



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Journal of Experimental Psychology: General

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa40773>

Paper:

Weidemann, C. (in press). Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *Journal of Experimental Psychology: General*

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

Neural activity reveals interactions between episodic and semantic memory systems during retrieval

Christoph T. Weidemann^{1,2}, James E. Kragel², Bradley C. Lega³, Gregory A. Worrell⁴, Michael R. Sperling⁵, Ashwini D. Sharan⁵, Barbara C. Jobst⁶, Fatemeh Khadjevand⁴, Kathryn A. Davis⁷, Paul A. Wanda², Allison Kadel², Daniel S. Rizzuto², and Michael J. Kahana²

¹Swansea University, Department of Psychology, Swansea, Wales, UK

²University of Pennsylvania, Department of Psychology, Philadelphia, PA, USA

³University of Texas Southwestern, Department of Neurosurgery, Dallas, TX, USA

⁴Mayo Clinic, Department of Neurology, Rochester, MN, USA

⁵Thomas Jefferson University Hospital, Department of Neurology, Philadelphia, PA, USA

⁶Dartmouth Medical Center, Department of Neurology, Lebanon, NH, USA

⁷Hospital of the University of Pennsylvania, Department of Neurology, Philadelphia, PA, USA

Manuscript in press at *Journal of Experimental Psychology: General*

Abstract

Whereas numerous findings support a distinction between episodic and semantic memory, it is now widely acknowledged that these two forms of memory interact during both encoding and retrieval. The precise nature of this interaction, however, remains poorly understood. To examine the role of semantic organization during episodic encoding and retrieval, we recorded intracranial encephalographic signals as 69 neurosurgical patients studied and subsequently recalled categorized and unrelated word lists. Applying multivariate classifiers to neural recordings, we were able to reliably predict encoding success, retrieval success, and temporal and categorical clustering during recall. By assessing how these classifiers generalized across list types, we identified specific retrieval processes that predicted recall of categorized lists and distinguished between recall transitions within and between category clusters. These results particularly implicate retrieval (rather than encoding) processes in the categorical organization of episodic memories.

Keywords: free recall, episodic memory, semantic memory, intracranial EEG, machine learning

The ability to remember events and facts depends on separate episodic and semantic memory systems respectively (Tulving, 1972; Schacter & Tulving, 1994; Tulving, 2002). Neuroimaging (Poldrack et al., 2001; Kapur et al., 1994) and neuropsychological (Rogers et al., 2004; Patterson, Nestor, & Rogers, 2007) evidence implicates distinct neural bases for episodic and semantic memory: Damage to the hippocampal formation selectively impairs episodic memory (Vargha-Khadem et al., 1997), whereas injury to the anterior temporal lobe results in semantic memory deficits (Simons, Graham, Galton, Patterson, & Hodges, 2001).

Despite clear differences between episodic and semantic memory systems, interactions between them are ubiquitous: The semantic structure of remembered items affects performance in episodic memory tasks, such as free recall, where recall of semantically similar items tends to occur in clusters (Bower, Clark, Lesgold, & Winzenz, 1969; Bousfield,

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). 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. CTW & JEK contributed equally to this work. CTW & JEK analyzed the data and wrote the paper; CTW, JEK, & MJK designed analyses and edited the paper; BCL, GAW, MRS, ADS, BCJ, & KAD recruited participants and provided general assistance; FK, PAW, & AK collected data; DSR & MJK designed experiments. We thank Youssef Ezzyat for insightful discussions. The authors declare that they have no competing financial interests. Part of this work was presented at the 2018 Context and Episodic Memory Symposium in Philadelphia, PA. All of the de-identified raw data and analysis code used in this study may be freely downloaded from the Cognitive Electrophysiology Data Portal (http://memory.psych.upenn.edu/Electrophysiological_Data). Correspondence concerning this article should be addressed to Christoph T. Weidemann, Swansea University, Department of Psychology, Singleton Park, Swansea SA2 8PP, e-mail: ctw@cogsci.info.

1953; Howard & Kahana, 2002). A striking demonstration of semantic structure affecting episodic memory is the finding that for lists of words that are strongly related (e.g., “pillow”, “night”, “moon”, “bed”), a missing associate (e.g., “sleep”) is often confidently mis-remembered as having been studied (H. L. Roediger III & McDermott, 1995). Likewise, in tasks asking participants to freely recall all studied items in any order (i.e., free recall), categorically structured study lists lead to increased performance and categorical organization of recall sequences (Bousfield, 1953; Bower et al., 1969).

Many studies have attempted to attribute such effects of semantic structure of the memoranda to either encoding or retrieval processes. Given that measures of memory performance reflect the joint effects of encoding and retrieval, any such distinction on the basis of these measures relies on strong assumptions about experimental manipulations or participant characteristics. Typical approaches include attempts to selectively disrupt or enhance either encoding or retrieval (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Fernandes & Grady, 2008) as well as examinations of neuropsychological populations with well-defined memory deficits (Gershberg & Shimamura, 1995; Greenberg, Keane, Ryan, & Verfaellie, 2009). To the extent that experimental manipulations or neuropsychological deficits are not selective, however, associated conclusions are open to alternative interpretations. Indeed, there is little consensus across studies about the relative contributions of encoding and retrieval processes in the interplay between episodic and semantic memory (see Greenberg & Verfaellie, 2010, for a review).

Here we pursue a novel route to link specific interactions between semantic and episodic memory to encoding and retrieval processes. Rather than relying solely on memory performance, we leverage ongoing neural activity during study and recall phases of a free-recall task to identify such interactions. In this effort, we build upon prior work that has characterized neural biomarkers of encoding and retrieval: Structures in prefrontal, temporal, and posterior parietal cortices respond differentially during the encoding of subsequently remembered and forgotten items (Wagner et al., 1998; Uncapher & Wagner, 2009)—a subsequent memory effect (Paller & Wagner, 2002). Direct measures of neural activity obtained from intracranial electroencephalography (iEEG) have implicated the same neural structures in memory processing, exhibiting increases in high frequency (i.e., > 70 Hz) power with concomitant decreases in power at lower frequencies during successful encoding (Burke, Long, et al., 2014; Long, Burke, & Kahana, 2014). Similar neural signatures during the recall period of free recall tasks predict successful retrieval (Kragel et al., 2017; Burke, Sharan, et al., 2014).

We also take advantage of the statistical power of multivariate (“machine learning”) classifiers to quantify the neural signals during encoding and retrieval periods that are predictive of different aspects of memory performance or recall organization (Norman, Polyn, Detre, & Haxby, 2006). Specifically, we compare performance of such classifiers trained on features from iEEG recordings while participants studied and recalled lists of semantically categorized or unrelated words during a free-recall task. This approach allowed us to directly compare neural signals during encoding and retrieval phases that predict successful memory, as well as categorical and temporal clustering of recall sequences (which respectively index effects of semantic and episodic memory). In addition to highlighting encoding and retrieval effects that generalize across memoranda with varying semantic structure, our findings particularly implicate retrieval processes as drivers of the interactions between

semantic and episodic memory systems.

