



Published in final edited form as:

Neuroimage. 2017 July 15; 155: 60–71. doi:10.1016/j.neuroimage.2017.03.042.

Similar patterns of neural activity predict memory function during encoding and retrieval

James E. Kragel^a, Youssef Ezzyat^a, Michael R. Sperling^b, Richard Gorniak^c, Gregory A. Worrell^d, Brent M. Berry^d, Cory Inman^e, Jui-Jui Lin^f, Kathryn A. Davis^g, Sandhitsu R. Das^g, Joel M. Stein^h, Barbara C. Jobstⁱ, Kareem A. Zaghloul^j, Sameer A. Sheth^k, Daniel S. Rizzuto^{a,1}, and Michael J. Kahana^{a,*,1}

^aDepartment of Psychology, University of Pennsylvania, Philadelphia PA 19104, USA

^bDepartment of Neurology, Thomas Jefferson University Hospital, Philadelphia, PA 19107, USA

^cDepartment of Radiology, Thomas Jefferson University Hospital, Philadelphia PA 19107, USA

^dDepartment of Neurology, Mayo Clinic, Rochester MN 55905, USA

^eDepartment of Neurosurgery, Emory School of Medicine, Atlanta GA 30322, USA

^fDepartment of Neurosurgery, University of Texas Southwestern, Dallas TX 75390, USA

^gDepartment of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA

^hDepartment of Radiology, Hospital of the University of Pennsylvania, Philadelphia PA 19104, USA

ⁱDepartment of Neurology, Dartmouth Medical Center, Lebanon NH 03756, USA

^jSurgical Neurology Branch, National Institutes of Health, Bethesda MD 20814, USA

^kDepartment of Neurosurgery, Columbia University Medical Center, New York NY 10032, USA

Abstract

Neural networks that span the medial temporal lobe (MTL), prefrontal cortex, and posterior cortical regions are essential to episodic memory function in humans. Encoding and retrieval are supported by the engagement of both distinct neural pathways across the cortex and common structures within the medial temporal lobes. However, the degree to which memory performance can be determined by neural processing that is common to encoding and retrieval remains to be determined. To identify neural signatures of successful memory function, we administered a delayed free-recall task to 187 neurosurgical patients implanted with subdural or intraparenchymal depth electrodes. We developed multivariate classifiers to identify patterns of spectral power across the brain that independently predicted successful episodic encoding and retrieval. During encoding and retrieval, patterns of increased high frequency activity in prefrontal, MTL, and

^{*}Corresponding author. kahana@psych.upenn.edu (M.J. Kahana).

¹These authors contributed equally to this work.

Conflict of interest

The authors declare no competing financial interests.

inferior parietal cortices, accompanied by widespread decreases in low frequency power across the brain predicted successful memory function. Using a cross-decoding approach, we demonstrate the ability to predict memory function across distinct phases of the free-recall task. Furthermore, we demonstrate that classifiers that combine information from both encoding and retrieval states can outperform task-independent models. These findings suggest that the engagement of a core memory network during either encoding or retrieval shapes the ability to remember the past, despite distinct neural interactions that facilitate encoding and retrieval.

Keywords

iEEG; MVPA; Free recall; Episodic memory

Introduction

Episodic memory requires both encoding operations that translate experiences into durable memories and retrieval operations that reactivate memories when prompted by a retrieval cue (Polyn and Kahana, 2008). Many neurocognitive models of episodic memory account for these processes using similar architectures, with distinct neural systems facilitating memory encoding and retrieval (Nyberg et al., 1996; Lepage et al., 1998; Kim, 2015). Neuroimaging studies provide evidence for these models, revealing dissociations between the neural correlates of episodic encoding and retrieval within the hippocampus (Zeineh et al., 2003; Eldridge et al., 2005), as well as functional networks spanning prefrontal, medial temporal, lateral temporal, and parietal cortical regions (Daselaar et al., 2009; Kim et al., 2010). In contrast to neural substrates that are specific to encoding and retrieval, structures within the hippocampal formation (Small et al., 2001; Stark and Okado, 2003; Prince et al., 2005) in addition to widespread cortical regions (Sederberg et al., 2007) contribute to both episodic encoding and retrieval. Understanding the relative contributions of neural processes engaged during encoding and retrieval is critical to understanding the mechanisms that ultimately determine whether information can be successfully retrieved at a later point in time. The goal of the present work is to determine whether memory performance is best predicted by distinct neural processes that occur during the encoding and retrieval of episodic information, or common electrophysiological states that may reflect the engagement of core mnemonic processes.

Theoretical models of episodic memory propose that successful memory retrieval involves the reinstatement of representations that were present during the initial experience of an event (McClelland et al., 1995; O'Reilly and Rudy, 2001). Using multivariate pattern analysis, it has been repeatedly demonstrated that patterns of cortical activation that are present during the encoding of an event are recapitulated during retrieval (Staresina et al., 2012; Ritchey et al., 2013; Wing et al., 2014). While considerable progress has been made linking the reinstatement of neural activity to the contents of retrieval (for a review, see Danker and Anderson (2010)), the neural states that facilitate the encoding and reactivation of this content may differ. Multivariate classification techniques allow for the estimation of latent cognitive states, such as attentive states that improve task performance (Rosenberg et al., 2015), and incorporate information from neural populations across the brain. By training

a multivariate classifier to identify neural signals that predict successful memory function, we can determine the relationship between the processes that mediate episodic encoding and retrieval.

If the success of episodic memory is determined by the recapitulation of the same processes that were engaged during encoding (Kolers, 1973; Tulving and Thompson, 1973; Kolers and Roediger, 1984), one would predict that engagement of neural systems that are generally predictive of memory success (e.g., the hippocampus; Prince et al. (2005)) should be able to identify when either successful encoding or retrieval operations occur. In contrast, if the function of neural systems that are specific to encoding or retrieval determine the fate of studied information, then classifiers trained to predict the success of memory formation should fail to identify when successful retrieval occurs. Recent electrophysiological investigations of human memory have demonstrated that shifts in the power spectrum, specifically increased high frequency activity (HFA) with concomitant decreases in low frequency activity (LFA) provide informative features for predicting successful memory formation (Burke et al., 2014a; Long et al., 2014; Greenberg et al., 2015) and retrieval (Burke et al., 2014b). These measures reflect a combination of oscillatory dynamics and broadband changes in HFA indexing the firing rate of local neuronal populations (Manning et al., 2009; Miller et al., 2014; Burke et al., 2015). Incorporating these measures of neural activity into multivariate classifiers enables the development of biomarkers of successful memory function, based on the dynamics of distributed neuronal populations.

To determine whether the same neural mechanisms support the encoding and retrieval of episodic memories and to determine the degree to which these measures predict memory performance, we examined intracranial electroencephalographic (iEEG) data from 187 patients who performed a free-recall task while they were undergoing clinical monitoring for the surgical treatment of their drug-resistant epilepsy. To classify successful encoding and retrieval states, we fit L2-penalized regression models to spectral activity recorded during the encoding of verbal stimuli, or in the moments preceding successful recall. Using a cross-classification approach, we tested whether the same changes in spectral power predict successful memory formation and retrieval. Finally, we estimated general models of episodic memory function by constructing a joint classifier that incorporated information from both encoding and retrieval phases of the task. These analyses aimed to uncover the degree to which latent mnemonic states can be measured from brain activity, and the degree to which variation in patterns of spectral power common to both encoding and retrieval can predict memory function.

Methods

Participants

187 patients with medication-resistant epilepsy underwent neurosurgical procedures to implant intracranial electrodes (subdural, depth, or both) to determine epileptogenic regions. Data reported were collected at Dartmouth Medical Center (Hanover, NH), Emory University Hospital (Atlanta, Georgia), Hospital of the University of Pennsylvania (Philadelphia, PA), Mayo Clinic (Rochester, MN), Thomas Jefferson University Hospital (Philadelphia, PA), Columbia University Medical Center (New York, NY), and University of

Texas Southwestern Medical Center (Dallas, TX). Prior to data collection, the research protocol was approved by the institutional review board at each hospital. Informed consent was obtained from either the participant or their guardians. Hemisphere dominance was determined by assessing handedness, Wada test (Wada and Rasmussen, 1960), or fMRI data collected during a verb generation task. Previous publications utilizing a subset of these data (88 patients) have characterized the electrophysiological signatures of successful memory encoding and retrieval (Burke et al., 2014a, 2014b; Long et al., 2014); however, the present analyses describing the relationship of successful encoding and retrieval are novel. Deidentified raw data and analysis code used in the present analyses is available at http://memory.psych.upenn.edu/Electrophysiological_Data.

Free-recall task

Each subject performed a variant of the delayed free-recall task in which they studied a list of words with the intention to commit the items to memory. The task was performed at the bedside on a laptop, using PyEPL software (Geller et al., 2007). Analog pulses were sent to available recording channels to enable alignment of experimental events (e.g., stimulus presentation) with the recorded iEEG signal.

