Human episodic memory retrieval is accompanied by a neural jump back in time

Sarah Folkerts[1], **Ueli Rutishauser**[2,3], **Marc W. Howard**[4]

***For correspondence:**
marc777@bu.edu (MWH)

[1]School of Medicine, University of Warwick; [2] Department of Neurosurgery, Cedars-Sinai Medical Center ; [3]Computation and Neural Systems, California Institute of Technology; [4]Psychological and Brain Sciences, Boston University

Abstract Cognitive psychologists have long hypothesized that experiences are encoded in a temporal context that changes gradually over time. When an episodic memory is retrieved, the state of context is recovered—a jump back in time. We recorded from single units in the MTL of epilepsy patients performing an item recognition task. The population vector changed gradually over minutes during presentation of the list. When a probe from the list was recollected, the population vector reset to the neighborhood of the original presentation of that probe during study—a neural "jump back in time." Probes that were not recollected did not cause a jump back in time. These results constitute the first direct evidence that recollection of an episodic memory in humans is associated with recovery of a gradually-changing state of temporal context—a neural "jump-back-in-time" that parallels the act of remembering.

Introduction

Episodic memory refers to our ability to vividly remember specific events from our own experience. The vividness of episodic memory, along with the specificity of the memory to a particular place and time has led researchers to characterize episodic memory as mental time travel (***Tulving, 1972***; ***Hassabis et al., 2007***; ***Schacter et al., 2007***). This verbal description has been operationalized in computational models of episodic memory in which the flow of time is described by a slowly and gradually-changing state of temporal context (***Howard and Kahana, 2002***; ***Sederberg et al., 2008***; ***Polyn et al., 2009***; ***Howard et al., 2015***). In these models episodic memory retrieval is accompanied by the recovery of a prior state of temporal context—a jump back in time—that accounts for the behavioral contiguity effect (***Kahana, 1996***; ***Schwartz et al., 2005***; ***Howard et al., 2008***; ***Unsworth, 2008***).

This computational hypothesis makes two predictions. First, in addition to stimulus-evoked activity, the activity of neurons involved in episodic memory should in addition also change gradually over time. This prediction aligns with a large body of animal work showing that neural ensembles in the hippocampus, amygdala and prefrontal cortex change slowly over time scales up to at least tens of minutes (***Manns et al., 2007***; ***Hyman et al., 2012***; ***Mankin et al., 2012***; ***MacDonald et al., 2011***; ***Salz et al., 2016***; ***Rubin et al., 2015***; ***Cai et al., 2016***; ***Rashid et al., 2016***). Second, during retrieval of an existing memory, the prior state (temporal context) associated with an episodic memory should be restored. Although some prior studies have attempted to measure this hypothesized phenomenon (***Manning et al., 2011***; ***Howard et al., 2012***; ***Yaffe et al., 2014***), due to methodological limitations of those studies there is presently no definitive study linking this phenomenon to

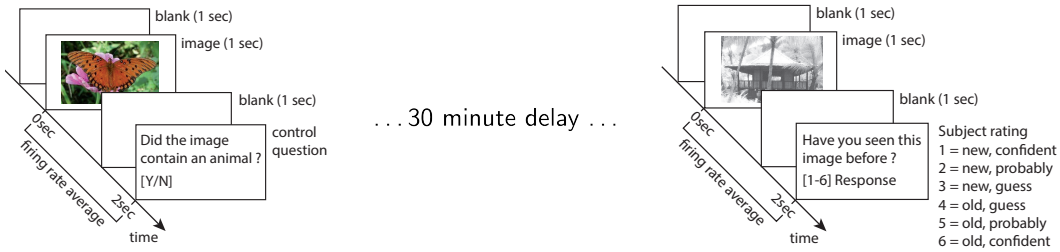

Figure 1. *The behavioral task*. During a study (learning) phase, participants were asked to learn set of pictures. In order to ensure that the patients were attending to the picture, they responded to an orienting task after each item. After a 30 minute delay, participants were presented with a test list that included both stimuli from the study session and also new probes. For each, they indicated whether they thought they had seen an item before or not on a 6-point confidence scale.

episodic memory in humans.

In the laboratory, episodic memory can be studied using the item recognition task. In item recognition, participants are presented with a study list of novel stimuli to remember (here we use pictures). After study, participants are provided with a set of probe stimuli one at a time, some of which were on the study list and some of which were not. The task of the subjects is to distinguish probe stimuli that were on the list from those that were not. Many authors have hypothesized that recognition memory is supported by two processes, recollection and familiarity (***Yonelinas et al., 2002***; ***Eichenbaum et al., 2007***; ***Staresina et al., 2013***; ***Wixted, 2007***). According to this viewpoint (which it should be noted is not universally accepted, ***Squire et al.*** (***2007***)) recollection corresponds to vivid episodic memory in which details of the study experience is recovered. When an old probe is recollected, triggering retrieval an episodic memory, this is believed to lead participants to endorse the probe as old with high confidence (***Yonelinas et al., 2002***; ***Diana et al., 2007***). High confidence old responses are associated with the recovery of detailed source information about the context in which a probe was studied (***Onyper et al., 2010***) are associated with a behavioral contiguity effect (***Schwartz et al., 2005***), and with the activation of neurons in the medial temporal lobe (***Rutishauser et al., 2015***).

In this study epilepsy patients performed an item recognition task, rating their confidence that probes were presented on a six-point scale (Figure 1). During both study and retrieval, single units were recorded from microelectrodes implanted in their medial temporal lobes. Population vectors were measured across units; consistent with previous results the population vectors changed gradually during study of the list. Comparing the population vector in response to an old probe at test to the population vectors during study will enable us to evaluate whether temporal context is recovered. We test the hypothesis that probes that were recollected—here operationalized as probes that receive a highest-confidence response—are accompanied by greater recovery of temporal context than probes that were not recollected—here operationalized as probes that did not receive a highest-confidence response.

Results

Behavioral results were consistent with episodic memory for some old probes.

Patients judged each item presented during the test phase as either old (seen before) or new (not seen before) together with a confidence rating (Figure 1). The behavioral results from patients were broadly consistent with canonical behavioral results from control participants (***Kahana, 2012***). Patients used all confidence ratings and used the highest confidence old response approximately five times more often for old probes than for new probes (Figure 2a). Patients also distinguished old from new probes above chance for lower confidence ratings (Figure 2a). High-confidence responses were therefore a good proxy for successful recovery of a veridical memory. In contrast, the lower confidence old responses (4-5) were less effective in discriminating old probes from new probes.

