

Oscillatory patterns in temporal lobe reveal context reinstatement during memory search

Jeremy R. Manning^a, Sean M. Polyn^b, Gordon H. Baltuch^c, Brian Litt^d, and Michael J. Kahana^{e,1}

^aNeuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA 19104; ^bDepartment of Psychology, Vanderbilt University, Nashville, TN 37240; ^cDepartments of ^cNeurosurgery and ^dNeurology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; and ^eDepartment of Psychology, University of Pennsylvania, Philadelphia, PA 19104

Edited* by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved June 8, 2011 (received for review October 14, 2010)

Psychological theories of memory posit that when people recall a past event, they not only recover the features of the event itself, but 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 recall neighboring list items successively. 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.

EEG | electrocorticography | oscillations | free recall | contiguity

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 time line.

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 nonlist 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 (8–10). 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 (11–14). 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 (15–20) and in the representations of individual stimuli (21). For each study and recall event, we analyzed these oscillatory components across all recording electrodes (Fig. 1*A* and *B*). We constructed a matrix containing, for each electrode, measurements of mean oscillatory power in five frequency bands (δ : 2–4 Hz, θ : 4–8 Hz, α : 8–12 Hz, β : 12–30 Hz, and γ : 30–99 Hz) during each study event (200–1,600 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 a smaller number of orthogonal components (Fig. 1*C*). 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

Author contributions: J.R.M., S.M.P., and M.J.K. designed research; J.R.M., G.H.B., and B.L. performed research; J.R.M. analyzed data; and J.R.M., S.M.P., and M.J.K. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence should be addressed. E-mail: kahana@psych.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015174108/-DCSupplemental.

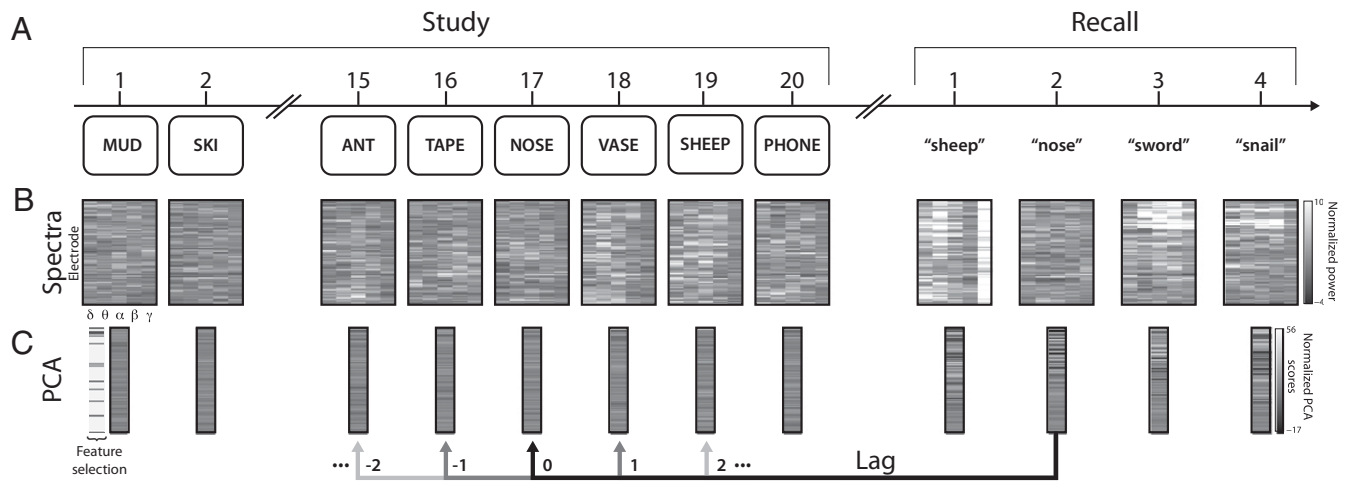


Fig. 1. Illustration of behavioral and electrophysiological methods. (A) After studying a list of 20 words and performing a brief distraction task, a participant recalls as many words as he can remember, in any order. (B) During each study presentation and just before each recall event, we calculate the z-transformed oscillatory power at each recording electrode in each of five frequency bands (δ : 2–4 Hz, θ : 4–8 Hz, α : 8–12 Hz, β : 12–30 Hz, and γ : 30–99 Hz). (C) We use PCA to find a smaller number of orthogonal components that jointly account for a large proportion of the variation in the data shown in 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.

