encoding and retrieval. This interpretation does not align with our results because coactivation of indexes of similar events/images would be expected to reduce recognition accuracy, rather than facilitate it. Second, it may be the case that feature-specific reactivation was detected within hippocampal subregions that contain overlapping representations. This interpretation is supported by studies

(Bakker et al. 2008; Rolls 2013) which suggest that pattern-separated representations are limited to the CA3 and dentate gyrus, whereas representations of similar events within the CA1 tend to overlap. Future work exploring feature-specific reactivation within individual hippocampal subfields will be required clarify how pattern-separation varies between subregions.

Multiple analyses, including image-specific and feature-specific reactivation both within- and between-subject, provided clear and consistent evidence that fine-grained visual features are represented within the pHC, whereas less detailed features are represented within the aHC. Although our results support and expand upon the majority of previous research using human and animal models (Kjelstrup et al. 2008; Evensmoen et al. 2015; Schlichting et al. 2015; Brunec et al. 2018; Sekeres et al. 2018; Grady 2020), not all findings in the literature are consistent with this dichotomy. In a recent longitudinal study using a recognition task paradigm similar to the current study's (except testing occurred 1 day or 28 days after encoding, and novel non-lure images were included in the recognition test), Dandolo and Schwabe (2018) found that detailed visual memories were represented within the aHC, whereas gist-like memories were represented within the pHC. The authors' conclusion was based upon two findings: a between-subject correlation between lure accuracy and univariate activity within the hippocampus, and a multivariate model comparing cross-image neural pattern similarity for old, lure and new images. For the between-subject univariate correlation, recognition accuracy on lure trials positively correlated with univariate activity within the aHC, but not the pHC. In contrast, we found that participants' average recognition accuracy on lure trials negatively correlated with reactivation within the aHC, and positively correlated with reactivation within the pHC. One potential explanation for this discrepancy is that memories may become more detailed within the aHC, relative to the pHC, over time (1 day). However, other work (Brunec et al. 2018) has found that representations within the aHC are less distinct than those within the pHC for memories encoded over periods much longer than 1 day. An alternative explanation is that univariate activity within the aHC during recognition may not reflect recollection of image-specific information, and instead reflects other functions such as context retrieval. The multivariate measure suffers from a similar issue of interpretation. Representational similarity analysis (RSA) was used to detect neural patterns shared between different images. Consequently, the detected effects do not represent retrieval of image-specific details, and instead might reflect the retrieval of the shared encoding context during recognition. Therefore, Dandolo and Schwabe's (2018) univariate and multivariate results do not address whether the pHC or aHC represent task-relevant event-specific details, whereas our results based upon image/feature-specific reactivation do—and support the idea that the pHC and aHC represent detailed and gist-like memories, respectively.

Clear individual differences were observed with respect to the visual feature levels recalled and the detail/location of those features. For the recollection task, results suggest that individuals with high lure accuracy relied upon detailed low- and high-level visual reactivation within the pHC and were likely hindered by semantic reactivation within the aHC, whereas low lure accuracy subjects ineffectually relied upon gist-like low-level reactivation within the aHC in an attempt to compensate for inaccurate low-level reactivation within the pHC. Our findings are consistent with and expand upon previous work by Sheldon et al. (2016) who found that people who tend to rely upon episodic memory (i.e., memories containing spatiotemporal and contextual details) have greater functional connectivity between the medial temporal lobes and posterior visual regions (i.e., neocortical regions predominately connected to the pHC; Poppenk et al. 2013). In contrast, people who tend to rely upon semantic memory were found to have greater functional connectivity between the medial temporal lobes and the prefrontal cortex (i.e., neocortical regions predominately connected to the aHC; Poppenk et al. 2013). Our findings are also consistent with Armson et al. (2019), who found that the association between eye fixation rate during free recall and the number of episodic details recalled was significantly greater for participants who tend to rely upon episodic memory (trait episodic and semantic memory in both studies was measured via questionnaire; Palombo et al. 2013). It is reasonable to assume that the tendency to rely upon episodic or semantic memories would have an impact on the current study's recognition task because the strong semantic similarity between the encoded and lure images greatly favors episodic memory (a fact the participants were aware of), forcing those who tend to use semantic memory to attempt to use episodic memory instead. It is therefore plausible that the individual differences observed within our study were driven by stable trait-like preferences for episodic or semantic memory. If this is the case, the current study suggests that those who favor semantic memory may do so because they are often unable to accurately recall detailed low-level perceptual features—possibly as a result of limited/impaired communication between the pHC and early visual cortex.