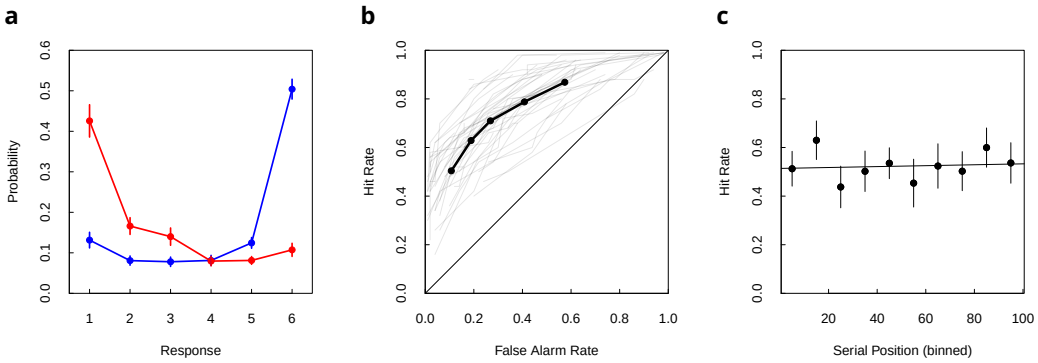

Figure 2. *Behavioral results.* **a.** Participants successfully distinguished repeated probes from new probes. Shown is the probability of each response (1-6) conditional on the ground truth, i.e. whether the stimulus is old (blue) or new (red). Note that responses (1,2,3) for new (red) stimuli and responses (4,5,6) for old (blue) stimuli are correct whereas the others are incorrect. Patients had good memory as demonstrated by using the highest confidence rating (1 or 6) for about half of the new and old probes, respectively. Error bars are SEM across n=49 sessions. **b.** Behavioral ROC curves for each participant included in this study (grey lines) and the average ROC (red dashed line). The ROC plots hit rate as a function of false alarm rate for each possible criterion; chance performance would be along the diagonal. These ROC curves are typical of item recognition studies, with a reliable asymmetry characteristic of episodic memories (see text for details). **c.** The 30 minute delay between study and test successfully eliminated behavioral recency effect. The hit rate—here the probability of an old probe receiving a highest-confidence response—is shown as a function of each probe's binned serial position during study. The slope of the regression line is not significantly different from zero, regression coefficient $(2 \pm 7) \times 10^{-4}$. Error bars are the 95% confidence interval.

This indicates that patients were able to differentiate weak from strong memories using subjective confidence ratings.

We next quantified each patient's behavior using an receiver operating characteristics (ROC) curve (Figure 2b). The ROC shows hit rate—probability of a yes response to an old probe—as a function of false alarm rate—probability of a yes response to a new probe—for each possible confidence criterion. To the extent the ROC points lie above the diagonal, memory is above chance. The ROC curves were asymmetric (Figure. 2b, z-ROC slope was significantly less than one, $.69 \pm .05$, mean $\pm$ SEM, $p < 10^{-6}$), which is commonly taken as a signature of episodic memory accompanied by successfull recollection (***Yonelinas, 2002***; ***Wixted, 2007***; ***Fortin et al., 2004***).

It is important for the later neural analysis that the probability of recognizing a stimulus as old does not depend on how long ago it was seen; some previous attempts to measure a neural jump-back-in-time (***Howard et al., 2012***), were confounded by a large behavioral recency effect. In this study, the study and test period were separated by a delay of approximately 30 min. This successfully eliminated the recency effect: the hit rate (probability of a yes response) was independent of the position an item was shown in during study (Figure 2c). There was no significant effect of serial position, as indicated by a regression coefficient $(2 \times \pm 7) \times 10^{-4}$, that was not reliably different from zero. This shows that the delay between study and test was effective in eliminating the recency effect at test.

The population vector changed gradually over at least a minute during study.

A key requirement for contextual reinstatement to occur is that neural activity changes gradually across multiple trials during learning. Such gradual changes would be a signature of temporal information because they are not imposed by the stimuli, which are randomized in order. We thus first tested whether neurons within the MTL exhibit signatures of a gradually changing temporal context. We constructed population vectors from the mean firing rate in a 2 s window following stimulus onset for all recorded units (see methods for details). We then tested whether the pairwise similarity between population vectors from study events differed systematically as a function of

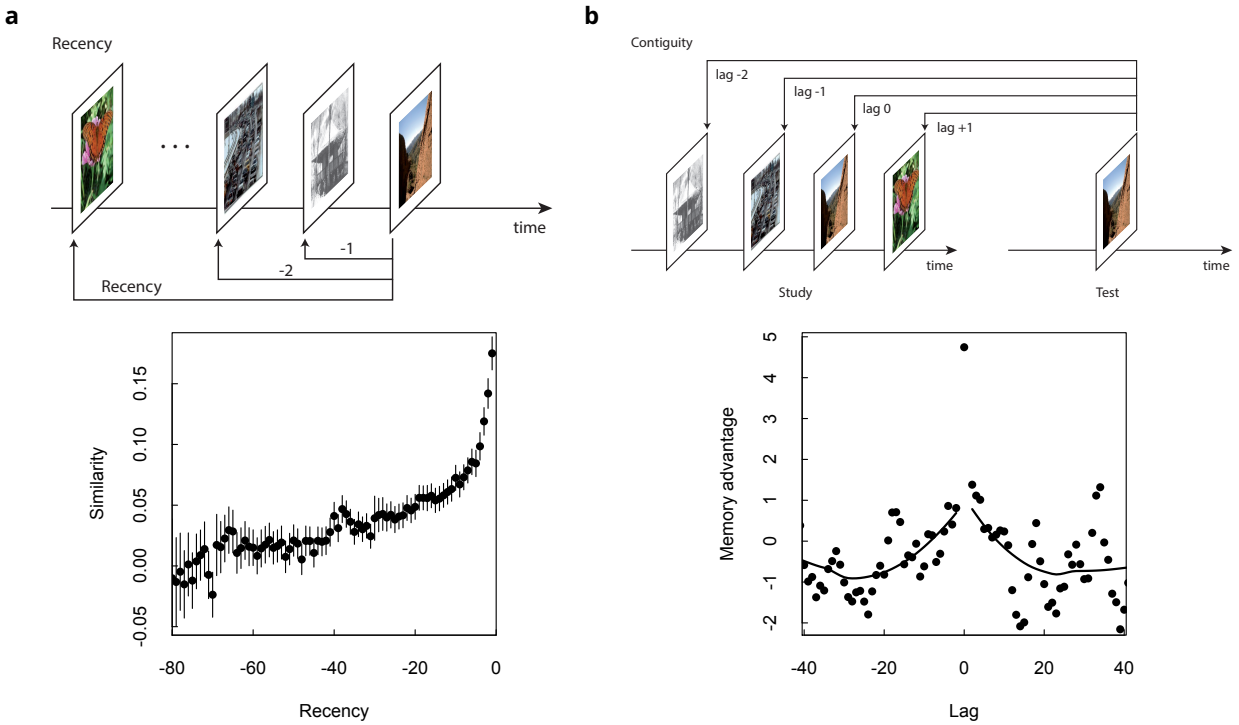

Figure 3. *Neural jump-back in time.* **a.** Neural recency effect. Top: Schematic describing the definition of recency. For each presentation of a stimulus, a population vector was computed for the 2 s following presentation of the stimulus. This vector was then compared to the population vector from all preceding stimulus presentations and the similarity was aggregated as a function of the recency between the comparisons. Bottom: The population vector shows a recency effect, changing (conservatively) to at least recency -30 during study, corresponding to about two minutes. Smoothed curves are from a LOESS regression. **b.** Neural contiguity effect showing a jump-back-in-time. Top: Schematic of the lag variable. For a test probe, similarity of the population vector after the test probe is compared to the population vectors of each study event. The similarity is aggregated as a function of lag, the difference between the original presentation of the probe stimulus and the other list stimulus; the lag to the repeated stimulus is zero. Bottom: To isolate the effect due to episodic memory, we took the difference between the similarity for pictures receiving a highest-confidence response and pictures that were not well-remembered (see methods for details). This "memory advantage" is in units of a paired t-statistic. For clarity, a sliding binning procedure was used to plot the results for lags other than zero. Critically, the memory advantage is peaked around zero, falling off gradually in both the forward and backward directions, indicating a neural jump-back-in-time associated with successful episodic memory retrieval.