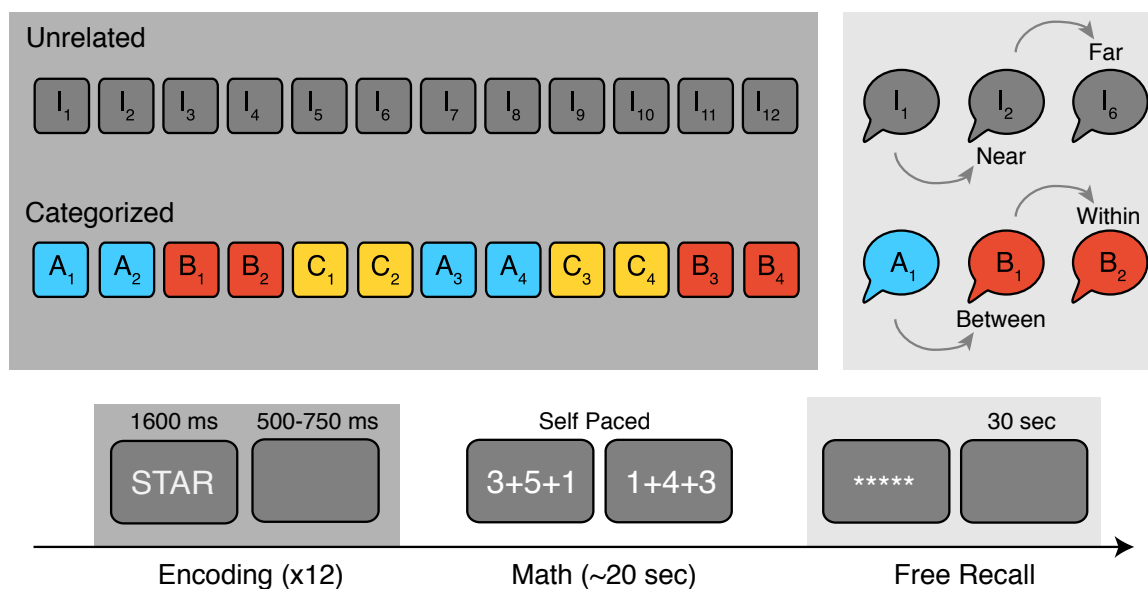


Figure 1. Schematic of the task structure. Each study list comprised of 12 unrelated items (I) or of 12 items drawn from three distinct categories (A, B, C, with 4 items drawn per category). Analyses of retrieval organization examined temporal clusters based on “near” or “far” transitions (i.e., those from items presented adjacently in the study list or not). We analyzed categorical clustering by contrasting between and within-category recall transitions (i.e., those from items that belonged to a different or the same category respectively).

Method

We analyzed direct neural recordings from cortical and deep brain structures from 69 participants as they performed free recall of categorized and unrelated words lists (Figure 1). As part of regular clinical procedure for the monitoring of epileptic seizures, patients were implanted with electrodes (i.e., a combination of subdural grids, strips, and/or depth electrodes) to directly measure iEEG activity. To examine the influence of semantic structure on the neurophysiological markers of memory function, we examined changes in the spectral power of the EEG time series recorded during the encoding and retrieval of categorized and unrelated word lists in a free-recall task.

Participants

We examined data from 69 neurosurgical patients (25 female, with an average age at implant of 36.3 years, ranging from 20 to 63 years of age) with medication-resistant epilepsy who had intracranial electrodes implanted for diagnostic reasons. The data were collected as part of a larger project in collaboration with Columbia University Medical Center (New York, NY), Dartmouth Medical Center (Hanover, NH), Emory Hospital (Atlanta, Georgia), Hospital of the University of Pennsylvania (Philadelphia, PA), Mayo Clinic (Rochester, MN), Thomas Jefferson University Hospital (Philadelphia, PA), and University

of Texas Southwestern Medical Center (Dallas, TX). Institutional review boards at the respective hospitals approved our research protocol and we obtained informed consent from each participant. We selected data from patients who participated in both standard and categorized free-recall tasks (see below). From a total of 94 participants who participated in both versions of the free-recall task, we selected those for which we had at least 0.8 power to detect classifier performance of at least 0.7 area under the receiver operating characteristic (ROC) curve (AUC) for subsequent recall encoding and retrieval decoding and cross-decoding classifiers (see below). We used permutation analyses to estimate chance classification performance for these power analyses and selected the sample of 69 patients on which all analyses presented here are based. For neural analysis of temporal clustering, we analyzed data from a subset of these patients ($n = 60$) who had a sufficient number of observations to train classification models. Some of the analyzed data have been used in other studies (Kragel et al., 2017; Ezzyat et al., 2017; Long et al., 2017; Solomon et al., 2017), but the analyses reported here are novel.

Free-recall tasks

During their time in the epilepsy monitoring unit, patients participated in a range of memory tasks including two versions of a delayed free-recall task that differed in the semantic structure of the study lists (described in detail below). Within each experimental session we only presented one list type and our selection criteria (described above) required that all included participants contributed data from at least one session with each list type (i.e., each participant contributed data from at least two sessions). Each session consisted of up to 26 cycles of encoding, delay, and free recall; the first of which was used as practice and not further analyzed. A 10 s countdown preceded each study list which consisted of a total of 12 words, presented for 1600 ms each, and separated by blank inter-stimulus intervals of 750-1000 ms (randomly sampled from a uniform distribution). The two versions of the task differed only in the composition of the study lists (described below). Following each study list (and prior to free recall), participants were asked to solve simple math problems of the form $A + B + C = ??$ where, A , B , and C , were random integers in $[1, 9]$. Individual math problems were presented one at a time until a response was entered on a keypad when either a new math problem was presented or a signal to initiate free recall was given (math problems were presented until a delay of at least 20 s was achieved). A central row of asterisks presented together with an 800 Hz tone signaled the start of each recall period. Participants were given 30 s to recall as many words from the most recent study list as possible in any order. Vocal responses were recorded and annotated offline.

Study list construction

Study lists consisted of either unrelated or categorized words with the same list type used throughout a given session. For unrelated word lists, a pool of 300 words was constructed by selecting words from a larger word pool used in a large-scale study of free recall on a separate set of participants. For this separate study, the effects of each individual word on recall performance were modeled while accounting for serial position, frequency, concreteness, imageability, and length. Words for which recall performance fell on either end of the resulting distribution were removed to yield 300 words with intermediate recall

performance. From this pool of 300 words, individual lists of 12 items were constructed such that the mean pairwise semantic similarity within list was relatively constant across lists (with a Latent Semantic Analysis cosine similarity of around .2). All 300 words were presented exactly once across the 25 experimental lists in each complete session, but individual participants did not see the same list twice across sessions.

Categorized lists were generated by first asking a set of 40 participants on Amazon Mechanical Turk to generate lists of exemplars from 28 categories. From these responses the 25 exemplars that were generated most frequently from each category were shown to a second set of online participants, who rated their typicality for the respective category. The 12 most prototypical exemplars, according to these ratings, were selected and the three categories with the lowest mean prototypicality ratings across those 12 exemplars were discarded to yield a word pool of 300 words consisting of 25 categories with 12 highly prototypical exemplars each. Each experimental study list was constructed by randomly selecting four exemplars from each of three randomly selected categories. Words were sequentially presented in pairs (two exemplars from a given category), but no two pairs from the same category were presented consecutively. Each session began with the presentation of a practice study list which was always composed of the same 12 (unrelated) words in random order regardless of whether the words in the experimental study lists were unrelated or categorized; recall performance from this practice list was not analyzed.

Recording and processing of electrophysiological signals

We recorded from subdural grids and strips (space between adjacent contacts: 10 mm) and from depth electrodes (space between adjacent contacts: 5–10 mm) on a variety of recording systems across clinical sites with sampling rates varying between 500 and 2000 Hz to accommodate the local recording environment. We re-referenced all recordings using a bipolar referencing scheme (Burke et al., 2013), applied a 4th order Butterworth filter with a 58-62Hz stop-band to remove line noise, and convolved the resulting signals with Morlet wavelets (wave number 5; 8 center frequencies, log-spaced between 3 and 180 Hz) to obtain spectral power. For the analysis of encoding data, we averaged power over the entire 1600 ms stimulus presentation interval. We constructed patterns of neural activity from the recall period of the task by averaging spectral power from 900 to 100 ms preceding a correct recall response. We used mirrored buffering of recall period data to prevent spreading of low frequency signals produced during recall vocalization from mixing with potential signals of interest. To discriminate between successful and unsuccessful retrieval, we additionally constructed failed-retrieval events. The onsets of these events within the recall period were matched to valid recalls in different lists, with the constraint that no valid recall occurred in the following 2000 ms. To ensure that response production from prior recalls did not influence our estimates of retrieval-related activity, we eliminated recalls from our analysis in which the onset of the prior recall occurred within 1500 ms. We applied the same constraint to failed-retrieval events.