According to prominent theories of the role of the hippocampus in episodic memory, communication between the hippocampus and neocortex is necessary for episodic retrieval, at least before consolidation (Teyler and DiScenna 1986; McClelland et al. 1995; Nadel et al. 2000; Sekeres et al. 2018; Barron et al. 2020). The theories posit that events are represented by unique spatiotemporal arrays of neocortical modules, and that sparse representations of these arrays are stored, or "indexed," by the hippocampus at encoding. During retrieval, the partial reactivation of an event's neocortical array reactivates the hippocampal index associated with the event, which in turn reactivates the associated neocortical array, thereby emulating the original experience. Joint reactivation within the hippocampus and neocortex is therefore required for successful episodic retrieval. While previous work has shown that reactivation of low-level features within the early visual cortex correlates with hippocampal univariate activity (Bosch et al. 2014), such findings indicate that the hippocampus is involved in neocortical reactivation, but do not establish the manner of its involvement. We found that detailed visual memory recall was only associated with low-level reactivation within the hippocampus (pHC) when reactivation of the same low-level features co-occurred within the early visual cortex, and vice-versa—thereby providing evidence supporting the role of the hippocampus in indexing neocortical activity. Moreover, because our measure of feature-specific reactivation requires that similar features be associated with similar activity patterns across events, this suggests that the hippocampus contains a functionally topographic mapping of the neocortex, as proposed by Teyler and DiScenna (1986) in their well-known formulation of hippocampal indexing. As our findings are limited to low-level visual features, future reactivation studies should explore whether the observed hippocampal-neocortical interaction is evident for different features in the context of different tasks (the current task was designed to favor low-level visual features).

To elaborate upon how our findings support neocortical indexing by the hippocampus, consider three alternative outcomes of our experiment: 1) A positive correlation between recognition accuracy and low-level visual reactivation is found within the hippocampus but not the early visual cortex when hippocampal low-level visual reactivation is controlled for (as it was in our model). This would be inconsistent with neocortical indexing as it would suggest that visual memories are stored entirely within the hippocampus (before consolidation), and that neocortical reactivation of recently stored visual memories is largely epiphenomenal, i.e., feedback from the hippocampus may result in neocortical reactivation but this reactivation does not directly contribute to memory recall; 2) No significant interaction between hippocampal and early visual reactivation is found, despite reactivation within each region correlating with recognition accuracy. This would be inconsistent with neocortical indexing as it would indicate that hippocampal memory representations are not dependent upon corresponding neocortical representations (and vice-versa); 3) An interaction is found between low-level reactivation within the early visual cortex and high-level, but not low-level, reactivation within the hippocampus. This would be inconsistent with hippocampal indexing as it would suggest that high-level information stored within the hippocampus is required to facilitate low-level cortical reactivation (with the image-specific low-level information being stored elsewhere, possibly within the visual cortex), rather than low-level cortical representations being directly indexed by the hippocampus.

Aside from our within-subject findings, individual differences provide additional support for hippocampus's role in indexing neocortical activity. We found that participants with low lure accuracy showed no interaction between low-level visual reactivation within the hippocampus and calcarine sulcus, i.e., there was no correlation between recognition accuracy and low-level reactivation within the early visual cortex, irrespective of hippocampal reactivation. This finding suggests that individual differences in recognition memory may be the consequence of a functional disconnection between the hippocampus and neocortex. Future work should investigate why some individuals may lack the hippocampal-neocortical functional connections that appear to be necessary for detailed and accurate episodic memory.

Due to the relatively small size of the hippocampus and the variety of information stored about multiple events at any one time, it is infeasible that all low-level visual features from a given event could be stored within the region. Therefore, some selection process must occur that aims to maximize information recall while minimizing the number of features stored in the hippocampus. To address this concern, we outline an expanded variant of the hippocampal indexing theory: Prediction error indexing (PEI). Aside from theories of hippocampal indexing (Teyler and DiScenna 1986; Nadel et al. 2000; Sekeres et al. 2018), PEI incorporates ideas drawn from the complementary learning systems (CLS) theory (Kumaran et al. 2016), the predictive coding theory (Rao and Ballard 1999; Bastos et al. 2012; Barron et al. 2020) and the predictive, interactive multiple memory systems (PIMMS) theory (Henson and Gagnepain 2010). In short, PEI assumes that hippocampal memory representations optimize compression by taking advantage of the event-specific information that overlaps with the statistical information stored within the neocortex. This is accomplished by biasing the features that are indexed by the hippocampus to those that are not predicted by the neocortex at encoding, with the rest of the features reconstructed at retrieval via neocortical inference. During episodic memory recall, a hippocampal index of higher level features activates the corresponding neocortical representations, which, in turn, are used to infer lower level features, while a hippocampal index of a sparse subset of lower level features serves to constrain this inference to be specific to the recalled event. Consequently, the indexing of higher level features takes precedence over lower level features because reactivation of lower level features is (usually) dependent upon reactivation of higher level features. Therefore, consistent with our results, PEI predicts that memory for lower level features will show greater variability, both within- and between-subjects. Future work could test PEI by associating feature-specific hippocampal reactivation with the predictability of a stimuli's features.