time between those events. We found a gradual increase in similarity for pairs of study events closer together in time (Figure 3a); the regression coefficient was $.00123 \pm .00008$, $F_{(1,78)} = 221.6$, $p < .001$. A similar recency effect was also evident during the test phase (Figure A4); the regression coefficient was $.00089 \pm .00006$, $F_{(1,78)} = 188.7$, $p < .001$. We thus found significant temporal context effects during both study and test.

We next tested whether this contiguity effect was also visible for specific subsets of visually selective (VS) units (***Rutishauser et al., 2015***). As ***Rutishauser et al.*** (***2015***) described previously, VS units respond shortly after the onset of a stimulus conditional on the visual category of a stimulus. For example, a subset of "animal selective" VS units change their firing rate only when the image contains an animal. 213 (out of 1286) recorded units qualified as VS units (see methods) and the analysis that follows is restricted to these units. If VS units only carry information about the stimulus that was jus presented, their activity should not vary gradually and would thus not show a temporal context effect. Contrary to this prediction, we found that VS units alone also exhibited a robust neural similarity effect similar to that observed to all recorded units (Fig. A3a). This was true both during study, $.0013 \pm .0001$, $F_{(1,78)} = 104.6$, $p < .001$, as well as test, $.0009 \pm .0001$, $F_{(1,78)} = 84.64$, $p < .001$. Consequently, the response of VS units is modulated by temporal context in addition to visual input. This suggests that feedforward visual input is modulated by temporal context, a critical prediction of the temporal context model.

Episodic recollection was associated with the recovery of temporal context.

We next computed the similarity in neural response between pairs of test and study items (see Figure 3b, top, for an illustration). The contextual reinstatement model predicts that the neural response to a recollected old probe that was originally presented at position i will be similar to the neural response to study events that were presented shortly before or after position i. The variable lag describes the difference between the serial position of the original presentation of an old probe and a study item; lag zero corresponds to the comparison between an old probe at test and its original presentation during study (when it was new). We would expect the neural pattern similarity at lag zero to be large to the extent the response is determined by visual input, which is similar for study and test of the same stimulus.

We first computed the neural similarity as a function of lag separately for recollected probes and unrecollected probes—operationalized as probes that did and did not receive a highest confidence old response (Figure A2). We found that similarity tended to decrease as a function of |lag|, for recollected probes. That is, for probes that were recollected, the population vector at test was more similar to the population vector for study events close in time—in both the forward and backward direction—to the time at which that probe stimulus was originally studied. In contrast, there was a tendency for similarity to *increase* as a function of |lag| for probes that were not recollected (see supplementary information for details). That is, for probes that did not receive a highest confidence response, the population vector at test was *less* similar to the population vectors for items studied close in time to the probe stimulus than for items studied further away. Both of these findings also held (see Fig. A2b, supplementary information) for VS units considered alone. This suggests that the degree of contextual reinstatement predicted the success or failure of episodic recollection.

To isolate the contribution to neural pattern similarity attributable to recollection, we calculated the difference between the neural pattern similarity as a function of lag for probes that were recollected and those that were not (see methods for details). In the following, we refer to this difference as 'memory advantage,' which is in units of a t-statistic comparing recollected to unrecollected probes. By examining the memory advantage as a function of lag, we can simply assess whether episodic memory retrieval results in a neural contiguity effect as predicted by retrieved context models.

The memory advantage index showed a robust contiguity effect (Figure 3b). This was also true for VS units considered alone (Figure A3b). The memory advantage at lag zero was significant, $t = 4.75$, $p < .001$. The effect at lag zero, however, does not indicate reinstatement of temporal

context. This is because similar visual features were present during study of stimulus i and test of stimulus i. To test for a neural jump-back-in-time, we asked whether the memory advantage changed systematically as a function of lag. A jump-back-in-time requires that the repeated image presentation triggers a retrieval of previous context, i.e. the reinstatement of the neural ensemble activity present before the first encounter with the probe stimulus. This would be expected to manifest in a decrease in the neural memory advantage as a function of lag in both the forward direction (lag increasing from zero) and in the backward direction (lag decreasing from zero).

To evaluate this prediction we performed an ANOVA on memory advantage with |lag| as regressor (excluding lag zero) and direction (backward or forward) as a categorical variable. There was a significant effect of |lag|, $F_{(1,56)} = 16.4$, $p < .001$ but no effect of direction, $F_{(1,56)} = .003$ nor an interaction $F_{(1,56)} = .01$. The effect of |lag| means that for recollected probes, the population vector was more similar to states of the population vector close together in time to the original presentation of the probe stimulus. This is as predicted by the hypothesis that recollection is accompanied by recovery of a gradually-changing state of temporal context in the human brain.

A decrease in the memory advantage extending to lags near zero but restricted to the forward direction (positive values of lag) would correspond to persistence of stimulus-specific features in memory such as in a short-term memory buffer. In contrast a jump-back-in-time would cause reconstruction of the pattern of activation *prior* to initial presentation of the probe. Thus, a jump-back-in-time would manifest as a advantage for lags near zero in both directions. To quantify the effect, we performed linear regressions of memory advantage onto lag separately for each direction. We found a regression coefficient of $-.06 \pm .02$, $F_{(1,28)} = 7.69$, $p < .01$ for the forward direction (lags 1 to 30), and $.07 \pm .02$ with $F_{(1,28)} = 8.694$ and $p < .01$ for the backward direction (lags -30 to -1). Thus, we found an effect of lag on the memory index separately in both the forward and backward directions as predicted by a neural jump-back-in-time.

We found similar evidence for a neural jump-back-in-time when considering only VS units (Figure A3b). The memory advantage at lag zero was significant, $t = 5.05$, $p < .001$. An ANOVA with |lag| as regressor (excluding lag zero) and direction (backward or forward) as a categorical variable showed a significant effect of |lag|, $F_{(1,56)} = 18.9$, $p < .001$ but no effect of direction, $F_{(1,56)} = .07$ nor an interaction, $F_{(1,57)} = .91$. Also, considering the forward and backward directions separately, we found a regression coefficient of $-.06 \pm .03$, $F_{(1,28)} = 5.4$, $p < .03$ for the forward direction (lag 1 to lag 30) and $.09 \pm .02$, $F_{(1,28)} = 15.03$, $p < .001$ for the backward direction (lag -30 to lag -1). Thus, a signal compatible with contextual reinstatement was visible even when only considering VS units that were sensitive to the category of the visual stimulus presented during study.