The recall task consisted of three distinct phases: encoding, delay, and retrieval. During encoding, lists of 12 words were presented in the native language (either English or Spanish) of the subject. Words were selected at random, without replacement, from a pool of nouns (<http://memory.psych.upenn.edu/WordPools>). Word presentation lasted for a duration of 1600 ms, followed by a blank inter-stimulus interval (ISI) of 750 to 1000 ms (see Fig 1a). Presentation of word lists was followed by a 20 s post-encoding delay. Subjects performed an arithmetic task during the delay in order to disrupt memory for end-of-list items. Math problems of the form $A+B+C=??$ were presented to the participant, with values of A, B, and C were set to random single digit integers. Responses were made on a keypad, with presentation of additional math problems following each response (i.e., a self-paced task). After the delay, a row of asterisks, accompanied by an 800 Hz auditory tone, was presented for a duration of 300 ms to signal the start of the recall period. Subjects were instructed to recall as many words as possible from the most recent list, in any order during the 30 s recall period. Vocal responses were digitally recorded and parsed offline using Penn TotalRecall (<http://memory.psych.upenn.edu/TotalRecall>). Subjects performed up to 25 recall trials in a single recall session. Across sessions, subjects performed an average of 38 trials in total (range 7–150).

A subset of patients ($n=88$) performed a variant (hereafter Experiment 2) of the previously described task. List presentation consisted of a total of 15 items. In addition, a green fixation cross served as a list-cue to signal an upcoming list of words. The list-cue was presented for a duration of 1600 ms, followed by the presentation of a blank screen for 800 to 1200 ms. The ISI in this variant of the task lasted from 800 to 1200 ms in duration. The recall period for this version of the task was 45 s in length.

Electrophysiological recordings and data processing

iEEG signal was recorded using subdural grids and strips (contacts spaced 10 mm apart) or depth electrodes (contacts spaced 5–10 mm apart) using recording systems at each clinical site. iEEG systems included DeltaMed & XITek (Natus), Grass Telefactor, and Nihon-Kohden EEG systems. Signals were sampled at 500, 512, 1000, 1024, or 2000 Hz, depending on hardware restrictions and considerations of clinical application. Signals recorded at individual electrodes were converted to a bipolar montage by computing the difference in signal between adjacent electrode pairs on each strip, grid, and depth electrode (Burke et al., 2013). Bipolar signal was notch filtered at 60 Hz with a fourth order 2 Hz stop-band butterworth notch filter in order to remove the effects of line noise on the iEEG signal. Electrodes determined to be within the epileptogenic zone were excluded from analysis.

Anatomical localization

Anatomical localization of electrode placement was accomplished using independent processing pipelines for depth and surface electrodes. For data collected as a part of Experiment 1, post-implant CT images were coregistered with presurgical T1 and T2 weighted structural scans with Advanced Normalization Tools (Avants et al., 2008). For patients with MTL depth electrodes, hippocampal subfields and MTL cortices were automatically labeled in a pre-implant, T2-weighted MRI using the automatic segmentation of hippocampal subfields (ASHS) multi-atlas segmentation method (Yushkevich et al., 2015). Subdural electrodes were localized by reconstructing whole-brain cortical surfaces from pre-implant T1-weighted MRIs using Freesurfer (Fischl et al., 2004), and snapping electrode centroids to the cortical surface using an energy minimization algorithm (Dykstra et al., 2012). The localization for data collected in Experiment 2 differed by using FMRIB's linear image registration tool (Jenkinson et al., 2002) for coregistration of CT and structural scans. In addition, MTL depth electrodes that were visible on CT scans were localized to either the hippocampus or PHG by neuroradiologists with expertise in MTL anatomy.

Each subject's T1-weighted MRI was additionally registered to an average T1 constructed from a sample of 101 patients, facilitating group-level comparisons of subdural electrodes on the cortical surface. Registration to MNI space further enabled group analysis of subdural and depth electrodes near MTL structures. Using this approach, we achieved extensive coverage across the cortical surface (Fig. 1b, left), and the MTL (Fig. 1b, right).

In order to rule out the possibility that classification of successful recall, which involves the vocalization of a verbal response, was being informed by neural processing in cortical regions involved in the planning and production of a vocal response, we constructed classifiers using an anatomically restricted set of neural features (Fig. 1b), by excluding electrodes localized within inferior frontal gyrus (IFG) pars triangularis, IFG pars orbitalis, precentral gyrus, postcentral gyrus, paracentral gyrus, superior temporal gyrus, the bank of the superior temporal sulcus, transverse temporal gyrus, and supramarginal gyrus as defined in the Desikan-Kiliany atlas (Desikan et al., 2006).

Spectral power

To compute spectral power during word encoding, we applied the Morlet wavelet transform (wave number 5) to all bipolar electrode EEG signals from the onset to the offset of stimulus presentation, across 50 logarithmically spaced frequencies from 3 to 180 Hz. Spectral power during recall was estimated from 500 ms preceding the onset of response vocalization, for correct recalls and recall errors alike. We also computed spectral power during unsuccessful periods of memory search, defined as 500 ms epochs in which the onset of any vocalization did not occur in the following 2000 ms. All events within the recall period (i.e., correct recalls and unsuccessful search) were required to be free of vocalization onsets in the preceding 2000 ms, to account for potential differences in response production across event types. Power estimates were log transformed and down sampled to 50 Hz. To avoid edge artifacts, we included buffers of 1000 ms surrounding events of interest during the computation of spectral power. For multivariate decoding analyses, features (i.e., log-transformed power from a specific electrode and frequency band) were standardized using the mean and standard deviation estimated from all training samples, for each individual session. Test data were normalized using session means and standard deviations estimated from the training data.

By focusing on patterns of neural activity in the moments preceding individual retrieval events, we excluded recalls that were preceded by the onset of any additional vocalization or the recall period within 2000 ms. This duration was determined in order to allow estimation of low frequency power without contamination of signal from response production earlier in the recall period. While this approach decreased power in the analysis of retrieval data, it allowed for low frequency features, which have been previously demonstrated to predict successful recall (Burke et al., 2014b), to inform classification of memory states. This procedure resulted in the removal of an average of 53.8% of recall events per subject (range 6.3% to 92.7%), which had an inter-response time of 1198 ± 15.4 ms (mean \pm SEM). As a result, some subjects (21 from Experiment 1, and 19 from Experiment 2) did not have a sufficient number of recall events to perform classification. For a given analysis, each subject was required to have at least 20 observations per condition. For the purposes of classification analyses, we treated unique montages (resulting from reimplants, or changes in recording electrodes) of individual patients as a single subject. This approach resulted in 202 subjects for analysis of encoding period data, and 162 subjects for analysis of retrieval period data.

Pattern classification

To identify neural features that were associated with latent memory states during the task, we trained L2-penalized logistic regression classifiers to distinguish between spectral power associated with memory success and failure. During encoding, power patterns were averaged across the 1600 ms stimulus duration at each bipolar electrode, resulting in a set of features to train the encoding classifier. For retrieval analysis, features were constructed from the average spectral power in the 500 ms preceding response vocalization of either correct recalls or periods of unsuccessful memory search. While we sought to identify neural markers of successful retrieval by comparing successful and unsuccessful memory search epochs, these events are confounded by when they occur in the recall period. As a result,

changes in neural processing that distinguish between these two periods (e.g., an indicator of decreased attention at the end of the recall period) may limit our ability to detect memory-related signals. To overcome this issue, we selected unsuccessful memory search events that were matched to the onset of individual recall events in separate lists. This resulted in an average of 69.2 ± 4.6 successful, and 169.9 ± 8.8 unsuccessful search events, per subject.

Classifier performance was evaluated using a cross-validation procedure. We implemented an n -fold cross-validation scheme, where n was equal to the number of completed lists. Data from one list was held out from training to test the ability of the classifier to generalize to novel data, and held out lists were matched across encoding and retrieval classifiers. The remaining $n-1$ lists were used to train the parameters of the classifier. We chose the value of the cost parameter, C , from one of 22 values, spaced from 10^{-4} to 10^6 . For each subject, C was determined by selecting the value that maximized the performance of the classifier as assessed by computing the area under the receiver operating characteristic curve using list-based cross validation in the remaining sample of subjects. This approach enabled optimization of the cost parameter on an independent set of data. We trained the classifier weights by minimizing a loss function over \mathbf{w} for a set of n training events:

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{w}}{2} + C \sum_{i=1}^n \log \left(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i} \right), \quad (1)$$

where \mathbf{x}_i is a set of features (i.e., the pattern of spectral power across all electrodes during encoding or retrieval), and y_i is the corresponding class label for each event. Class labels during encoding were defined as 1 for recalled items, and -1 for forgotten items. During retrieval, correct recalls were labeled 1 and unsuccessful search periods were labeled -1 . A bias term was included by appending a feature that was fixed to 1 for each observation. The loss function was weighted for each observation proportionally to the number of observations in each class, to prevent class imbalance from influencing classification results. We performed L2-logistic regression as implemented by the `liblinear` package (Fan et al., 2008).

In addition to constructing separate classifiers to identify successful memory states during either the encoding or retrieval periods of the task, we developed joint classifiers that were trained on observations from both encoding and retrieval phases of the task. The loss function was weighted by an additional scaling parameter, w_{enc}/w_{rec} which controlled the relative contribution of encoding and retrieval observations to training of the classifier. We examined generalization of this classifier to both encoding and recall periods of the task over a range of 15 log-spaced values from $10^{-1.4}$ to $10^{1.4}$. This range of parameter values allowed us to parametrically modulate the influence of different task phases on memory performance.

To determine whether individual classifiers performed above chance, we first constructed the receiver operating characteristic and computed the area under the curve (AUC) across left-out samples (i.e., across the test set). We implemented a permutation test by constructing a null distribution of classifier performance by permuting the class labels within the training

data. Subjects with an observed AUC in the top 95% of the null distribution were considered significant, corresponding to a one-tailed test. To assess group performance, we used a binomial test to determine whether the proportion of subjects with significant classification would exceed the false positive rate of 0.05.

