Published in final edited form as:
Neuroimage. 2019 November 01; 201: 116001. doi:10.1016/j.neuroimage.2019.07.014.

Decoding the tradeoff between encoding and retrieval to predict memory for overlapping events

Nicole M. Long[1], **Brice A. Kuhl**[1]

[1:]Department of Psychology, University of Oregon 97403

Abstract

When new events overlap with past events, there is a natural tradeoff between encoding the new event and retrieving the past event. Given the ubiquity of overlap among memories, this tradeoff between memory encoding and retrieval is of central importance to computational models of episodic memory (O'Reilly & McClelland, 1994; Hasselmo, 2005). However, prior studies have not directly linked neural markers of encoding/retrieval tradeoffs to behavioral measures of how overlapping events are remembered. Here, by decoding patterns of scalp electroencephalography (EEG) from male and female human subjects, we show that tradeoffs between encoding and retrieval states are reflected in distributed patterns of neural activity and, critically, these neural tradeoffs predict how overlapping events will later be remembered. Namely, new events that overlapped with past events were more likely to be subsequently remembered if neural patterns were biased toward a memory encoding state–or, conversely, away from a retrieval state. Additionally, we show that neural markers of encoding vs. retrieval states are surprisingly independent from previously-described EEG predictors of subsequent memory. Instead, we demonstrate that previously-described EEG predictors of subsequent memory are better explained by task engagement than by memory encoding, per se. Collectively, our findings provide important insight into how the memory system balances memory encoding and retrieval states and, more generally, into the neural mechanisms that support successful memory formation.

Keywords

EEG; MVPA; episodic memory; encoding; retrieval; subsequent memory effect

Introduction

Most of our experiences have some overlap with past events and this overlap can function as a cue that triggers memory retrieval (Kuhl, Shah, DuBrow, & Wagner, 2010; Zeithamova, Dominick, & Preston, 2012). However, retrieval of the past potentially comes at the expense of encoding new memories (Huijbers, Pennartz, Cabeza, & Daselaar, 2009; Duncan,

Corresponding Author: Nicole Long (niclong@virginia.edu); Brice Kuhl (bkuhl@uoregon.edu).

Conflict of Interest: The authors declare no competing financial interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Sadanand, & Davachi, 2012; Patil & Duncan, 2018). For example, upon encountering an acquaintance, you may find yourself remembering a previous conversation with this acquaintance only to realize that you have not encoded the current conversation. This tradeoff between encoding and retrieval that arises whenever memories overlap is of central importance to computational models of episodic memory (O'Reilly & McClelland, 1994) and may reflect a fundamental opposition between the neural states that support encoding vs. retrieval (Hasselmo, 2005). Yet, there is surprisingly little evidence showing that neural measures of encoding/retrieval states predict how overlapping events will later be remembered.

Evidence from rodent and human studies collectively motivates the idea that encoding/retrieval tradeoffs are reflected in electrophysiological measures (Hasselmo, Bodelon, & Wyble, 2002; Rizzuto, Madsen, Bromfield, Schulze-Bonhage, & Kahana, 2006; Griffin, Eichenbaum, & Hasselmo, 2007; Manns, Zilli, Ong, Hasselmo, & Eichenbaum, 2007; Colgin et al., 2009; Hasselmo & Stern, 2014). Much of this work has focused on relatively rapid alternations between encoding and retrieval states (theta phase), with each state lasting 100 ms or less (Hasselmo, 2005). However, there is also evidence from human and rodent studies of more sustained oscillatory signals (including theta amplitude) that reflect the state of the memory system (Kirov, Weiss, Siebner, Born, & Marshall, 2009; Molter, O'Neill, Yamaguchi, Hirase, & Leinekugel, 2012). These sustained oscillatory signals potentially reflect neuromodulatory effects (in particular, acetylcholine levels), which exert an influence on a relatively slow timescale (on the order of seconds, Meeter, Murre, & Talamini, 2004; Hasselmo & McGaughy, 2004). Indeed, evidence from human behavioral studies indicates that biases toward encoding vs. retrieval states can last at least several seconds (Duncan et al., 2012; Patil & Duncan, 2018), consistent with the timescale at which acetylcholine is thought to influence the memory system (Meeter et al., 2004). Moreover, these behavioral studies also indicate that biases toward encoding vs. retrieval states influence how *new events* are remembered (Duncan et al., 2012). Collectively, these findings motivate the idea that relatively long-timescale biases toward encoding vs. retrieval states are reflected in electrophysiological measures and that these biases may critically determine how overlapping events are subsequently remembered.

To the extent that biases toward encoding vs. retrieval are reflected in sustained electrophysiological activity patterns, an important secondary question is whether these electrophysiological patterns mirror classic subsequent memory effects (SMEs). Numerous scalp and intracranial EEG studies have identified neural predictors of subsequent memory (Paller, Kutas, & Mayes, 1987; Fernandez et al., 1999; Friedman & Johnson, 2000; Otten & Rugg, 2001; Sederberg, Kahana, Howard, Donner, & Madsen, 2003; Gruber, Tsivilis, Montaldi, & Müller, 2004). In particular, subsequent memory is consistently predicted by increases in high frequency activity and decreases in low frequency activity (Osipova et al., 2006; Sederberg et al., 2006; Burke et al., 2014; Long, Burke, & Kahana, 2014; Greenberg, Burke, Haque, Kahana, & Zaghloul, 2015). On the one hand, this pattern of high frequency increases and low frequency decreases may reflect a neural state that is biased toward memory encoding–and away from memory retrieval. On the other hand, however, it is possible that this neural pattern reflects other dimensions of cognitive processing (vigilance, elaborative processing, arousal, etc.) that tend to be correlated with subsequent remembering but that do not directly map to whether the memory system is in an encoding vs. retrieval state. Thus, comparing neural markers of encoding vs. retrieval states to canonical SMEs is important for understanding, in mechanistic terms, the dimensions that contribute to successful memory formation.

Here, we report a human scalp EEG study in which subjects first studied an initial set of object images and then studied highly similar (overlapping) images (Bakker, Kirwan, Miller, & Stark, 2008). Critically, however, during study of the overlapping images, we explicitly biased subjects toward encoding vs. retrieval states. Afterward, subjects completed a final recognition test that probed memory for all of the previously-presented object images. We address three primary questions. First, can biases toward encoding vs. retrieval states be decoded from distributed electrophysiological activity patterns? Second, do these electrophysiological biases predict how overlapping events will later be remembered? Finally, how do neural markers of encoding vs. retrieval states relate to canonical SMEs?

Materials and Methods

Subjects

Forty (30 female; mean age = 21 years) right-handed, native English speakers from the University of Oregon community participated. All subjects had normal or corrected-to-normal vision. Informed consent was obtained in accordance with the University of Oregon Institutional Review Board. Four subjects were excluded from the final dataset: one who felt ill during set up and subsequently exited the experiment, two who failed to respond to over 50% of recognition phase trials, and one who likely inverted the recognition phase response mappings (accuracy of 10%). Thus, data are reported for the remaining 36 subjects. The raw, de-identified data as well as associated experimental and analysis codes used in this study can be accessed via the Kuhl Lab website (http://kuhllab.com/publications/).

Materials

Stimuli consisted of 576 object pictures, drawn from an image database with multiple exemplars per object category (Konkle, Brady, Alvarez, & Oliva, 2010). From this database, we chose 144 unique object categories and 4 exemplars from each category. For each subject, one exemplar in a set of four served as a List 1 item, one as a List 2 item, and the two remaining exemplars served as lures for the recognition phase. Object condition assignment was randomly generated for each subject.

Experimental Design and Statistical Analysis

Procedure

General Overview.: In each of eight runs, subjects viewed two lists containing object images. For the first list, each object was new (List 1 objects). For the second list (List 2 objects), each object was again new, but was categorically related to an object from the first list. For example, if List 1 contained an image of a bench, List 2 would contain an image of a different bench. During List 1, subjects were instructed to encode each new object. During List 2, however, each trial contained an instruction to either encode the current object (e.g., the new bench) or to retrieve the corresponding item from List 1 (the old bench). Following eight runs, subjects completed a two-alternative forced-choice recognition test that separately assessed memory for List 1 and List 2 objects.

List 1.: On each trial, subjects saw a single object presented for 3000 ms followed by a 1000 ms inter-stimulus interval (ISI; Figure 1A). Subjects were instructed to study the presented object in anticipation for a later memory test.

List 2.: On each trial, subjects saw a cue word, either “OLD” or “NEW” for 2000 ms. The cue was followed by presentation of an object for 2000 ms, which was followed by a 2000 ms ISI (Figure 1A). All objects in List 2 were non-identical exemplars drawn from the same category as the objects presented in the immediately preceding List 1. That is, if a subject saw a bench and a fan during List 1, a different bench and a different fan would be presented during List 2. On trials with a “NEW” instruction (encode trials), subjects were to encode the presented object. On trials with an “OLD” instruction (retrieve trials), subjects tried to retrieve the categorically related item from the preceding List 1. Importantly, this design prevented subjects from completely ignoring List 2 items following “OLD” instructions in that they could only identify the to-be-retrieved object category by processing the List 2 item.