Discussion

We found that the population activity of human MTL units changed gradually over minutes (Fig. 3a). This replicates, in humans, prior evidence for a gradually-changing temporal context signal found previously in the MTL of animals (***Naya and Suzuki, 2011***; ***MacDonald et al., 2011***; ***Manns et al., 2007***; ***Hyman et al., 2012***; ***Mankin et al., 2012***; ***Cai et al., 2016***; ***Rashid et al., 2016***; ***Rubin et al., 2015***) and humans (***Howard et al., 2012***; ***Yaffe et al., 2014***; ***Manning et al., 2011***; ***Hsieh et al., 2014***; ***Hsieh and Ranganath, 2015***). Crucially, visually selective category cells also showed this gradual change (Fig. A3a). This is important because it suggests that the population vector does not merely change gradually over time but also carries information about the identity of the stimuli presented.

The critical new insight that this paper contributes is a first demonstration that the retrieval of human episodic memory is associated with the recovery of a gradually-changing state of temporal context—a neural jump back in time. This analysis measures the similarity between a population vector caused by an old probe and population vectors during study of the neighbors of the original presentation of the probe stimulus. The difference between the population similarity calculated for probes that were recollected and the population similarity calculated for probes that were not recollected was greater for the neighbors of the original presentation and fell off reliably with distance from the original presentation of the probe stimulus in both directions (Fig. 3b). Notably,

robust contiguity effect associated with recollection was observed when considering population vectors constructed from only the visual selective units (Fig. A3b). Examination of the raw contiguity measures (Fig. A2) suggests that the VS units were largely responsible for the effect (the analysis with all units includes the VS units).

We did not merely observe contiguity effect for recollected probes, but also an anti-contiguity effect for old probes that were not recollected (Fig. A2). Methodologically, this means that had we averaged over all old probes, it would have been more difficult to observe a neural contiguity effect. Had we observed that the probes that did not evoke recollection resulted in a contiguity effect that was weaker, this would suggest a continuity between recollection and familiarity. Because there was not such an effect, and indeed a tendency towards an anti-contiguity effect it suggests that successful retrieval of preceding temporal context is only observed for probes that are recollected.

In this study, we observed that the brain state in the moments after recollection of an episodic memory resembled the gradually-changing temporal context at the time that memory was encoded. However, this does not imply that this recovered context persists long after the recollection of that probe. If context retrieved by a probe persisted long after the presentation of the probe, one would expect the rate of contextual drift during the test list to be very different than the rate of drift during study. While there is some difference (Figure A4), the discrepancy is modest compared to what one would expect if retrieved context persisted long after the presentation of a probe. One possibility is that retrieved context only becomes available for a short time after presentation of the probe and then dissipates. This is analogous with "awake replay" events in the rodent hippocampus ***Carr et al.*** (***2011***); ***Pfeiffer and Foster*** (***2015***), in which hippocampal place cells briefly fire as if the animal is in a remote location during sharp-wave-ripple events. Perhaps the neural jump-back-in-time is a transient discontinuity in the stream of temporal context much like the transient discontinuity in the representation of position.

Methodological advantages of the present study compared to previous attempts to measure a neural jump-back-in-time.

This study avoids methodological pitfalls of previous papers that addressed whether human episodic memory is associated with a jump back in time. A previous study with human single units ***Howard et al.*** (***2012***) used continuous recognition, in which probes are intermixed with study items in a continuous stream of experience. In that study there was a robust behavioral recency effect. Because recency is confounded with lag in the backward direction, it was necessary to statistically decouple recency from contiguity in that study. Here, the 30 minute delay between study and test and the absence of a behavioral recency effect eliminated any confound due to recency. Note also that we would expect a neural recency effect to be present for old probes that were not recollected as well as old probes that were recollected. Because the neural contiguity effect was observed in the difference between these suggests it is not due to a confound between recency and contiguity.

Another prior study used autocorrelated features from ECoG in a free recall study (***Manning et al., 2011***). They found that the features during recall of the word studied at serial position i in the list resembled the features during study of nearby list items. However, because free recall is extended in time and exhibits a robust behavioral contiguity effect (***Kahana, 1996***; ***Sederberg et al., 2010***) that finding does not establish that recall of word i is associated with *recovery* of a gradually changing temporal context. Because of the behavioral contiguity effect, the recall of word i is likely to have been preceded by neighbors of word i, so that the neural contiguity effect could have been due to the persistence of item representations from previous recalls. Similar concerns apply to a human single unit study that argued for recovery of spatial context during a free recall task (***Miller et al., 2013***). Another ECoG study used cued recall to establish that successful recovery of a word was associated with recovery of temporally-varying features from the list (***Yaffe et al., 2014***). Although this study was able to establish a correlation between successful memory and a contiguity effect, the analyses included lag zero in the measurement of the neural contiguity effect in the backward direction. Because the neural contiguity effect in the backward direction is a distinctive

signature of a jump back in time, whereas similarity at lag zero may be attributed to repeated items, the analyses reported in that paper did not clearly establish a neural jump back in time.

Recent studies showing the importance of temporal context in human memory.

This study adds to a growing body of work from human cognitive neuroscience that suggests that a gradually-changing state of temporal context affects memory in a range of tasks. A free recall study using fMRI showed that the content of lingering item representations during study predicted free recall transitions during retrieval (***Chan et al., In press***). A recent fMRI study showed that the amount of drift in the right entorhinal cortex between two events in a radio program predicted participants' judgment of the duration between the two events (***Lositsky et al., 2016***). Similarly, when participants rate the relative recency of two probes, hippocampal pattern similarity predicted the order judgment (***DuBrow and Davachi, 2014***, ***2016***). Moreover, in that same study, successful judgments were associated with reinstatement of stimuli that intervened between the two probes. A recent study with patients with MTL damage showed that patients were impaired at their ability to perform temporal ordering, as if an intact MTL was required for recovering temporal context (***Dede et al., 2016***).

Finally, a pair of recent studies suggest that the recovery of temporal context we observed in the laboratory could also reflect a mechanism for memory in more natural settings. In natural settings, the visual features the participant experiences are autocorrelated in both time and space, unlike the randomly-assembled list of visual stimuli experienced in a fixed location used in the present study. A recent study of natural memory automatically recorded pictures as participants went about their daily lives for several weeks (***Nielson et al., 2015***). After a delay, participants were brought into the scanner and shown images from their own lives. The pattern similarity between pairs of images that were well-remembered was computed. The pattern similarity in the anterior hippocampus predicted the distance in both time and space between pairs of remembered images, on the scale of hours to weeks for time and tens of meters to kilometers for space. Another recent study, adding to work in virtual reality environments (***Chadwick et al., 2010***; ***Copara et al., 2014***) observed similar results for episodic memory in a well-controlled virtual environment in which spatial and temporal proximity could be deconfounded (***Deuker et al., 2016***). In light of this growing body of evidence and modeling work suggesting a deep connection between temporal context and spatial context (***Howard et al., 2014***), the present study suggests that recovery of a gradually-changing state of spatiotemporal context is an essential aspect of human episodic memory that depends crucially on the function of the MTL.