In order to rule out confounding effects of serial position and output position, we used linear mixed effects models to determine whether classifier output varied as a function of memory success (i.e., subsequent memory status or recall success). During encoding, we constructed a model with subject treated as a random effect, and memory status, list position, and the interaction between the two treated as fixed effects. Significance was determined by constructing nested models with a fixed effect term constrained to zero. After fitting the restricted model, a parametric bootstrap ($n=1000$) was performed to determine whether the variance explained by the additional parameter was greater than fitting residual error with an additional parameter. We utilized a similar procedure to evaluate potential confounds in retrieval classifiers, with the fixed effects of memory status and output position.

Identifying neural features predictive of memory success

In order to determine which spectral features (i.e., which frequencies and regions) facilitated discrimination between successful and unsuccessful memory states, we constructed a forward model of these latent states from our linear classifiers (Haufe et al., 2014). For each subject, we computed a set of activations, A , based on the learned coefficients in \mathbf{w} :

$$A_x = \sum_{\mathbf{x}} \mathbf{w} \sum_{\hat{\mathbf{s}}}^{-1}, \quad (2)$$

where $\Sigma_{\mathbf{x}}$ is the covariance matrix of the full set of features, with each feature standardized within each experimental session, and $\sum_{\hat{\mathbf{s}}}$ is the covariance matrix of the output of the classifier for each predicted class. For each subject and frequency, we computed an estimate of the activation at each electrode. Features were aggregated into LFA (<10 Hz) and HFA (>60 Hz) by averaging estimated activity across the frequency dimension. As a preprocessing step to the estimation of forward models, individual events were scanned for spikes and discarded if the kurtosis value for any feature exceeded a threshold of 2.5.

We projected the associated activations of each subdural electrode to all vertices on the cortical surface within 10 mm of the midpoint of electrode centroids in each bipolar pair. For electrodes localized within the MTL, statistical maps were constructed for each subject by assigning voxels within a 3 mm radius of each bipolar midpoint a value equal to its corresponding activation (overlapping regions were assigned the average of all activation values within 3 mm). Each subject's activation map was smoothed with a 4 mm gaussian filter before subsequent analysis.

Group inference was performed by nonparametric statistical testing (Maris and Oostenveld, 2007). For each test, we performed a one-sample t-test on estimated levels of activation across subjects. Resultant statistical maps were thresholded for statistical significance by constructing a null distribution obtained by random sign flipping ($n=1000$), and identifying

positive and negative thresholds at 2.5th and 97.5th percentiles of the null distribution, ensuring a false discovery rate of 0.05.

Results

Behavioral results

To validate patient performance, we examined characteristic behavior on the free-recall task. Subjects recalled an average of 3.26 ± 0.10 (mean \pm SEM) items per list, yielding an overall $24.27 \pm 0.008\%$ of items recalled. By interleaving encoding and recall with a distracting arithmetic task, we observed primacy effects in both recall initiation and overall recall rates (Fig. 2a,b). These findings are consistent with typical performance on the delayed-free recall task (Glanzer and Cunitz, 1966), and suggest that the neural mechanisms associated with memory performance reflect the encoding and retrieval of long-term memories (Strange et al., 2002). An additional behavioral marker of episodic memory function is the temporal contiguity effect (Kahana, 1996; Sederberg et al., 2010), in which subjects recall items in clusters based on their temporal proximity in the study list. Subjects recalled items with temporal organization (Fig. 2c), indicated by an average temporal factor score of 0.66 ± 0.01 , which reflects significantly greater temporal clustering than chance levels of 0.5 ($t_{191} = 22.46$, $p < 10^{-10}$; see Polyn et al. (2009)). These findings indicate normal task performance, despite relatively low levels of overall recall, consistent with impaired memory performance in epilepsy (Hermann et al., 2008).

Multiple factors contribute to performance on the recall task, including the ability to target and retrieve individual items from the previous list. When memory fails due to the inability to retrieve task-relevant information, studied items are omitted from recall. As shown in Fig. 2d, the time to first recall was inversely related to the number of items yet to be recalled, and the amount of time between successive recalls increased exponentially in the recall sequences, consistent with search dynamics commonly observed in single-trial free recall (Murdock and Okada, 1970; Rohrer and Wixted, 1994). Of relevance for analysis of neural activity during memory search, inter-response times greater than 2 seconds were distributed throughout the majority of recall sequences, except early on in the recall sequence for subjects with a high rate of recall.

Memory failure during retrieval can also occur when items not studied on the previous list can be endorsed as targets, resulting in false recall. We observed false memories in the form of extra-list intrusions (ELIs), prior list intrusions (PLIs) and repetitions of previously recalled items. Recall errors were primarily ELIs, with an average rate of 0.88 ± 0.10 ELIs per list. PLIs were observed with less frequency, with an average rate of 0.57 ± 0.03 intrusions per list. The remaining errors were repetitions, which we rarely observed, with an average of 0.17 ± 0.03 of these errors made per list.

Multivariate classification results

In order to determine whether the same neural mechanisms support episodic memory encoding and retrieval, we constructed subject-specific classifiers that were capable of predicting latent cognitive states from patterns of spectral power across multiple electrodes

and frequencies. First, we developed independent classifiers of memory success during either the encoding or retrieval of individual items. We tested the degree to which these models generalized across different phases of the task, allowing us to determine whether common neural features predict successful encoding and retrieval. Finally, we constructed classification models that incorporated information from both the encoding and retrieval phases of the task, allowing us to determine which electrophysiological features were specific to either encoding or retrieval, or are common indicators of episodic memory function.

Decoding of successful memory formation and retrieval

We first constructed a classifier to identify patterns of spectral power associated with successful encoding. This classifier provides an item level estimate of the probability that subsequent remembering will occur. To relate the output of this classifier to trial-level variability in recall performance, we separated each subject's encoding events into terciles based on classifier output. By computing the probability of recall for items encoded in each of these terciles, we translated the output of the classifier to a measure of actual memory performance. We observed an $80.53 \pm 2.97\%$ difference in the proportion of items recalled in the upper, relative to the lower tercile of classifier output (Fig. 3c). The difference in recall performance was evident across all list positions, although smaller in magnitude for items encoded at the beginning of the list (Fig. 3a). We observed significant classification ($p < 0.05$, permutation test) in 168 out of 202 subjects (treating changes in electrode coverage as independent subjects), with an average AUC of 0.64 ± 0.005 . The number of subjects with significant classification performance was more than one would expect by chance ($p < 10^{-10}$, binomial test).

As the probability of recalling an item decreased across the encoding interval, we conducted a control analysis to rule out the possibility that processes that varied as a function of list position, but not memory specific processing, were primarily informing the classifier. We constructed a linear mixed effects model, predicting the evidence of each item being subsequently remembered (i.e., the exponential term in Eq. (1)) from both subsequent memory status and the serial position of each item. We tested whether unique variance in predicted memory states (i.e., classifier output) was associated with each factor by comparing fitness of the full model to models in which the effect of each factor was restricted to zero. We observed a significant effect of list position ($\chi^2_1 = 122.59$, $p < 0.001$ bootstrap test) and subsequent memory status ($\chi^2_1 = 188.43$, $p < 0.001$ bootstrap test). These findings confirm that the ability to identify encoding periods with a high probability of encoding success was independent of simple serial position effects that influence recall behavior.

Next, we developed classifiers to discriminate between patterns of neural activity during epochs of memory search associated with successful and unsuccessful retrieval (i.e., periods of memory search that did produce any form of recall response). The relationship between successful retrieval and changes in spectral power was observed across different periods of memory search (Fig. 3b), with the biggest difference observed later in the recall period. Of the 162 classifiers constructed to predict retrieval success for individual subjects, we

observed above chance classification ($p < 0.05$, permutation test) in 155, indicative of significant group-level performance ($p < 10^{-10}$, binomial test). Classifier performance for held out lists of items was evaluated using AUC, with average of 0.83 ± 0.01 , indicating robust performance. We tested whether variability in memory states estimated by the retrieval success classifier could be predicted from the period in which an event occurred (early, middle, or late epochs defined by terciles of event onsets within the search period) and whether retrieval was successful or not. We observed significant effects for both search period ($\chi^2_1=14.32$, $p < 0.001$ bootstrap test) and retrieval status ($\chi^2_1=340.8$, $p < 0.001$ bootstrap test). These findings establish that retrieval success classifiers are informative across the duration of the search period, and are more sensitive as the retrieval period progresses. When sorting memory search epochs into terciles by classifier output (Fig. 3c), subjects were $189.3 \pm 5.0\%$ more likely to recall an item when in the upper, compared to the lower tercile of classifier output.

These results suggest that memory-related processing across cortical and subcortical sites can be used to reliably classify when retrieval will occur during memory search; however, such classification may be informed by neural features predictive of processes that occur subsequent to retrieval itself, including phonological retrieval and the planning of a vocal response. In order to rule out the possibility that the ability to decode successful retrieval was not primarily driven by difference in response-related activity, we constructed additional classifiers that excluded ROIs associated with vocal response production and planning (Hermes et al., 2014). Performance of classifiers constructed from this restricted feature space was observed to be significantly above chance ($p < 0.05$, permutation test) in 151 out of 162 subjects. Furthermore, we observed an average AUC of 0.79 ± 0.01 , demonstrating a significant reduction from models with an unrestricted feature space ($t_{161} = -10.3$, $p < 10^{-18}$). These findings suggest that classification of retrieval states is not driven by response related features alone.