Subjects completed eight runs with two lists in each run (List 1, List 2). Subjects viewed 18 objects per list, yielding a total of 288 object stimuli from 144 unique object categories. Subjects did not make a behavioral response during either List 1 or 2.

Recognition phase.: Following the eight runs, subjects completed the recognition phase. On each trial, subjects saw two exemplars from the same object category (e.g. two benches; Figure 1A). One object had previously been encountered either during List 1 or 2. The other object was a lure and had not been presented during the experiment. Subjects selected (via mouse click) the previously presented object. Subjects had 4000 ms to respond. If the subject failed to respond in time, the trial was counted as incorrect. Trials were separated by a 1000 ms ISI. There were a total of 288 recognition trials (corresponding to the 288 total List 1 and 2 items presented in the experiment). Note: List 1 and List 2 items never appeared in the same trial together, thus subjects never had to choose between two previously presented items. List 1 and List 2 items were presented randomly throughout the recognition phase.

EEG Data Acquisition and Preprocessing

EEG recordings were collected using a BrainAmp system (Brain Products, Inc.) and an ActiCap equipped with 32 Ag-AgCl active electrodes positioned according to the extended 10-20 system, with electrodes placed on both the left and right mastoids. We additionally included six passive electrodes for recording eye movements and blinks: two each above and below the left and right eyes, plus two in the eye cannulas. The two mastoid electrodes and six passive electrodes were used for recording and artifact detection purposes only and are not included in any analyses. All electrodes were digitized at a sampling rate of 1000 Hz and were referenced to a right-mastoid electrode. Offline, electrodes were later converted to an average reference. Impedances of all electrodes were kept below 50 kΩ. Electrodes that demonstrated high impedance or poor contact with the scalp were excluded from the average reference. Bad electrodes were determined by voltage thresholding (see below). A combination of EEGLAB (Delorme & Makeig, 2004) and custom Matlab codes were used to process the EEG data. We used an automatic artifact correction algorithm based on (Nolan, Whelan, & Reilly, 2010) using Independent Components Analysis (ICA; Bell & Sejnowski, 1995; Onton & Makeig, 2006) to detect and correct for eye blinks and saccades.

Using raw EEG signals, we performed three preprocessing steps to identify and correct electrodes with severe artifacts separately for each subject. First, we calculated the mean correlation between each electrode and all other electrodes as electrodes should be moderately correlated with other electrodes due to volume conduction. We z-scored these means across electrodes and rejected electrodes with z-scores less than −3. Second, we calculated the variance for each electrode as electrodes with very high or low variance across a session are likely dominated by noise or have poor contact with the scalp. We then z-scored variance across electrodes and rejected electrodes with a $|z| >= 3$. Finally, we expect many electrical signals to be autocorrelated, but signals generated by the brain versus noise are likely have different forms of autocorrelation. Therefore, we calculated the Hurst exponent, a measure of long-range autocorrelation, for each electrode and rejected electrodes with a $|z| > = 3$. Electrodes that were marked as bad by this procedure were interpolated using EEGLAB's (Delorme & Makeig, 2004) spherical spline interpolation algorithm. On average one electrode was interpolated per subject (M = 1.25, SD = 0.5542, range = 0-3).

We next ran ICA on this artifact-corrected data. The maximum number of independent components (ICs) that can be reliably estimated depends on the number of samples recorded for each electrode. Following (Nolan et al., 2010) we extracted $c = floor(\sqrt{L / k})$ ICs where L is the number of samples in the session and k is a constant set to 25 (Onton & Makeig, 2006) or the number of non-interpolated electrodes, whichever was smaller. We then ran EEGLAB's implementation of infomax ICA (Delorme & Makeig, 2004; Bell & Sejnowski, 1995) on the first c principal components of the EEG matrix to decompose it into ICs.

ICs that capture blinks or saccades should be highly correlated with the raw signal from the passive electrodes located around the eyes. Therefore, for each IC we computed the absolute value of its correlation with each of the 4 EOG electrodes positioned above and below the eyes. We retained the maximum of those values and z-scored the maximum correlations across ICs. We rejected ICs with $|z| >= 3$. ICs that capture artifacts isolated to single electrodes (e.g., an electrode shifting) should have high weights for the implicated electrodes but low weights for other electrodes. To identify such ICs, we calculated the kurtosis of the weights across electrodes and excluded any IC with $|z| >= 3$. Finally, ICs capturing white noise should have a nearly flat power spectrum (vs. the 1/f spectrum expected for neural signals). Therefore, we calculated the absolute value of the slope of the power spectrum for the frequencies included in the analyses (2–100 Hz) and rejected ICs with $z >= -3$ (i.e., the ones closest to zero slope). Rejected ICs were removed from the matrix and the remaining IC activation time courses were projected back into electrode space. Finally, a fourth order 2 Hz stopband butterworth notch filter was applied at 60 Hz to eliminate electrical line noise. All subsequent analyses were carried out on this corrected EEG data.

EEG data analysis

We applied the Morlet wavelet transform (wave number 6) to all electrode EEG signals from 2500 ms preceding to 4000 ms following object presentation, across 46 logarithmically spaced frequencies (2–100 Hz; Long & Kahana, 2015). We included a 1000 ms buffer on both sides of the data to minimize edge effects. After log-transforming the power, we downsampled the data by taking a moving average across 100 ms time intervals and sliding the window every 25 ms, resulting in 257 time intervals (65 non-overlapping). Power values were then *z*-transformed by subtracting the mean and dividing by the standard deviation power. Mean and standard deviation power were calculated across all List 1 and List 2 items and across time points for each frequency.

Univariate analyses

To test effects specific to high frequency activity (HFA) and low frequency activity (LFA), we divided the *z*-transformed power into two distinct frequency bands (HFA, 28 - 100 Hz; LFA, 2 - 26 Hz) by taking the mean of the *z*-power in each frequency band. The cutoff of 28 Hz was derived from an independently collected dataset (Burke et al., 2014). We additionally averaged *z*-power across the stimulus interval (0-3000 ms for List 1; 0-2000 ms for List 2) and then averaged *z*-power across our conditions of interest. For the subsequent memory effect (SME) analyses, in order to reduce the influence of small bin sizes subjects were required to have a minimum of five events per condition of interest in order to be included in analyses (Long & Kahana, 2015). For the List 1 SME, subjects had on average 120 remembered and 24 forgotten items (SD = 14); one subject was excluded. For the List 2 SMEs, subjects had on average 58 remembered encode items and 14 forgotten encode items (SD = 8), and 54 remembered retrieve items and 18 forgotten retrieve items (SD = 9); three subjects were excluded. For the List 1 SME based on List 2 retrieve trials, subjects had on average 58 remembered items and 14 forgotten items (SD = 14); 12 subjects were excluded.

Pattern classification analyses

Pattern classification analyses were performed using penalized (L2) logistic regression (penalty parameter = 1), implemented via the Liblinear toolbox (Fan, Chang, Hsieh, Wang, & Lin, 2008) and custom MATLAB code. Classifier performance was assessed in two ways. “Classification accuracy” represented a binary coding of whether the classifier successfully guessed the instruction condition. We used classification accuracy for general assessment of classifier performance (i.e., whether instructions could be decoded). “Classifier evidence” was a continuous value reflecting the logit-transformed probability that the classifier assigned the correct instruction for each trial. Classifier evidence was used as a trial-specific, continuous measure of state information, which was used to assess the degree to which evidence for a given state predicted subsequent memory performance.

We used leave-one-run-out cross validation classification to test whether encode/retrieve instructions could be decoded. For each subject, a classifier was trained to discriminate encode from retrieve instructions during List 2 using the average *z*-power across the 0-2000 ms stimulus interval, 46 logarithmically spaced frequencies from 2 to 100 Hz, and all 30 electrodes.

Statistical analyses

We used repeated measures ANOVAs and paired-sample *t*-tests to assess the effect of encode/retrieve instruction on behavioral memory performance. We used repeated measures ANOVAs to assess all SMEs.

We used paired-sample *t*-tests to compare classification accuracy across subjects to chance decoding accuracy, as determined by permutation procedures. Namely, for each subject we shuffled the condition labels of interest (e.g., “encode” and “retrieve” for the List 2 instruction classifier) and then calculated classification accuracy. We repeated this procedure 1000 times for each subject and then averaged the 1000 shuffled accuracy values for each subject. These mean values were used as subject-specific empirically derived measures of chance accuracy. Paired samples *t*-tests compared the observed (unshuffled) accuracy values to the shuffled accuracy values.

We used logistic regression to assess whether classifier-based encoding evidence predicted subsequent memory. For each logistic regression analysis, regressors included: encoding evidence, instruction (encode, retrieve), run number, and serial position. We used one-sample *t*-tests to compare the logistic regression beta values to zero.