Methods and Materials

Patients

54 recording sessions were made from 35 patients who were evaluated for possible surgical treatment of epilepsy using implantation of depth electrodes. All patients volunteered for the study and gave informed consent. Protocols were approved by the institutional review boards of the Cedars-Sinai Medical Center, Huntington Memorial Hospital and the California Institute of Technology. Out of the 54 recording sessions, 44 were previously reported by ***Rutishauser et al.*** (***2015***) and 10 are unpublished. Five sessions were rejected because memory accuracy was not sufficiently high ($d' < .5$).

Electrophysiology and spike sorting

The recording methods and single-unit data analyses for this dataset have been described in detail before (***Rutishauser et al., 2015***). Briefly, the recordings analyzed here were obtained from depth electrodes implanted bilaterally within the hippocampus and amygdala (8 microwires each, 32 channels per patient in total). Broadband extracellular recordings were filtered .1 Hz to 9 kHz and sampled at 32 kHz (Neuralynx Inc). Electrodes were localized based on post-operative MRI images.

Electrode locations were chosen according to clinical criteria alone. Spikes were detected and sorted as previously described (***Rutishauser et al., 2006***).

Behavioral task

The task (Figure 1) consisted of two parts: a study (learning) followed by a test phase. During study, patients viewed a list of 100 photographs of natural scenes. There were 25 instances each from five different visual categories (animals, people, cars/vehicles, outdoor scenes/houses and flowers/food items; see Figure 3 for examples). Each image appeared on the screen for 1 s, followed by a delay of 1 s, followed by an orienting task in which participants answered whether the image they had just seen contained an animal or not. During the test phase, subjects were shown 100 images, half of which were identical to those seen previously ("old") and half were novel ("new"). After each image, subjects indicated whether they saw the image before or not together with how confident they were about their decision (1= “new, certain”, 2 = “new, probably”, 3 = “new, guess”, 4 = “old, guess”, 5 = “old, probably”, 6 = “old, certain”). There was no response deadline.

Artifact rejection

We excluded 96 units that contributed no spikes to the firing rate vectors or that had a bimodal firing rate distribution. This left a total number of 1286 units used in this report. In addition, we excluded trials during which there was an abrupt signal loss in several simultaneously recorded units (Figure A1. Such loss is likely attributable to recording problems and we thus excluded such periods of time. To achieve this, time periods during which a fraction of $\geq .25$ of the units ceased firing for $\geq .05$ of the total trial duration were classified as artifacts. We identified such artifacts in two study and four test sessions.

Population vectors

Population vectors were computed from the average firing rates within a 2 s window starting at stimulus onset. To control for changes in baseline firing rate for different units, the mean firing rates for each unit were z-scored with respect to the average firing rate of that unit across all events. After z-scoring, all statistics reported were computed across all recorded units across all sessions. Trials with reaction times that exceeded 2.5 standard deviations of the reaction time distribution of a given patient were excluded (136/10439 were excluded based on this criterion).

Recency analysis

In order to evaluate whether the ensemble changed gradually over time we analyzed how the similarity between population vectors changed as a function of recency, the difference between the serial position of the two events. For instance, the comparison between the population vector from presentation of the seventh stimulus in the list to the population vector from the fourth stimulus in the list is associated with a recency of -3 (see Figure 3a, top). A normalized inner product of z-scored population vectors (the inner product normalized by the number of units) was used to characterise the similarity between the ensemble response between a pair of events as a function of recency. In order to avoid any possible confounding influence of a primacy effect on the analysis, only events after the first 20 item presentations were included.

Neural contiguity analyses

In order to evaluate whether memory for an event caused reconstruction of the gradually-changing neural state during study of that event we compared ensemble similarity as a function of lag, defined as follows. Given an old test probe that was originally (during learning) presented at serial position i, and a study event presented at serial position j, lag is defined as $j - i$. To be concrete, consider the population vector from the test of an old probe originally presented at serial position seven. We compared this population vector to each of the population vectors from a study event. The lag associated with comparison of the test event to the study event at serial position seven—the

same stimulus—is zero. The lag associated with the comparison to the population vector from study at serial position eight—which immediately followed study of the stimulus—is $+1$; the lag associated with the comparison to the event from serial position six is -1. For each old probe, lag defines a number line across the study serial positions with zero at the original study location of the probe stimulus. Ensemble similarity between each pair of events consisting of an old test probe and a study stimulus was aggregated as a function of lag.

Note that the number of data points entering into the contiguity analyses changes as a function of lag. For instance, there are many more combinations of serial positions that result in a lag of $+1$ than there are combinations that lead to a lag of $+50$. For statistical tests, we restricted our attention to lags between -30 and $+30$. Because lag zero is a special case (similarity could be boosted simply because the same visual features present during study and test of the same stimulus), lag zero was not included in statistical analyses of lag.

Isolating the neural signature of episodic memory—memory advantage index

A goal of this work is to identify the neural correlates of episodic recollection. Here, we used confidence ratings to compare between memories with large *vs* small episodic contributions. A large body of work has argued that responses to old items that do not receive the highest confidence response rely on familiarity, whereas highest-confidence old responses rely on a mixture of familiarity and recollection (***Yonelinas et al., 2002***; ***Diana et al., 2007***). In order to isolate the contribution due to episodic recollection, we computed a difference between old probes that received a highest-confidence response and old probes that did not receive a highest-confidence response.

In order to compute this difference due to recollection we started by taking the product for each pair of stimuli that entered into the contiguity analysis aggregated by lag. However, rather than averaging over units, as in taking the normalized inner product we computed a matrix with each possible lag corresponding to the columns and each unit corresponding to the rows. Separate matrices were computed using the similarity for low-confidence and high confidence trials. To estimate a difference attributable to episodic recollection, we took a paired t-test (over units) for each lag as a measure of "memory advantage". The use of the paired t-statistic minimizes variability due to difference in the units.

The t-statistic can be used to evaluate the null hypothesis directly (values greater than 1.96 are statistically different from zero), but also the t-statistic can be compared for different lags. If there was no recovery of temporal context, the memory advantage would be the same across lags; a systematic change in the memory advantage as a function of lag must reflect recovery of some form of information that was present during study of the list.

Visually-selective (VS) units

In order to determine if the gradually-changing temporal context representations examined using the recency and contiguity analyses were distinct from visual representations we repeated these analyses restricting the analysis to visually-selective (VS) units. VS units are those that responded differently to the different categories of images, as assessed by an ANOVA on their firing rate. Methods for identifying VS units were identical to those reported in detail previously (***Rutishauser et al., 2015***). 213/1286 units were classified as visually-selective.

References

Cai DJ, Aharoni D, Shuman T, Shobe J, Biane J, Song W, Wei B, Veshkini M, La-Vu M, Lou J, Flores SE, Kim I, Sano Y, Zhou M, Baumgartel K, Lavi A, Kamata M, Tuzsynski M, Mayford M, Golshani P, et al. A shared neural ensemble links distinct contextual memories encoded close in time. Nature. 2016; 534(7605):115–118.