Univariate analyses

To account for heterogeneous electrode coverage across participants, we aggregated electrodes into 9 regions of interest (ROIs; Figure 3a, top). For each ROI (and participant

who contributed to that ROI) we calculated differences in z -transformed (by session) log-power during either the 1600 ms encoding period or the 800 ms retrieval period between event types of interest (e.g., subsequently recalled vs. forgotten items). We then divided the mean differences across participants by their standard errors to obtain the t -values shown in Figure 3. Preliminary analyses did not indicate significant differences as a function of hemisphere and we thus collapsed across hemispheres for each ROIs to simplify presentation and increase power.

Multivariate classification

We trained L2-regularized logistic regression classifiers to distinguish patterns of brain activity associated with different memory states. The first classifier discriminated between neural activity associated with successful and unsuccessful memory function. During encoding, we determined successful memory function by whether or not each studied item was subsequently recalled. Likewise, during retrieval, we contrasted intervals in the recall period that lead to successful recall with those that did not in order to assess successful memory function. We used the second classifier to discriminate neural states that predicted whether recall of items from a categorized list would follow a recall from a same-category item (i.e., within-category transition) or a recall from a different-category item (i.e., between-category transition; see Figure 1). We constructed a third classifier to identify items that would be recalled within a temporal cluster (i.e., the preceding recall was presented in an adjacent serial position) or not (see Figure 1). We used these classifiers to decode patterns of spectral power observed during encoding (i.e., identifying biomarkers of subsequent clustering) and during recall periods.

For all classifiers, we fixed the regularization parameter C , as implemented in scikit-learn (Pedregosa et al., 2011), to 0.0007 based on explorations on a larger data set. We fit these classifiers separately to data from each participant using the z -transformed (by session) log-power in the 8 frequencies across all electrodes as features. For classifiers trained and tested on the same list type (i.e., decoding classifiers), we cross-validated classifier performance by holding out each list (across all sessions using that list type) once. For classifiers trained and tested on different list types (i.e., cross-decoding classifiers) we trained the classifier on all data from the respective list type and evaluated its performance on all data from the other list type. By examining the degree to which each classifier would generalize to held out lists with a different semantic structure (i.e., a cross-decoding approach; Kragel & Polyn, 2015; Kragel et al., 2017) compared to its performance on held-out test lists that matched the semantic structure of the training lists (i.e., a decoding approach), we were able to quantify neural signals that were specifically sensitive to the semantic structure of the memoranda. We used each classifier’s AUC as measure of its accuracy, with 0.5 corresponding to chance performance and 1.0 indicating perfect classification (Fawcett, 2006).

Statistical analysis

We assessed significance of group level effects by treating participants as random effects. We evaluated differences in recall performance and classifier generalization across list types using repeated-measures ANOVAs, testing for significance at $p < .05$. For tests

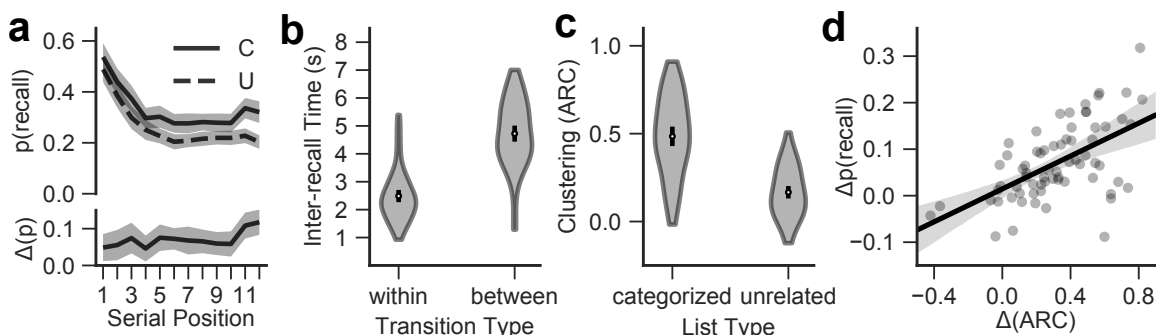


Figure 2. Recall performance and organization. (a) Probability of recall for categorized (C) and unrelated (U) word lists as a function of serial position (top panel) and associated differences (bottom panel). Shaded regions denote 95% confidence intervals. (b) Violin plots illustrating the distributions of inter-recall times for transitions within and between category clusters. Means and 95% confidence intervals are indicated. (c) Violin plots illustrating the distributions of adjusted ratio of clustering (ARC) scores for categorized and unrelated lists (we determined clustering for unrelated lists relative to the categorical structure of matched categorized lists). Means and 95% confidence intervals are indicated. (d) Scatter plot showing differences in ARC for categorized vs. unrelated word lists against corresponding differences in the probability of recall for each participant and associated regression line. Shaded region denotes the 95% confidence region of the regression line.

of a single factor (e.g., category clustering), we performed two-tailed one sample *t*-tests except in cases where we tested whether classifier performance exceeded chance. For tests of univariate differences in activity across multiple ROIs and frequencies (see Figure 3), we adjusted the resultant *p*-values using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995) to control the false discovery rate (FDR, $q < .05$).

Results

As described above, the semantic structure of study lists profoundly affects subsequent recall with categorized lists generally producing increased rates of recall (relative to similar lists of unrelated items), semantically organized recalls, and reduced inter-recall times when items are recalled within a semantic cluster (Bower et al., 1969; Bousfield, 1953; Howard & Kahana, 2002; Wingfield, Lindfield, & Kahana, 1998). We first demonstrate these effects in our data and then use analyses of neural signals to test the relative contributions of encoding and retrieval processes to recall performance and organization.

To verify that categorically organized lists improved recall performance in our patient population, we conducted a 2 (list type) \times 12 (serial position) repeated measures ANOVA on the probability of recall (see Figure 2a). Participants recalled a higher percentage of items from categorized than unrelated word lists ($F(1, 68) = 52.46$, $MSE = 0.04$, $p < .001$), with a significant main effect of list position reflecting enhanced recall of early list items across both list types (primacy effect; $F(11, 748) = 56.79$, $MSE = 0.02$, $p < .001$). The distractor-filled retention interval limited the benefit of recency on memory performance to little (on categorized lists) or none (on unrelated lists), resulting in a significant interaction between list type and serial position ($F(11, 748) = 1.93$, $MSE = 0.01$, $p = .03$). These

results replicate previous findings of improved recall performance for semantically structured memoranda (Bousfield, 1953; Bower et al., 1969).

We next confirmed that presentation of categorized lists influenced recall dynamics by causing subjects to recall bursts of same-category items (Wingfield et al., 1998). As in previous studies (Pollio, Richards, & Lucas, 1969; Patterson, Meltzer, & Mandler, 1971), within-category recall transitions were significantly faster than between-category transitions ($t(68) = 14.54$, $SE = 0.15$, $p < .001$; see Figure 2b). This difference remained significant when eliminating transitions that were faster than 1500 ms and thus did not contribute to our analyses of brain activity (see Method; $t(68) = 8.36$, $SE = 0.20$, $p < .001$). We quantified the degree of clustering with the adjusted ratio of clustering (ARC) statistic (Roemaker, Thompson, & Brown, 1971),¹ which was significantly greater for categorized than unrelated lists (using the category structure of matched categorized lists to quantify baseline levels of clustering; $t(68) = 10.08$, $SE = 0.03$, $p < .001$; Figure 2c). This measure confirms that category clustering exceeded what one would expect from reinstatement of the encoding sequence. Indeed, participants were also more likely to cluster same-category items that were not presented in adjacent serial positions in the study list (ARC calculated only for items in non-adjacent positions in the study lists was significantly greater for categorized lists than for matched unrelated lists; $t(68) = 6.03$, $SE = 0.033$, $p < .001$). Participants who exhibited greater category clustering also recalled a higher percentage of items on categorized as compared to unrelated word lists (Figure 2d, $r = .56$, $p < .001$). Having established that our participants exhibited the classic effects associated with the recall of categorized lists, we next analyzed neural activity to determine the relative contributions of encoding and retrieval processes to this behavior.