Identifying general neural signatures of successful memory function

Having constructed independent classifiers that were capable of estimating the probability of memory success during either encoding or retrieval, we next sought to determine the degree to which these models reflect common or distinct processes during each phase of the free-recall task. We evaluated the degree to which the previously trained classifiers generalized across different phases of the free-recall task by evaluating classifier AUC on held out lists. As shown in Fig. 4a, while decoding of memory states was less reliable when performed across phases of the task, we observed above chance classifier performance using this cross-decoding approach, providing evidence that common neural features predict successful memory function during both encoding and retrieval phases of the task. When we tested the encoding success classifier on periods of memory search associated with successful and unsuccessful retrieval in held out lists of data, we observed an average AUC of 0.68 ± 0.01 . Application of the retrieval success classifier to average patterns of activity during encoding yielded an average AUC of 0.59 ± 0.01 .

While at first glance our cross-decoding results may suggest that the retrieval success model has greater specificity than the encoding success model, differences in classifier

generalization may result from a higher degree of separability in underlying neural states in different phases of the task. We assessed the degree to which model generalization varied as a function of classifier (i.e., encoding or retrieval success) and the phase of the task on which the classifier was tested using a linear mixed effects model. We observed significant effects of classifier type ($\chi^2=178.69, p < 0.001$ bootstrap test), task phase ($\chi^2=295.25, p < 0.001$ bootstrap test), and a significant interaction between the two ($\chi^2=295.25, p < 0.001$ bootstrap test). A post-hoc test confirmed that the decrease in classifier performance when decoding memory states across phases of the task was greater during the retrieval period ($t_{161} = 11.71, p < 10^{-22}$).

These findings suggest the presence of common neural predictors of mnemonic success during both encoding and retrieval, in addition to neural features that are specific to each phase of the task. In order to distinguish between these two types of neural activity, we implemented classifiers that were trained on both encoding and retrieval events, and weighted the relative contribution of each phase of the task to learning. As shown in Fig. 4b, identification of successful encoding states, as determined by subsequent memory, was modulated by the incorporation of neural activity associated with successful retrieval. When learning was biased towards retrieval processing (i.e., $w_{enc}/w_{ret} < 1$), we observed significant ($p < 0.05$, FDR corrected) decreases in prediction of subsequent memory, consistent with our cross-decoding results. Remarkably, performance of joint classifiers were significantly better than the encoding classifier alone at identifying optimal encoding states (Fig. 4b, red line). We next examined the ability of joint classifiers to distinguish between successful and unsuccessful memory search. We observed a significant ($p < 0.05$, FDR corrected) decrease in classifier performance compared to our retrieval classifier (Fig. 4c, blue line) when learning was biased towards patterns of neural activity present during the encoding period. These results demonstrate that the profiles of neural activation that indicate successful retrieval can be leveraged to build a more accurate description of memory states during learning, based on common changes in the local field potential.

Common and distinct neural signatures of encoding and retrieval success

In order to identify which regions of the brain exhibit changes in activity that are generally predictive of successful memory function, we reconstructed patterns of activity that covaried with memory outcomes from joint models with the most extreme weightings (i.e., $w_{enc}/w_{rec} = 0.04$ and $w_{enc}/w_{rec} = 25.12$), and performed conjunction analyses to identify neural features common and specific to the encoding and retrieval models ($p < 0.05$ FDR corrected, permutation procedure for each independent contrast). Increases in HFA and concomitant decreases in low frequency activity (LFA) have been previously identified as neural signatures of successful memory formation using both iEEG in epileptic patients and scalp EEG in healthy individuals (Long et al., 2014). As a result, we focused on these frequency bands when identifying neural features common or specific to episodic encoding and retrieval.

As shown in Fig. 5a, we observed common increases in HFA within inferior prefrontal, temporal, hippocampal, and parietal cortices. In addition to these common features, the observed differences in classification performance across phases of the task suggest the

existence of features that are specific to the encoding and retrieval of list items. We observed significantly increased HFA within lateral occipital regions that was specific to successful encoding (Fig. 5b). In contrast to the typical marker of increased HFA supporting memory outcomes, decreased HFA within the right anterior PFC was additionally found to be a marker of successful encoding. Specific to retrieval, bilateral regions of the anterior hippocampus exhibited increased HFA, in addition to bilateral dorsolateral prefrontal cortex, and pre- and postcentral gyri (Fig. 5c). Activation of motor cortex likely reflects movement of the throat and mouth prior to vocalization, as well as potential smearing of signal from the beginning of vocalization.

Decreases in low frequency activity commonly predictive of successful encoding and retrieval were observed across multiple cortical regions, including bilateral prefrontal, temporal, and inferior parietal regions (Fig. 6a). Within the MTL, robust decreases in LFA were associated with goodness of memory in left lateralized hippocampus, and bilateral PHG. Changes in LFA specific to encoding were primarily observed in lateral and medial temporal cortices, as well as in additional hippocampal regions (Fig. 6b). LFA effects specific to retrieval were found across widespread cortical sites, including lateral prefrontal, inferior parietal, and visual cortices (Fig. 6c).

These findings replicate previous studies that show increased HFA, concurrent with decreased low frequency power across prefrontal, temporal, and hippocampal sites predicts the formation of episodic memories (Burke et al., 2014a; Long et al., 2014). By incorporating these signals into multivariate classifiers, we demonstrate the ability to predict trial-to-trial variability in encoding and retrieval processes based on global brain dynamics.

Discussion

Through the multivariate classification of iEEG recorded neural activity, we identified neural states that predicted changes in memory encoding and retrieval. By implementing a cross-decoding approach, we tested the correspondence between the processes that occur during encoding and retrieval, and were able to determine a high degree of overlap between neural predictors of successful encoding and recall. At a broad level, our findings suggest that successful encoding and retrieval rely upon the function of common neural substrates, a putative core episodic memory network, and that variability in the activation of this network predicts the mnemonic fate of processed information.

Correspondence between neural processes engaged during encoding and retrieval

Recent approaches to understanding memory function in the human brain have used multivariate classification techniques to characterize neural mechanisms involved in the formation and retrieval of episodic memories (e.g., Kuhl et al., 2012; Kuhl and Chun, 2014). While there is theoretical consensus regarding how the contents of memories are represented in the brain (Rissman and Wagner, 2012), it remains an outstanding question whether activation of common neural pathways can facilitate both episodic encoding and retrieval. Consistent with multiple neuroimaging studies that have compared the formation and retrieval of episodic memories (Zeineh et al., 2003; Eldridge et al., 2005), we observed activation in neural systems that were specific to either the encoding or recall of verbal

information. By using a multivariate approach to estimating successful memory processing, we demonstrate that in spite of these dissociations, the same patterns of neural activity observed across multiple recording sites and frequencies were generally predictive of memory function. In doing so, we have identified a putative general memory network spanning the lateral prefrontal cortex, lateral temporal cortex, and MTL whose activation is associated with enhanced memory performance.

The common increases in HFA within this network may reflect the engagement of similar cognitive operations across different phases of the recall task, utilizing HFA as a localizer of neuronal activity (Burke et al., 2015). One potential cognitive process that may account for variability in memory performance is goal-direction attention (Corbetta and Shulman, 2002), which would modulate processing of relevant information during the memory task, including the selection of contextually valid information during recall (Cabeza et al., 2008). As an alternative to an attention-based account, common changes in spectral power may reflect cognitive operations supporting the maintenance and integration of episodic content (Polyn and Kahana, 2008). Recent work examining the electrophysiological correlates of episodic encoding (Long and Kahana, 2015) argues that increased neural activity within left prefrontal, lateral temporal, and MTL sites reflects processing that supports the formation of episodic memories (i.e., the association of information within a spatiotemporal context), as it predicts subsequent temporal organization of learned information. Our findings build upon this work, and suggest that the operations supported by this network are not specific to the encoding of memories, as they facilitate the retrieval of previously learned content.

The features identified by our multivariate models to be predictive of successful memory function match univariate studies of the electrophysiological correlates of episodic encoding (Burke et al., 2014a; Greenberg et al., 2015) and retrieval (Sederberg et al., 2007; Burke et al., 2014b). These studies identified increases in HFA with concomitant decreases in low frequency power across the brain, including recordings from prefrontal and temporal lobe sites, that were indicative of successful memory function. Our results also reveal differences between the neural underpinnings of encoding and retrieval, supporting theoretical models of brain function that emphasize a division of neural resources during each operation. Specific to retrieval, we observed increased HFA within right prefrontal cortex and the MTL, consistent with longstanding neuroimaging findings of asymmetric hemispheric activation during episodic encoding and retrieval (Nyberg et al., 1996). While neural processing may facilitate memory function on a specific phase of the task, such as the observed activation of ventral visual regions during encoding (see also, Burke et al., 2014a), our cross-classification findings demonstrate that similar neural states give rise to successful memory encoding and retrieval. One caveat to this correspondence results from our choice to restrict our analysis to changes in spectral power, leveraging its utility to serve as a marker of local neuronal firing. In addition to changes in levels of local neural activity, functional connectivity between neuronal populations has been shown to predict the formation (Ranganath et al., 2005; Fell et al., 2008) and retrieval (Watrous et al., 2013; Kragel and Polyn, 2015) of episodic memories. This raises the possibility that while changes in spectral activity are conserved across encoding and retrieval, connectivity states that predict performance during encoding and retrieval may differ (Huijbers et al., 2011; Duncan et al., 2014).