Results

Behavior

We first tested whether instructions influenced performance on the recognition test. While encode/retrieve instructions only appeared during List 2, we also considered whether memory for List 1 items was influenced by List 2 instructions (e.g., whether memory for the old bench was influenced by whether the new bench was associated with an encode vs. retrieve instruction). An ANOVA with factors of list (1,2) and instruction (encode, retrieve; Figure 1B) revealed a list by instruction interaction ($F_{1,35}$ = 8.1981, p = 0.0070). For List 1 items, memory was comparable for encode (M = 83.68%, SD = 10.12%) and retrieve items (M = 84.41%, SD = 10.00%; difference between encode vs. retrieve: t_{35}= −0.7343, p = 0.4677). For List 2 items, however, memory was better for encode (M = 80.86%, SD = 11.51%) than retrieve items (M = 75.96%, SD = 12.77%; difference between encode vs. retrieve: t_{35}= 3.0398, p = 0.0045). While subtle, these results confirm that subjects were able to shift between encoding and retrieval states in a goal-directed manner.

EEG markers of encoding vs. retrieval states

We next assessed potential neural differences between encoding vs. retrieval states (Tulving et al., 1994; Lepage, Habib, & Tulving, 1998). Of particular interest was whether biases toward encoding vs. retrieval states could be decoded on a trial-by-trial basis. To this end, we conducted a multivariate pattern classification analysis (Richter, Chanales, & Kuhl, 2016). Specifically, we trained a classifier to discriminate encoding vs. retrieval trials based on a feature space comprised of all 30 electrodes × 46 logarithmically spaced frequencies ranging from 2 to 100 Hz. For this analysis, spectral power was averaged over the stimulus interval. Using within-subject, leave-one-run-out classifiers, mean classification accuracy was 55.71% (SD = 9.11%), which was reliably greater than chance, as determined by permutation tests (t_{35}= 3.7476, $p = 0.0006$; Figure 2A). Excluding the two subjects who had the highest classification accuracies (accuracies > 80%, z scores > 3; see Figure 2A), classification accuracy remained reliably greater than chance (mean accuracy = 54.04%, SD = 6.04%, t_{33}= 3.8939, $p = 0.0004$). At the level of individual subjects, classification accuracy was reliably above chance (observed accuracy >95% of accuracies from the permuted distribution) in 12 out of the 36 subjects. To visualize state effects over time, we measured encoding evidence across 100 ms intervals during the List 2 trials, separately for encode and retrieve trials (Figure 2B). In a 2 (encode, retrieve) × 20 (time interval) repeated measures ANOVA, we found a reliable interaction between instruction and time interval ($F_{19,665}$ = 2.092, $p = 0.0043$). As we did not have a *priori* predictions about specific time intervals, all subsequent analyses are based on data averaged across the entire time window. Figure 2C displays the mean difference in spectral power between encode vs. retrieve trials for each of the 30 electrode × 46 frequency bins.

Critically, we next tested whether trial-level evidence of memory states (derived from the classifiers) predicted subsequent memory for List 2 items (Figure 2D). We predicted that greater evidence for an encoding state (or, conversely, less evidence for a retrieval state) would be associated with better subsequent memory. Logistic regression analyses (which included factors of trial instruction and encoding evidence) revealed a significant, positive relationship between encoding evidence and subsequent memory (mean $\beta = 0.0561$, SD = 0.1343, one-sample *t*-test vs. 0, t_{34}= 2.4709, $p = 0.0186$). As a control analysis, we tested whether List 2 encoding evidence also predicted memory for List 1 items, but this relationship was not significant (mean $\beta = -0.0122$, SD = 0.1250, one-sample *t*-test vs. 0, t_{34}= −0.5767, $p = 0.5680$; Figure 2D). This null result argues against a non-selective relationship between encoding evidence and particular stimulus categories–for example, that some categories of images (e.g., benches) tend to elicit stronger encoding evidence and tend to be better remembered.

While the observed relationship between List 2 encoding evidence and subsequent List 2 memory controlled for the actual instruction on each trial, we next considered a stronger test: whether evidence for an encoding state predicted subsequent memory when specifically considering only List 2 trials where the instruction was to retrieve. Strikingly, when only considering retrieve trials, classifier evidence for an encoding state reliably predicted subsequent memory ($\beta = 0.0951$, SD = 0.1776, *t*34= 3.1670, $p = 0.0032$). This result indicates that the classifier generated a meaningful index of how mnemonic processing was oriented, as opposed to simply indexing whether subjects complied with the task instructions. We also tested whether evidence for an encoding state predicted subsequent memory for List 2 trials where the instruction was to encode. For these trials, the relationship between classifier evidence for an encoding state and subsequent memory was numerically positive, but did not approach significance ($\beta = 0.0240$, SD = 0.2604, t_{32}= 0.5284, $p = 0.6009$). However, this null result should be interpreted with some caution since memory for encode items was high overall, resulting in relatively few encode items in the ‘forgotten’ bin. As a complementary–and higher powered analysis–we applied the classifiers trained on List 2 data to all of the List 1 trials. The List 1 trials can effectively be thought of as trials with an ‘encode’ instruction, and there were twice as many List 1 trials as List 2 encode trials. Notably, the classifier was more likely to label List 1 trials as ‘encode’ than it was to label the List 2 trials (combining across instruction conditions) as encode (List 1: M = 54.46%, SD = 7.20%; List 2: M = 47.57%, SD = 6.04%; difference between List 1 and List 2: t_{35} = 3.7112, p = 0.0007). More importantly, stronger encoding evidence during List 1 predicted better subsequent memory for List 1 items (mean β = 0.0745, SD = 0.2111; t_{34}= 2.0875, p = 0.0444). This result, which replicates the List 2 results (Figure 2D), provides additional validation that the classifier indexed meaningful variability in memory states even when controlling for the instructions that subjects received.

One potential concern about the observed relationship between encoding evidence and subsequent memory is that the number of subsequently remembered trials was not balanced across the encode and retrieve conditions–namely, there were more subsequently remembered items in the encode condition than the retrieve condition. Although this imbalance was very small (see Figure 1B), it is possible that instead of, or in addition to, learning to discriminate memory states, the classifier learned to predict subsequent memory. To address this, we re-ran the main classification analyses such that for each iteration of classifier training (i.e., for each fold for each subject) the encode and retrieve trials in the training set contained an equal number of subsequently remembered List 2 items (as well as an equal number of subsequently forgotten List 2 items). This was accomplished by randomly dropping trials from conditions with the higher counts. This process was repeated 10 times per classification fold and classifier performance was averaged across the 10 iterations. With this balancing, the classifiers could not learn to discriminate between the encode and retrieve conditions based on subsequent memory status. Using evidence from these balanced classifiers, we replicated the key result relating List 2 encoding evidence to List 2 subsequent memory (mean β = 0.0646, SD = 0.1385, one-sample t-test vs. 0, t_{34}= 2.7601, p = 0.0092). The relationship also remained significant when just considering retrieve trials (mean β = 0.1046, SD = 0.1824, one-sample t-test vs. 0, t_{34}= 3.3945, p = 0.0018) and numerically positive, but not significant, for encode trials (mean β = 0.0340, SD = 0.2773, one-sample t-test vs. 0, t_{34}= 0.7046, p = 0.4862). We also applied these balanced classifiers, trained on List 2 data, to all of the List 1 trials. As we observed with the ‘unbalanced’ classifiers, List 1 trials were more likely to be labeled as ‘encode’ than List 2 trials (List 1: M = 53.40%, SD = 6.61%; List 2: M = 47.39%, SD = 5.85%; difference between List 1 and List 2: t_{35} = 4.0566, p = 0.0003) and stronger encoding evidence during List 1 predicted better subsequent memory for List 1 items (mean β = 0.0763, SD = 0.2191, t_{34}= 2.0598, p = 0.0471).

Finally, although our inclusion of frequencies above 40 Hz was motivated by prior evidence that subsequent memory effects in these high frequency ranges are generally similar across scalp and intracranial EEG (Long et al., 2014), one potential concern is that high frequency effects may have been influenced by eye movements and that these ‘contaminated’ EEG effects contributed to classification accuracy. To address this concern, we re-ran the main classification analyses excluding frequencies above 40 Hz (resulting in 35 frequency features available for classification). With this approach, we again observed above-chance classification performance (M = 55.88%, SD = 9.53%, t_{35} = 3.7060, p = 0.0007). Further, encoding evidence during List 2 trials reliably predicted List 2 memory (mean β = 0.0649, SD = 0.1386, t_{34} = 2.7704, p = 0.0090) and did not predict List 1 memory (mean β = −0.0103, SD = 0.1260, t_{34}= −0.4848, p = 0.6309). When List 2 trials were separated into retrieve and encode trials, encoding evidence on retrieve trials reliably predicted List 2 memory (mean $\beta = 0.1028$, SD = 0.1615, $t_{34} = 3.7673$, $p = 0.0006$) whereas encoding evidence on encode trials did not (mean $\beta = 0.0375$, SD = 0.2583, $t_{32} = 0.8332$, $p = 0.4109$). Finally, when the classifier was trained on List 2 trials and tested on List 1 trials, there was a trend toward significant prediction of List 1 memory (mean $\beta = = 0.0657$, SD = 0.2118, $t_{34} = 1.8355$, $p = 0.0752$). Thus, the observed results were virtually identical with the exclusion of high frequency activity.

