

# Oscillatory patterns in temporal lobe reveal context reinstatement during memory search

Jeremy R. Manning <sup>\*</sup>, Sean M. Polyn <sup>†</sup>, Gordon Baltuch <sup>‡</sup>, Brian Litt <sup>§</sup>, and Michael J. Kahana <sup>¶</sup>

<sup>\*</sup> Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, USA, <sup>†</sup> Department of Psychology, Vanderbilt University, Nashville, USA, <sup>‡</sup> Department of Neurosurgery, University of Pennsylvania School of Medicine, Philadelphia, USA, <sup>§</sup> Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, USA, and <sup>¶</sup> Department of Psychology, University of Pennsylvania, Philadelphia, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America.

**Psychological theories of memory posit that when people recall a past event, they not only recover the features of the event itself, but they also recover information associated with other events that occurred nearby in time. The events surrounding a target event, and the thoughts they evoke, may be considered to represent a context for the target event, helping to distinguish that event from similar events experienced at different times. The ability to reinstate this contextual information during memory search has been considered a hallmark of episodic, or event-based, memory. We sought to determine whether context reinstatement may be observed in electrical signals recorded from the human brain during episodic recall. Analyzing electrocorticographic recordings taken as 69 neurosurgical patients studied and recalled lists of words, we uncovered a neural signature of context reinstatement. Upon recalling a studied item, we found that the recorded patterns of brain activity were not only similar to the patterns observed when the item was studied, but were also similar to the patterns observed during study of neighboring list items, with similarity decreasing reliably with positional distance. The degree to which individual patients displayed this neural signature of context reinstatement was correlated with their tendency to successively recall neighboring list items. These effects were particularly strong in temporal lobe recordings. Our findings show that recalling a past event evokes a neural signature of the temporal context in which the event occurred, thus pointing to a neural basis for episodic memory.**

episodic memory | oscillations | contiguity

Abbreviations: ECoG, electrocorticography

The pivotal distinction between memory for facts (semantic memory) and memory for episodes or experiences (episodic memory) has been argued to reflect, at least in part, the reinstatement of a gradually changing context representation that reflects not only external conditions, but also an ever-changing internal context state [1, 2]. According to this view, the unique quality of episodic memory is that in remembering an episode we partially recover its associated mental context, and that this context information conveys some sense of when the experience took place, in terms of its relative position along our autobiographical timeline.

A number of laboratory memory tasks rely on episodic memory, including experimenter-cued tasks (e.g. item recognition and cued recall) and self-cued tasks (e.g. free recall). Performing these episodic memory tasks requires distinguishing the current list item from the rest of one's experience. According to early theories of episodic memory (e.g. [3, 4]) context representations are composed of many features that fluctuate from moment to moment, gradually drifting through a multidimensional feature space. These contextual features may reflect environmental cues, recently studied items, participants' internal mental states, or may evolve randomly over time. During recall, the context representation forms part of the retrieval cue, enabling us to distinguish list items from non-list items. Understanding the role of context in memory processes is particularly important in tasks such as free recall, where the retrieval cue is "context" itself.

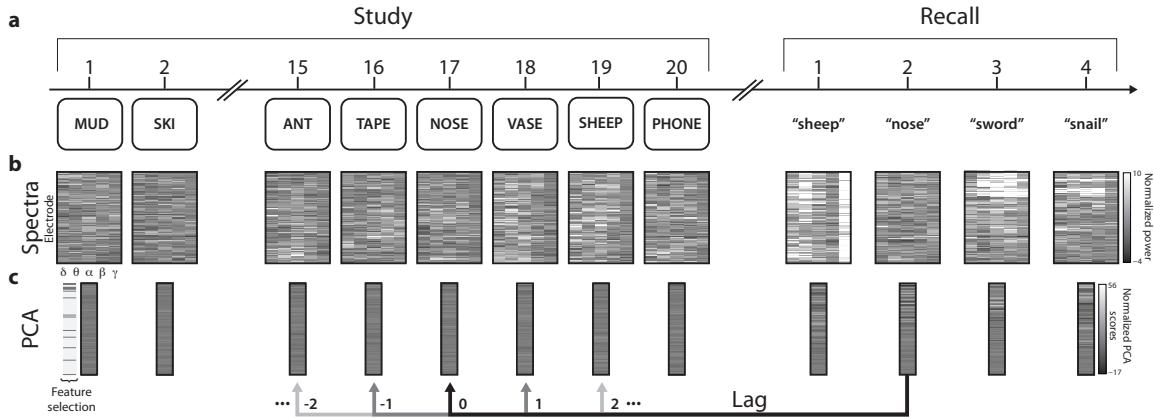
Recent neurocomputational models of episodic memory [5, 6] suggest that contextual reinstatement underlies the *contiguity effect* — people's tendency to successively recall items that were presented in nearby positions on a studied list [7]. Behavioral studies of memory show that, for a given class of memories, the contiguity effect can span many other intervening memories [8, 9, 10]. This result is difficult to explain according to the view that contiguity arises from direct item-to-item associations that are established within a few seconds, as suggested by other classes of psychological and neurobiological theories [11, 12, 13, 14]. The contiguity effect is an example of temporal clustering, which is perhaps the dominant form of organization in free recall.

Although this behavioral evidence provides indirect support for context-based theories of memory, there is no direct neurophysiological evidence for contextual reinstatement. To test the context reinstatement hypothesis, we studied 69 neurosurgical patients who were implanted with subdural electrode arrays and depth electrodes during treatment for drug-resistant epilepsy. As electrocorticographic (ECoG) signals were recorded, the patients volunteered to participate in a *free recall* memory experiment, in which they studied lists of common nouns and then attempted to recall them verbally in any order following a brief delay.

## Results

The recorded ECoG signals simultaneously sample local field potentials throughout the brain, and can be analyzed in terms of specific time-varying oscillatory components of neural activity. Such components have been implicated in memory encoding and retrieval processes [15, 16, 17, 18, 19, 20] and in the representations of individual stimuli [21]. For each study and recall event we analyzed these oscillatory components across all recording electrodes (Fig. 1a,b). We constructed a matrix containing, for each electrode, measurements of mean oscillatory power in five frequency bands ( $\delta$ : 2 – 4 Hz,  $\theta$ : 4 – 8 Hz,  $\alpha$ : 8 – 12 Hz,  $\beta$ : 12 – 30 Hz, and  $\gamma$ : 30 – 99 Hz) during each study event (200 to 1600 ms relative to each word's appearance on screen) and recall event (-600 to 200 ms relative to vocalization). We then used principal components analysis (PCA) to distill these highly correlated neural features into

## Reserved for Publication Footnotes



**Fig. 1.** Illustration of behavioral and electrophysiological methods. **a**, After studying a list of 20 words and performing a brief distraction task, a participant recalls as many words as he can remember, in any order. **b**, During each study presentation and just prior to each recall event, we calculate the  $z$ -transformed oscillatory power at each recording electrode in each of five frequency bands ( $\delta$ : 2–4 Hz,  $\theta$ : 4–8 Hz,  $\alpha$ : 8–12 Hz,  $\beta$ : 12–30 Hz, and  $\gamma$ : 30–99 Hz). **c**, We use principal components analysis (PCA) to find a smaller number of orthogonal components that jointly account for a large proportion of the variation in the data shown in panel **b**. We select those components that show significant positive autocorrelation (a defining feature of temporal context) during the study phase of the experiment. We then compute the similarity (normalized dot product) between the feature vectors of each recall event (e.g., “nose”) and the feature vectors associated with the corresponding study event (lag = 0), as well as the similarity of the recall event to surrounding study events with varying lags.

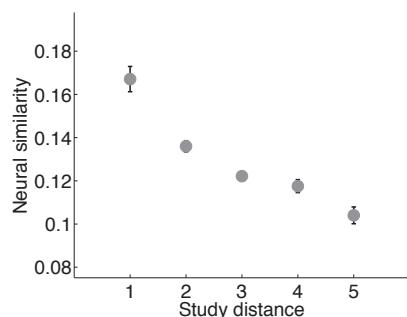
a smaller number of orthogonal components (Fig. 1c). In this way, each component reflects a linear combination of the power in each frequency band across all recording electrodes such that the pairwise neural similarities between events are preserved.