The models of latent memory states that we have developed can inform the development of closed-loop systems for memory enhancement. Applications of real-time classification have demonstrated the utility in the real-time detection of attentional lapses to enhance learning by manipulating task difficulty (deBettencourt et al., 2015). Attempts to enhance memory through invasive and noninvasive modulation of neural activity (for a review, see Kim et al., 2016) may prove more efficacious by accounting for the state of the brain when stimulation is applied. Recent work has demonstrated that the ability to enhance memory performance via direct brain stimulation during encoding is dependent upon the state of the brain (Ezzyat et al., in press). Given the correspondence between electrophysiological states and proper memory function, our findings suggest that perturbation of this network during retrieval, in addition to encoding, is likely to modulate memory function.

Decoding memory states from patterns of neural activity

A challenge to the study of human memory is to determine the processes that are responsible for both the transformation of experience into lasting memories as well as the processes involved in retrieving encoded information. Neuroscientific endeavors to better elucidate these processes often rely on contrasting markers of neural function based upon the success of an attempt to remember. For example, the subsequent memory paradigm (Paller and Wagner, 2002) has been utilized to reveal the contributions of prefrontal, MTL, and parietal cortex to the process of memory formation in fMRI (Wagner et al., 1998; Davachi et al., 2001; Kim, 2011) and electrophysiological studies of human memory (Sederberg et al., 2007; Long et al., 2014). While these approaches have established a foundation with which to understand the neural processes that underpin human memory, the correspondence between neural states and general memory performance is often overlooked (cf., Hariri et al., 2003). In contrast, our classification models provide estimates of the probability of memory succeeding or failing based on the observed electrophysiologic state of the brain, effectively providing an estimate of latent cognitive states related to memory processing. This approach enabled us to determine the contributions of processing during encoding and retrieval periods to memory performance.

We observed the greatest correspondence between patterns of neural activity and behavior during successful retrieval, as compared to successful encoding. While it is tempting to interpret these results as indicating that neural processing during retrieval, rather than encoding, plays a greater role in determining the ability to remember (for a theoretical perspective on the role of retrieval processes in remembering, see Tulving (1974)), we do not believe this to be the case. If one assumes our classifiers are influenced by the engagement of multiple cognitive operations (e.g., the maintenance of memory representations during encoding and retrieval), then classifier performance will be determined by the degree to which these processes are differentially engaged during successful or unsuccessful memory function. In this light, improved classification of successful retrieval results from a greater correspondence between the outcome of memory search (i.e., successful recall or failure) and the processes that generate changes in spectral activity during recall. The inability to achieve comparable classifier performance during encoding, as well as the relatively limited cortical regions found to be specific to episodic encoding, results from items that were forgotten despite sufficient processing at the time of encoding. During the free-recall task,

wherein subjects are free to determine which cues they use to probe memory, memory can fail due to ineffective self-generated retrieval cues. As it is challenging to determine the retrieval cues used by a subject during a specific recall period (cf., Polyn et al., 2005), future work should examine the relative contributions of encoding processes when retrieval cues are under experimental control. This could rule out variability in memory performance due to inefficient use of retrieval cues during recall, a phenomenon observed in patients with prefrontal damage (Stuss et al., 1994).

An alternative interpretation of our findings is that additional episodic encoding occurs in the moments leading up to recall. Indeed, the act of retrieval is known to enhance learning, with activation of inferior prefrontal and inferior temporal regions (Buckner et al., 2001) as well as structures within the MTL (Stark and Okado, 2003) predicting the degree to which novel recognition probes are learned during retrieval tasks. We believe it is unlikely that the present results solely reflect incidental encoding processes engaged during free recall. It has been demonstrated that processing within the prefrontal cortex and MTL that predicts subsequent remembering is sensitive to the novelty of learned information (Kirchhoff et al., 2000), with MTL activation attenuating in response to the repetition of well encoded stimuli (Turk-Browne et al., 2006). As retrieval of encoded information results from the reactivation of previously formed item representations, processes associated with the encoding of novel stimuli into long-term memory are unlikely to be engaged during memory search (Kragel and Polyn, 2016).

Conclusion

The processes involved in encoding an event into a lasting memory and remembering that event at a later point in time rely upon inherently distinct neural mechanisms. We demonstrated that multivariate models of memory success can generalize across the encoding and recall of verbal information. These findings suggest that in the presence of neural processes that are specific to the encoding and recall of episodic memories, engagement of a putative core memory system generally shapes the ability to remember the past. Determining the contribution of this system to experimental paradigms that involve different aspects of episodic memory (e.g., the spatial and autobiographical nature of remembering) is a critical next step in identifying neural systems that are generally predictive of the ability to remember.

Acknowledgments

We thank Blackrock Microsystems for providing neural recording equipment. This work was supported by the DARPA Restoring Active Memory (RAM) program (Cooperative Agreement N66001-14-2-4032), as well as National Institutes of Health Grant MH55687. We are indebted to all patients who have selflessly volunteered their time to participate in our study. The views, opinions, and/or findings contained in this material are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008; 12:26–41. [PubMed: 17659998]

- Buckner R, Wheeler M, Sheridan M. Encoding processes during retrieval tasks. *J Cogn Neurosci*. 2001; 13:406–415. [PubMed: 11371316]
- Burke JF, Long NM, Zaghoul KA, Sharan AD, Sperling MR, Kahana MJ. Human intracranial high-frequency activity maps episodic memory formation in space and time. *NeuroImage*. 2014a; 85:834–843. [PubMed: 23827329]
- Burke JF, Ramayya AG, Kahana MJ. Human intracranial high-frequency activity during memory processing: neural oscillations or stochastic volatility? *Curr Opin Neurobiol*. 2015; 31:104–110. [PubMed: 25279772]
- Burke JF, Sharan AD, Sperling MR, Ramayya AG, Evans JJ, Healey MK, Beck EN, Davis KA, Lucas TH, Kahana MJ. Theta and high-frequency activity mark spontaneous recall of episodic memories. *J Neurosci*. 2014b; 34:11355–11365. [PubMed: 25143616]
- Burke JF, Zaghoul KA, Jacobs J, Williams RB, Sperling MR, Sharan AD, Kahana MJ. Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J Neurosci*. 2013; 33:292–304. [PubMed: 23283342]
- Cabeza R, Ciaramelli E, Olson IR, Moscovitch M. The parietal cortex and episodic memory: an attentional account. *Nat Rev Neurosci*. 2008; 9:613–625. [PubMed: 18641668]
- Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 2002; 3:201–215. [PubMed: 11994752]
- Danker JF, Anderson JR. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychol Bull*. 2010; 136:87–102. [PubMed: 20063927]
- Daselaar SM, Prince SE, Dennis NA, Hayes SM, Kim H, Cabeza R. Posterior midline and ventral parietal activity is associated with retrieval success and encoding failure. *Front Human Neurosci*. 2009; 3:13.
- Davachi L, Maril A, Wagner AD. When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. *J Cogn Neurosci*. 2001; 13:1059–1070. [PubMed: 11784444]
- deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. *Nat Neurosci*. 2015; 18:470–475. [PubMed: 25664913]
- Desikan R, Segonne B, Fischl B, Quinn B, Dickerson B, Blacker D, Buckner RL, Dale A, Maguire A, Hyman B, Albert M, Killiany N. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006; 31:968–980. [PubMed: 16530430]
- Duncan K, Tompary A, Davachi L. Associative encoding and retrieval are predicted by functional connectivity in distinct hippocampal area CA1 pathways. *J Neurosci*. 2014; 34:11188–11198. [PubMed: 25143600]
- Dykstra AR, Chan AM, Quinn BT, Zepeda R, Keller CJ, Cormier J, Madsen JR, Eskandar EN, Cash SS. Individualized localization and cortical surface-based registration of intracranial electrodes. *NeuroImage*. 2012; 59:3563–3570. [PubMed: 22155045]
- Eldridge L, Engel S, Zeineh M, Bookheimer S, Knowlton B. A dissociation of encoding and retrieval processes in the human hippocampus. *J Neurosci*. 2005; 25:3280–3286. [PubMed: 15800182]
- Ezzyat Y, Kragel JE, Burke JF, Levy DF, Lyalenko A, Wanda P, O’Sullivan L, Hurley K, Busygin S, Pedisich I, Sperling MR, Worrell GA, Kucewicz MT, Davis KA, Lucas TH, Inman CS, Lega BC, Jobst BC, Sheth S, Zaghoul K, Jutras M, Stein JM, Das S, Gorniak R, Rizzuto DS, Kahana MJ. Direct brain stimulation modulates encoding states and memory performance in humans. *Current Biology*. In press.
- Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. Liblinear: a library for large linear classification. *J Mach Learn Res*. 2008; 9:1871–1874.
- Fell J, Ludwig E, Rosburg T, Axmacher N, Elger C. Phase-locking within human mediotemporal lobe predicts memory formation. *NeuroImage*. 2008; 43:410–419. [PubMed: 18703147]
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004; 14:11–22. [PubMed: 14654453]
- Geller AS, Schleifer IK, Sederberg PB, Jacobs J, Kahana MJ. PyEPL: a cross-platform experiment-programming library. *Behav Res Methods*. 2007; 39:950–958. [PubMed: 18183912]