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: $t(68) = -9.31$, $P < 10^{-10}$], 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 as a result of 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 (*Materials and Methods*), a defining feature of temporal context, for further analysis. In the remainder of this paper, 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

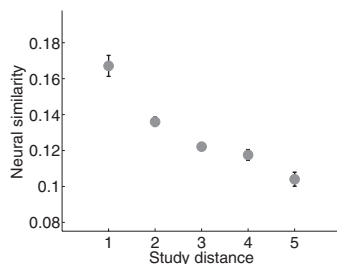


Fig. 2. Evolution of ECoG activity as participants study lists of words. Mean neural similarity is shown as a function of study distance (difference in serial position) between pairs of presented words. Error bars denote ± 1 SEM.

for the same word than during study events for other words (Fig. 3 E and 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. 14), 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 [paired-sample t test across 39 participants with at least 5 autocorrelated features: $t(38) = 3.10$, $P = 0.004$; Fig. 44]. 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 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 (β 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. 44).

We conducted three neural network simulations to contrast the predictions of three models of the observed feature vector dynamics (Fig. 3 and Fig. S1). In the autocorrelated noise model (Fig. 3 A and D), neural activity evolves randomly over time, irrespective of what is happening in the experiment. In the content reinstatement model (Fig. 3 B and E), each neuron represents a different word; a neuron is activated if its associated word

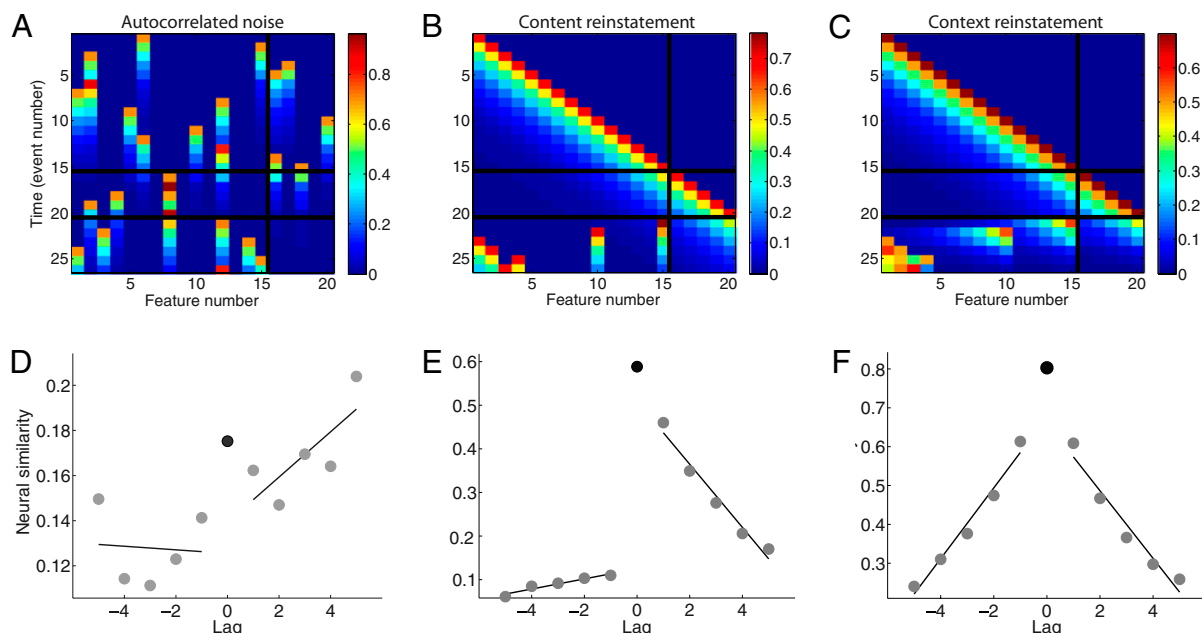


Fig. 3. Predicted neural similarity as a function of lag according to three models. (A–C) Pattern of activations for a simulated 20-neuron neural network as a 15-item list is studied. Events 1–15 of each matrix show activations after each item is presented. Events 16–20 show activations as distracting items are presented. Events 21–26 show activations as items 15, 10, 1, 2, 4, and 3 are recalled (in that order). 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 and neurons 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 B, but here we simulate context reinstatement during recall. (D–F) 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). Further details on the simulations are presented in *SI Materials and Methods*.

is presented or recalled. In the context reinstatement model (Fig. 3 *C* and *F*), each neuron also represents a different word. We simulate context reinstatement by activating not only the neuron associated with the word being recalled, but other neurons that were active at the time the recalled word was studied. Of these three simulations, 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 (further details and discussion are provided in *SI Materials and Methods*).