Context-based models conceive of context as a representation that integrates incoming information with a long time constant [22], leading to the prediction that the representation of temporal context evolves gradually as the experiment progresses [23]. We asked whether the neural recordings supported a gradually changing representation of context by regressing, for each participant, the mean similarity between the neural vectors (in principal component space) on their positional distance in the studied list (Fig. 2). The similarity in recorded activity during each pair of word presentations decreased with the positional distance between the presentations ( $t$ -test on distribution of  $t$ -values from the regressions:  $t(64) = -9.31, p < 10^{-10}$ ), indicating that the ECoG recordings evolve gradually over the course of the studied lists. Whereas this gradually changing neural representation is consistent with context-based models, such a result would also be expected to arise due to other autocorrelated neural processes that lack the rich dynamics implied by context-based theories of memory. To determine whether this gradually changing neural representation reflects the contexts in which list items

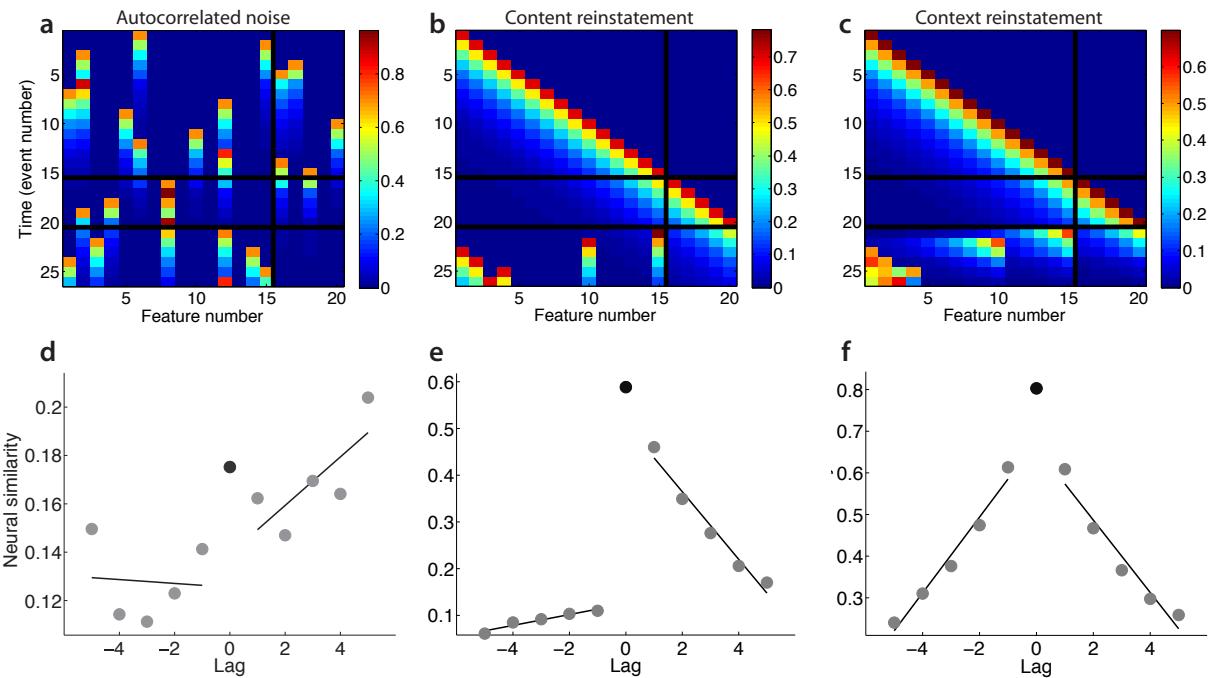
were studied, we selected PCA-derived features from study events that showed significant positive autocorrelations (see *Methods*) – a defining feature of temporal context – for further analysis. In the remainder of this manuscript, we refer to the set of autocorrelated principal components as *feature vectors*.

To test whether the gradually changing neural representation we identified is reinstated during recall, we compared feature vectors recorded during each study and recall event. First we identified the serial position (on the studied list) of each correctly recalled word. If neural activity during study is reinstated during recall, then the neural activity recorded during a given recall event should be more similar to activity recorded during the study event for the same word than during study events for other words (Figs. 3e,f). This finding would not be expected if the neural activity we measured did not contain content or context information (Fig. 3d). For each correctly recalled word (e.g., “nose” in Fig. 1a), we calculated the similarity between the feature vector associated with the recall event and the feature vectors associated with each of the studied items (e.g., ANT, TAPE, NOSE, VASE, SHEEP), where similarity is defined as the normalized dot product between the feature vectors (the vectors were normalized to have unit length before the dot product was performed). We assigned each studied item a lag (positional distance) relative to the recalled item (e.g., VASE has a lag of +1 to “nose,” ANT has a lag of -2 to “nose,” and NOSE has a lag of 0 to “nose”). We found that the mean neural similarity at lag = 0 was significantly greater than the mean neural similarity at other lags (Fig. 4a, paired-sample  $t$ -test across 39 participants with at least 5 autocorrelated features;  $t(38) = 3.10, p = 0.004$ ). This result would arise if the signal represents either content (the list words themselves) or context (the cues surrounding the items).

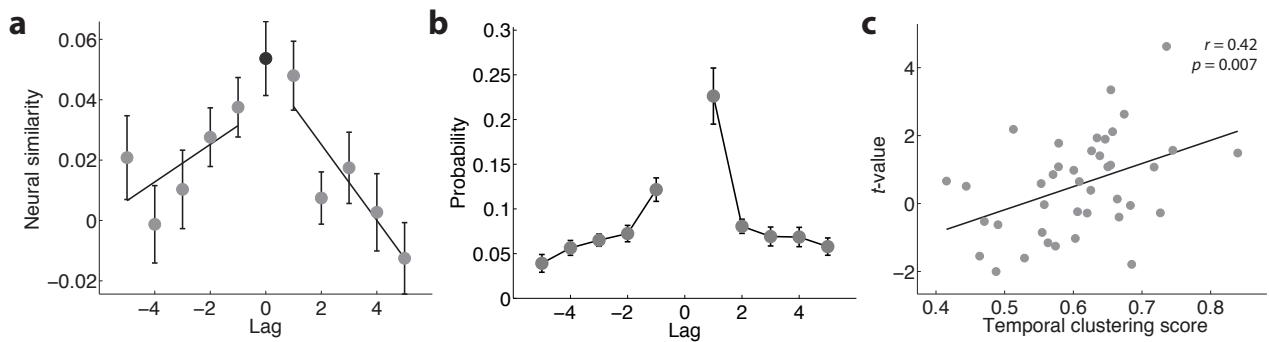
To distinguish between content and context reinstatement we compared the feature vectors associated with each recall event with the feature vectors associated with the neighbors of the recalled word in the study sequence. Context-based models predict that similarity between feature vectors should decrease as a function of absolute lag in both the forward (positive) and backward (negative) directions [22]. For each participant we regressed the mean neural similarity between feature



**Fig. 2.** Evolution of ECoG activity as participants study lists of words. Mean neural similarity is shown as a function of study distance (difference in serial position) between pairs of presented words. Error bars denote  $\pm 1$  SEM.



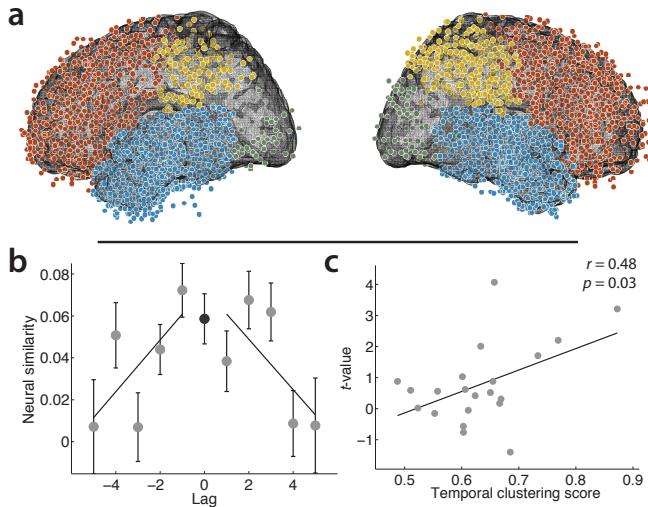
**Fig. 3.** Predicted neural similarity as a function of lag according to three models. **a-c.** These panels show the pattern of activations for a simulated 20-neuron neural network as a 15 item list is studied. Events 1–15 of each matrix show activations after each item is presented. Events 16–20 show activations as distracting items are presented. Events 21–26 show activations as items 15, 10, 1, 2, 4, and 3 are recalled. In each simulation, a single neuron is activated during each experimental event. Once activated, a neuron's activity decays gradually; thus multiple neurons may be active at a given time. **a.** For the *Autocorrelated noise* simulation, each experimental event activates a random neuron, irrespective of which item is being presented or recalled. **b.** For the *Content reinstatement* simulation, each neuron is activated by a single item or distractor (neurons 1–15 represent items; 16–20 represent distractors). Only content information (specific to a single item) is reinstated during recall. **c.** The *Context reinstatement* simulation is similar to that shown in panel **b**, but here we simulate context reinstatement during recall. **d-f.** These panels show the average expected neural similarity between the pattern of activity during study and recall as a function of lag. Each simulation used the same presented and recalled items that were included in our data analyses (Fig. 4). See *Supplemental materials* for further details on the simulations.