- Glanzer M, Cunitz AR. Two storage mechanisms in free recall. *J Verbal Learn Verbal Behav.* 1966; 5:351–360.
- Greenberg JA, Burke JF, Haque R, Kahana MJ, Zaghoul KA. Decreases in theta and increases in high frequency activity underlie associative memory encoding. *NeuroImage.* 2015; 114:257–263. [PubMed: 25862266]
- Hariri AR, Goldberg TE, Mattay VS, Kolachana BS, Callicott JH, Egan MF, Weinberger DR. Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *J Neurosci.* 2003; 23:6690–6694. [PubMed: 12890761]
- Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage.* 2014; 87:96–110. [PubMed: 24239590]
- Hermann B, Seidenberg M, Jones J. The neurobehavioural comorbidities of epilepsy: can a natural history be developed? *Lancet Neurol.* 2008; 7:151–160. [PubMed: 18207113]
- Hermes D, Miller KJ, Vansteensel MJ, Edwards E, Ferrier CH, Bleichner MG, van Rijen PC, Aarnoutse EJ, Ramsey NF. Cortical theta wanes for language. *NeuroImage.* 2014; 85:738–748. [PubMed: 23891904]
- Huijbers W, Pennartz CM, Cabeza R, Daselaar SM. The hippocampus is coupled with the default network during memory retrieval but not during memory encoding. *PLoS One.* 2011; 6:e17463. [PubMed: 21494597]
- Jenkinson M, Bannister P, Brady M, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage.* 2002; 17:825–841. [PubMed: 12377157]
- Kahana MJ. Associative retrieval processes in free recall. *Mem Cogn.* 1996; 24:103–109.
- Kim H. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *NeuroImage.* 2011; 54:2446–2461. [PubMed: 20869446]
- Kim H. Encoding and retrieval along the long axis of the hippocampus and their relationships with dorsal attention and default mode networks: the HERNET model. *Hippocampus.* 2015; 25:500–510. [PubMed: 25367784]
- Kim H, Daselaar SM, Cabeza R. Overlapping brain activity between episodic memory encoding and retrieval: roles of the task-positive and task-negative networks. *NeuroImage.* 2010; 49:1045–1054. [PubMed: 19647800]
- Kim K, Ekstrom AD, Tandon N. A network approach for modulating memory processes via direct and indirect brain stimulation: Toward a causal approach for the neural basis of memory. *Neurobiology of Learning and Memory.* 2016
- Kirchhoff BA, Wagner AD, Maril A, Stern CE. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J Neurosci.* 2000; 20:6173–6180. [PubMed: 10934267]
- Kolers PA. Remembering operations. *Mem Cogn.* 1973; 1:347–355.
- Kolers PA, Roediger HL. Procedures of mind. *J Verbal Learn Verbal Behav.* 1984; 23:425–449.
- Kragel JE, Polyn SM. Functional interactions between large-scale networks during memory search. *Cereb Cortex.* 2015; 25:667–679. [PubMed: 24084128]
- Kragel JE, Polyn SM. Decoding episodic retrieval processes: frontoparietal and medial temporal lobe contributions to free recall. *J Cogn Neurosci.* 2016; 28:125–139. [PubMed: 26401811]
- Kuhl BA, Bainbridge W, Chun M. Neural reactivation reveals mechanisms for updating memory. *J Neurosci.* 2012; 32:3453–3461. [PubMed: 22399768]
- Kuhl BA, Chun MM. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J Neurosci.* 2014; 34:8051–8060. [PubMed: 24899726]
- Lepage M, Habib R, Tulving E. Hippocampal PET activations of memory encoding and retrieval: the HIPER model. *Hippocampus.* 1998; 8:313–322. [PubMed: 9744418]
- Long NM, Burke JF, Kahana MJ. Subsequent memory effect in intracranial and scalp EEG. *NeuroImage.* 2014; 84:488–494. [PubMed: 24012858]
- Long NM, Kahana MJ. Successful memory formation is driven by contextual encoding in the core memory network. *NeuroImage.* 2015; 119:332–337. [PubMed: 26143209]

- Manning JR, Jacobs J, Fried I, Kahana MJ. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J Neurosci*. 2009; 29:13613–13620. [PubMed: 19864573]
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*. 2007; 164:177–190. [PubMed: 17517438]
- McClelland JL, McNaughton BL, O'Reilly RC. 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*. 1995; 102:419–457. [PubMed: 7624455]
- Miller KJ, Honey CJ, Hermes D, Rao RP, den Nijs M, Ojemann JG. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *NeuroImage*. 2014; 85:711–720. [PubMed: 24018305]
- Murdock BB, Okada R. Interresponse times in single- trial free recall. *J Verbal Learn Verbal Behav*. 1970; 86:263–267.
- Nyberg L, Cabeza R, Tulving E. Pet studies of encoding and retrieval: the hera model. *Psychon Bull Rev*. 1996; 3:135–148. [PubMed: 24213861]
- O'Reilly RC, Rudy JW. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol Rev*. 2001; 108:311–345. [PubMed: 11381832]
- Paller KA, Wagner AD. Observing the transformation of experience into memory. *Trends Cogn Sci*. 2002; 6:93–102. [PubMed: 15866193]
- Polyn SM, Kahana MJ. Memory search and the neural representation of context. *Trends Cogn Sci*. 2008; 12:24–30. [PubMed: 18069046]
- Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. *Science*. 2005; 310:1963–1966. [PubMed: 16373577]
- Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. *Psychol Rev*. 2009; 116:129–156. [PubMed: 19159151]
- Prince S, Daselaar S, Cabeza R. Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. *J Neurosci*. 2005; 25:1203–1210. [PubMed: 15689557]
- Ranganath C, Cohen MX, Brozinsky CJ. Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *J Cogn Neurosci*. 2005; 17:994–1010. [PubMed: 16102232]
- Rissman J, Wagner AD. Distributed representations in memory: insights from functional brain imaging. *Annu Rev Psychol*. 2012; 63:101–128. [PubMed: 21943171]
- Ritchey M, Wing EA, LaBar KS, Cabeza R. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb Cortex*. 2013; 23:2818–2828. [PubMed: 22967731]
- Rohrer D, Wixted JT. An analysis of latency and interresponse time in free recall. *Mem Cogn*. 1994; 22:511–524.
- Rosenberg MD, Finn ES, Constable RT, Chun MM. Predicting moment-to-moment attentional state. *NeuroImage*. 2015; 114:249–256. [PubMed: 25800207]
- Sederberg PB, Miller JF, Howard WH, Kahana MJ. The temporal contiguity effect predicts episodic memory performance. *Mem Cogn*. 2010; 38:689–699.
- Sederberg PB, Schulze-Bonhage A, Madsen JR, Bromfield EB, McCarthy DC, Brandt A, Tully MS, Kahana MJ. Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cereb Cortex*. 2007; 17:1190–1196. [PubMed: 16831858]
- Small S, Nava A, Perera G, DeLaPaz R, Mayeux R, Stern Y. Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nat Neurosci*. 2001; 4:442–449. [PubMed: 11276237]
- Staresina BP, Henson RN, Kriegeskorte N, Alink A. Episodic reinstatement in the medial temporal lobe. *J Neurosci*. 2012; 32:18150–18156. [PubMed: 23238729]
- Stark CE, Okado Y. Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. *J Neurosci*. 2003; 23:6748–6753. [PubMed: 12890767]

- Strange BA, Otten LJ, Josephs O, Rugg MD, Dolan RJ. Dissociable human perirhinal, hippocampal, and parahippocampal roles during verbal encoding. *J Neurosci.* 2002; 22:523–528. [PubMed: 11784798]
- Stuss DT, Alexander MP, Palumbo CL, Buckle L, Sayer L, Pogue J. Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks. *Neuropsychology.* 1994; 8:355.
- Tulving E. Cue-dependent forgetting. *Am Sci.* 1974; 62:74–82.
- Tulving E, Thompson DM. Encoding specificity and retrieval processes in episodic memory. *Psychol Rev.* 1973; 80:352–373.
- Turk-Browne NB, Yi DJ, Chun MM. Linking implicit and explicit memory: common encoding factors and shared representations. *Neuron.* 2006; 49:917–927. [PubMed: 16543138]
- Wada J, Rasmussen T. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. *J Neurosurg.* 1960; 17:266–282.
- Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science.* 1998; 281:1188–1191. [PubMed: 9712582]
- Watrous AJ, Tandon N, Conner CR, Pieters T, Ekstrom AD. Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nat Neurosci.* 2013; 16:349–356. [PubMed: 23354333]
- Wing EA, Ritchey M, Cabeza R. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *J Cogn Neurosci.* 2014; 27:679–691. [PubMed: 25313659]
- Yushkevich PA, Pluta JB, Wang H, Xie L, Ding SL, Gertje EC, Mancuso L, Kliot D, Das SR, Wolk DA. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapp.* 2015; 36:258–287.
- Zeineh M, Engel S, Thompson P, Bookheimer S. Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science.* 2003; 299:577. [PubMed: 12543980]