Relation to univariate subsequent memory effects

An important aspect of the preceding analyses is that we predicted subsequent memory from decoded memory state evidence, as opposed to directly predicting subsequent memory from EEG activity patterns. However, many prior studies have directly compared EEG activity for subsequently remembered vs. forgotten events (Paller et al., 1987; Sederberg et al., 2006; Osipova et al., 2006; Burke et al., 2014; Long et al., 2014). These studies have consistently observed that subsequently remembered events are associated with an increase in high frequency activity (HFA_i, > 28 Hz) and a decrease in low frequency activity (LFA_d, < 28 Hz). This raises the important question of whether the memory state classifier described here exploited the same information that has previously been associated with successful encoding (i.e., the $\text{HFA}_i/\text{LFA}_d$ pattern) or whether the classifier tracked a distinct dimension of memory formation. To formally address this question, we conducted several additional analyses.

First, we sought to replicate the canonical $\text{HFA}_i/\text{LFA}_d$ pattern that has previously been associated with subsequent remembering both in intracranial and scalp EEG (Sederberg et al., 2006; Osipova et al., 2006; Burke et al., 2014; Long et al., 2014). We did this using List 1 items, which served as an independent data set. First, we calculated the difference in spectral power for items that were subsequently remembered vs. forgotten–i.e., a subsequent memory effect (SME). Qualitatively–and consistent with prior studies–subsequent memory was associated with increases in HFA and decreases in LFA (Figure 3). To simplify subsequent analyses, we first reduced the 46 frequencies into an HFA band (> 28 Hz) and an LFA band (< 28 Hz); the cutoff of 28 Hz was derived from an independent, prior study (Burke et al., 2014). We then created a functional region of interest (ROI) that was comprised of electrodes that exhibited the predicted pattern: significantly more positive SMEs in HFA than LFA. Using a threshold of $p < 0.01$, uncorrected, this resulted in five electrodes being included in the ROI (CP2, TP10, Pz, P4, Oz; Figure 3, bold electrode labels). This ROI, which was defined based only on List 1 SMEs, was specifically used to assess the (independent) HFA/LFA effects from the List 2 data.

To assess whether the $\text{HFA}_i/\text{LFA}_d$ pattern was related to encoding vs. retrieval states, we ran an ANOVA with factors of electrode (the five electrodes in the functional ROI) and frequency band (HFA vs. LFA), with the dependent variable being the contrast of List 2 encode vs. retrieve trials (Figure 4A). This ANOVA did not reveal a significant main effect of frequency band ($F_{1,35} = 0.108$, $p = 0.7438$) or an interaction between electrode and frequency band ($F_{4,140} = 1.335$, $p = 0.2600$). This was also true if all 30 electrodes were included (p's > 0.80). Thus, the pattern of HFA increases and LFA decreases that has repeatedly been associated with subsequent remembering–here and elsewhere–did not reflect whether subjects were in an encoding vs. retrieval state. As a more direct test of the relationship between the $\text{HFA}_i/\text{LFA}_d$ pattern and memory state evidence, we correlated trial-level classifier evidence for an encoding state with the trial-level difference in HFA vs. LFA (averaged across the five electrodes in the functional ROI). This analysis was separately performed for List 2 encode and retrieve trials, with each *r* value Fisher *z*-transformed and averaged, resulting in a single measure (mean *z* score) per subject. The mean *z* score was −0.0026 (SD = 0.1378), which was not different from 0 ($t_{35} = -0.114$, $p = 0.9099$), confirming that the electrophysiological pattern that signaled an encoding state was not the same as the $\text{HFA}_i/\text{LFA}_d$ pattern that has typically been associated with subsequent remembering. As a further test of whether the difference between encoding vs. retrieval states was related to the difference between subsequently remembered vs. forgotten trials, we correlated the frequency × electrode spectrogram from the List 2 encode vs. retrieve contrast with the spectrogram from the List 1 remembered vs. forgotten contrast (Figure 4B). A separate correlation was computed for each subject and resulting correlation values were compared to zero. Again, there was no evidence for a relationship between these measures (mean zRho = −0.0266, SD = 0.2047; t-test vs. 0: $t_{35} = -0.7792$, $p = 0.4411$). Finally, we trained a pattern classifier on the List 1 data to predict subsequent memory (remembered, forgotten) for List 1 items. We found modest, but reliably above chance classification accuracy (M = 52.09%, SD = 5.30%, $t_{34} = 2.3481$, $p = 0.0248$). We then applied this classifier to List 2 trials in order to test whether encode trials tended to be labeled as ‘remembered’ and retrieve trials as ‘forgotten’. Using this definition of accuracy, the classifier was not significantly above chance (M = 49.56%, SD = 3.56%, $t_{34} = -0.7297$, $p = 0.4706$), again consistent with the idea that the pattern of spectral activity associated with subsequent memory was distinct from the pattern of spectral activity that distinguished encode vs. retrieve trials.

While the $\text{HFA}_i/\text{LFA}_d$ pattern did not reflect the tradeoff between encoding vs. retrieval states, a separate question is whether this pattern predicted subsequent memory independent of task instructions. As described above, classifier evidence for an encoding state predicted subsequent memory for List 2 items even when subjects were instructed to retrieve the List 1 item. If the $\text{HFA}_i/\text{LFA}_d$ pattern reflects successful memory formation, per se, then the pattern should generalize across trials, regardless of instruction. To test this, we derived separate SMEs for List 2 encode and retrieve trials (Figure 5B-C). In considering SMEs for retrieve trials, it is important to emphasize that our task was deliberately designed such that subjects could not strategically ignore List 2 items altogether–rather, subjects needed to first encode each List 2 item before retrieving the corresponding List 1 item. Indeed, behavioral results confirmed that memory for List 2 retrieve items was well above chance ($t_{35} = 12.20$, $p < 0.0001$) and only subtly worse than memory for List 2 encode items.

Averaging the SMEs across the functional ROI, we found a significant interaction between instruction (encode, retrieve) and frequency band (HFA vs. LFA; $F_{1,32} = 5.160$, $p = 0.0300$). For encode trials, the SME was numerically more positive for HFA than LFA ($t_{32} = 1.7392$, $p = 0.0916$), and was qualitatively similar to the List 1 SME (Figure 3) and to the typical $\text{HFA}_{i,}/\text{LFA}_d$ pattern. For retrieve trials, however, the SME was numerically more positive for

LFA than HFA ($t_{32} = -1.5351$, $p = 0.1346$), qualitatively *opposite* to the typical HFA$_i$,/LFA$_d$ pattern.

Why might the relationship between the HF$_i$/LFA$_d$ pattern and subsequent memory qualitatively flip when the instruction was to retrieve? One possibility is that instead of reflecting memory formation, per se, or a tradeoff between encoding vs. retrieval states, the HFA$_i$,/LFA$_d$ pattern reflects the degree to which subjects are 'on task.' In the vast majority of subsequent memory studies, memory encoding (or processing that supports memory encoding) is task relevant, whereas memory retrieval would be task irrelevant or 'off task.' Thus, due to the nature of most subsequent memory paradigms, it is hard to know whether the HFA$_i$/LFA$_d$ pattern reflects memory formation or task engagement (which, in turn, leads to better subsequent memory). Notably, the HFA$_i$/LFA$_d$ pattern has also been observed across other cognitive tasks and domains (Crone, Miglioretti, Gordon, Sieracki, et al., 1998; Crone, Miglioretti, Gordon, & Lesser, 1998; Crone, Boatman, Gordon, & Hao, 2001; Miller et al., 2007; Jensen, Kaiser, & Lachaux, 2007), suggesting that this pattern is not memory specific.

Importantly, a task engagement account of the HFA$_i$/LFA$_d$ pattern makes a testable prediction: on List 2 retrieve trials, the HFA$_i$/LFA$_d$ pattern should reflect *successful retrieval* of corresponding *List 1 items* and, therefore, should predict better subsequent memory for List 1 items. Indeed, HFA increases and LFA decreases have previously been associated with successful retrieval when retrieval is task relevant (Hanslmayr, Staudigl, & Fellner, 2012). To test this prediction, we again measured spectral activity during List 2 retrieve trials (using the functional ROI), but now separated these List 2 EEG responses according to subsequent memory for corresponding *List 1 items* (Figure 5D). Strikingly, better subsequent memory for to-be-retrieved List 1 items was associated with a clear HFA$_i$/LFA$_d$ pattern (t_{26}= 2.3741, $p = 0.0253$)–precisely the pattern that is typically associated with successful encoding. Thus, when subjects were instructed to retrieve a similar item from the past, the HFA$_i$/LFA$_d$ pattern predicted relatively worse memory for the new item (List 2 item) but better memory for the old item (List 1 item). Collectively, these results strongly argue against the idea that the HFA$_i$/LFA$_\mathrm{d}$ pattern reflects memory formation, per se–or even an optimal state for memory encoding–and instead argue that this pattern reflects some aspect of task engagement, which has generally been confounded with processing that leads to memory formation.