Carr MF, Jadhav SP, Frank LM. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nature Neuroscience. 2011; 14(2):147–153.

Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA. Decoding individual episodic memory traces in the human hippocampus. Current Biology. 2010; 20(6):544–547.

Chan SCY, Applegate MC, Morton NW, Polyn SM, Norman KA. Lingering representations of stimuli influence recall organization. Neuropsychologia. In press; .

Copara MS, Hassan AS, Kyle CT, Libby LA, Ranganath C, Ekstrom AD. Complementary roles of human hippocampal subregions during retrieval of spatiotemporal context. Journal of Neuroscience. 2014; 34(20):6834–6842.

Dede AJ, Frascino JC, Wixted JT, Squire LR. Learning and remembering real-world events after medial temporal lobe damage. Proceedings of the National Academy of Sciences. 2016; 113(47):13480–13485.

Deuker L, Bellmund JL, Schröder TN, Doeller CF. An event map of memory space in the hippocampus. eLife. 2016; 5:e16534.

Diana RA, Yonelinas AP, Ranganath C. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends in Cognitive Science. 2007; 11(9):379–86.

DuBrow S, Davachi L. Temporal memory is shaped by encoding stability and intervening item reactivation. Journal of Neuroscience. 2014 Oct; 34(42):13998–4005. doi: 10.1523/JNEUROSCI.2535-14.2014.

DuBrow S, Davachi L. Temporal binding within and across events. Neurobiology of learning and memory. 2016; 134:107–114.

Eichenbaum H, Yonelinas AP, Ranganath C. The Medial Temporal Lobe and Recognition Memory. Annual Review of Neuroscience. 2007; 30:123–152.

Fortin NJ, Wright SP, Eichenbaum H. Recollection-like memory retrieval in rats is dependent on the hippocampus. Nature. 2004; 431(7005):188–91.

Hassabis D, Kumaran D, Vann SD, Maguire EA. Patients with hippocampal amnesia cannot imagine new experiences. Proceedings of the National Academy of Sciences USA. 2007; 104(5):1726–31. doi: 10.1073/pnas.0610561104.

Howard MW, Kahana MJ. A Distributed Representation of Temporal Context. Journal of Mathematical Psychology. 2002; 46(3):269–299.

Howard MW, Shankar KH, Aue W, Criss AH. A distributed representation of internal time. Psychological Review. 2015; 122(1):24–53.

Howard MW, Viskontas IV, Shankar KH, Fried I. Ensembles of Human MTL Neurons "Jump Back in Time" in Response to a Repeated Stimulus. Hippocampus. 2012; 22(9):1833–1847.

Howard MW, Youker TE, Venkatadass V. The persistence of memory: Contiguity effects across several minutes. Psychonomic Bulletin & Review. 2008; 15(PMC2493616):58–63.

Howard MW, MacDonald CJ, Tiganj Z, Shankar KH, Du Q, Hasselmo ME, Eichenbaum H. A unified mathematical framework for coding time, space, and sequences in the hippocampal region. Journal of Neuroscience. 2014; 34(13):4692–707. doi: 10.1523/JNEUROSCI.5808-12.2014.

Hsieh LT, Gruber MJ, Jenkins LJ, Ranganath C. Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. Neuron. 2014; 81(5):1165–1178.

Hsieh LT, Ranganath C. Cortical and subcortical contributions to sequence retrieval: Schematic coding of temporal context in the neocortical recollection network. NeuroImage. 2015; 121:78–90.

Hyman JM, Ma L, Balaguer-Ballester E, Durstewitz D, Seamans JK. Contextual encoding by ensembles of medial prefrontal cortex neurons. Proceedings of the National Academy of Sciences USA. 2012; 109:5086–91. doi: 10.1073/pnas.1114415109.

Kahana MJ. Associative retrieval processes in free recall. Memory & Cognition. 1996; 24:103–109.

Kahana MJ. Foundations of human memory. OUP USA; 2012.

Lositsky O, Chen J, Toker D, Honey CJ, Shvartsman M, Poppenk JL, Hasson U, Norman KA. Neural pattern change during encoding of a narrative predicts retrospective duration estimates. eLife. 2016 nov; 5:e16070. doi: 10.7554/eLife.16070.

MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H. Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. Neuron. 2011; 71(4):737–749.

Mankin EA, Sparks FT, Slayyeh B, Sutherland RJ, Leutgeb S, Leutgeb JK. Neuronal code for extended time in the hippocampus. Proceedings of the National Academy of Sciences. 2012; 109:19462–7. doi: 10.1073/pnas.1214107109.

Manning JR, Polyn SM, Litt B, Baltuch G, Kahana MJ. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proceedings of the National Academy of Science, USA. 2011; 108(31):12893–7.

Manns JR, Howard MW, Eichenbaum HB. Gradual changes in hippocampal activity support remembering the order of events. Neuron. 2007; 56(3):530–540.

Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, Hefft S, Merkow M, Polyn SM, Jacobs J, Kahana MJ, Schulze-Bonhage A. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science. 2013; 342(6162):1111–4. doi: 10.1126/science.1244056.

Naya Y, Suzuki WA. Integrating What and When Across the Primate Medial Temporal Lobe. Science. 2011; 333(6043):773–776.

Nielson DM, Smith TA, Sreekumar V, Dennis S, Sederberg PB. Human hippocampus represents space and time during retrieval of real-world memories. Proceedings of the National Academy of Sciences. 2015; 112(35):11078–11083.

Onyper SV, Zhang YX, Howard MW. Some-or-none recollection: Evidence from item and source memory. Journal of Experimental Psychology: General. 2010; 139(2):341–64.

Pfeiffer BE, Foster DJ. Autoassociative dynamics in the generation of sequences of hippocampal place cells. Science. 2015; 349(6244):180–183.

Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. Psychological Review. 2009; 116:129–156.

Rashid AJ, Yan C, Mercaldo V, Hsiang HLL, Park S, Cole CJ, De Cristofaro A, Yu J, Ramakrishnan C, Lee SY, et al. Competition between engrams influences fear memory formation and recall. Science. 2016; 353(6297):383–387.

Rubin A, Geva N, Sheintuch L, Ziv Y. Hippocampal ensemble dynamics timestamp events in long-term memory. eLife. 2015; 4:e12247.

Rutishauser U, Mamelak AN, Schuman EM. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron. 2006; 49(6):805–813.

Rutishauser U, Ye S, Koroma M, Tudusciuc O, Ross IB, Chung JM, Mamelak AN. Representation of retrieval confidence by single neurons in the human medial temporal lobe. Nature neuroscience. 2015; 18(7):1041–1050.

Salz DM, Tiganj Z, Khasnabish S, Kohley A, Sheehan D, Howard MW, Eichenbaum H. Time cells in hippocampal area CA3. Journal of Neuroscience. 2016; 36:7476–7484.

Schacter DL, Addis DR, Buckner RL. Remembering the past to imagine the future: the prospective brain. Nature Reviews, Neuroscience. 2007; 8(9):657–661.