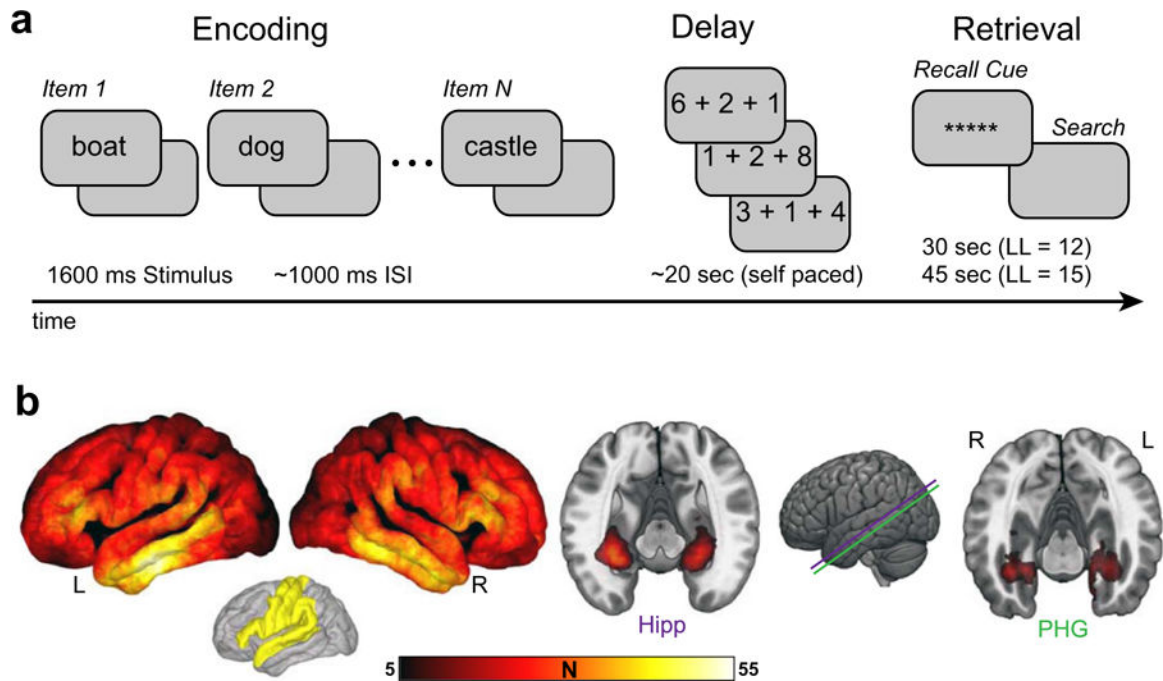


Fig. 1.

Recall task and electrode coverage. a. Experimental paradigm. On each trial, patients studied a list of words, performed a self paced arithmetic task, and finally recalled the items studied on the most recent list in any order. b. Electrode coverage. Left, the number of subjects with bipolar electrode centers within 10 mm of each vertex of the average cortical surface. Right, cross sections along the longitudinal axis of the MTL showing the number of subjects with bipolar electrode coverage localized within hippocampus (Hipp) or parahippocampal gyrus (PHG), extending 3 mm from electrode centroids. A priori anatomical regions of interest excluded from a subset of analyses are depicted in yellow. The depicted coverage discounts electrodes within the epileptogenic zone. L, left; R, right.

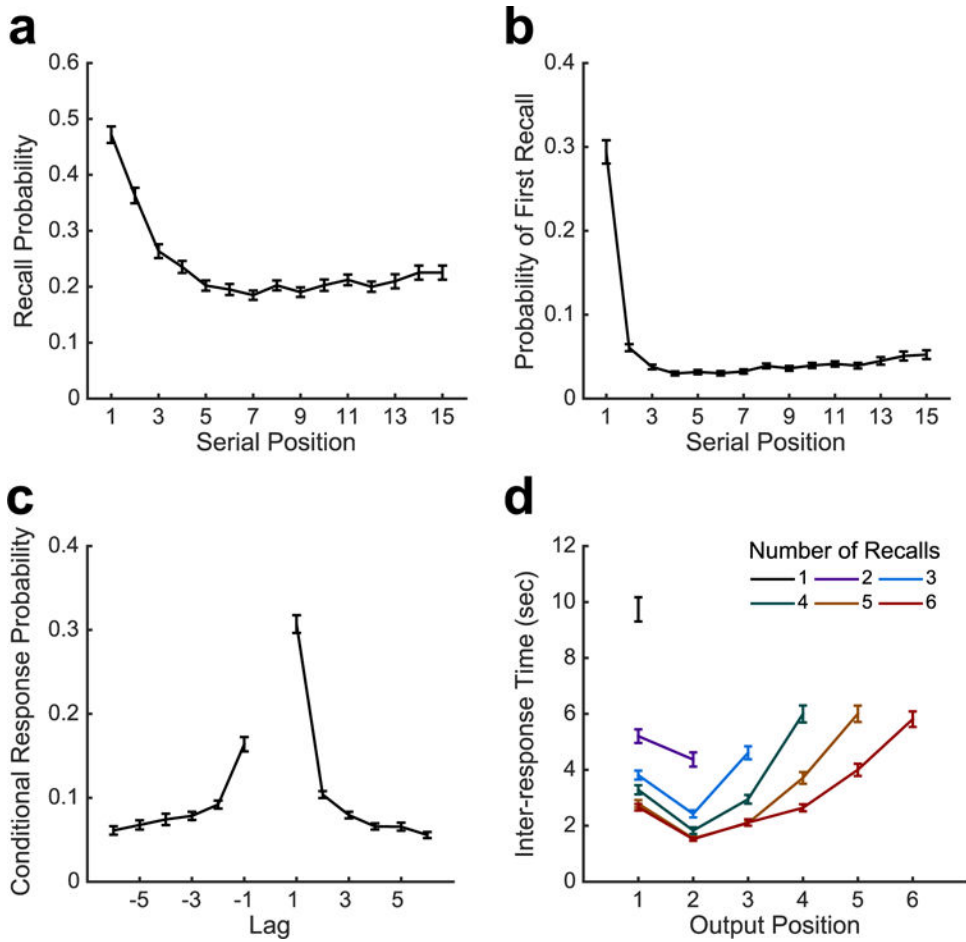
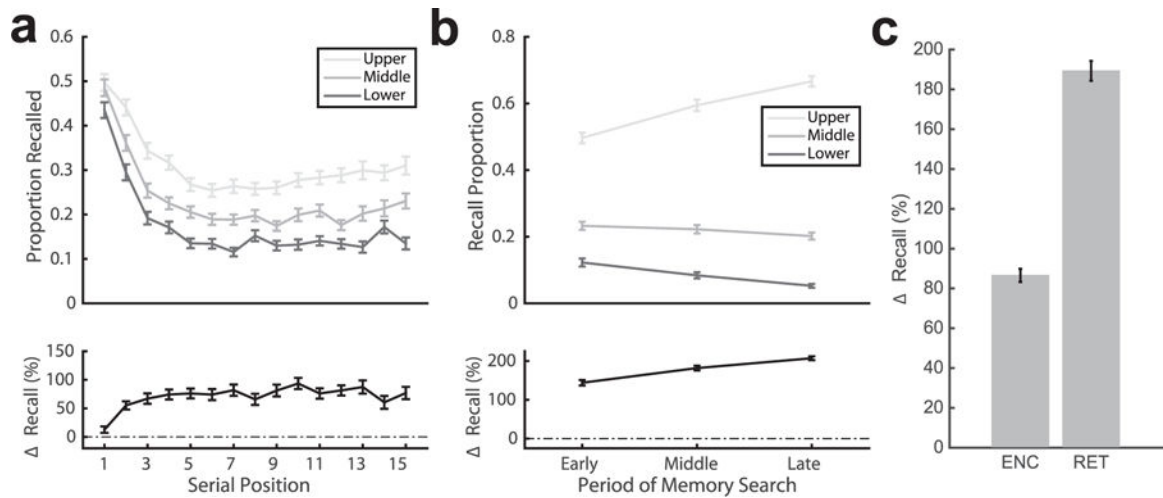
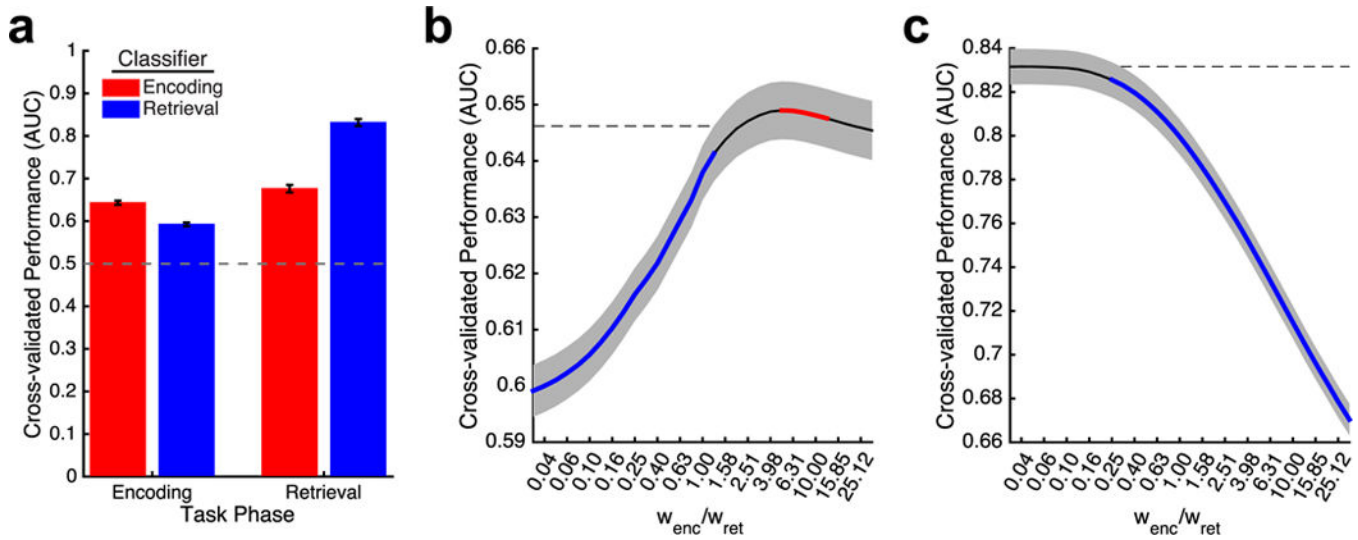


Fig. 2. Behavioral results. a. The overall probability that items from each list position were recalled. b. The probability that an item from a given list position was the first item recalled, as a function of serial position. c. The probability that a recall transition will come from a temporally proximal list position, given it is available for recall. d. Inter-response time (measured from onset to onset of response vocalization) as a function of recall sequence length. Error bars represent standard error of the mean.