The decrease in neural similarity with absolute lag mirrors the contiguity effect: people’s striking tendency to make transitions to neighboring items rather than remote ones, as seen in be-

havioral data for the same participants (Fig. 4B). However, the forward asymmetry in the contiguity effect (Fig. 4B) is evident in neither the neural data (Fig. 4A) nor our simulation of context reinstatement (Fig. 3F). Models in which recall of an item retrieves both context and content information (e.g., 24) account for the forward asymmetry by virtue of the content information providing a boost in similarity to subsequently studied items (Fig. 3E). In this way, if our feature selection framework filters out content information, one would not expect to see asymmetry in the neural data (further discussion is provided in *SI Materials and Methods*).

Consistent with the hypothesis that the contiguity effect arises due to context reinstatement (5, 6, 24), participants with stronger

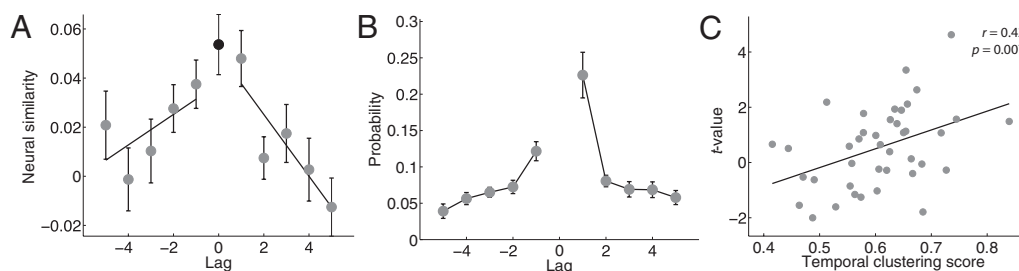


Fig. 4. A neural signature of temporal context reinstatement. (A) Neural similarity between the feature vector corresponding to recall of a word from serial position i and study of a word from serial position $i + \text{lag}$ [black dot denotes study and recall of the same word (i.e., lag = 0)]. (B) 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 A and B denote ± 1 SEM. (C) Participants exhibiting greater context reinstatement also exhibited more pronounced contiguity effects. Here, the t value associated with the regressions in A serves as a measure of the degree context reinstatement for each participant. (Only the regressions for negative lags were used, because 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 (*SI Materials and Methods, Analysis Methods*).

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 (23, 27–32) and prefrontal cortex (22, 32, 33) 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 [positive: $t(20) = -2.20$, $P = 0.04$; negative: $t(20) = -2.82$, $P = 0.01$; Fig. 5B]. 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 ($r = 0.48$, $P = 0.03$; Fig. 5C). The frontal lobe exhibited a weak neural signature of context reinstatement that trended toward 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. Fur-

thermore, 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, 5, 6, 24). 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 providing 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 words are more likely to be rehearsed than temporally distant words. Rehearsal-based models have been shown to be difficult to distinguish from context-based models (34, 35), 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, activating the representation of an item before 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 correlation 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 (36–40)]. 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 ($r = 0.13$, $P = 0.42$; *SI Materials and Methods, Analysis Methods* and Fig. S2), suggesting that rehearsal during study, per se, is unlikely to account for our findings. Although rehearsal is one of the hypothesized mechanisms underlying primacy, we recognize that other factors, such as enhanced attention to early list items (39, 41), 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, 5, 6) posit that continuity is provided by a context representation that changes gradually over time as a consequence of the integration