Schwartz G, Howard MW, Jing B, Kahana MJ. Shadows of the past: Temporal retrieval effects in recognition memory. Psychological Science. 2005; 16(11):898–904.

Sederberg PB, Howard MW, Kahana MJ. A context-based theory of recency and contiguity in free recall. Psychological Review. 2008; 115:893–912.

Sederberg PB, Miller JF, Howard MW, Kahana MJ. The temporal contiguity effect predicts episodic memory performance. Memory & Cognition. 2010; 38:689–699.

Squire LR, Wixted JT, Clark RE. Recognition memory and the medial temporal lobe: a new perspective. Nature Reviews, Neuroscience. 2007; 8(11):872–83.

Staresina BP, Fell J, Dunn JC, Axmacher N, Henson RN. Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. Proceedings of the National Academy of Sciences. 2013; 110(8):3119–24. doi: 10.1073/pnas.1215710110.

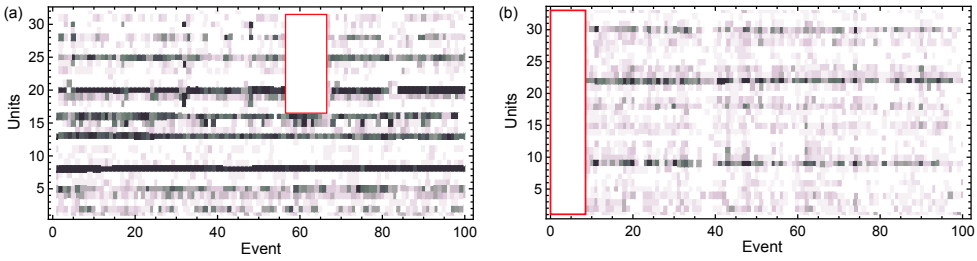

Figure A1. *Artifact rejection.* The raster plots show the activity of each unit (row) as a function of time. The method for artifact rejection described in the text identified the red squares as an artifact. The rejected units were all located in the same brain region.

Tulving E. Episodic and semantic memory. In: Tulving E, Donaldson W, editors. *Organization of Memory.* New York: Adademic Press; 1972.p. 381–403.

Unsworth N. Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. Journal of Memory and Language. 2008; 59:223–236.

Wixted JT. Dual-process theory and signal-detection theory of recognition memory. Psychological Review. 2007; 114(1):152–76.

Yaffe RB, Kerr MSD, Damera S, Sarma SV, Inati SK, Zaghloul KA. Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. Proceedings of the National Academy of Sciences. 2014; 111(52):18727–32. doi: 10.1073/pnas.1417017112.

Yonelinas AP, Kroll NE, Quamme JR, Lazzara MM, Sauvé MJ, Widaman KF, Knight RT. Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. Nature Neuroscience. 2002; 5(11):1236–41.

Yonelinas AP. The Nature of Recollection and Familiarity: A Review of 30 Years of Research. Journal of Memory and Language. 2002; 46(3):441–517.

Appendix

Artifact rejection results

Figure A1 shows examples of artifacts that were rejected by the artifact rejection algorithm. There were a total of 1286 units across each potentially present in 200 events Of these, a total of 664 points were rejected. These artifacts were found and rejected from six of 49 sessions.

Raw contiguity analyses

The results for the analysis of neural similarity as a function of lag for both all units and for the VS units are shown in Figure A2. The similarity at lag zero was much higher than other lags for highest-confidence responses for both all units taken together and for VS units. For all units (Fig. A2a), the similarity value at lag zero for recollected probes (filled circles) was greater than for other lags (sign test 40/40, $p < .001$). Similar results were found for VS units taken alone (Fig. A2b). For old probes that were not recollected (open circles), there was no discernible advantage for lag zero for all units taken together. For VS units there was a reliable advantage for lag zero over other lags (39/40, $p < .001$ by a sign test) although the numerical value was much smaller than for recollected probes.

At more remote lags, there was a tendency for a contiguity effect for probes that received a highest-confidence response and a tendency towards an anti-contiguity effect for probes that did not receive a highest-confidence response. We quantified this by performing an ANOVA with |lag| as regressor (excluding lag zero) and direction (backward or forward) as a categorical variable separately for recollected and unrecollected old probes. We did this ANOVA for both all units (Fig. A2a) and restricting our attention to VS units (Fig. A2b). Both set of analyses led to similar conclusions. Considering all units, the neural similarity for recollected probes showed

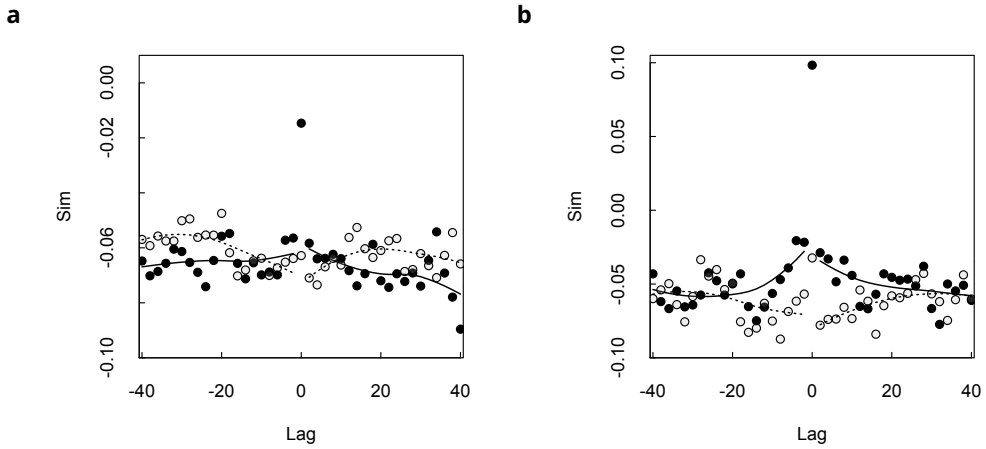

Figure A2. *Neural similarity as a function of lag for old probes that were recollected (received a highest-confidence yes response, filled circles) and old probes that were not recollected (grey open circles).* Statistical analyses confirm that there was a contiguity effect (inverted-V centered around zero) for recollected probes but an anti-contiguity effect (V-shaped centered around zero) for unrecollected probe. All data points except lag zero were binned. A LOESS curve was fitted for each data set. **a.** All units. **b.** Analysis restricted to units categorized as visually-selective.