If the semantic relatedness of memoranda interacts with episodic encoding and retrieval processes, one would expect corresponding differences in associated neural biomarkers of memory function. To identify these differences, we partitioned electrode locations into the 9 ROIs shown in Figure 3a and contrasted average log-transformed power (z -transformed by session) from electrodes within each ROI as a function of subsequent recall (Figure 3b).² Specifically, we separately averaged power during encoding periods (i.e., over the 1600 ms study word presentation intervals) for words that were subsequently recalled and those that were not. To identify biomarkers of successful memory function during retrieval, we averaged power during 800 ms intervals ranging from 900 ms to 100 ms prior to successful recalls and contrasted these with matched intervals not preceding recalls (these intervals putatively index unsuccessful retrieval attempts). Despite the clear differences in recall performance and organization as a function of list type (Figure 2), these encoding and retrieval biomarkers of successful memory were remarkably similar across both list types (Figure 3b): Increased power in higher frequencies and decreased power in lower frequencies indicated subsequent memory and imminent retrieval across a wide range of ROIs for both categorized and unrelated lists. We found no significant differences across list types even

¹ARC = $\frac{R - E(R)}{R_{\max} - E(R)}$, where R is the observed number of category repetitions (i.e., the number of times two consecutively recalled items shared the same category), R_{\max} is the largest possible number of category repetitions, and $E(R)$ is the number of category repetitions that are expected by chance (see Stricker, Brown, Wixted, Baldo, & Delis, 2002, for a detailed description of how to calculate these measures).

²The number of observations across ROIs varied depending on the number of participants contributing electrodes to each ROI (Figure 3a, bottom).

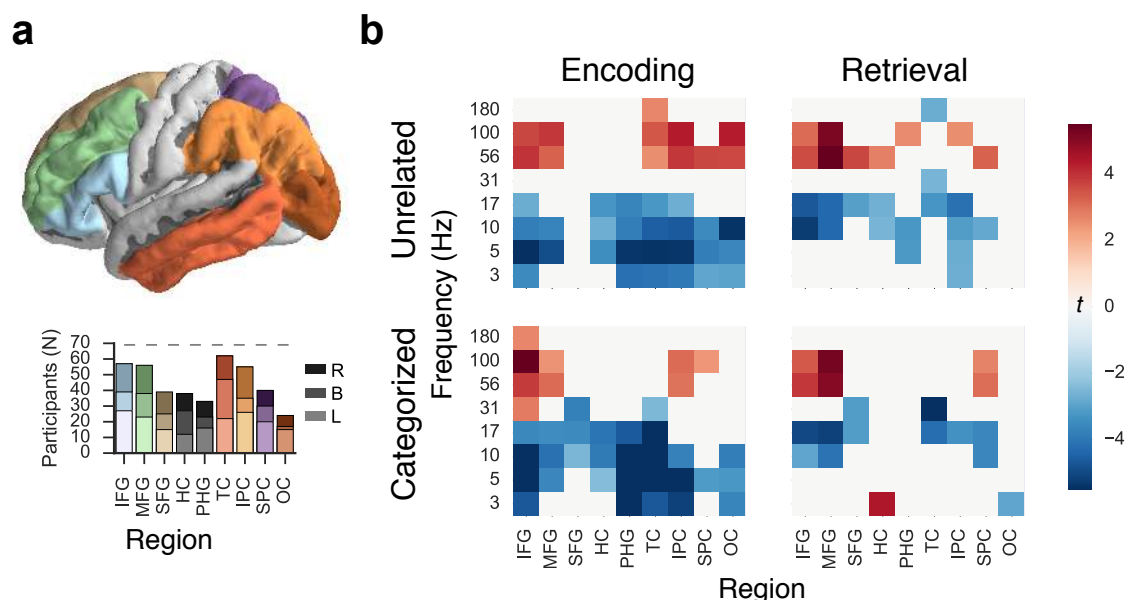


Figure 3. Encoding and retrieval biomarkers from categorized and unrelated word lists. (a) Regions of interest (ROIs) are displayed on an average cortical surface template. The number of participants contributing electrodes to each ROI is shown below, separately for left (L) hemisphere, right (R) hemisphere, or bilateral (B) coverage. Dashed line indicates the total number of participants. IFG: inferior frontal gyrus; MFG: middle frontal gyrus; SFG: superior frontal gyrus; HC: hippocampus; PHG: parahippocampal gyrus; TC: temporal cortex; IPC: inferior parietal cortex; SPC: superior parietal cortex; OC: occipital cortex. (b) Statistical maps for changes in spectral power for successful versus unsuccessful encoding and retrieval. Statistically significant (FDR corrected, $q < .05$) increases and decreases are shown in red and blue, respectively.

when we did not correct for multiple comparisons. While these findings may suggest that the same neural processes underlie successful memory encoding and retrieval independent of the categorical structure of the list, differences in neural signal may have been obscured due to aggregation across ROIs and differences in electrode placement across participants.

To increase our power to detect differences in the influence of semantic structure on episodic encoding and retrieval, we turned to multivariate classification of memory states within individuals. We trained L2-penalized logistic regression models on spectral iEEG features during study and recall of categorized and unrelated word lists. These models learned to discriminate between neural activity corresponding to differences in (subsequent) recall performance (e.g., between neural activity during encoding periods that predicted whether the studied word was subsequently recalled). We evaluated the ability of classifiers to generalize to lists with differing semantic structure, allowing us to determine whether the neural biomarkers of memory encoding or retrieval were specifically influenced by categorical list structure (see Figure 4 for an illustration of our cross-decoding approach).

Figure 4a depicts average ROC functions showing decoding and cross-decoding performance for classifiers trained on neural activity during either encoding or retrieval periods