**Fig. 4.** A neural signature of temporal context reinstatement. **a.** Neural similarity between the feature vector corresponding to recall of a word from serial position  $i$  and study of a word from serial position  $i + \text{lag}$  (black dot denotes study and recall of the same word, i.e., lag = 0). **b.** Participants tend to successively recall neighboring study items (the contiguity effect). Here, we show the probability of recalling an item from serial position  $i + \text{lag}$  immediately following an item from serial position  $i$ , conditional on the availability of an item in that list position for recall. Error bars in panels **a** and **b** denote  $\pm 1$  SEM. **c.** Participants exhibiting greater context reinstatement also exhibited more pronounced contiguity effects. Here the  $t$ -value associated with the regressions in panel **a** serves as a measure of the degree context reinstatement for each participant. (Only the regressions for negative lags were used, as the regressions for positive lags are not expected to distinguish between content and context reinstatement; Fig. 3.) The temporal clustering score measures the degree to which responses were clustered on the basis of their temporal contiguity at study (see *Supplemental analysis methods*).

vectors on lag, separately for positive and negative lags (two regressions were performed for each participant). Each regression yielded a  $t$ -value associated with the slope ( $\beta$  coefficient) of the fitted line. Consistent with the context-reinstatement hypothesis,  $t$ -tests on the distributions of  $t$ -values across participants indicated that similarity decreased with absolute lag in both the positive ( $t(38) = -3.63, p = 0.0008$ ) and negative ( $t(38) = -2.42, p = 0.02$ ) directions (Fig. 4a).

We conducted three neural network simulations to contrast the predictions of three models of the observed feature vector dynamics (Figs. 3, S2). In the *autocorrelated noise* model (Fig. 3a,d) neural activity evolves randomly over time, irrespective of what is happening in the experiment. In the *content reinstatement* model (Fig. 3b,e) each neuron represents a different word; a neuron is activated if its associated word is presented or recalled. In the *context reinstatement* model (Fig. 3c,f) each neuron also represents a different



**Fig. 5.** Evidence for context reinstatement in the temporal lobe. **a.** Each dot marks the location of a single electrode from our dataset in Montreal Neurological Institute (MNI) space. We divided our dataset into four regions of interest: temporal lobe (blue; 1,815 electrodes), frontal lobe (red; 1,737 electrodes), parietal lobe (yellow; 512 electrodes), and occipital lobe (green; 138 electrodes). **b – c.** These panels are in the same format as Figures 4a and c, but reflect data from temporal lobe electrodes only. Of the temporal lobe electrodes, 13.9% were located in hippocampus, 27.0% were located in medial temporal lobe structures (excluding hippocampus), and the remaining 59.1% were located in other temporal lobe regions.

word. We simulate context reinstatement by activating not only the neuron associated with the word being recalled, but also other neurons that were active at the time the recalled word was studied. Our simulations show that only the context reinstatement model predicts that neural similarity will decrease substantially with absolute lag in both the positive and negative directions, as observed in the neural data (see *Supplemental materials* for further details and discussion).

The decrease in neural similarity with absolute lag elegantly mirrors the contiguity effect – people’s striking tendency to make transitions to neighboring items rather than remote ones, as seen in behavioral data for the same participants (Fig. 4b). Consistent with the hypothesis that the contiguity effect arises due to context reinstatement [24, 6, 5], participants with stronger neural signatures of context reinstatement exhibited more pronounced contiguity effects than did participants with weaker reinstatement effects ( $r = 0.42$ ,  $p = 0.007$ ; Fig. 4c). In recalling a list item, people not only reinstate that item’s representation, as has been recently documented [25, 26] but they also revive the brain activity associated with neighboring items. Further, the degree of this neural context reinstatement effect predicts the tendency of an individual participant to recall neighboring list items successively during memory search.

Having identified a neural signature of context reinstatement, we next asked whether this phenomenon could be localized to one or more brain regions. For example, recent work has given rise to the hypothesis that the medial temporal lobe [27, 28, 29, 30, 23, 31, 32, 33] and prefrontal cortex [32, 22, 34] are critically involved in the maintenance and updating of temporal context. To test for regional specificity of context reinstatement, we repeated our test for neural context reinstatement using electrodes from each of the following regions of interest (Fig. 5a): temporal lobe (including the hippocampus and medial temporal lobe), frontal lobe (including prefrontal cortex), parietal lobe, and occipital lobe. We found

that neural activity recorded from temporal lobe electrodes exhibited a decrease in similarity with increasing absolute lag in both the positive and negative directions (Fig. 5b; positive:  $t(20) = -2.20$ ,  $p = 0.04$ ; negative:  $t(20) = -2.82$ ,  $p = 0.01$ ). As in the whole brain analysis, the neural signature of context reinstatement in the temporal lobe was significantly correlated with the temporal clustering of participants’ recalls (Fig. 5c;  $r = 0.48$ ,  $p = 0.03$ ). The frontal lobe exhibited a weak neural signature of context reinstatement that trended towards significance (positive:  $t(20) = -2.85$ ,  $p = 0.01$ ; negative:  $t(20) = -1.54$ ,  $p = 0.14$ ). However, this frontal signature of context reinstatement was not correlated with temporal clustering of participants’ recalls ( $r = -0.08$ ,  $p = 0.73$ ). Our findings in the parietal and occipital lobes were inconclusive due to insufficient data.

Our inability to find neural signatures of context reinstatement in extratemporal brain regions does not necessarily mean that those regions are not involved in context reinstatement. Furthermore, our analysis does not distinguish between structures contained within the regions of interest we examined. Thus an important goal of future work will be to more precisely localize the neural machinery underlying context reinstatement.

## Discussion

The preceding analyses demonstrate that when recalling an item, the pattern of neural activity exhibits graded similarity to the neural activity measured during the encoding of items studied in neighboring list positions. Furthermore, the strength of this neural contiguity effect tracks the behavioral contiguity effect in free recall: participants who exhibit a stronger tendency to make transitions among neighboring items during recall also exhibit a stronger relation between neural similarity and absolute lag. This pattern of results is exactly what one would predict on the basis of retrieved context theories of episodic memory [1, 24, 5, 6]. These theories posit that a gradually changing contextual state becomes associated with each experienced event, and that recalling an event revives the contextual state associated with the original experience. This retrieved context, in turn, activates other memories that were associated with similar contexts, producing the contiguity effect seen in recall tasks (Fig. 4b). The present findings provide critical neurobiological evidence in support of context reinstatement by showing that remembering an item reinstates the patterns of distributed oscillatory activity associated with surrounding (contextual) items from the original study episode. This neural signature of context reinstatement was observed both for the whole brain analysis and for recordings taken only from the temporal lobe.

Retrieved context models are one of a broader class of episodic memory models that provide insight into our finding that patterns of neural activity are reinstated prior to recall. For example, by rehearsal-based models, words are rehearsed after they are presented, and more recently presented items are more likely to be rehearsed than temporally distant items. Rehearsal-based models have been shown to be difficult to distinguish from context-based models [35, 36], likely because a context-based mechanism is necessary to explain the pattern of rehearsals made in a free-recall task. If associations are formed between items that are rehearsed successively, then activating the representation of an item prior to recall would be expected to activate the representations of other items rehearsed nearby in time (consistent with the neural signature of context reinstatement we observe).