Discussion

Here we show that biases toward memory encoding vs. retrieval states can be decoded from spectral EEG patterns and, critically, these decoded biases predict how overlapping events will later be remembered. These findings are consistent with computational models emphasizing encoding/retrieval tradeoffs during mnemonic processing (O'Reilly & McClelland, 1994; Hasselmo, 2005) and build on behavioral evidence of encoding/retrieval tradeoffs when new events overlap with past events (Duncan et al., 2012). Surprisingly, however, the spectral EEG patterns that reflected these tradeoffs between encoding and retrieval states were unrelated to canonical spectral EEG patterns that have previously been associated with successful memory encoding. Collectively, these findings provide important insight into the neural mechanisms that determine how overlapping events are subsequently remembered.

Before considering our EEG findings, it is important to emphasize several features of our behavioral task and results. By design, our critical manipulation between encode and retrieve trials was subtle, reflecting an effort to minimize perceptual, motor, or decision-related differences between these trials that might inflate classifier performance. Of particular importance, subjects were required to attend to each List 2 object image because this image either functioned as the to-be-encoded item (encode trials) or as the retrieval cue for the to-be-retrieved item (retrieve trials). Critically, subsequent memory was significantly better for List 2 items in the encode condition than the retrieve condition–confirming that subjects successfully modulated their cognitive processing according to instructions. However, this difference was modest in magnitude and List 2 items from both conditions were generally well remembered indicating that subjects did not simply ignore List 2 items in the retrieve condition.

In contrast to memory for List 2 items, memory for List 1 items was not significantly influenced by List 2 instructions. In particular, List 2 retrieve instructions did not benefit memory for corresponding List 1 items. This null result raises a potential concern that retrieve instructions did not, in fact, successfully drive subjects toward a retrieval state. However, we were not altogether surprised by this null result given the nature of the recognition test. We used a forced-choice recognition test that required subjects to discriminate between highly similar exemplars. We used this challenging recognition test in order to avoid ceiling effects in recognition accuracy, with the tradeoff being that this particular test was likely not optimal for revealing retrieval-based strengthening. Namely, retrieval-based strengthening may preferentially occur at the level of generic category-level information, as opposed to strengthening of the idiosyncratic details that would benefit performance on the forced-choice recognition test. Indeed, this is precisely what we observed in another recent study (Lee, Samide, Richter, & Kuhl, 2018). Moreover, it is worth emphasizing that, for retrieve trials, the success rate of retrieval was likely to be variable. Indeed, when considering the List 2 retrieve trials alone, we found that HFA increases and LFA decreases predicted whether corresponding List 1 items would later be remembered or not (Figure 5D). This result suggests that subjects did actively engage in retrieval on retrieve trials, but that retrieval success was variable. This result highlights an important distinction between a retrieval state and successful retrieval: successful adoption of a retrieval state is not the same thing as successful retrieval (Lepage, Ghaffar, Nyberg, & Tulving, 2000; Richter et al., 2016). Related to this point, it is worth noting that a retrieval state can subsume multiple different kinds of processing that depend on specific task demands–so called ‘retrieval orientations’ (Rugg & Wilding, 2000). Here, while we cannot tease apart the components of the retrieval state that are specific to the particular retrieval task we used, we believe–for the reasons described above–that the retrieve trials successfully induced a retrieval state.

In order to classify encoding vs. retrieval states, we used subject-specific pattern classifiers that could exploit idiosyncratic (subject unique) information contained within distributed patterns of spectral power across electrodes. As such, it was not our goal to draw conclusions about how specific frequencies and/or electrodes relate to encoding vs. retrieval states. That said, our use of spectral information as a feature dimension was motivated by prior studies relating encoding and retrieval to distinct spectral signals (Rizzuto et al., 2006; Hasselmo & Stern, 2014). In particular, there has been interest in the relationship between memory states and theta. While much of this interest has focused on theta phase established by the hippocampus (Hasselmo et al., 2002; Siegle & Wilson, 2014; Kerrén, Linde-Domingo, Hanslmayr, & Wimber, 2018), there is also considerable evidence, both from human and non-human animals, that sustained increases in theta power are related to successful learning or memory encoding (Berry & Seager, 2001; Seager, Johnson, Chabot, Asaka, & Berry, 2002; Guderian, Schott, Richardson-Klavehn, & Duzel, 2009; Kirov et al., 2009; Molter et al., 2012). Qualitatively, our results are consistent with these prior studies in that we observed relative increases in theta power—centered at approximately 8 Hz—when comparing encode vs. retrieve trials. However, these differences were very modest in magnitude and, on their own, do not support conclusions about the relative contributions of different frequencies. Moreover, it should be emphasized that, here, we recorded scalp EEG whereas much of the existing literature relating theta to memory states has focused on *hippocampal theta*. While beyond the scope of the present study, how or whether scalp EEG (including decoded information about memory states) relates to hippocampal theta is an interesting question for future research.

Regardless of the specific spectral components that were associated with encoding vs. retrieval states, our findings suggest that biases between these states can operate over relatively long timescales (seconds). This result is consistent with prior evidence that encoding vs. retrieval states can be decoded from temporally-coarse fMRI activity patterns (Richter et al., 2016) and also with behavioral evidence that biases between encoding vs. retrieval states can linger across trials (Duncan et al., 2012; Patil & Duncan, 2018) and with the timescale at which neuromodulatory influences on the hippocampus are thought to occur (Meeter et al., 2004; Hasselmo & McGaughy, 2004). Thus, considering the current findings in relation to prior evidence for encoding/retrieval tradeoffs, it seems likely that these tradeoffs can occur at multiple timescales and that different timescales for these biases may reflect distinct neural mechanisms (Honey, Newman, & Schapiro, 2017).

Highly consistent with prior scalp and intracranial EEG studies (Osipova et al., 2006; Sederberg et al., 2006; Long et al., 2014; Greenberg et al., 2015), we found that subsequent memory was predicted by a pattern of HFA increases and LFA decreases, but with the important caveat that this pattern was only evident when encoding was goal relevant. Namely, when subjects were trying to encode a new stimulus, this pattern predicted subsequent memory for the new stimulus; but, when subjects were trying to retrieve a past stimulus, this same pattern predicted subsequent memory for the old (retrieved) stimulus. This novel dissociation clearly indicates that the $\text{HFA}_i/\text{LFA}_d$ pattern is better explained by some dimension of task engagement–which may involve encoding a new stimulus or retrieving an old stimulus–than by memory formation, per se. Additionally, the $\text{HFA}_i/\text{LFA}_d$ pattern was not evident when comparing encode vs. retrieve trials, indicating that this pattern does not reflect an encoding/retrieval dimension. Thus, our results replicate prior reports linking the $\text{HFA}_i/\text{LFA}_d$ pattern to subsequent remembering, but provide critical insight into what this pattern actually reflects. Importantly, the significance of our findings are not limited to scalp EEG studies, as the HFA_i/LFA_d pattern is thought to be closely related to both intracranial EEG and fMRI BOLD responses that predict subsequent remembering (Kim, 2011; Burke et al., 2014; Long et al., 2014).

Ultimately, we show that both the HFA_i/LFA_d pattern and decoded evidence for an encoding state have utility for predicting subsequent memory. But which approach to predicting subsequent memory is more generalizable? We found that decoded evidence for an encoding state predicted subsequent memory not only in a ‘standard’ encoding task (List 1 trials) but also when subjects were specifically instructed to retrieve past events (List 2 retrieve trials). This generalization across task demands provides critical evidence that the relationship between encoding evidence and subsequent memory was not dependent on an explicit encoding task. This finding is important when considering that, in the real world, new memories are typically formed in the absence of explicit encoding demands. Moreover, real world memories are often formed as attention alternates between external stimuli and internal thoughts or memories. In fact, the specific situation we consider here–where a new event is similar to a past event–is more likely the norm than the exception in the real world. As such, successful memory formation in real world contexts is potentially well explained by fluctuations between encoding vs. retrieval states. In contrast, despite the ubiquity of the HFA_i/LFA_d pattern in lab-based subsequent memory studies, it may be of limited relevance to explaining memory encoding in the real world to the extent that its expression is strongly task dependent.

In summary, our findings reveal that encoding/retrieval tradeoffs described by computational models (O’Reilly & McClelland, 1994) are reflected in–and can be decoded from–distributed patterns of neural activity. Critically, these tradeoffs determine whether overlapping events will later be remembered. These findings are highly relevant to the broad literature related to neural factors that promote successful memory formation and can potentially inform efforts to use neurofeedback or neurostimulation to ‘boost’ memory encoding (Ezzyat et al., 2017) or to bias how overlapping events are remembered. More generally, our findings are relevant to understanding interactions and relationships between attention and memory (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008) in that tradeoffs between encoding vs. retrieval states potentially reflect a broader tradeoff between externally-oriented attention (allowing for memory encoding) vs. internally-oriented attention (directed toward thoughts or memories; Chun & Johnson, 2011; Honey et al., 2017).

Acknowledgments

This work was supported by a grant from the National Institutes of Health (NINDS, 1R01NS089729 - PI: Brice Kuhl). We thank Sarah Sweigart for assistance with data collection.