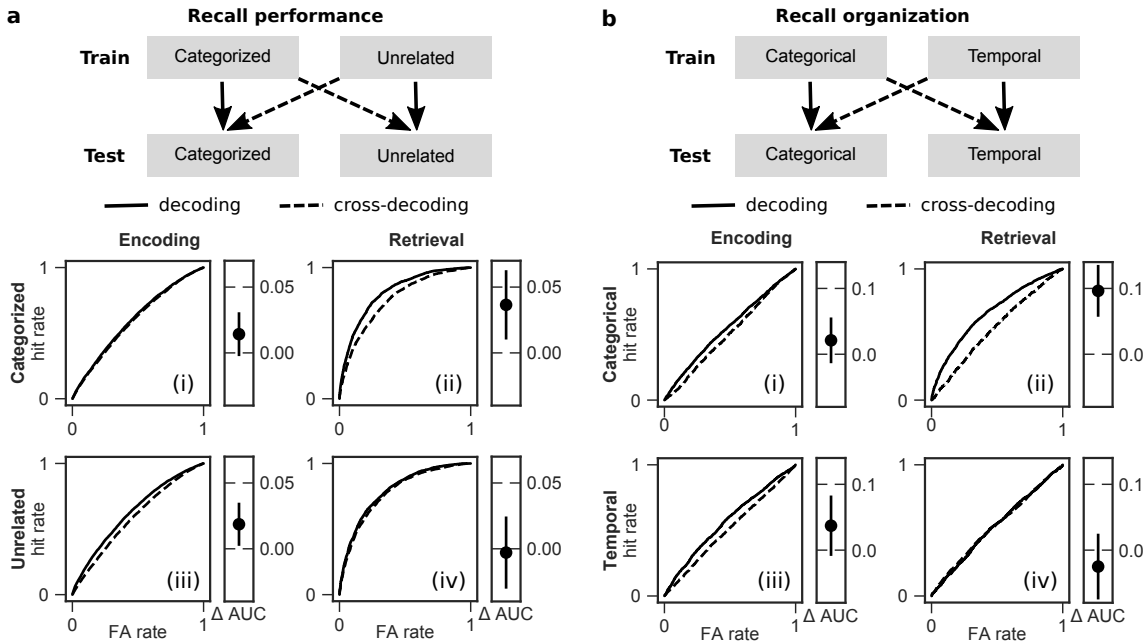


Figure 4. An illustration of decoding and cross-decoding approaches with associated ROC functions and AUC differences (ΔAUC , indexing classifier specificity) for recall performance (a) and recall organization (b) classifiers. Subpanels are numbered using roman numerals (i–iv) to facilitate discussion of the results. Subpanels (i) and (iii) show results for encoding classifiers with (ii) and (iv) showing performance of retrieval classifiers. Subpanels (i)–(ii) and (iii)–(iv) show results from classifiers trained on categorized and unrelated lists respectively. Subsequent clustering classifiers distinguished within- from between-category recall transitions in categorized lists and near from far recall transitions in unrelated lists (see Figure 1). Error bars around the mean ΔAUC values indicate the 95% confidence intervals.

in each list type to predict (subsequent) recall. To the extent that these ROC functions overlap, the corresponding classifier generalizes beyond the semantic structure of the training lists, whereas differences in these ROC functions indicate that the corresponding classifier capitalizes on neural processes that are specific to the semantic structure of the memoranda. To measure this specificity, we calculated the difference in the AUCs between the decoding and cross-decoding performance (ΔAUC), depicted alongside the corresponding ROC functions in Figure 4a.

The curvature of all ROC functions in Figure 4a along with the substantial degree of overlap between corresponding decoding and cross-decoding ROC functions suggests that both encoding and retrieval classifiers predicted (subsequent) recall reliably for both list types. Indeed, decoding performance was significantly above chance for all classifiers (all $ts(68) > 14.16$, $SEs \leq 0.01$, $ps < .001$) with similar results for cross-decoding performance (all $ts(68) > 13.18$, $SEs < 0.02$, $ps < .001$). These findings, together with the remarkably similar univariate (subsequent) recall contrasts across both list types (Figure 3b), implicate general memory processes that determine recall performance irrespective of the semantic

structure of the memoranda.

The difference between decoding and cross-decoding performance was significant in the encoding classifier for unrelated lists ($t(68) = 2.27$, $SE = 0.01$, $p = .03$; Figure 4a (iii)) and in the retrieval classifier for categorized lists ($t(68) = 2.76$, $SE = 0.01$, $p = .007$; Figure 4a (ii)), with no other Δ AUCs deviating significantly from zero (all $t_s(68) < 1.70$, $SEs \leq 0.01$, $ps > .09$). A 2 (training list type) \times 2 (task phase; i.e., encoding vs. retrieval) repeated measures ANOVA on Δ AUC values revealed a significant interaction ($F(1, 68) = 5.12$, $MSE = 0.006$, $p = .03$) but no significant main effects of training list type ($F(1, 68) = 1.47$, $MSE = 0.01$, $p = .23$) or task phase ($F(1, 68) = 0.002$, $MSE = 0.007$, $p = .97$). Direct post-hoc comparisons between Δ AUC values for encoding and retrieval classifiers did not show significant differences for either list type ($t_s(68) < 1.59$, $SEs \leq 0.01$, $ps > .12$). These findings are consistent with prior demonstrations of common neural signatures indexing episodic and semantic memory (Rajah & McIntosh, 2005), while also highlighting processes during retrieval of categorized lists that are sensitive to the semantic structure of the list.

Having identified specificity of the retrieval classifier when it is trained on categorized lists and (to a lesser extent) the encoding classifier when it is trained on unrelated lists, we next aimed to link neural activity during encoding and retrieval to the organization of recall sequences. The most prominent difference in recall behavior between the two list types is the tendency to recall categorized lists in category clusters (Figure 2c). As described in the introduction, category clustering is a reliable marker of interactions between episodic and semantic memory systems. The tendency to cluster recalls of items that were studied nearby in time (e.g., in adjacent positions in the study list), on the other hand, is a general feature of episodic memory (Kahana, 1996). To establish the contributions of encoding and retrieval processes to these types of recall organization, we trained multivariate classifiers to distinguish recall within and between such clusters based on neural activity during encoding and retrieval periods. Specifically, for categorized lists, we trained classifiers to distinguish neural signals for words that were recalled immediately following the recall of another word in the same category (i.e., within-category recall) from those that were recalled immediately following the recall of a word from a different category (i.e., between-category recall; see Figure 1). Likewise for unrelated lists, we trained classifiers to distinguish brain activity for words that were recalled immediately following the recall of an adjacently presented item (i.e., near recall) from those that followed the recall of an item that was presented further away in the study list (i.e., far recall; see Figure 1). The fact that we presented pairs of items from the same category in categorized lists partially confounded temporal distance in the study list with category membership. Our cross-decoding approach, however, is able to determine to what extent each classifier is specific to the categorical or temporal clustering it has been trained to identify.

Figure 4b shows the ROC functions for decoding and cross-decoding performance for classifiers trained to predict categorical or temporal clustering on the basis of neural activity during encoding and retrieval periods. Decoding performance was significantly above chance for all classifiers ($t(59) = 2.5$ – 11.93 , $SE = 0.01$ – 0.03 , $p \leq .02$) with the exception of the temporal clustering retrieval classifier ($t(59) = 0.41$, $SE = 0.02$, $p = .68$; Figure 4b (iv)). Additionally, decoding performance for the categorical clustering retrieval classifier was higher than that for the corresponding encoding classifier ($t(59) = 6.65$, $SE = 0.03$, $p < .001$) with no significant difference in decoding performance as a function of

task phase for the temporal clustering classifiers ($t(59) = 1.50$, $SE = 0.03$, $p = .13$). A 2 (classifier type, predicting temporal or categorical clustering) \times 2 (task phase; i.e., encoding vs. retrieval) repeated measures ANOVA exhibited significant main effects of classifier type ($F(1, 59) = 10.72$, $MSE = 0.02$, $p = 0.002$) and task phase ($F(1, 59) = 4.50$, $MSE = 0.01$, $p = 0.04$), as well as a significant interaction ($F(1, 59) = 21.95$, $MSE = 0.01$, $p < .001$). In line with our previous result that retrieval classifiers are specifically sensitive to the semantic structure of categorized lists, these findings particularly implicate retrieval processes in the production of categorical clusters in episodic recall.