To assess the extent to which variability in rehearsal strategies across participants might explain the observed cor-

relation between the neural and behavioral contiguity effects (Fig. 4c) we performed an analysis of the neural correlates of the primacy effect in our data. It has been well established that rehearsal is associated with enhanced recall for early list items (i.e., the primacy effect [37, 38, 39, 40, 41]). Thus if our basic findings were driven by rehearsal, one might expect that participants exhibiting strong neural contiguity should also show a strong primacy effect. We observed no significant correlation between primacy and neural contiguity ( $r = 0.13, p = 0.42$ ; see *Supplemental analysis methods*), suggesting that rehearsal during study *per se* is unlikely to account for our findings. While rehearsal is one of the hypothesized mechanisms underlying primacy, we recognize that other factors, such as enhanced attention to early list items [40, 42] may also contribute to primacy. Nonetheless, it is clear that the mechanisms underlying the primacy effect are unrelated to the neural contiguity effect we observe.

Modern psychological and neuroscientific investigations are still grappling with basic questions regarding how the human brain establishes continuity in a rapidly changing environment, and how our memory system revives prior states of the world. Recent neurocomputational models of human memory [1, 6, 5] posit that continuity is provided by a context representation that changes gradually over time as a consequence of the integration of present and past events. The current state of context is assumed to become associated with each newly experienced event, such that reminders of the event retrieve the event's associated context. This notion is consistent with Tulving's contention that episodic memory retrieval is like mental time travel, in that when we remember the past, many details of the prior experience are retrieved along with the desired material [2]. By showing that a component of the neural activity retrieved during memory search shows graded similarity to the brain states observed during the study of neighboring stimuli, we provide neural evidence for temporal context reinstatement in humans.

## Materials and Methods

**Participants.** We tested 69 patients with drug-resistant epilepsy who had arrays of subdural and/or depth electrodes surgically implanted for one to four weeks to localize the sites of seizure onset (see Tab. S1). The clinical team determined

the placement of these electrodes with the goal of localizing suspected epileptogenic foci and identifying functional regions to be avoided in surgery (see **Supplemental recording methods** for details on recording methods). Our research protocol was approved by the appropriate institutional review boards and informed consent was obtained from the participants and their guardians. Data were collected as part of a long-term multicenter study with previously published articles describing separate analyses conducted on subsets of these data [43, 44, 19].

**Behavioral methods.** Participants studied lists of 15 or 20 high-frequency nouns for a delayed free recall task. Following a fixation cue, the computer displayed each word for 1600 ms followed by a 800–1200 ms blank inter-stimulus interval. Each word was displayed at most once within a single testing session. For 18 s following list presentation, participants solved a series of single-digit addition problems of the form  $A + B + C = X$ . Participants were then given 45 s to recall list items in any order. Vocal responses, digitally recorded during the trial, were scored for subsequent analysis. Participants recalled  $22.7 \pm 1.0\%$  (mean  $\pm$  SEM) of the studied words. Repetitions and incorrect recalls ( $32.4 \pm 2.6\%$  of all responses) were excluded from our analyses, as were responses that occurred within 1 s of a prior vocalization.

**Data analysis.** We measured oscillatory power in the ECoG recordings by applying a Hilbert transform to the Butterworth-bandpassed signal in each of five frequency bands (see **Results**). To reduce edge artifacts, we computed power at each frequency for the entire recording session before parsing the recordings into experimental events. Before applying PCA to the frequency  $\times$  electrode matrices (Fig. 1b) we  $z$ -transformed power values relative to the distribution of all events in the recording session (the  $z$ -transformation was performed independently for each frequency-electrode pair). We used the Kaiser criterion to choose, for each participant, the principal components that explained a substantial proportion of the variance [45]. We identified principal components that changed gradually during the study period (see **Supplemental analysis methods**). Features that met these criteria were identified in 132 (out of 144) recording sessions. We further excluded an additional 68 sessions in which fewer than five candidate context features were identified. This threshold was chosen to balance two factors: first, we wanted to ensure that the feature vectors were of high enough dimensionality that it would be possible to observe neural signatures of context reinstatement (Fig. 4a) for each participant; second, we wanted to maximize the amount of data included in our analysis. We report the mean number of features selected for each participant in Table S1. To prevent selection bias, recall events were not used in the feature selection process.

**ACKNOWLEDGMENTS.** We thank D. Brainard, M. Howard, J. Jacobs, L. Lohnas, L. Ungar, and C. Weidemann for useful discussions. We also thank E. Rosenberg for help testing patients and J. Miller for helping to organize patient information. This work was supported by National Institutes of Mental Health grants MH55687 and F31MH088118, and a grant from the Dana Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of our supporting organizations.

1. Bower, G. H. (1972) in *Coding Processes in Human Memory*, eds. Melton, A. W & Martin, E. (John Wiley and Sons, New York), pp. 85–121.
2. Tulving, E. (1983) *Elements of Episodic Memory*. (Oxford, New York).
3. Estes, W. K. (1955) Statistical theory of spontaneous recovery and regression. *Psychological Review* 62, 145–154.
4. Anderson, J. R & Bower, G. H. (1972) Recognition and retrieval processes in free recall. *Psychological Review* 79, 97–123.
5. 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.
6. Sederberg, P. B, Howard, M. W, & Kahana, M. J. (2008) A context-based theory of recency and contiguity in free recall. *Psychological Review* 115, 893–912.
7. Kahana, M. J. (1996) Associative retrieval processes in free recall. *Memory & Cognition* 24, 103–109.
8. Howard, M. W & Kahana, M. J. (1999) Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25, 923–941.
9. Howard, M. W, Youker, T. E, & Venkatadass, V. (2008) The persistence of memory: Contiguity effects across several minutes. *Psychonomic Bulletin & Review* 15, 58–63.
10. Polyn, S. M, Erlikhman, G, & Kahana, M. J. (2011) Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal Experimental Psychology: Learning, Memory and Cognition* 37, 766–775.
11. Raaijmakers, J. G. W & Shiffrin, R. M. (1981) Search of associative memory. *Psychological Review* 88, 93–134.
12. Jensen, O & Lisman, J. E. (1998) An oscillatory short-term memory buffer model can account for data on the Sternberg task. *J. Neuroscience* 18, 10688–10699.
13. Lisman, J. (1999) Relating hippocampal circuitry to function: recall of memory sequences by reciprocal Dentate-CA3 interactions. *Neuron* 22, 233–242.
14. Rolls, E. T & Kesner, R. P. (2006) A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology* 79, 1 – 48.
15. Fell, J, Klaver, P, Lehnertz, K, Grunwald, T, Schaller, C, Elger, C. E, & Fernandez, G. (2001) Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience* 4, 1259–1264.
16. Paller, K. A & Wagner, A. D. (2002) Observing the transformation of experience into memory. *Trends in Cognitive Sciences* 6, 93–102.
17. Buzsáki, G. (2006) *Rhythms of the Brain*. (Oxford University Press, New York).
18. Osipova, D, Takashima, A, Oostenveld, R, Fernández, G, Maris, E, & Jensen, O. (2006) Theta and gamma oscillations predict encoding and retrieval of declarative memory. *J Neurosci* 26, 7532–7531.
19. Sederberg, P. B, Schulze-Bonhage, A, Madsen, J. R, Bromfield, E. B, Litt, B, Brandt, A, & Kahana, M. J. (2007) Gamma oscillations distinguish true from false memories. *Psychological Science* 18, 927–932.
20. Rutishauser, U, Ross, I, Mamelak, A, & Schuman, E. (2010) Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* 464, 903–907.
21. Jacobs, J & Kahana, M. J. (2009) Neural representations of individual stimuli in humans revealed by gamma-band ECoG activity. *Journal of Neuroscience* 29, 10203–10214.
22. Polyn, S. M & Kahana, M. J. (2008) Memory search and the neural representation of context. *Trends in Cognitive Sciences* 12, 24–30.
23. Manns, J. R, Howard, M. W, & Eichenbaum, H. (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56, 530–540.
24. Howard, M. W & Kahana, M. J. (2002) A distributed representation of temporal context. *Journal of Mathematical Psychology* 46, 269–299.
25. Polyn, S. M, Natu, V. S, Cohen, J. D, & Norman, K. A. (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* 310, 1963–1966.

26. Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 3, 96–101.
27. Kinsbourne, M. & Wood, F. (1975) Short-Term Memory and the Amnesic Syndrom, eds. Deutsch, D. D & Deutsch, J. A. (Academic Press, New York), pp. 258 – 291.
28. O'Keefe, J & Nadel, L. (1978) The hippocampus as a cognitive map. (Oxford University Press, New York).
29. Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172.
30. Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003) Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci U S A* 100, 2157 – 2162.
31. Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321, 1322 – 1327.
32. Jenkins, L. J. & Ranganath, C. (2010) Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience* In press.
33. Howard, M. W., Viskontas, I. V., Shankar, K. H., & Fried, I. (Submitted) Human medial temporal lobe ensembles maintain and retrieve a gradually-changing state of temporal context. Submitted.
34. Schacter, D. L. (1987) Memory, amnesia, and frontal lobe dysfunction. *Psychobiology* 15, 21–36.
35. Davelaar, E. J., Usher, M., Haarmann, H. J., & Goshen-Gottstein, Y. (2008) Postscript: Through TCM, STM shines bright. *Psychological Review* 115, 1116–1118.
36. Laming, D. (2006) Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32, 1146–1163.
37. Welch, G. B. & Burnett, C. T. (1924) Is primacy a factor in association-formation. *American Journal of Psychology* 35, 396–401.
38. Postman, L. & Phillips, L. W. (1965) Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology* 17, 132–138.
39. Glanzer, M. & Cunitz, A. R. (1966) Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior* 5, 351–360.
40. Atkinson, R. C & Shiffrin, R. M. (1968) In *The psychology of learning and motivation*, eds. Spence, K. W & Spence, J. T. (Academic Press, New York) Vol. 2, pp. 89–105.
41. Rundus, D. (1971) An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology* 89, 63–77.
42. Neath, I & Crowder, R. G. (1990) Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 316–327.
43. Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003) Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience* 23, 10809–10814.
44. Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., Tully, M. S., & Kahana, M. J. (2007) Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex* 17, 1190–1196.
45. Kaiser, H. F. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20, 141 – 151.

*Supplemental materials for*  
**Oscillatory patterns in temporal lobe reveal context  
reinstatement during memory search**

Jeremy R. Manning<sup>1</sup>, Sean M. Polyn<sup>2</sup>, Gordon Baltuch<sup>3</sup>, Brian Litt<sup>4</sup>,  
and Michael J. Kahana<sup>5</sup>

<sup>1</sup>Neuroscience Graduate Group,  
University of Pennsylvania, Philadelphia, USA

<sup>2</sup>Department of Psychology,  
Vanderbilt University, Nashville, USA

<sup>3</sup>Department of Neurosurgery and <sup>4</sup>Department of Neurology  
University of Pennsylvania School of Medicine, Philadelphia, USA

<sup>5</sup>Department of Psychology,  
University of Pennsylvania, Philadelphia, USA

## Overview

This document is comprised of four main sections. First we include a table containing basic information about each neurosurgical patient whose data we analyzed. We next provide a table summarizing several details of the recording setup for each of our collaborating hospitals, as well as a description of the recording methods used in our study. In the third section we provide additional details on several analyses referenced in the main paper. Finally, we provide a list of supplemental references cited in this document.

## Supplemental patient information

### A note on intracranial recordings

Despite the high data quality provided by human intracranial recordings, there are several factors one should consider when interpreting the results of any intracranial study of human epilepsy patients, including those we report in the present manuscript. First, whereas in animal studies electrodes are placed according to researchers' needs, the placements of implanted human electrodes are determined solely by clinical teams with the goal of localizing the seizure focus in order to ensure the best possible outcome for the patient. This means that, for some patients, the brain areas most relevant to a particular research question may receive little or no electrode coverage. In order to obtain adequate coverage of all relevant brain areas, we have analyzed data from many patients (Tab. S1). A second concern is that medications or recent seizures might change the electrophysiological properties of the brain. For this reason, we refrained from collecting data while the patients were on high dosages of pain medications or anti-epileptic drugs, or during the 6 hour period following any clinically significant seizure. A third issue is that the brain is known to re-wire itself to compensate for damage, including damage caused by epilepsy (Ribak and Dashtipour, 2002), which could lead to cognitive remapping. While we cannot control for cognitive remapping that may have occurred in individual participants, we have averaged our anatomical analyses over many patients; thus results due to remapping in one patient will average out in the population analyses. A fourth concern is that severe epilepsy can lead to cognitive impairment. To address this issue, we have analyzed data only from patients with scores on the Wechsler Intelligence and Wechsler Memory Scales within 1.5 standard deviations of the mean for their age group.

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
1	BW	33	F	R	64	3	20	1	15	55	2	26	12
2	BW	51	F	R	40	5	20	1	15	66	2	8	1
3	BW	32	M	R	32	2	15	3	39	212	5	9	4
4	BW	40	M	R	96	5.5	15	2	20	82	5	38	28
5	BW	44	M	R	16	1	15	2	20	58	1	12	19
6	BW	27	M	R	64	1.5	15	2	20	76	49	12	3
7	BW	38	M	R	104	9.33	15	3	30	136	3	30	15
8	CH	13	F	R	64	9	20	1	12	59	0	2	1
9	CH	12	F	R	104	19	20	1	15	39	0	0	2
10	CH	15	M	L	128	15.67	20	3	30	90	1	12	8
11	CH	17	M	R	64	4	20	3	45	178	20	27	17
12	CH	15	M	R	123	13	20	1	15	86	3	6	3
13	CH	11	M	R	104	0	20	2	30	104	3	2	3
14	CH	14	F	R	72	0	20	1	15	104	2	7	5
15	CH	8	F	R	86	6.5	20	2	30	159	5	18	10
16	CH	17	M	R	84	12	20	1	14	30	2	9	12
17	CH	17	M	L	124	10.5	20	4	60	116	2	104	13
18	CH	20	F	R	128	8.5	15	2	24	114	2	9	4
19	CH	14	M	R	94	6.67	15	3	30	94	0	18	13
20	CH	17	M	L	80	7	15	2	20	14	0	10	22
21	CH	19	F	R	125	8.5	15	2	17	47	2	10	1
22	CH	16	M	R	156	13	15	1	16	76	1	4	1

Continued on next page ...

ID	HOSP	AGE	SEX	HAND	ELC	FEA	L LEN	SES	LST	REC	REP	PLI	ELI
23	CH	12	M	L	83	5	15	2	20	52	9	12	25
24	CH	13	M	R	72	4.75	15	4	40	200	3	4	2
25	FR	33	M	R	98	8	20	1	9	43	0	0	0
26	FR	25	M	R	85	21	20	1	9	45	18	1	2
27	FR	31	M	L	56	4	20	1	9	24	0	4	0
28	FR	41	F	R	63	9	20	1	7	23	0	9	6
29	FR	34	F	L	40	3	20	1	7	38	3	5	3
30	FR	45	F	L	100	10	20	1	8	27	0	20	3
31	FR	46	F	L	14	0	20	1	1	4	3	0	5
32	FR	20	M	R	84	4	20	1	15	42	6	6	2
33	FR	53	F	L	41	5	20	1	15	49	26	21	21
34	FR	50	M	R	68	4	20	2	30	116	5	36	15
35	FR	28	M	L	112	11	20	1	15	37	1	1	1
36	FR	30	F	R	60	7	20	1	15	67	9	10	2
37	FR	37	F	L	30	6	15	1	20	65	84	73	42
38	FR	18	M	L	30	2	15	1	20	121	7	12	13
39	FR	23	M	L	58	3	15	4	56	281	98	32	29
40	FR	21	M	L	93	4	15	1	10	49	2	10	3
41	FR	28	F	R	86	7	15	1	10	36	4	4	6
42	FR	35	F	L	122	3	15	2	20	54	0	6	6
43	FR	37	F	L	52	2.5	15	4	35	161	38	29	28
44	FR	19	M	L	74	2	15	2	30	148	19	14	35
45	FR	41	F	R	30	3	15	1	15	15	0	3	38
46	FR	21	F	R	64	6	15	1	15	50	0	3	7
47	FR	43	F	R	56	0	15	1	15	23	5	11	59
48	FR	19	M	R	30	3	15	2	25	120	2	5	27
49	FR	21	M	R	70	5	15	5	53	408	41	8	194
50	FR	35	F	R	62	6	15	1	15	44	24	16	114
51	FR	25	M	R	84	5	15	2	30	145	5	3	96
52	FR	47	M	L	82	4	15	1	4	13	1	3	35
53	FR	45	F	R	88	4	15	1	10	43	11	3	0
54	TJ	25	M	R	62	0.67	15	3	48	232	3	6	1
55	TJ	40	F	R	94	3.25	15	4	64	164	8	54	43
56	TJ	39	M	L	56	2	15	1	16	53	1	8	20
57	TJ	34	F	R	111	4.3	15	10	154	513	7	110	24
58	TJ	44	M	R	125	7	15	1	13	31	1	7	6
59	UP	38	M	R	62	4.75	15	4	40	135	3	68	24
60	UP	30	M	R	86	3	15	2	20	54	5	24	21
61	UP	43	M	R	66	2.33	15	3	18	31	22	12	33
62	UP	36	M	R	88	5.75	15	4	40	70	6	114	50
63	UP	25	M	R	62	3	15	4	40	135	2	1	2
64	UP	18	F	R	76	7.33	15	3	30	104	5	6	3
65	UP	27	F	R	48	2	15	2	32	104	2	43	20
66	UP	55	F	L	80	1.5	15	2	32	81	11	61	24
67	UP	18	M	A	100	2.33	15	3	48	253	7	8	3
68	UP	38	F	R	86	6	15	1	16	48	14	3	73
69	UP	40	M	R	58	5.75	15	4	64	304	1	14	9