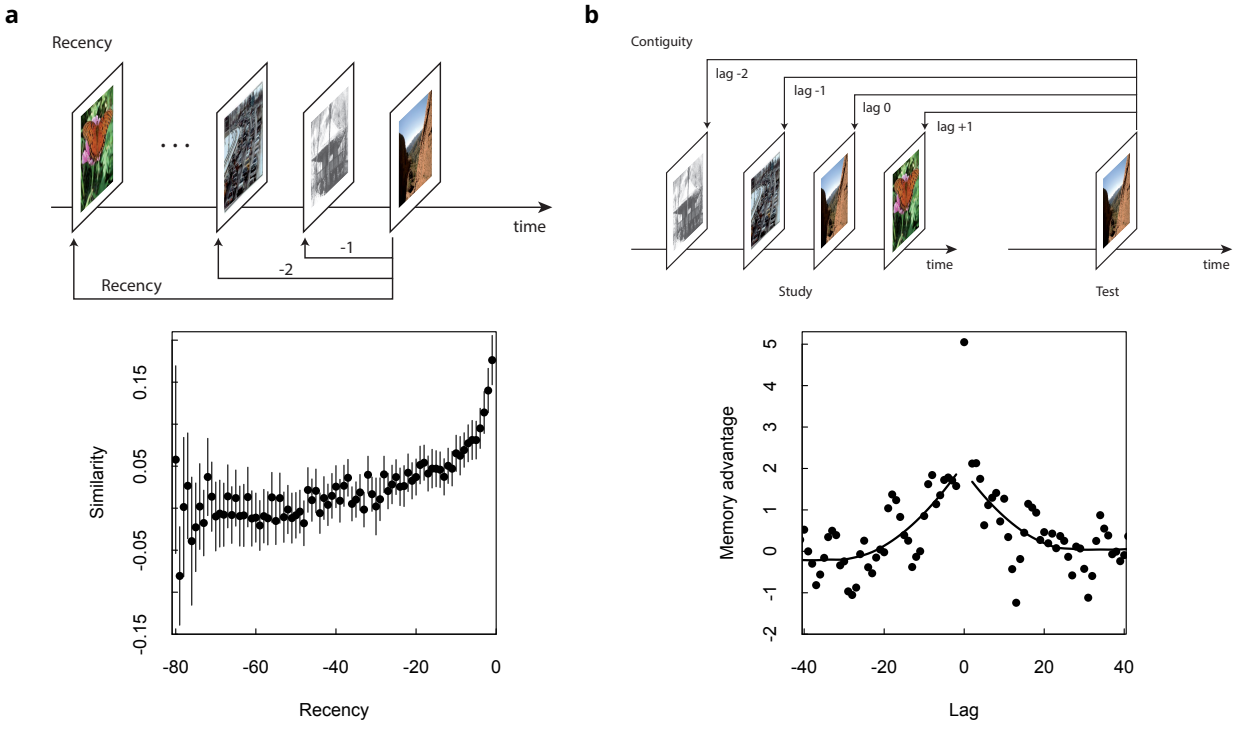

Figure A3. *Visual-category sensitive units showed a neural contiguity effect.* Similar to Fig. 3b VS units display a robust contiguity effect. The memory effect is enhanced around lag 0, falling off both in the backward and forward direction. This was confirmed by statistical analysis.

a significant effect of |lag| $F_{(1,56)} = 8.09$, $p < .01$ and no effect of direction $F_{(1,56)} = 2.15$, nor an interaction $F_{(1.56)} = 1.40$. For probes that were not recollected, the effect of |lag| was again significant $F_{(1,56)} = 9.59$, $p < .005$, and there was neither an effect of direction $F_{(1,56)} = 1.35$, nor an interaction $F_{(1.56)} = 1.79$. However, the effect of |lag| was in different directions for recollected and unrecollected probes. For recollected probes, the effect of |lag| was positive in both the forward and backward directions (like an inverted-V); for unrecollected probes, the effect of |lag| was negative in both the forward and backward directions (a V-shaped curve around zero). For recollected old probes, the effect of lag on neural similarity in the forward direction reached significance, $(-.46 \pm .15) \times 10^{-3}$, $p < .005$ and the contiguity effect in the backward direction did not reach significance $(.19 \pm .18) \times 10^{-3}$, $p > .2$. For unrecollected probes there was a reliable negative effect of lag in the backward direction, $(-.52 \pm .17) \times 10^{-3}$, $p < .005$. There was a similar trend in the forward direction, although the trend did not reach significance $(.21 \pm .16) 10 \times^{-3}$.

Results were similar when considering only VS units (Fig. A2b). For recollected probes, an ANOVA on neural similarity with |lag| as regressor (excluding lag zero) and direction (backward or forward) as a categorical variable resulted in a significant effect of |lag| $F_{(1,56)} = 5.39$ $p < 0.05$ but no effect of direction $F_{(1,56)} = .02$, nor an interaction $F_{(1.56)} = .93$. For unrecollected probes the effect of |lag| was again significant $F_{(1,56)} = 11.89$ $p < 0.005$, but there was no effect of direction $F_{(1,56)} = .07$, nor an interaction $F_{(1.56)} = .08$. As with all units, the effect of |lag| on neural similarity was in different directions for the recollected and unrecollected probes such that for recollected probes the effect of contiguity is shaped like an inverted-V centered around zero but shaped like a V for unrecollected probes. For recollected probes, there was a reliable contiguity effect in the backward direction, $(.99 \pm .43) \times 10^{-3}$, $p < .03$ and a tendency toward a contiguity effect in the forward direction $(-.41 \pm .42) \times 10^{-3}$, $p < .35$. In contrast, for the unrecollected probes there was a reliable anti-contiguity effect in both the forward and backward directions, $(0.91 \pm .40) \times 10^{-3}$, $p < .05$ and $(-1.08 \pm .41) \times 10^{-3}$, $p < .05$.

Taken together these findings suggest that episodic recollection was associated with recovery of gradually-changing information. Moreover, the VS units robustly exhibited the effect taken alone, suggesting that the retrieval of temporal context was primarily driven by units that were sensitive to the content of the stimuli (note also the change of scale between the panels in Figure A2).

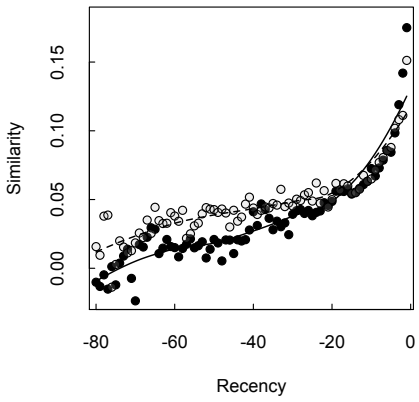

Figure A4. *Enhanced neural similarities during test.* The increasing retrieval of previous temporal context leads to an enhancement of population vector similarities during test session. During test, the similarity was reliably higher at widely separated points, consistent with the predictions of retrieved temporal context. Smoothed curves are from a LOESS regression.

Additional evidence for a jump back in time

If some old probes caused a jump back in time, we would expect this to result in greater pattern similarity between pairs of test vectors and pairs of study vectors. If two test probes recover information from the temporal context during presentation of those items, then this would be expected to result in additional similarity between these two test events if those two test probes were close together in the list. Consistent with this hypothesis, although both study and test lists showed a reliable recency effect, for the test list the similarity stablized at a higher baseline value. At recency less than about -20, the test similarity was reliably higher (Fig. A4) at almost all values of recency.

As a simple quantitative measure the similarity for test is greater than the similarity for study at $13/20$ of the points for recency -1 through -20. In contrast, for recency -21 through -80 the simi-

larity is greater for $55/60$ of the values. These two proportions differ from one another, $\chi^2(1) = 6.41$, $p < .05$.