The contributions of this study were 5-fold. First, we found evidence for low-level visual representations within the hippocampus. Second, results from multiple analyses indicated that fine-grained and gist-like features are represented within the pHC and aHC, respectively. Third, we found that feature-specific representations within the hippocampus are functionally topographic and dependent upon feature-specific representations within the neocortex, thereby providing compelling evidence that the hippocampus facilitates episodic memory by storing a compressed representation, or "index," of the neocortical activity that occurred during perception. Fourth, individual differences in recognition lure accuracy were associated with striking differences in the relation between trial-by-trial memory accuracy and feature-specific reactivation within and between the hippocampus and neocortex, indicating that those with high (low) lure accuracy retrieved consistently accurate (inaccurate) detailed representations of low-level features. Fifth, we outlined a variant of the hippocampal indexing theory, PEI, that is consistent with our results and incorporates recent theoretical advances in the fields of memory and perception. Overall, the current study's results stress the importance of conceptualizing and measuring hippocampal function as part of an extended hippocampal-neocortical network, rather than in isolation. By doing so, future studies may refine our understanding of the mechanisms underlying memory and reveal the causes individual differences within healthy and clinical populations.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex Communications* online.

Funding

The Natural Sciences and Engineering Research Council of Canada (488937 to B.R.B.); the Canadian Institutes of Health Research (152879 to B.R.B).

Notes

We thank Asaf Gilboa for his helpful input.

Conflict of Interest: The authors declare no competing interests.

Author Contributions

Conceptualization, M.B.B., B.R.B.; methodology, M.B.B., B.R.B.; software, M.B.B., B.R.B.; formal analysis, M.B.B.; data curation, M.B.B., B.R.B.; writing—original draft, M.B.B.; writing—review and editing, M.B.B., B.R.B.; visualization, M.B.B.; supervision, B.R.B.

Data Availability

Data for all analyses covered in the article are available at https://github.com/MichaelBBone/ConcurrentReacHippoEV.

Code Availability

Code for all analyses covered in the article are available at https://github.com/MichaelBBone/ConcurrentReacHippoEV.

References

Argall BD, Saad ZS, Beauchamp MS. 2006. Simplified intersubject averaging on the cortical surface using SUMA. *Hum Brain Mapp*. **27**(1):14–27.

Armson MJ, Diamond NB, Levesque L, Ryan JD, Levine B. 2019. Vividness of recollection is supported by eye movements in individuals with high, but not low trait autobiographical memory. *Cognition*. **206**:104487.

Bakker A, Kirwan CB, Miller M, Stark CE. 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*. **319**(5870):1640–1642.

Barron HC, Auksztulewicz R, Friston K. 2020. Prediction and memory: A predictive coding account. *Prog Neurobiol*. **192**:101821.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012. Canonical microcircuits for predictive coding. *Neuron*. **76**(4):695–711.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw*. **67**:1–48.

Berron D, Schütze H, Maass A, Cardenas-Blanco A, Kuijf HJ, Kumaran D, Düzel E. 2016. Strong evidence for pattern separation in human dentate gyrus. *J Neurosci*. **36**(29):7569–7579.

Bone MB, St-Laurent M, Dang C, McQuiggan DA, Ryan JD, Buchsbaum BR. 2019. Eye-movement reinstatement and neural reactivation during mental imagery. *Cereb Cortex*. **29**(3):1075–1089.

Bone MB, Ahmad F, Buchsbaum BR. 2020. Feature-specific neural reactivation during episodic memory. *Nat Commun*. **11**(1): 1–13.

Bosch SE, Jehee JF, Fernández G, Doeller CF. 2014. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *J Neurosci*. **34**(22):7493–7500.

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Liu ZX, Grady C, Rosenbaum RS, Winocur G, Barense MD *et al*. 2018. Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Curr Biol*. **28**(13):2129–2135.

Buchsbaum BR, Lemire-Rodger S, Fang C, Abdi H. 2012. The neural basis of vivid memory is patterned on perception. *J Cogn Neurosci*. **24**:1867–1883.

Catenoix H, Magnin M, Mauguiere F, Ryvlin P. 2011. Evoked potential study of hippocampal efferent projections in the human brain. *Clin Neurophysiol*. **122**(12):2488–2497.

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. **29**:162–173.

Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*. **9**(2):179–194.

Dandolo LC, Schwabe L. 2018. Time-dependent memory transformation along the hippocampal anterior–posterior axis. *Nat Commun*. **9**(1):1–1.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser JO, Cox R, Miller D, Neitz J. 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci*. **93**: 2382–2386.