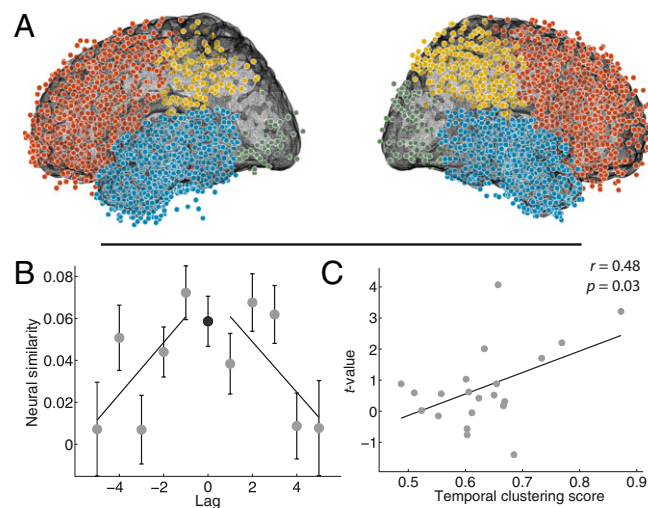


Fig. 5. Evidence for context reinstatement in the temporal lobe. (A) Each dot marks the location of a single electrode from our dataset in Montreal Neurological Institute space. We divided our dataset into four regions of interest: temporal lobe (blue, 1,815 electrodes), frontal lobe (red, 1,737 electrodes), parietal lobe (yellow, 512 electrodes), and occipital lobe (green, 138 electrodes). (B and C) Same format as in Fig. 4A and C but reflecting 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.

PNAS Early Edition | 5 of 5

Supporting Information

Manning et al. 10.1073/pnas.1015174108

SI Materials and Methods

Intracranial Recordings. *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 study. 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 to ensure the best possible outcome for the patient. For some patients, this means that the brain areas most relevant to a particular research question may receive little or no electrode coverage. To obtain adequate coverage of all relevant brain areas, we have analyzed data from many patients (Table 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 antiepileptic drugs, or during the 6-h period following any clinically significant seizure. A third issue is that the brain is known to rewire itself to compensate for damage, including damage caused by epilepsy (1), which could lead to cognitive remapping. Although we cannot control for cognitive remapping that may have occurred in individual participants, we have averaged our anatomical analyses over many patients; thus, results attributable 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 SDs of the mean for their age group.

Recording methods. Subdural grids or depth electrodes (Ad-Tech, Inc.) were implanted by neurosurgical teams solely for clinical purposes. The locations of the electrodes were determined by means of coregistered postoperative computed tomography and preoperative MRI scans, or from postoperative MRI scans, by an indirect stereotactic technique and converted into Montreal Neurological Institute coordinates. ECoG signals were recorded referentially using a Telefactor, Bio-Logic, XLTek, Neurofile, or Nicolet electroencephalographic digital video-EEG system. Depending on the amplifier, signals were sampled at 200, 256, 500, 512, or 1,024 Hz. Several hospitals applied band-pass filters to the recorded signals before writing to disk (Table 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).

Analysis Methods. Quantifying the contiguity effect. Fig. 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 (2). 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 ab-

solute value of the difference between the serial position of the just-recalled word and the set of words that have not yet been recalled. 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, because 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), the temporal clustering score would yield a value of 1 (because 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.

Quantifying the primacy and recency effects. The primacy and recency effects refer to an enhancement in memory for early and late list items, respectively, compared with memory for intermediate list items (3, 4). The number of items that show a boost in memorability attributable to primacy or recency is relatively invariant to changes in list length; the primacy effect generally affects the first few items, whereas the recency effect generally affects the last six or so items (4). 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, items 4–14 for 20-word lists). We then measured the strength of the primacy effect for each participant by dividing his or her mean probability of recalling items from primacy positions by his or her 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 his or her mean probability of recalling items from recency positions by his or her 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 Fig. S2. 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 (Fig. 3 and Fig. S1) 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 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 th experimental event (i.e., study presentation, distracting task, recall). Our main analysis entails selecting autocorrelated components of neural activity as the candidate context representation (*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 β is a constant; ρ_i is a function of $\mathbf{f}_{i-1}, \mathbf{w}_i$, and β (with $0 \leq \rho_i, \beta \leq 1$); and \mathbf{w}_i is the pattern of neural activity specifically evoked by the i th experimental event [details are presented by Polyn and Kahana (5)]. In this way, the neural activity measured after a given experimental event (e.g., presentation of the fifth 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 Fig. S14, 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, because 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 th presented item, we set $\mathbf{w}_i = \mathbf{w}_j$. As shown in Fig. S1B, for $\beta < 0.5$, similarity increases as a function of lag. Because β is small, \mathbf{f}_i is dominated by \mathbf{f}_{i-1} rather than \mathbf{w}_i . Because 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 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 attributable to the fact that when $\beta = 1$, $\mathbf{f}_i = \mathbf{w}_i$; thus, the neural activity evoked by the i 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 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 Fig. S1C, 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 de-