References

Bakker A, Kirwan CB, Miller M, & Stark CEL (2008). Pattern separation in the human hippocampal ca3 and dentate gyrus. Science, 319(5870), 1640–1642. [PubMed: 18356518]

Bell AJ, & Sejnowski TJ (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129–1159. [PubMed: 7584893]

Berry SD, & Seager MA (2001). Hippocampal theta oscillations and classical conditioning. Neurobiology of Learning and Memory, 76(3), 298–313. [PubMed: 11726239]

Burke JF, Long NM, Zaghloul KA, Sharan AD, Sperling MR, & Kahana MJ (2014). Human intracranial high-frequency activity maps episodic memory formation in space and time. NeuroImage, 85 Pt. 2, 834–843. [PubMed: 23827329]

Cabeza R, Ciaramelli E, Olson IR, & Moscovitch M (2008). The parietal cortex and episodic memory: an attentional account. Nature Reviews Neuroscience, 9(8), 613–625. [PubMed: 18641668]

Chun MM, & Johnson MK (2011). Memory: enduring traces of perceptual and reflective attention. Neuron, 72(4), 520–535. [PubMed: 22099456]

Colgin L, Denninger T, Fyhn M, Hafting T, Bonnevie T, Jensen O, ... Moser E (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. Nature, 462(7271), 353–357. [PubMed: 19924214]

Crone NE, Boatman D, Gordon B, & Hao L (2001). Induced electrocorticographic gamma activity during auditory perception. Clinical Neurophysiology, 112(4), 565–582. [PubMed: 11275528]

Crone NE, Miglioretti DL, Gordon B, & Lesser RP (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain, 121(12), 2301–2315. [PubMed: 9874481]

Crone NE, Miglioretti DL, Gordon B, Sieracki JM, Wilson MT, Uematsu S, & Lesser RP (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. Brain, 121, 2271–2299. [PubMed: 9874480]

Delorme A, & Makeig S (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. Journal of Neuroscience Methods, 134, 9–21. [PubMed: 15102499]

Duncan K, Sadanand A, & Davachi L (2012). Memory’s penumbra: Episodic memory decisions induce lingering mnemonic biases. Science, 337(6093), 485–487. [PubMed: 22837528]

Ezzyat Y, Kragel JE, Burke JF, Levy DF, Lyalenko A, Wanda P, ... others (2017). Direct brain stimulation modulates encoding states and memory performance in humans. Current Biology, 27(9), 1251–1258. [PubMed: 28434860]

Fan RE, Chang KW, Hsieh CJ, Wang XR, & Lin CJ (2008). Liblinear: A library for large linear classification. The Journal of Machine Learning Research, 9, 1871–1874.

Fernandez G, Effern A, Grunwald T, Pezer N, Lehnertz K, Dumpelmann M, ... Elger CE (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. Science, 285, 1582–1585. [PubMed: 10477525]

Friedman D, & Johnson R (2000). Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. Microscopy Research and Technique, 51, 6–28. [PubMed: 11002349]

Greenberg JA, Burke JF, Haque R, Kahana MJ, & Zaghloul KA (2015). Decreases in theta and increases in high frequency activity underlie associative memory encoding. Neuroimage, 114, 257–263. [PubMed: 25862266]

Griffin A, Eichenbaum H, & Hasselmo M (2007). Spatial representations of hippocampal CA1 neurons are modulated by behavioral context in a hippocampus-dependent memory task. The Journal of Neuroscience, 27(9), 2416–2423. [PubMed: 17329440]

Gruber T, Tsivilis D, Montaldi D, & Müller M (2004). Induced gamma band responses: An early marker of memory encoding and retrieval. Neuroreport, 15, 1837–1841. [PubMed: 15257158]

Guderian S, Schott B, Richardson-Klavehn A, & Duzel E (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. Proceedings of the National Academy of Sciences of the United States of America, 106(13), 5365. [PubMed: 19289818]

Hanslmayr S, Staudigl T, & Fellner M (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. Frontiers in Human Neuro-science, 6 (74).

Hasselmo ME (2005). What is the function of hippocampal theta rhythm? Linking behavioral data to phasic properties of field potential and unit recording data. Hippocampus, 15, 936–949. [PubMed: 16158423]

Hasselmo ME, Bodelon C, & Wyble BP (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. Neural Computation, 14, 793–817. [PubMed: 11936962]

Hasselmo ME, & McGaughy J (2004). High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation In Acetylcholine in the cerebral cortex (Vol. 145, pp. 207 – 231). Elsevier.

Hasselmo ME, & Stern CE (2014). Theta rhythm and the encoding and retrieval of space and time. NeuroImage, 85, 656–666. [PubMed: 23774394]

Honey CJ, Newman EL, & Schapiro AC (2017). Switching between internal and external modes: A multiscale learning principle. Network Neuroscience, 1(4), 339–356. [PubMed: 30090870]

Huijbers W, Pennartz CM, Cabeza R, & Daselaar SM (2009). When learning and remembering compete: a functional mri study. PLoS Biol, 7(1), e1000011.

Jensen O, Kaiser J, & Lachaux J (2007). Human gamma-frequency oscillations associated with attention and memory. Trends in Neurosciences, 30(7), 317–324. [PubMed: 17499860]

Kerrén C, Linde-Domingo J, Hanslmayr S, & Wimber M (2018). An optimal oscillatory phase for pattern reactivation during memory retrieval. Current Biology.

Kim H (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. NeuroImage, 54(3), 2446–2461. [PubMed: 20869446]

Kirov R, Weiss C, Siebner H, Born J, & Marshall L (2009). Slow oscillation electrical brain stimulation during waking promotes EEG theta activity and memory encoding. Proceedings of the National Academy of Sciences of the United States of America, 106(36), 15460. [PubMed: 19706399]

Konkle T, Brady TF, Alvarez GA, & Oliva A (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology: General, 139(3), 558. [PubMed: 20677899]

Kuhl BA, Shah AT, DuBrow S, & Wagner AD (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nature Neuroscience, 13(4), 501–506. [PubMed: 20190745]

Lee H, Samide R, Richter FR, & Kuhl BA (2018). Decomposing parietal memory reactivation to predict consequences of remembering. Cerebral Cortex, 1–14. [PubMed: 29253248]

Lepage M, Ghaffar O, Nyberg L, & Tulving E (2000). Prefrontal cortex and episodic memory retrieval mode. Proceedings of the National Academy of Sciences of the United States of America, 97(1), 506–511. [PubMed: 10618448]

Lepage M, Habib R, & Tulving E (1998). Hippocampal pet activations of memory encoding and retrieval: The hiper model. Hippocampus, 8, 313–322. [PubMed: 9744418]

Long NM, Burke JF, & Kahana MJ (2014). Subsequent memory effect in intracranial and scalp EEG. NeuroImage, 84, 488–494. [PubMed: 24012858]

Long NM, & Kahana MJ (2015). Successful memory formation is driven by contextual encoding in the core memory network. NeuroImage, 119, 332–337. [PubMed: 26143209]

Manns JR, Zilli EA, Ong KC, Hasselmo ME, & Eichenbaum H (2007). Hippocampal CA1 spiking during encoding and retrieval: Relation to theta phase. Neurobiology of learning and memory, 87(1), 9–20. [PubMed: 16839788]

Meeter M, Murre JMJ, & Talamini LM (2004). Mode shifting between storage and recall based on novelty detection in oscillating hippocampal circuits. Hippocampus, 14(6), 722–741. [PubMed: 15318331]

Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, ... Ojemann JG (2007). Spectral changes in cortical surface potentials during motor movement. The Journal of Neuroscience, 27, 2424–2432. [PubMed: 17329441]

Molter C, O'Neill J, Yamaguchi Y, Hirase H, & Leinekugel X (2012). Rhythmic modulation of theta oscillations supports encoding of spatial and behavioral information in the rat hippocampus. Neuron, 75(5), 889–903. [PubMed: 22958828]

Nolan H, Whelan R, & Reilly R (2010). Faster: fully automated statistical thresholding for eeg artifact rejection. Journal of neuroscience methods, 192(1), 152–162. [PubMed: 20654646]

Onton J, & Makeig S (2006). Information-based modeling of event-related brain dynamics. Progress in brain research, 159, 99–120. [PubMed: 17071226]

O'Reilly RC, & McClelland JL (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus, 4(6), 661–682. [PubMed: 7704110]

Osipova D, Takashima A, Oostenveld R, Fernndez G, Maris E, & Jensen O (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. The Journal of Neuroscience, 26(28), 7523–7531. [PubMed: 16837600]

Otten LJ, & Rugg MD (2001). Electrophysiological correlates of memory encoding are task-dependent. Cognitive Brain Research, 12(1), 11–18. [PubMed: 11489604]

Paller KA, Kutas M, & Mayes AR (1987). Neural correlates of encoding in an incidental learning paradigm. Electroencephalography and Clinical Neurophysiology, 67, 360–371. [PubMed: 2441971]

Patil A, & Duncan K (2018). Lingering cognitive states shape fundamental mnemonic abilities. Psychological Science, 29(1), 45–55. [PubMed: 29116882]

Richter FR, Chanales AJ, & Kuhl BA (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. NeuroImage, 124, 323–335. [PubMed: 26327243]