**Table S1: Patient and task information.** This table provides the hospital (HOSP) at which each patient's data were collected (see Tab. S2), as well as each patient's age (AGE), gender (SEX), handedness or language mapping (HAND), number of implanted electrodes (ELC), and mean number of features selected for analysis across all sessions for that patient (FEA). Information about the task includes the list length (L LEN) used for each participant, number of testing sessions (SES), and the number of lists each participant encountered across all sessions (LST). Performance information includes the total number of correct recalls across all lists (REC), the total number of repeated recalls (REP), and the total number of incorrect recalls, which includes recalls of previously presented items (prior list intrusions, or PLIs) and recalls of items which were never presented (extra-list intrusions, or ELIs). In total the 69 patients contributed 5,299 electrodes and 739 selected features, studying 29,030 items presented in 1,790 lists.

## Supplemental recording methods

Subdural grids or depth electrodes (Ad-Tech Inc., Racine Wisconsin) were implanted by neurosurgical teams solely for clinical purposes. The locations of the electrodes were determined by means of co-registered post-operative computed tomography and pre-operative magnetic resonance imaging (MRI), or from post-operative MRIs, by an indirect stereotactic technique and converted into Montreal Neurological Institute (MNI) coordinates. Electrocorticographic (ECoG) signals were recorded referentially using a Bio-Logic, XLTek, Neurofile, or Nicolet EEG digital video-EEG system. Depending on the amplifier, signals were sampled at 200, 256, 500, 512, or 1024 Hz. Several hospitals applied band-pass filters to the recorded signals prior to writing to disk (Tab. S2). Where applicable, frequencies outside of the filtered range were excluded from further analysis. Data were subsequently notch-filtered with a Butterworth filter with zero phase distortion at 50 or 60 Hz to eliminate electrical line and equipment noise. ECoG signals and behavioral events were aligned using synchronization pulses sent from the testing computer (mean precision < 4 ms).

HOSPITAL	HOSP	LOWER	UPPER
Brigham & Women's Hospital, Boston, MA	BW	0.5 Hz	60 Hz
Children's Hospital Boston, Boston, MA	CH	0.3 Hz	50 Hz
University Hospital of Freiburg, Freiburg, Germany	FR	0.1 Hz	100 Hz
Thomas Jefferson University Hospital, Philadelphia, PA	TJ	$-\infty$	$\infty$
Hospital of the University of Pennsylvania, Philadelphia, PA	UP	$-\infty$	$\infty$

Table S2: **Bandpass filters employed by our collaborating hospitals.** Hospital codes (HOSP) are referenced in Table S1. The LOWER and UPPER columns denote the lower and upper limits of the bandpass filters, respectively. Frequencies outside of the bandpassed range were excluded from further analysis.

## Supplemental analysis methods

### Quantifying the contiguity effect

Figure 4c depicts an analysis relating the neural reinstatement effect to the recall behavior of the participants. Specifically, we show that participants showing stronger neural reinstatement effects tend to exhibit a stronger contiguity effect (whereby neighboring list items tend to be recalled successively). The contiguity effect is measured using the temporal clustering score, an analysis technique described previously (Polyn et al., 2009). The temporal clustering score is calculated as follows.

For each recall transition we create a distribution of temporal distances between the just-recalled word and the set of words that have not yet been recalled. These distances are simply the absolute value of the difference between the serial position of the just-recalled word and the set of not-yet-recalled words. A percentile score is generated by comparing the temporal distance value corresponding to the next item in the recall sequence with the rest of the distribution. Specifically, we calculate the proportion of the possible distances that the observed value is less than, since strong temporal clustering will cause observed lags to be smaller than average. As is often the case, when there is a tie, we score this as the percentile falling halfway between the two items. If the participant always chose the closest temporal associate (which is only possible for pure serial recall in the forward or backward direction), then the temporal clustering score would yield a value of 1 (as there would never be an opportunity for a tie). A value of 0.5 indicates no effect of temporal clustering. Each patient was assigned a temporal clustering score by taking the average of the percentile scores across all observed recall transitions.

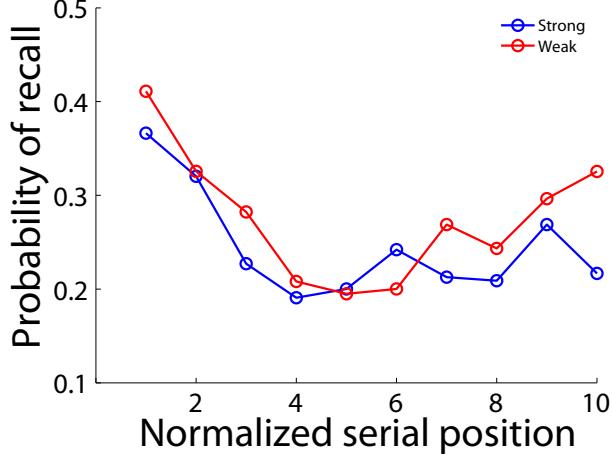


Figure S1: Serial position curves. The average serial position curve for participants exhibiting the top 50% strongest neural signatures of context reinstatement (by  $t$ -value) is shown in blue. The average serial position curve for participants exhibiting the bottom 50% strongest neural signatures of context reinstatement is shown in red. The serial positions have been normalized such that all participants, regardless of whether they studied 15- or 20-word lists, are shown in the same curves. Normalized serial positions 1 – 3 (i.e., primacy positions) correspond to absolute serial positions 1 – 3. Normalized serial position 4 (i.e., intermediate position) reflects the average probabilities of recall for items in serial positions 4 – 9 (15-word lists) or 4 – 14 (20-word lists). Normalized serial positions 5 – 10 (i.e., recency positions) correspond to the last 6 list items.

## Quantifying the primacy and recency effects

The primacy and recency effects refer to an enhancement in memory for early and late list items, respectively, as compared with memory for intermediate list items (Deese and Kaufman, 1957; Murdock, 1962). The number of items that show a boost in memorability due to primacy or recency is relatively invariant to changes in list length; the primacy effect generally affects the first few items, while the recency effect generally affects the last six or so items (Murdock, 1962). In order to measure the strength of the primacy effect, we labeled the first three serial positions on each list as primacy positions and the last six serial positions as recency positions. The remaining positions were labeled as intermediate list positions (i.e., items 4 – 9 for 15-word lists, or items 4–14 for 20-word lists). We then measured the strength of the primacy effect for each participant by dividing their mean probability of recalling items from primacy positions by their mean probability of recalling items from intermediate list positions. The main text reports that the neural signature of context reinstatement ( $t$ -value) is not correlated with the strength of the primacy effect ( $r = 0.13, p = 0.42$ ).

We also performed an analogous analysis to test whether the neural signature of context reinstatement was influenced by the factors underlying the recency effect. We measured the strength of the recency effect for each participant by dividing their mean probability of recalling items from recency positions by their mean probability of recalling items from intermediate list positions. The neural signature of context reinstatement is not correlated with the strength of the recency effect ( $r = 0.13, p = 0.40$ ). Mean serial position curves for participants showing strong (top 50%) and weak (bottom 50%) neural signatures of context reinstatement (by  $t$ -value) are shown in Figure S1. As shown in the figure, the primacy effect, recency effect, and overall probability of recall are roughly conserved across the two groups of participants.