creases 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. Although we have not ruled out every possible model that does not include some form of context reinstatement, neither autocorrelated noise (Fig. S1A) nor content reinstatement alone (Fig. S1B) 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 in Fig. 4C), why is the neural signature of context reinstatement symmetrical, whereas the contiguity effect is forward asymmetrical?

Consistent with the neural data, our simulations show that context reinstatement, per se, implies a symmetrical decrease in neural similarity with lag (Fig. 3F). Thus, the forward asymmetry in the behavioral data must arise as the result of 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 in the backward direction (Fig. 3E). In this way, the behavioral data might reflect both context and content reinstatement [e.g., figure 6 in the article by Howard and Kahana (6)]. However, because we examine only autocorrelated components of neural activity, our neural analysis is (intentionally) biased toward 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} :

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

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

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

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

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

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 (*Results*).

Identifying time interval of the recall event. Our main analysis (Fig. 44) compares the neural activity elicited by a studied word with 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 on-screen and the processing of the word by the participant.

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 1,000 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 his or her 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 (*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 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 principal components 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 among frequency bands [repeated measures ANOVA: $F(4,37) = 0.57$, $P = 0.69$]. This finding suggests that the selected features are composed of oscillatory activity at a broad range of frequencies (Fig. S34).

- Ribak CE, Dashtipour K (2002) Neuroplasticity in the damaged dentate gyrus of the epileptic brain. *Prog Brain Res* 136:319–328.
- Polyn SM, Norman KA, Kahana MJ (2009) A context maintenance and retrieval model of organizational processes in free recall. *Psychol Rev* 116:129–156.
- Deese J, Kaufman RA (1957) Serial effects in recall of unorganized and sequentially organized verbal material. *J Exp Psychol* 54:180–187.
- Murdock BB (1962) The serial position effect of free recall. *J Exp Psychol* 64:482–488.
- Polyn SM, Kahana MJ (2008) Memory search and the neural representation of context. *Trends Cogn Sci* 12:24–30.
- Howard MW, Kahana MJ (2002) A distributed representation of temporal context. *J Math Psychol* 46:269–299.

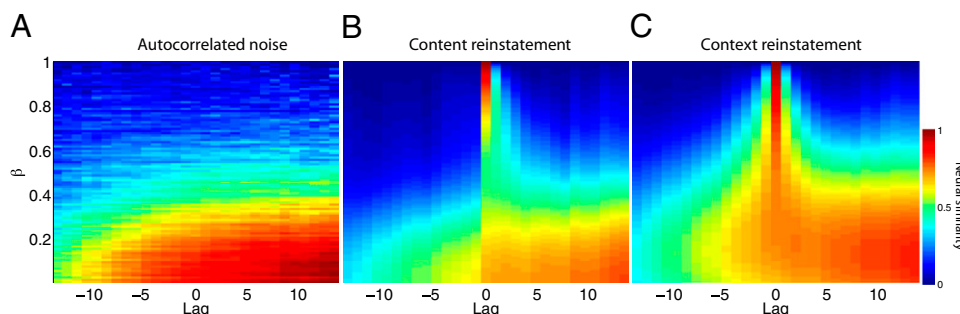


Fig. S1. Simulated neural similarity as a function of lag and drift rate (β), given no content or context information in the neural recordings (*A*; Fig. 3 *A* and *D*), content reinstatement without context reinstatement (*B*; Fig. 3 *B* and *E*), and context reinstatement (*C*; Fig. 3 *C* and *F*). Similarity is computed as the normalized dot product between the simulated feature vector after the recall of the i th word and the feature vector corresponding to presentation of word $i + \text{lag}$. The first dimension (initialized to 1 before the start of the simulation) was ignored for the similarity calculations. Simulation results in Fig. 3 used $\beta = 0.7$ [this choice was motivated by previously reported simulation results (2)].