Rizzuto D, Madsen JR, Bromfield EB, Schulze-Bonhage A, & Kahana MJ (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. NeuroImage, 31(3), 1352–1358. [PubMed: 16542856]

Rugg MD, & Wilding EL (2000). Retrieval processing and episodic memory. Trends in Cognitive Sciences, 4(3), 108–115. [PubMed: 10689345]

Seager MA, Johnson LD, Chabot ES, Asaka Y, & Berry SD (2002). Oscillatory brain states and learning: Impact of hippocampal theta-contingent training. Proceedings of the National Academy of Sciences of the United States of America, 99, 1616–20. [PubMed: 11818559]

Sederberg PB, Gauthier LV, Terushkin V, Miller JF, Barnathan JA, & Kahana MJ (2006). Oscillatory correlates of the primacy effect in episodic memory. Neuroimage, 32(3), 1422–1431. [PubMed: 16814568]

Sederberg PB, Kahana MJ, Howard MW, Donner EJ, & Madsen JR (2003). Theta and gamma oscillations during encoding predict subsequent recall. The Journal of Neuroscience, 23(34), 10809–10814. [PubMed: 14645473]

Siegle JH, & Wilson MA (2014). Enhancement of encoding and retrieval functions through theta phase-specific manipulation of hippocampus. Elife, 3, e03061. [PubMed: 25073927]

Tulving E, Kapur S, Markowitsch HJ, Craik FIM, Habib R, & Houle S (1994). Neuroanatomical correlates of retrieval in episodic memory: auditory sentence recognition. Proceedings of the National Academy of Sciences of the United States of America, 91, 2012–2015. [PubMed: 8134341]

Zeithamova D, Dominick AL, & Preston AR (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron, 75(1), 168–179. [PubMed: 22794270]

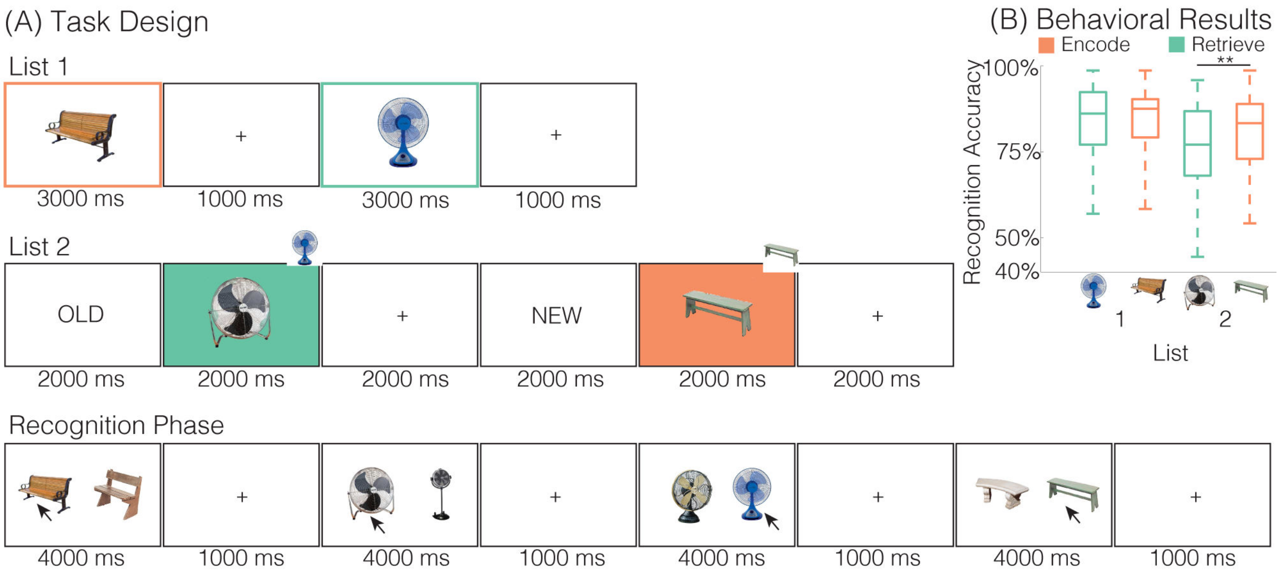

Figure 1. Task Design and Behavioral Results.

(A) During List 1, subjects studied individual objects (e.g. bench, fan). During List 2, subjects saw novel objects that were from the same categories as the items shown in List 1 (e.g., a new bench, a new fan). Preceding each List 2 object was an "OLD" instruction cue or "NEW" instruction cue. The "OLD" cue signaled that subjects were to *retrieve* the corresponding item from List 1 (e.g., the old fan). The "NEW" cue signaled that subjects were to *encode* the current item (e.g., the new bench). Colored boxes are shown here for illustrative purposes and were not present during the actual experiment. Each run of the experiment contained a List 1 and List 2; object categories (e.g., bench) were not repeated across runs. After eight runs, subjects completed a two alternative force choice recognition test that tested memory for each List 1 and List 2 object. On each trial, a previously presented object, *either* from List 1 or List 2, was shown alongside a novel lure from the same category. The subject's task was to choose the previously presented object. List 1 and List 2 objects were never presented together. **(B)** Behavioral results. Recognition accuracy is shown separated by list (1,2) and instruction condition (encode, orange; retrieve, teal). There was a significant interaction between list and instruction, primarily driven by greater accuracy for List 2 items presented with an encode instruction relative to a retrieve instruction. Error bars denote SEM; ** $p < 0.01$.

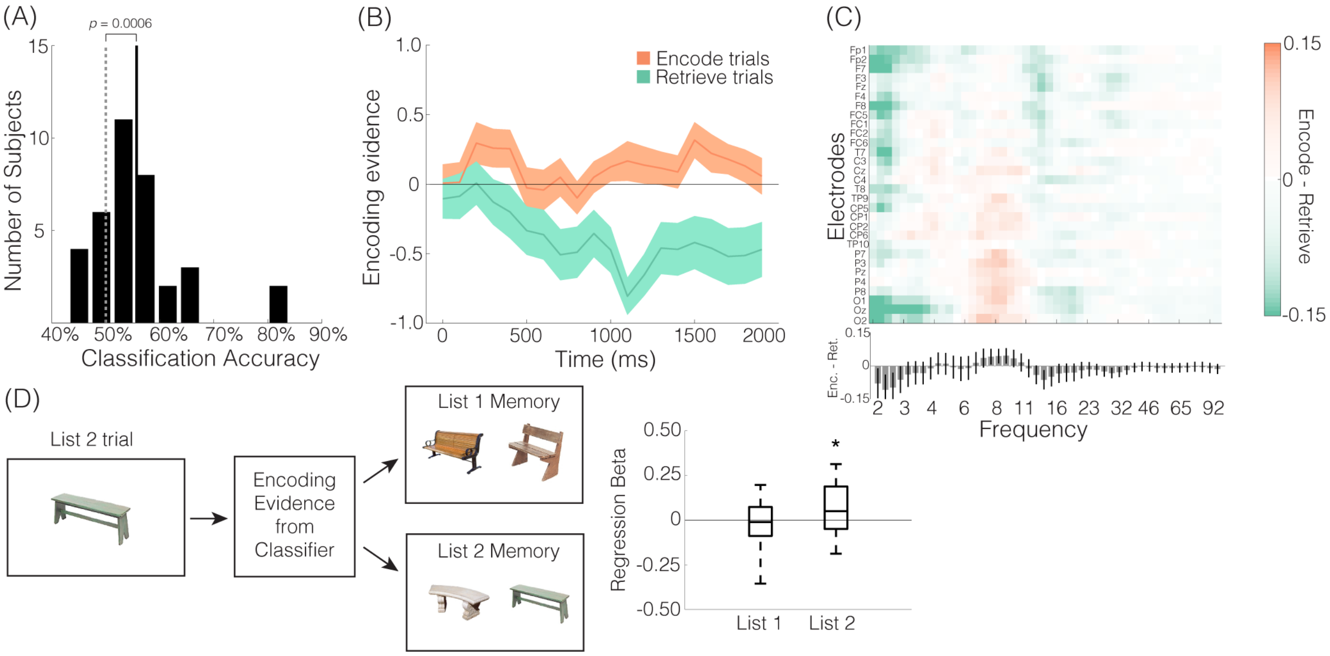

Figure 2. Decoding memory states.