## Simulations

We conducted three neural network simulations (Figs. 3, S2) that predict the expected outcome of our test for context reinstatement under various model assumptions. As described in the main text, the *autocorrelated noise* model has neural activity evolve randomly over time, irrespective of what is happening in the experiment. The *content reinstatement* model has each neuron represent a different word; a neuron is activated if

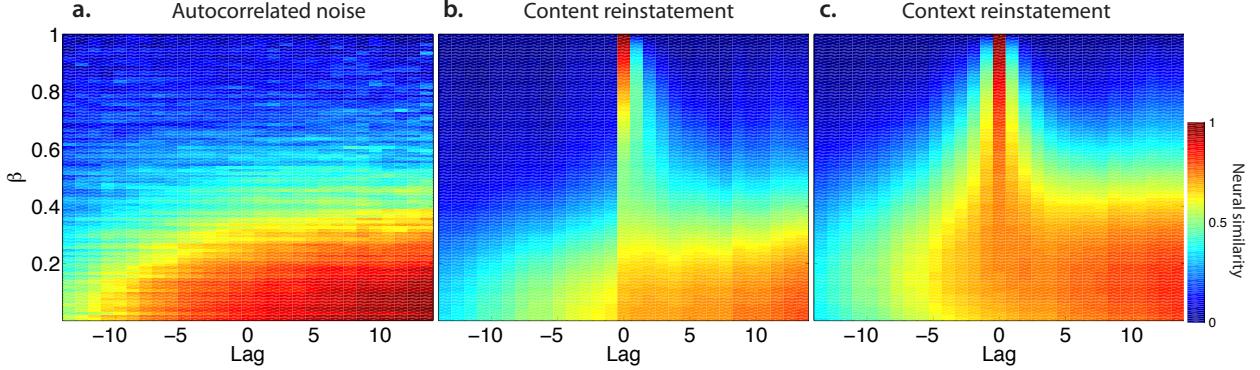


Figure S2: Simulated neural similarity as a function of lag and drift rate ( $\beta$ ) given (a) no content or context information in the neural recordings (Fig. 3a,d), (b) content reinstatement without context reinstatement (Fig. 3b,e), and (c) context reinstatement (Fig. 3c,f). Similarity is computed as the normalized dot product between the simulated feature vector after the recall of the  $i^{\text{th}}$  word and the feature vector corresponding to presentation of word  $i + \text{lag}$ . The first dimension (initialized to 1 prior to the start of the simulation) was ignored for the similarity calculations. Simulation results in Figure 3 in the main text used  $\beta = 0.7$  (this choice was motivated by previously reported simulation results (Polyn et al., 2009)).

its associated word is presented or recalled. The *context reinstatement* model also has each neuron represent a different word. We simulate context reinstatement by activating not only the neuron associated with the word being recalled, but also other neurons that were active at the time the recalled word was studied.

For all three simulations we define an activity vector,  $\mathbf{f}$ , that defines the pattern of activation across the network. Each neuron in the network takes on a value between 0 (inactive) and 1 (maximally active). Let  $\mathbf{f}_i$  denote the state of  $\mathbf{f}$  after the  $i^{\text{th}}$  experimental event (i.e., a study presentation, distracting task, or recall). Our main analysis entails selecting autocorrelated components of neural activity as the candidate context representation (see *Results*). After this feature selection, the feature vectors we analyze are autocorrelated – a property we need to take into account in our simulations. In particular,

$$\mathbf{f}_i = \rho_i \mathbf{f}_{i-1} + \beta \mathbf{w}_i,$$

where  $\beta$  is a constant;  $\rho_i$  is a function of  $\mathbf{f}_{i-1}, \mathbf{w}_i$ , and  $\beta$  (with  $0 \leq \rho_i, \beta \leq 1$ ); and  $\mathbf{w}_i$  is the pattern of neural activity specifically evoked by the  $i^{\text{th}}$  experimental event (see Polyn and Kahana, 2008, for details). In this way, the neural activity measured after a given experimental event (e.g. presentation of the 5<sup>th</sup> list item) is a recency-weighted blend of the activity evoked by previous experimental events (e.g. activity evoked by presentations of items 5, 4, 3, 2, and 1). We initialize  $\mathbf{f}_0$  by setting the activation of the first neuron to 1 and the activations of the other neurons to 0. We then simulate different experimental events by adjusting  $\mathbf{w}_i$  according to the particular rules of each model. We ensure that  $\mathbf{f}_i$  is always of unit length by setting

$$\rho_i = \sqrt{1 + \beta^2[(\mathbf{f}_{i-1} \cdot \mathbf{w}_i)^2 - 1]} - \beta(\mathbf{f}_{i-1} \cdot \mathbf{w}_i).$$

For the autocorrelated noise model, each  $\mathbf{w}_i$  is set to a vector of 0's, plus a 1 in a single random position. In this way, each  $\mathbf{w}_i$  activates one of the neurons in the network at random. As shown in Figure S2a, for  $\beta < 0.5$ , similarity between  $\mathbf{f}_i$  during presentation and  $\mathbf{f}_j$  during recall increases as a function of  $i$ . This is because, by definition, an autocorrelated signal measured at times  $t$  and  $t + \Delta$  becomes more similar as  $\Delta \rightarrow 0$ . For  $\beta > 0.5$ , similarity as a function of lag flattens out, since as  $\mathbf{f}_i$  is dominated by  $\mathbf{w}_i$  the average similarity between  $\mathbf{f}_i$  and  $\mathbf{f}_j$  approaches the expected similarity between two independent draws of  $\mathbf{w}_i$ .

For the content reinstatement model,  $\mathbf{w}$  is set differently depending on the type of experimental event. In this model, each neuron is assigned a different word or distractor. During presentation of study items or distractors,  $\mathbf{w}_i$  is set to a vector of all 0's except for a 1 in the position of the neuron representing the item or distractor being presented. During recall of the  $j^{\text{th}}$  presented item, we set  $\mathbf{w}_i = \mathbf{w}_j$ . As shown in Figure S2b, for  $\beta < 0.5$  similarity increases as a function of lag. Since  $\beta$  is small,  $\mathbf{f}_i$  is dominated by  $\mathbf{f}_{i-1}$  rather than  $\mathbf{w}_i$ .

Since the specifics of the experimental event contribute only minimally to  $\mathbf{f}$ , the simulation approximates the autocorrelated noise simulation. For  $0.5 < \beta < 1$ , neural similarity is roughly constant as a function of lag for negative lags, but decreases as a function of lag for positive lags. This is because the pattern of activation during the  $i^{\text{th}}$  presentation will only contain traces of  $\mathbf{w}_j$  if  $i > j$ . Finally, for  $\beta = 1$ , similarity is 1 when lag = 0 and is 0 everywhere else. This is due to the fact that when  $\beta = 1$ ,  $\mathbf{f}_i = \mathbf{w}_i$ , and so the neural activity evoked by the  $i^{\text{th}}$  item will be present only during its presentation or recall.

The context reinstatement model is identical to the content reinstatement model during the presentation of study items and distractors. To simulate context reinstatement during recall of the  $j^{\text{th}}$  presented item, we set  $\mathbf{w}_i = \mathbf{f}_j$  (recall that  $\mathbf{f}_j$  will contain a recency-weighted average of the activations associated with the previously-presented items). As shown in Figure S2c, for  $\beta < 0.5$ , similarity increases as a function of lag, just as in the other simulations. Importantly, for  $0.5 < \beta < 1$ , neural similarity decreases with absolute lag in *both* the positive and negative directions, as seen in the neural data (Fig. 4a). Finally, as in the content reinstatement simulation, for  $\beta = 1$  similarity is 1 when lag = 0 and is 0 everywhere else.

These simulations show that regardless of the precise rate at which neural activity evolves over time, the simplest model consistent with our neural results (Fig. 4a) is one in which the temporal context in which an item is studied is reinstated when the item is recalled. While we have not ruled out every possible model that does not include some form of context reinstatement, neither autocorrelated noise (Fig. S2a) nor content reinstatement alone (Fig. S2b) can account for the neural signature of context reinstatement we observed in our ECoG recordings.