**Fig. 3.**

Decoding successful memory formation and retrieval. **a.** Probability of encoding success as a function of list position. The top panel depicts the probability of recalling an item as a function of serial position, for each tercile of classifier output. Below, the difference in percent recall change (relative to average percent recall) between the upper and lower tercile of classifier output. **b.** Probability of retrieval success increases across the memory search period. The top panel shows the average proportion of recalls made in epochs decoded across the recall period, sorted into terciles based upon classifier output. Below, the difference in the proportion of recalls made relative to the mean between upper and lower terciles of classifier output. **c.** Overall change in memory performance during encoding (ENC) and retrieval (RET) between high and low classifier terciles. Error bars represent standard error of the mean.

**Fig. 4.**

Common neural signatures of successful memory function. a. Classifier performance as a function of task phase. Average area under the curve (AUC) for classifiers trained to detect successful encoding and retrieval processing. Error bars reflect standard error of the mean, across subjects. b. Predicting successful encoding using a joint classifier. AUC for joint classifiers applied to encoding period data, with significant ($p < 0.05$, FDR corrected) increases and decreases in performance, compared to the encoding classifier (dashed line) are shown in red and blue, respectively. Shaded area depicts standard error of the mean. c. Predicting successful retrieval using a joint classifier. AUC for joint classifiers applied to retrieval period observations, with significant decreases in performance compared to the retrieval classifier alone (dashed line) are depicted in blue ($p < 0.05$, FDR corrected).

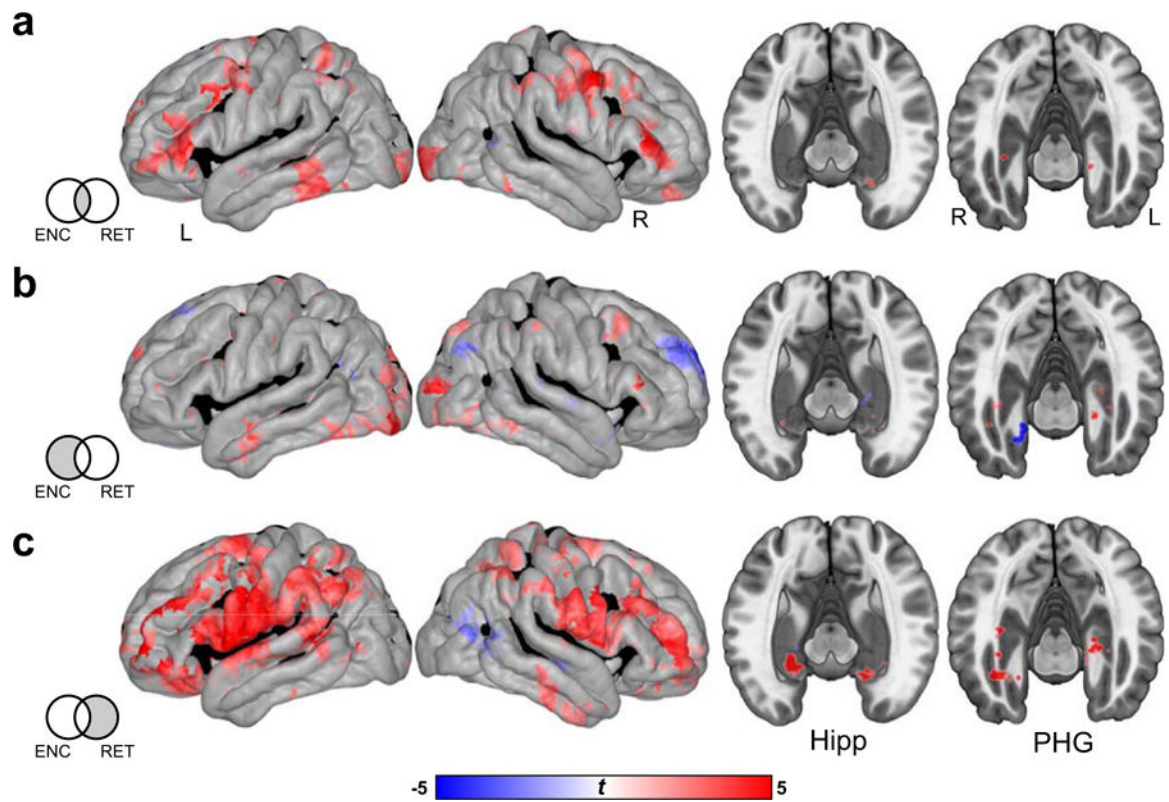


Fig. 5. High frequency activity (HFA) common and specific to successful episodic encoding and retrieval. a. Regions common to successful encoding and retrieval. b. Regions with HFA that significantly differs between subsequently remembered and forgotten items, exclusively masking out any regions that exhibit retrieval related effects. c. Regions showing significant differences in HFA, specific to episodic retrieval. Significant ($p < 0.05$, FDR corrected) increases and decreases in HFA depicted on the cortical surface (left) and MTL subregions (right) are shown in red and blue, respectively.

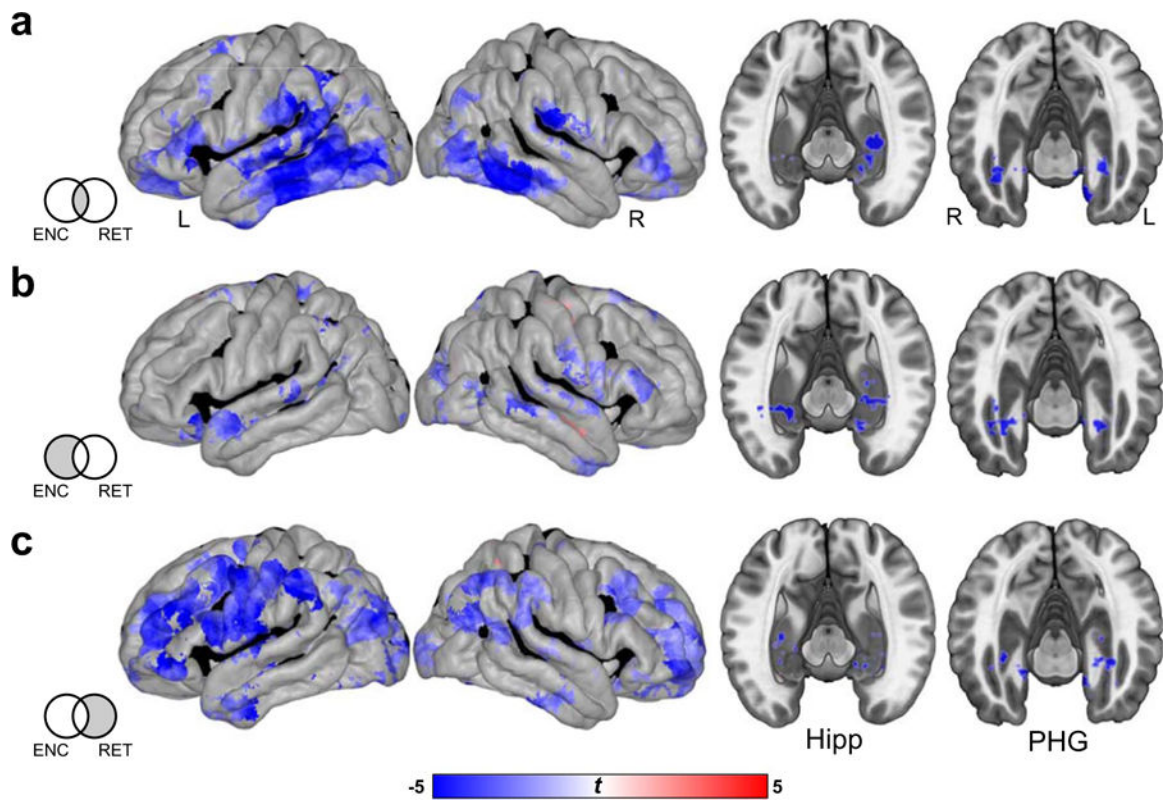


Fig. 6.

Low frequency activity (LFA) features common and specific to successful episodic encoding and retrieval. a. Regions exhibiting significant differences in LFA during successful memory processing common to successful encoding and retrieval. b. Regions with changes in LFA that significantly differs between subsequently remembered and forgotten items, exclusively masking out any regions that exhibit retrieval related effects are shown. c. Regions showing significant differences in LFA with memory success, specific to episodic retrieval. Significant ($p < 0.05$, FDR corrected) increases and decreases in LFA are depicted on the cortical surface (left) and MTL subregions (right) are shown in red and blue, respectively.