Our cross-decoding approach further highlights the specific role of retrieval processes in producing categorically clustered recall sequences. Figure 4b shows the ΔAUC values for the clustering classifiers alongside the corresponding decoding and cross-decoding ROC functions. Only the retrieval classifier trained to identify categorical clustering exhibited significantly better decoding than cross-decoding performance ($t(59) = 7.07$, $SE = 0.01$, $p < .001$; Figure 4b (ii)). We performed a 2 (classifier type) \times 2 (task phase) repeated measures ANOVA on the ΔAUC values and found a non-significant effect of task phase ($F(1, 59) = 0.08$, $MSE = 0.03$, $p = .77$), a significant effect of classifier type ($F(1, 59) = 7.59$, $MSE = 0.02$, $p = .007$), and a significant interaction between the two ($F(1, 59) = 15.46$, $MSE = 0.02$, $p < .001$). A post-hoc test confirmed that the categorical clustering classifier was more sensitive to the semantic structure of the memoranda at retrieval than at encoding ($t(59) = 3.83$, $SE = 0.02$, $p < .001$), an effect which appears to be mostly driven by the superior decoding performance for the retrieval classifier described above. Indeed, given that the temporal clustering retrieval classifier’s decoding performance was not above chance, we would not expect the cross-decoding performance of the categorical clustering retrieval classifier to exceed chance. While encoding and retrieval processes clearly both contribute to category clustering, the categorical clustering retrieval classifier’s particularly high decoding performance corroborates our earlier finding that only processes at retrieval uniquely predict semantic clustering of episodic memory.³

Discussion

Prior knowledge about the world has profound effects upon how we encode and retrieve experiences, leading us to organize episodic memories along semantic dimensions (Bower et al., 1969; Bousfield, 1953). The relative contributions of encoding and retrieval processes that specifically drive this interaction between episodic and semantic memory systems, however, are poorly understood. By directly comparing recall performance and associated neural activity during encoding and retrieval periods for lists of categorized and unrelated words, we characterized how semantic structure of memoranda affects encoding and retrieval of events. Multivariate classifiers that predicted encoding and retrieval success generalized across categorized and unrelated lists (Figure 4a), identifying shared

³Because participants tended to recall clusters of same-category items in bursts (Figure 2b), it is possible that retrieval prior to initiating a cluster enabled participants to make multiple recalls without the need for retrieval to occur within the cluster (e.g., as if items were read out from a queue). Examination of inter-recall times suggests that to the extent this was the case, it did not disproportionately affect one transition type over another: the proportion of rapid responses (inter-response times less than 1 s) following between-category recalls (40%) and within-category recalls (43%) did not significantly differ (mean difference = 3.82%, $SE = 0.02$, $t(59) = -1.07$, $p = .29$).

processes that support episodic memory irrespective of semantic content. By contrasting decoding and cross-decoding performance, we also found that specific processes support successful retrieval from categorized lists (Figure 4a, ii), suggesting that retrieval mechanisms are critical for producing semantically organized recall sequences. Our temporal and categorical clustering analyses supported this view: Neural signals during retrieval contained significantly more information about category clustering than corresponding signals during encoding (Figure 4b, i–ii). These classifiers indexed processes that were specific to categorical clustering; attempts to predict temporal clustering were less successful, even for classifiers trained to detect temporal clustering (i.e., not just for cross-decoding, but also for decoding classifiers; Figure 4b, iv). Taken together, these findings suggest that the semantic structure of the memoranda affects recall performance mainly through changes in retrieval (rather than encoding) processes.

Our use of multivariate classification to estimate cognitive states during encoding and retrieval allowed us to overcome a central challenge to understanding human memory: even though encoding and retrieval processes are both critical determinants of performance in memory tests, only their joint effects are observable in overt behavior. By quantifying neural signals that predict recall performance and organization, we were able to effectively track associated encoding and retrieval processes as they occurred. Previous studies of the relation between episodic and semantic memory systems (Vargha-Khadem et al., 1997; Greenberg et al., 2009; Wingfield et al., 1998), especially those attempting to resolve the relative contributions of encoding and retrieval processes (Gershberg & Shimamura, 1995; Greenberg & Verfaellie, 2010; Johnston, Griffith, & Wagstaff, 1972; Park, Smith, Dudley, & Lafronza, 1989), have yielded equivocal results. Our finding that neural signals at both encoding and retrieval predict recall performance and organization may help explain why different studies have identified either encoding or retrieval processes as responsible for the semantic organization of episodic memory. This finding is also consistent with previous neuroimaging work showing a high degree of overlap in brain regions engaged during tasks probing episodic and semantic memory (Burianova, McIntosh, & Grady, 2010) as well as the recruitment of episodic information even in putatively semantic tasks (e.g., category fluency; Ryan, Cox, Hayes, & Nadel, 2008; Westmacott & Moscovitch, 2003; Sheldon & Moscovitch, 2012; see Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006, for a review). In line with these results, our findings contribute to the growing evidence for a general system that subserves memory function irrespective of task demands. Our separate identification of neural signals indexing retrieval processes that are specifically tuned to the semantic structure of the memoranda highlights the benefit of distinguishing processes associated with general memory function from those specifically responsible for categorical organization of recall.

Theories on the relation between episodic and semantic memory typically focus on the question of how semantic knowledge arises from individual episodes (McClelland, McNaughton, & O'Reilly, 1995; Nelson & Shiffrin, 2013; Mack, Love, & Preston, 2017; Howard, Shankar, & Jagadisan, 2011). Our work addresses the flip-side of this question: How does established semantic knowledge affect memory for specific events? In our study, categorized lists led to improved rates of recall and categorical clustering of recall sequences. There are at least three common approaches to modeling such effects: they can result from overlapping representations of individual memory traces (Abbott, Austerweil, & Griffiths, 2015;

Katkov, Romani, & Tsodyks, 2017), processes at encoding that activate memory representations that are related to studied items (Collins & Loftus, 1975; Watkins & Gardiner, 1979; Cooke, Durso, & Schvaneveldt, 1986; Underwood, 1965), or adaptive retrieval mechanisms that differentially make use of categorical or contextual information as needed (Shiffrin, 1970; Patterson et al., 1971; Morton & Polyn, 2016). Below we consider the ability of each of these approaches to explain our data, even though they are clearly related (e.g., encoding and retrieval processes can leverage similarities in the representation of memory traces as discussed below).

The idea that the similarity between memory traces drives recall dynamics can explain basic patterns of behavior in free recall (Katkov et al., 2017) and semantic fluency tasks (Abbott et al., 2015). Contemporary models of free recall assume that each encoded item is associated with context information that reflects semantic information and recent history (Polyn, Norman, & Kahana, 2009; Lohnas, Polyn, & Kahana, 2015). Because context serves as a cue during retrieval, associating each item with such context information can result in categorically structured recall sequences and improved recall performance for categorized lists. While our analyses do not directly examine the neural activity evoked by individual items, previous electrophysiology studies have identified neural signals during encoding that reflected semantic information and were reinstated during retrieval, driving semantic organization of recall (Manning, Sperling, Sharan, Rosenberg, & Kahana, 2012; Morton et al., 2013). Our finding of neural signals during retrieval that specifically predict category clustering, however, suggest that additional retrieval processes guide the organization of recall sequences.

It seems natural to assume that the details of how items are represented in memory should also affect encoding. Spreading activation (Collins & Loftus, 1975), for example, refers to the idea that processing of any item also activates nearby items in a semantic network. This mechanism can explain a range of findings in semantic priming (Meyer & Schvaneveldt, 1971, 1976) and recall (Cooke et al., 1986) tasks. Similarly, a classic explanation for false memories of semantic associates to studied items is an implicit associate response during encoding, which creates a representation of the semantic associate that is later retrieved, just like memories for studied items (Underwood, 1965). To the extent that encoding processes adapt to the semantic structure of the memoranda in ways that result in differential recall performance or organization, we would have expected associated neural signals to specifically predict category clustering. Instead, neural signals during encoding periods generalized across different list types, suggesting that encoding processes are more sensitive to features of individual items rather than the semantic structure of a study list.