Evensmoen HR, Ladstein J, Hansen TI, Møller JA, Witter MP, Nadel L, Håberg AK. 2015. From details to large scale: The representation of environmental positions follows a granularity gradient along the human hippocampal and entorhinal anterior–posterior axis. *Hippocampus*. **25**(1):119–135.

Fukushima K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*. **36**(4):193–202.

Grady CL. 2020. Meta-analytic and functional connectivity evidence from functional magnetic resonance imaging for an anterior to posterior gradient of function along the hippocampal axis. *Hippocampus*. **30**(5):456–471.

Güçlü U, van Gerven MA. 2015. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci*. **35**(27):10005–10014.

Henson RN, Gagnepain P. 2010. Predictive, interactive multiple memory systems. *Hippocampus*. **20**(11):1315–1326.

Johnson MR, Johnson MK. 2014. Decoding individual natural scene representations during perception and imagery. *Front Hum Neurosci*. **8**:59.

Kier EL, Staib LH, Davis LM, Bronen RA. 2004. MR imaging of the temporal stem: anatomic dissection tractography of the uncinate fasciculus, inferior occipitofrontal fasciculus, and Meyer's loop of the optic radiation. *Am J Neuroradiol*. **25**(5):677–691.

Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser MB. 2008. Finite scale of spatial representation in the hippocampus. *Science*. **321**(5885):140–143.

Kuhl BA, Bainbridge WA, Chun MM. 2012. Neural reactivation reveals mechanisms for updating memory. *J Neurosci*. **32**(10):3453–3461.

Kumaran D, Hassabis D, McClelland JL. 2016. What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends Cogn Sci*. **20**(7):512–534.

Mandal BN, and Ma J. 2016. *Non-negative lasso and elastic net penalized generalized linear models*. https://CRAN.R-project.org/package=nnlasso.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*. **102**(3):419.

McCormick C, St-Laurent M, Ty A, Valiante TA, McAndrews MP. 2015. Functional and effective hippocampal–neocortical connectivity during construction and elaboration of autobiographical memory retrieval. *Cereb Cortex*. **25**(5):1297–1305.

Mumford JA, Turner BO, Ashby FG, Poldrack RA. 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*. **59**(3):2636–2643.

Nadel L, Samsonovich A, Ryan L, Moscovitch M. 2000. Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*. **10**(4):352–368.

Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL. 2015. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*. **105**:215–228.

Palombo DJ, Williams LJ, Abdi H, Levine B. 2013. The survey of autobiographical memory (SAM): A novel measure of trait mnemonics in everyday life. *Cortex*. **49**(6):1526–1540.

Poppenk J, Moscovitch M. 2011. A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. *Neuron*. **72**(6): 931–937.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. 2013. Long-axis specialization of the human hippocampus. *Trends Cogn Sci*. **17**(5):230–240.

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. **2**(1):79–87.

Rolls E. 2013. The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*. **7**:74.

Rubin DC, Schrauf RW, Greenberg DL. 2003. Belief and recollection of autobiographical memories. *Mem Cognit*. **31**(6):887–901.

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. 2009. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage*. **44**(3):839–848.

Schlichting ML, Mumford JA, Preston AR. 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat Commun*. **6**(1):1–0.

Sekeres MJ, Winocur G, Moscovitch M. 2018. The hippocampus and related neocortical structures in memory transformation. *Neurosci Lett*. **680**:39–53.

Sheldon S, Farb N, Palombo DJ, Levine B. 2016. Intrinsic medial temporal lobe connectivity relates to individual differences in episodic autobiographical remembering. *Cortex*. **74**:206–216.

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 13 April 2015, preprint: not peer reviewed.

St-Laurent M, Abdi H, Bondad A, Buchsbaum BR. 2014. Memory reactivation in healthy aging: evidence of stimulus-specific dedifferentiation. *J Neurosci*. **34**(12):4175–4186.

Teyler TJ, DiScenna P. 1986. The hippocampal memory indexing theory. *Behav Neurosci*. **100**(2):147.

Tulving E. 1993. What is episodic memory? *Curr Dir Psychol Sci*. **2**(3):67–70.

Wen H, Shi J, Zhang Y, Lu KH, Cao J, Liu Z. 2018. Neural encoding and decoding with deep learning for dynamic natural vision. *Cereb Cortex*. **28**(12):4136–4160.

Wing EA, Ritchey M, Cabeza R. 2015. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *J Cogn Neurosci*. **27**(4):679–691.

Yassa MA, Stark CE. 2011. Pattern separation in the hippocampus. *Trends Neurosci*. **34**(10):515–525.