## Neural symmetry vs. behavioral asymmetry

The neural data (Fig. 4a) show that the decrease in neural similarity with absolute lag falls off symmetrically in the forward (positive) and backward (negative) directions. A natural question, then, concerns why the behavioral data exhibit a clear forward asymmetry in the conditional response probability as a function of lag (Fig. 4b). In particular: if the neural signature of context reinstatement we observe is truly related to participants' behavior (as implied by Fig. 4c), why is the neural signature of context reinstatement symmetric, whereas the contiguity effect is forward asymmetric?

Consistent with the neural data, our simulations show that context reinstatement *per se* implies a symmetric decrease in neural similarity with lag (Fig. 3f). Thus, the forward asymmetry in the behavioral data must arise due to some additional process that is not captured by our neural analysis. One possibility is that, in addition to reinstating the recalled item's context, the representation of the recalled item itself receives an additional "boost." As described above, reinstating the representation of an item (without its associated context) implies a decrease in neural similarity as a function of lag in the forward direction, but not the backward direction (Fig. 3e). In this way, the behavioral data might reflect both context *and* content reinstatement (e.g. see Howard and Kahana, 2002, Fig. 6). However, because we examine only autocorrelated components of neural activity, our neural analysis is (intentionally) biased towards examining neural features related to context rather than neural features related to item representations. An interesting question for future studies will be to clarify the extent to which the context and item representations overlap.

## Selecting autocorrelated features

Context-based theories of memory posit the existence of a gradually-changing pattern of neural activity that becomes associated with each studied word during study, and is reinstated during recall. To identify candidate components of the context representation for a given recording session, we selected autocorrelated PCA-derived features of the neural representation (Fig. 1c) as follows. Separately for each feature  $x$ , we computed the Pearson's lag 1 autocorrelation coefficient ( $r$ ) and associated  $p$ -value for the values of  $x$  within each list. We then combined the autocorrelation coefficients across lists into a summary autocorrelation measure,  $\bar{r}$ :

$$\bar{r} = F^{-1}\left(\sum_{i=1}^L F(r_i)\right),$$

where  $r_i$  was the Pearson's lag 1 autocorrelation coefficient for the values of  $x$  measured during list  $i$ ,  $F()$  was the Fisher z-prime transformation:

$$F(r) = \frac{\ln(1+r) - \ln(1-r)}{2},$$

and  $F^{-1}()$  was the inverse of  $F()$ :

$$F^{-1}(z) = \frac{e^{2z-1}}{e^{2z+1}}.$$

In this way, if  $r_i$  has large positive values across all lists,  $\bar{r}$  will have a large positive value. Similarly, if  $r_i$  is negative across all lists,  $\bar{r}$  will have a large negative value. If  $r_i$  is sometimes positive and sometimes negative (with approximately equal probability),  $\bar{r}$  will take on a value near zero. (Note that  $-1 \leq r_i, \bar{r} \leq 1$ .)

We also obtained a *p*-value,  $\bar{p}$ , associated with  $\bar{r}$  by applying the inverse Normal transformation to the *p*-values associated with the Pearson's lag 1 autocorrelation coefficients for each list. We then summed across the transformed *p*-values and evaluated the cumulative normal distribution function at this sum to obtain  $\bar{p}$ . We selected features with  $\bar{r} > 0$  and  $\bar{p} < 0.1$  for further analysis (see *Results*).

## Identifying time interval of the recall event

Our main analysis (Fig. 4a) compares the neural activity elicited by a studied word to the neural activity elicited by a word's retrieval during the recall period. We restrict our analysis of the study period to ECoG activity beginning 200 ms after the appearance of a word and ending when the word disappears from the screen. Here the 200 ms delay was meant to account for the lag between the word's appearance onscreen and the processing of the word by the participant.

In order to search for the optimal time interval for the recall event, we tested for context reinstatement while varying both the duration and onset of the time interval for the recall event. We tested time intervals ranging in duration from 100 to 1000 ms (in increments of 100 ms), and onsets ranging from -1000 to 0 ms (in increments of 100 ms) relative to the time the participant began their vocalized recall. This analysis indicates that the context reinstatement effect is strongest for the recall interval ranging from -600 to 200 ms relative to vocalization.

To account for the possibility that different brain regions reinstate context at different times relative to vocalization, we repeated this optimization analysis separately for each region of interest. The best time interval for the temporal lobe was from -400 to -300 ms (Fig. 5b). The time interval that gave the strongest frontal lobe effect was from -900 to -400 ms; however, the frontal effect was not statistically reliable (see *Results*).

## Additional details of selected features

In addition to asking whether specific brain regions contribute to the representation of context (Fig. 5), a natural question is whether the principal components comprising the feature vectors tend to weight particular oscillatory components of ECoG activity more heavily than others. Because principal components analysis (PCA) performs a linear mapping from the  $n$ -dimensional space of the original set of activity vectors onto the  $m$ -dimensional PCA space (where  $m \leq n$ ), we can use the PCA coefficients to perform the inverse mapping of the feature vectors back onto the original  $n$ -dimensional space. The PCA coefficients tell us how much each of the elements in the original PC vectors contributes to each of the principal components in the feature vectors. This allowed us to determine the degree to which each oscillatory component recorded from each electrode contributes to each element of the feature vectors. For a given frequency band, we assessed the degree to which that frequency band contributed to the feature vectors across all study and recall events by examining the distribution of PCA coefficients assigned to that frequency band across all participants. An analysis of PCA coefficients across frequency bands revealed no significant differences between frequency bands (repeated measures analysis of variance (ANOVA):  $F(4, 37) = 0.57, MSE = 10^{-5}, p = 0.69$ ). This finding suggests that the selected features are comprised of oscillatory activity at a broad range of frequencies (Fig. S3a).

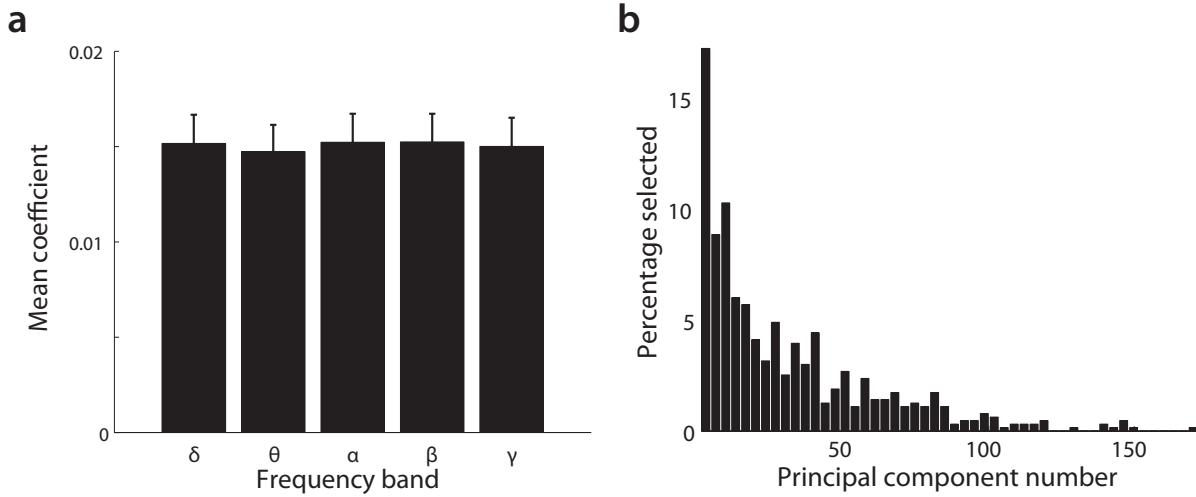


Figure S3: Selected features. **a.** Mean contributions of each frequency band to selected features. Error bars indicate 95% confidence intervals (Loftus and Masson, 1994). **b.** Percentage of selected features by principal component number. Smaller principal component numbers explain a higher proportion of the variance in the raw neural data.

## Supplemental references

- Deese, J. and Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, 54:180–187.
- Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46:269–299.
- Loftus, G. R. and Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1:476–490.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64:482–488.
- Polyn, S. M. and Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, 12:24–30.
- Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1):129–156.
- Ribak, C. E. and Dashtipour, K. (2002). Neuroplasticity in the damaged dentate gyrus of the epileptic brain. *Prog Brain Res*, 136:319 – 328.