Despite the success of theories relying on semantic networks and encoding processes to account for a wide range of data, several theorists consider flexible retrieval mechanisms necessary to explain human recall performance. Modern theories of false memories, for example, invoke both encoding and retrieval processes (H. L. Roediger III, Balota, & Watson, 2001; Meade, Watson, Balota, & Roediger, 2007), and some accounts of semantic fluency conceptualize retrieval from semantic memory as a dynamic, controlled process rather than a random walk on a semantic network (Hills, Jones, & Todd, 2012; Jones, Hills, & Todd, 2015). Various memory tests require the match of a probe item to memory. In these cases responses to new probe stimuli are often affected by the experience with other stimuli. Associated theories generally explain such context effects either by assuming changes in the

mapping between the match signal and responses (i.e., a response bias) or by assuming that the match signal itself adapts to the context (an explanation akin to the flexible retrieval processes for which we present evidence here). For example, a range of short-term priming experiments use a perceptual identification task requiring the match of a probe to a briefly presented target while ignoring primes presented in the vicinity of the target (Huber, Shiffrin, Lyle, & Ruys, 2001; Pecher, Zeelenberg, & Raaijmakers, 2002; Weidemann, Huber, & Shiffrin, 2005, 2008). Some of these studies have found strong effects of prime-target contingencies (e.g., how likely the prime is to match the target). Accordingly, trials that match in all aspects except for the history of previous trials (and thus the learned prime-target contingencies) can produce large differences in performance (Pecher et al., 2002; Weidemann et al., 2008). Whereas such effects are generally assumed to be the result of response biases (Pecher et al., 2002), there is evidence that the match signal adapts to these contingencies (Weidemann et al., 2008). Likewise, in tests of recognition memory requiring the classification of a probe stimulus as either previously studied (i.e., a target) or not (i.e., a lure), even responses to lures are affected by how targets are studied. Experimentally manipulating strength of encoding (e.g., by increasing study time; Ratcliff, Clark, & Shiffrin, 1990; Stretch & Wixted, 1998) not only increases hit rates, but also decreases the likelihood of false alarms to lures (i.e., a strength-based mirror effect; Glanzer & Adams, 1985). Whereas some accounts for this effect attribute it to response biases (Starns, White, & Ratcliff, 2012; Stretch & Wixted, 1998), it has been successfully modeled by assuming that the match signal aggregates evidence across the full set of targets, leading to a poorer match for lures when they are compared to more strongly encoded targets (Criss, 2010; Criss & McClelland, 2006; Kılıç, Criss, Malmberg, & Shiffrin, 2017; Shiffrin, Ratcliff, & Clark, 1990; Shiffrin & Steyvers, 1997). Our results provide converging evidence that the structure of encoded material can have profound effects that are specific to processes involved in probing memory (i.e., retrieval, rather than encoding, processes).

Whereas neural activity during retrieval predicted categorical clustering best, neural activity during the encoding of unrelated lists best predicted temporal clustering. Previous studies have interpreted increased activity within a putative “core memory network” during encoding as reflecting item to context associations, a central process in several models of episodic memory (Long & Kahana, 2015). Our ability to decode subsequent temporal clustering on the basis of encoding activity provides converging evidence for such encoding processes. Our attempts to decode temporal clustering on the basis of neural activity during retrieval failed, however, despite robust temporal clustering in the recall sequences for unrelated lists (see Kragel et al., 2017, for measures of temporal clustering in a superset of these data). Given the inherent limitations in our ability to intracranially record brain activity in humans (Parvizi & Kastner, 2018) and our focus on spectral power for a small set of frequencies as sole indices for this activity, it is likely that we missed relevant neural signals that could predict temporal clustering during retrieval. Nevertheless, the specificity of the categorical clustering classifier at retrieval implicates distinct processes in the semantic and episodic organization of memories.

The question about the relative contributions of encoding and retrieval processes to our ability to remember is often difficult to answer, because changes in encoding can be counteracted by changes in retrieval and vice versa. Using a multivariate cross-decoding approach, we tackled this problem by linking neural activity to encoding and retrieval pro-

cesses that either generalized across different list types or specifically predicted recall performance and organization for categorized lists. We identified retrieval signals that specifically predicted categorical recall organization, whereas signals at encoding did not. The lack of encoding signals that specifically predict categorical recall organization is particularly striking, given that our experimental setting imposed a clear categorical structure at the time of encoding and allowed for the full anticipation of the memory test. These findings provide converging evidence for the importance of flexible retrieval mechanisms in models of human memory and challenge the notion that similarity of long-term memory representations are the primary drivers of recall performance (Katkov et al., 2017). Indeed, our results suggest that flexible retrieval is a basic feature of human memory that adapts memory search to task demands even when these are readily apparent at encoding.

References

- Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*, 558–569. doi: 10.1037/a0038693
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society, Series B*, *57*, 289–300.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*(2), 229–240. doi: 10.1080/00221309.1953.9710088
- Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343. doi: 10.1016/S0022-5371(69)80124-6
- Burianova, H., McIntosh, A. R., & Grady, C. L. (2010). A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *NeuroImage*, *49*, 865–874. doi: 10.1016/j.neuroimage.2009.08.066
- Burke, J. F., Long, N. M., Zaghoul, K. A., Sharan, A. D., Sperling, M. R., & Kahana, M. J. (2014). Human intracranial high-frequency activity maps episodic memory formation in space and time. *NeuroImage*, *85 Pt. 2*, 834–843. doi: 10.1016/j.neuroimage.2013.06.067
- Burke, J. F., Sharan, A. D., Sperling, M. R., Ramayya, A. G., Evans, J. J., Healey, M. K., . . . Kahana, M. J. (2014). Theta and high-frequency activity mark spontaneous recall of episodic memories. *Journal of Neuroscience*, *34*, 11355–11365. doi: 10.1523/jneurosci.2654-13.2014
- Burke, J. F., Zaghoul, K. A., Jacobs, J., Williams, R. B., Sperling, M. R., Sharan, A. D., & Kahana, M. J. (2013). Synchronous and asynchronous theta and gamma activity during episodic memory formation. *Journal of Neuroscience*, *33*(1), 292–304. doi: 10.1523/JNEUROSCI.2057-12.2013
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428. doi: 10.1037/0033-295x.82.6.407
- Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 538–549. doi: 10.1037/0278-7393.12.4.538
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180. doi: 10.1037/0096-3445.125.2.159
- Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 484–499. doi: 10.1037/a0018435
- Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*, 447–460. doi: 10.1016/j.jml.2006.06.003

- Ezzyat, Y., Kragel, J. E., Burke, J. F., Levy, D. F., Lyalenko, A., Wanda, P., ... Kahana, M. J. (2017). Direct Brain Stimulation Modulates Encoding States and Memory Performance in Humans. *Current Biology*, *27*(9), 1251–1258. doi: 10.1016/j.cub.2017.03.028
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874. doi: 10.1016/j.patrec.2005.10.010
- Fernandes, M. A., & Grady, C. (2008). Age differences in susceptibility to memory interference during recall of categorizable but not unrelated word lists. *Experimental Aging Research*, *34*, 297–322. doi: 10.1080/03610730802273860
- Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia*, *33*, 1305–1333. doi: 10.1016/0028-3932(95)00103-a
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. doi: 10.3758/bf03198438
- Greenberg, D. L., Keane, M. M., Ryan, L., & Verfaellie, M. (2009). Impaired Category Fluency in Medial Temporal Lobe Amnesia: The Role of Episodic Memory. *Journal of Neuroscience*, *29*, 10900–10908. doi: 10.1523/JNEUROSCI.1202-09.2009
- Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological Society*, *16*, 748–753. doi: 10.1017/S1355617710000676
- Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 431–440. doi: 10.1037/a0027373
- Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*, 85–98. doi: 10.1006/jmla.2001.2798
- Howard, M. W., Shankar, K. H., & Jagadisan, U. K. K. (2011). Constructing semantic representations from a gradually-changing representation of temporal context. *Topics in Cognitive Science*. doi: 10.1111/j.1756-8765.2010.01112.x
- Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, *108*, 149–182. doi: 10.1037/0033-295x.108.1.149
- Johnston, W. A., Griffith, D., & Wagstaff, R. R. (1972). Speed, accuracy, and ease of recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 512–520. doi: 10.1016/S0022-5371(72)80034-3
- Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, *122*, 570–574. doi: 10.1037/a0039248
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. doi: 10.3758/BF03197276
- Kapur, S., Craik, F. I., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: levels of processing effect. *Proceedings of the National Academy of Sciences*, *91*(6), 2008–2011. doi: 10.1073/pnas.91.6.2008
- Katkov, M., Romani, S., & Tsodyks, M. (2017). Memory retrieval from first principles. *Neuron*, *94*, 1027–1032. doi: 10.1016/j.neuron.2017.03.048
- Kılıç, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology*, *92*, 65–86. doi: 10.1016/j.cogpsych.2016.11.005
- Kragel, J. E., Ezzyat, Y., Sperling, M. R., Gorniak, R., Worrell, G. A., Berry, B. M., ... Kahana, M. J. (2017). Similar patterns of neural activity predict memory function during encoding and retrieval. *NeuroImage*, *155*, 60–71. doi: 10.1016/j.neuroimage.2017.03.042
- Kragel, J. E., & Polyn, S. M. (2015). Decoding Episodic Retrieval Processes: Frontoparietal and Medial Temporal Lobe Contributions to Free Recall. *Journal of Cognitive Neuroscience*, *28*, 125–139. doi: 10.1162/jocn_a_00881
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*, 337–363. doi:

- 10.1037/a0039036
- Long, N. M., Burke, J. F., & Kahana, M. J. (2014). Subsequent memory effect in intracranial and scalp EEG. *NeuroImage*, *84*, 488–494. doi: 10.1016/j.neuroimage.2013.08.052
- Long, N. M., & Kahana, M. J. (2015). Successful memory formation is driven by contextual encoding in the core memory network. *NeuroImage*, *119*, 332–337. doi: 10.1016/j.neuroimage.2015.06.073
- Long, N. M., Sperling, M. R., Worrell, G. A., Davis, K. A., Gross, R. E., Lega, B. C., . . . Kahana, M. J. (2017). Contextually mediated spontaneous retrieval is specific to the hippocampus. *Current Biology*, *27*, 1–6. doi: 10.1016/j.cub.2017.02.054
- Mack, M. L., Love, B. C., & Preston, A. R. (2017). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*. doi: 10.1016/j.neulet.2017.07.061
- Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., & Kahana, M. J. (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *Journal of Neuroscience*, *32*(26), 8871–8878. doi: 10.1523/JNEUROSCI.5321-11.2012
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. doi: 10.1037//0033-295x.102.3.419
- Meade, M. L., Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, *56*, 305–320. doi: 10.1016/j.jml.2006.07.007
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234. doi: 10.1037/h0031564
- Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, *192*, 27–33. doi: 10.1126/science.1257753
- Morton, N. W., Kahana, M. J., Rosenberg, E. A., Sperling, M. R., Sharan, A. D., & Polyn, S. M. (2013). Category-specific neural oscillations predict recall organization during memory search. *Cerebral Cortex*, *23*(10), 2407–2402. doi: 10.1093/cercor/bhs229
- Morton, N. W., & Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, *86*, 119–140. doi: 10.1016/j.jml.2015.10.002
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, *16*, 179–190. doi: 10.1016/j.conb.2006.03.013
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*, 356–394. doi: 10.1037/a0032020
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430. doi: 10.1016/j.tics.2006.07.005
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*, 93–102. doi: 10.1016/s1364-6613(00)01845-3
- Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1185–1191. doi: 10.1037/0278-7393.15.6.1185
- Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, *21*, 474–483. doi: 10.1038/s41593-018-0108-2
- Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Inter-response times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 417–426. doi: 10.1016/s0022

-5371(71)80041-5

- Patterson, K. E., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987. doi: 10.1038/nrn2277
- Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (2002). Associative priming in a masked perceptual identification task: Evidence for automatic processes. *The Quarterly Journal of Experimental Psychology*, 55A, 1157–1173. doi: 10.1080/02724980244000143
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Poldrack, R. A., Clark, J., ParÁl-Blagoev, E. J., Shohamy, D., Creso, J. M., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414. doi: 10.1038/35107080
- Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, 8, 529–536. doi: 10.1016/s0022-5371(69)80099-x
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116, 129-156. doi: 10.1037/a0014420
- Rajah, M., & McIntosh, A. (2005). Overlap in the functional neural systems involved in semantic and episodic memory retrieval. *Journal of Cognitive Neuroscience*, 17(3), 470–482. doi: 10.1162/0898929053279478
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 163–178. doi: 10.1037//0278-7393.16.2.163
- Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association. doi: 10.1037/10394-006
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 76, 45–48. doi: 10.1037/h0031355
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111, 205–235. doi: 10.1037/0033-295X.111.1.205
- Ryan, L., Cox, C., Hayes, S. M., & Nadel, L. (2008). Hippocampal Activation during Episodic and Semantic Memory Retrieval: Comparing Category Production and Category Cued Recall. *Neuropsychologia*, 46(8), 2109–2121. doi: 10.1016/j.neuropsychologia.2008.02.030
- Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. Cambridge, Ma: MIT Press.
- Sheldon, S., & Moscovitch, M. (2012). The nature and time-course of medial temporal lobe contributions to semantic retrieval: An fMRI study on verbal fluency. *Hippocampus*, 22, 1451–1466. doi: 10.1002/hipo.20985
- Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, 168, 1601-1603. doi: 10.1126/science.168.3939.1601
- Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16(2), 179-195. doi: 10.1037//0278-7393.16.2.179
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4, 145–166. doi: 10.3758/bf03209391

- Simons, J. S., Graham, K. S., Galton, C. J., Patterson, K., & Hodges, J. R. (2001). Semantic knowledge and episodic memory for faces in semantic dementia. *Neuropsychology*, *15*, 101–114. doi: 10.1037//0894-4105.15.1.101
- Solomon, E. A., Kragel, J. E., Sperling, M. R., Sharan, A., Worrell, G., Kucewicz, M., . . . Kahana, M. J. (2017). Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nature Communications*, *8*, 1704. doi: 10.1038/s41467-017-01763-2
- Starns, J. J., White, C. N., & Ratcliff, R. (2012). The strength-based mirror effect in subjective strength ratings: The evidence for differentiation can be produced without differentiation. *Memory & Cognition*, *40*, 1189–1199. doi: 10.3758/s13421-012-0225-1
- Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1379–1396. doi: 10.1037/0278-7393.24.6.1379
- Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test—second edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, *8*, 425–435. doi: 10.1017/s1355617702813224
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. (p. 381–403). New York: Academic Press.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. doi: 10.1146/annurev.psych.53.100901.135114
- Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of learning and memory*, *91*(2), 139–154. doi: 10.1016/j.nlm.2008.10.011
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129. doi: 10.1037/h0022014
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connely, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376–380. doi: 10.1126/science.277.5324.376
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., . . . Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191. doi: 10.1126/science.281.5380.1188
- Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognition theory of recall. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 687–704. doi: 10.1016/S0022-5371(79)90397-9
- Weidemann, C. T., Huber, D. E., & Shiffrin, R. M. (2005). Confusion and compensation in visual perception: Effects of spatiotemporal proximity and selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 40–61. doi: 10.1037/0096-1523.31.1.40
- Weidemann, C. T., Huber, D. E., & Shiffrin, R. M. (2008). Prime diagnosticity in short-term repetition priming: Is primed evidence discounted, even when it reliably indicates the correct answer? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 257–281. doi: 10.1037/0278-7393.34.2.257
- Westmacott, R., & Moscovitch, M. (2003). The contribution of autobiographical significance to semantic memory. *Memory & Cognition*, *31*(5), 761–774. doi: 10.3758/BF03196114
- Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology and Aging*, *13*, 256–266. doi: 10.1037/0882-7974.13.2.256