(A) We trained subject-specific L2-logistic regression classifiers to discriminate encode vs. retrieve trials during List 2. The classifiers were trained and tested on average spectral power across the 0-2000 ms stimulus interval with all electrodes and frequencies used as features. Mean classification accuracy across all subjects (solid vertical line) is shown along with a histogram of mean classification accuracies for individual subjects (black bars) and mean classification accuracy for permuted data across all subjects (dashed vertical line). Mean classification accuracy for permuted data ranged from 49.79% to to 50.31% across individual subjects (1000 permutations per subject). **(B)** Time-course of encoding evidence across the 2000 ms stimulus interval (i.e., the time window when the object image was on screen). Here, the classifier was trained on the full 2000 ms interval, as described in (A), but tested on 100 ms time windows. **(C)** Mean spectrogram of differences in spectral power for encode vs. retrieve trials as a function of electrode (y-axis) and frequency (x-axis). Orange indicates greater power for encode trials, teal indicates greater power for retrieve trials. Spectrograms were generated for each subject and then averaged across subjects. Bar graph below the spectrogram illustrates the mean spectral difference, averaging across electrodes and then across subjects, between encode vs. retrieve trials at each frequency. Error bars denote SEM. **(D)** Subject-specific logistic regression analyses tested whether trial-level encoding evidence derived from the classifiers during List 2 predicted accuracy on the subsequent recognition memory test. Separate regressions were performed to predict memory for List 1 items and List 2 items. Box and whisker plots show a positive relationship between encoding evidence during List 2 trials and subsequent memory for List 2 items but no relationship between encoding evidence during List 2 trials and subsequent memory for List 1 items. * $p < 0.05$, ** $p < 0.01$

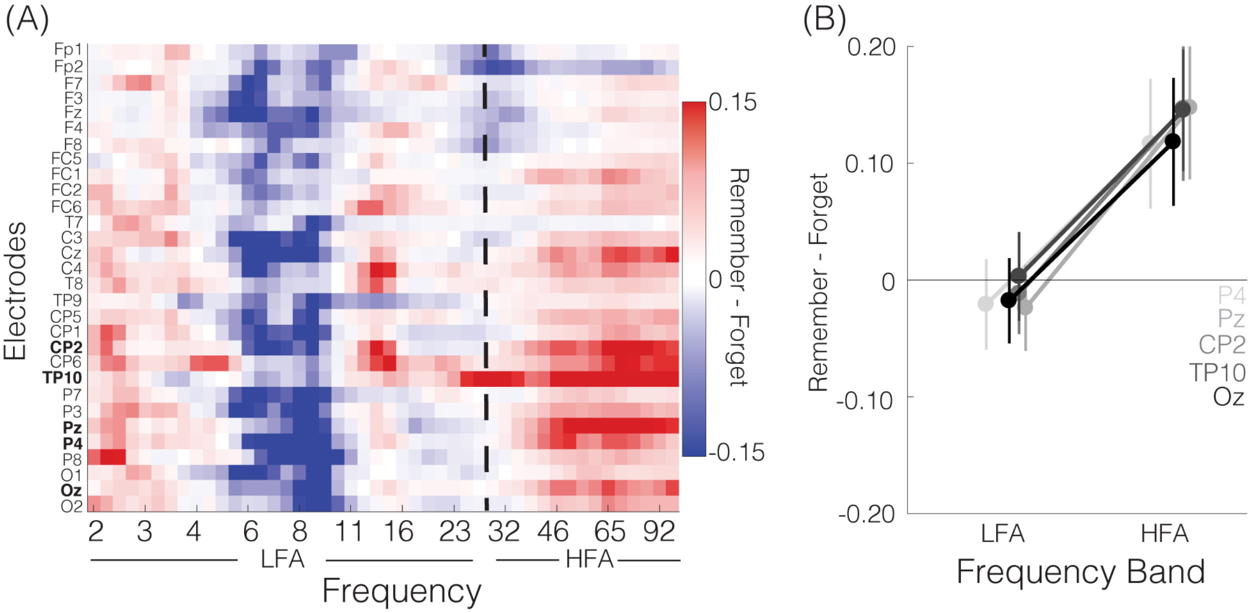

Figure 3. List 1 Univariate Subsequent Memory Effects.

(A) Mean spectrogram shows differences in spectral power for remembered vs. forgotten List 1 objects as a function of electrode (y-axis) and frequency (x-axis). Red indicates greater power for subsequently remembered items, blue indicates greater power for subsequently forgotten items. Spectrograms were generated for each subject and then averaged across subjects. Electrode names in bold text are the five electrodes that exhibited a reliable effect of frequency band (HFA vs. LFA; $p < 0.01$). These electrodes served as a functional region of interest (ROI) for subsequent analyses. **(B)** Subsequent remembering was associated with decreases in low frequency activity (LFA, < 28 Hz) and increases in high frequency activity (HFA, > 28 Hz), consistent with previous findings. Error bars denote SEM.

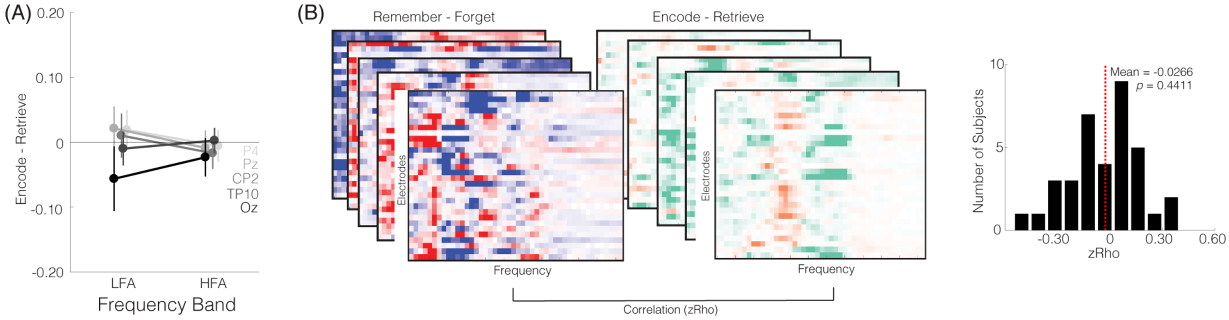

Figure 4. List 2 encode/retrieve and List 1 SME comparison.

(A) The difference in spectral power between List 2 Encode and List 2 Retrieve trials in the functional ROI, separately for HFA and LFA bands. Error bars denote SEM. **(B)** Correlation between List 1 SME and List 2 encode/retrieve contrast. For each subject, we correlated the instruction contrast (encode - retrieve) and the subsequent memory contrast (remember - forget) at each electrode and frequency. The left and middle spectrograms illustrate this procedure. The right panel shows a histogram of zRho values across subjects. The average zRho value did not reliably differ from zero ($t_{35} = -0.7792$, $p = 0.4411$).

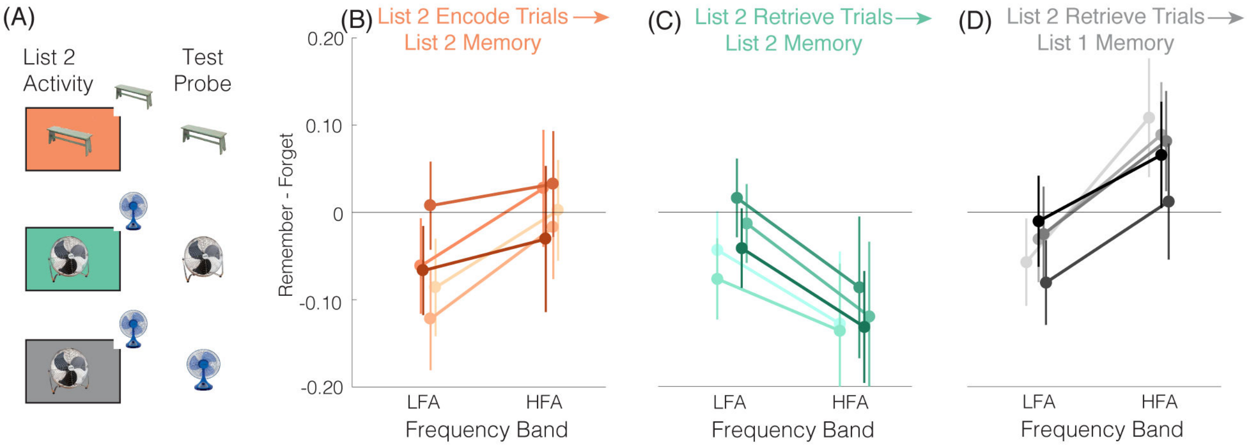

Figure 5. List 2 Univariate Subsequent Memory Effects.

Subsequent memory effects for the functional ROI from Figure 3. Each title describes the condition from which the EEG data were drawn (List 2 encode trials or List 2 retrieve trials) and the items from the recognition test that are included in the subsequent memory analysis (List 1 or List 2 items; note: the schematic shown in (A) also illustrates these relationships). For each plot in (B-D), each line reflects data from one of the five electrodes from the functional ROI. Subsequent memory effects for List 2 items significantly differed for encode vs. retrieve trials (panel C compared to panel D; $p = 0.0300$). Namely, when the goal was to encode (C), subsequent memory was predicted by relative HFA increases and LFA decreases, qualitatively identical to the pattern for List 1 items shown in (A). However, when the goal was to retrieve (D), a qualitatively opposite pattern was observed, with relative *decreases* in HFA and *increases* in LFA. Strikingly, for List 2 retrieve trials, HFA increases and LFA decreases predicted subsequent memory for to-be-retrieved List 1 *items* (E), similar to the canonical subsequent memory pattern as shown in (A). Thus, on retrieve trials, HFA increases and LFA decreases predicted relatively *worse* memory for the new List 2 item, but relatively *better* memory for the old List 1 item. Error bars denote SEM.