Oscillatory patterns in temporal lobe reveal context reinstatement during memory search

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

Submitted to Proceedings of the National Academy of Sciences of the United States of America. DRAFT - DO NOT CITE!

Psychological theories of memory posit that when people recall a past event, they not only recover the features of the event itself, but they also recover contextual features representing information about neighboring events and the thoughts they evoke. The ability to associate items with their temporal context and to recover contextual information during recall has been termed episodic memory. We sought to determine whether contextual reinstatement in human memory may be observed in electrical signals recorded from the human brain. By analyzing electrocorticographic recordings taken as 69 neurosurgical patients studied and recalled lists of words, we were able to uncover a neural signature of context reinstatement. Upon recalling a studied item, we found that the pattern of brain activity was not only similar to the pattern observed when the item was studied, but was also similar to neighboring list items, with similarity decreasing reliably with positional distance. The degree to which individual patients displayed this neural signature of context reinstatement was correlated with their tendency to successively recall neighboring list items. These effects were particularly strong in temporal lobe recordings. Our findings show that recalling a past episode evokes a neural signature of the temporal context in which the episode occurred, thus pointing to a neural basis for episodic

episodic memory oscillations contiguity

Abbreviations: ECoG, electrocorticography

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

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

Recent neurocomputational models of episodic memory [5, 6] suggest that contextual reinstatement underlies the *contiguity effect* — people's tendency to successively recall items that were presented in nearby positions on a studied list [7]. Behavioral studies of memory show that, for a given class of

memories, the contiguity effect can span many other intervening memories [8, 9]. 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 [10, 11, 12, 13, 14]. The contiguity effect is an example of temporal clustering, which is perhaps the dominant form of organization in free recall.

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

Results

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

Context-based models conceive of context as a representation that integrates incoming information with a long time constant [21], leading to the prediction that the representation of temporal context evolves gradually as the experi-

Reserved for Publication Footnotes

^{*}Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, USA,†Department of Psychology, Vanderbilt University, Nashville, USA,‡Department of Neurosurgery, University of Pennsylvania School of Medicine, Philadelphia, USA, ®Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, USA, and Department of Psychology, University of Pennsylvania, Philadelphia, USA

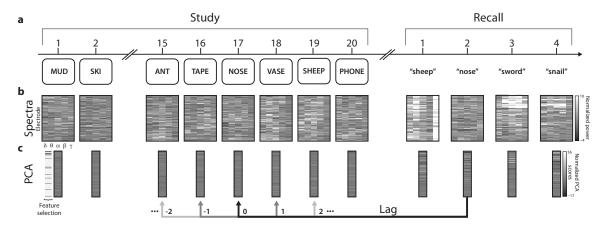


Fig. 1. Illustration of behavioral and electrophysiological methods. a. After studying a list of 20 words and performing a brief distraction task, a participant recalls as many words as he can remember, in any order. b. During each study presentation and just prior to each recall event, we calculate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

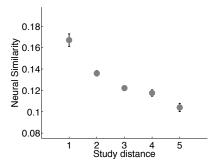


Fig. 2. Evolution of ECoG activity as participants study lists of words. Mean neural similarity is shown as a function of study distance (difference in serial position) between pairs of presented words. Error bars denote ± 1 SEM.

ment progresses [22]. We asked whether the neural recordings supported a gradually changing representation of context by regressing, for each participant, the mean similarity between the principal component vectors on their positional distance in the studied list (Fig. 1). The similarity in recorded activity during each pair of word presentations decreased with the positional distance distance between the presentations (t-test on distribution of t-values from the regressions: $t(64) = -9.31, p < 10^{-10}$), indicating that the ECoG recordings evolve gradually over the course of the studied lists. Whereas this gradually changing neural representation is consistent with context-based models, such a result would also be expected to arise due to other autocorrelated neural processes that lack the rich dynamics implied by context-based theories of memory. To determine whether this gradually changing neural representation reflects the contexts in which list items were studied, we selected the autocorrelated PCA-derived features (henceforth referred to as feature vectors) for further analysis (see *Methods*).

To test whether the gradually changing neural representation is reinstated during recall, we compared feature vectors recorded during each study and recall event. First we identified the serial position (on the presentation list) of each correctly recalled word. If neural activity during study is reinstated during recall, then the neural activity recorded during

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

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 [21]. For each participant we regressed the mean neural similarity between feature vectors on lag, seperately 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(39) = -3.63, p = 0.0008) and negative (t(39) = -2.42, p = 0.02) directions. The decrease in similarity with absolute lag in both directions cannot be explained solely by content reinstatement or autocorrelated noise in the recordings (Figs. 3, S1). Mean similarity as a function of lag across participants is shown in Figure 4a.

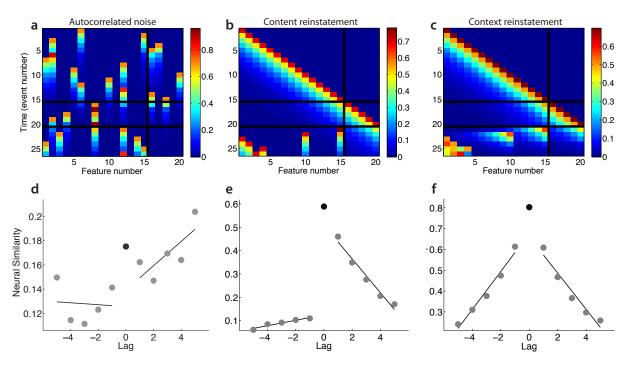


Fig. 3. Predicted neural similarity as a function of lag according to three models. a-c. These panels show the pattern of activations for a simulated 20 neuron neural network as a 15 item list is studied. Events 1 - 15 of each matrix show activations after each item is presented. Events 16-20 show activations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

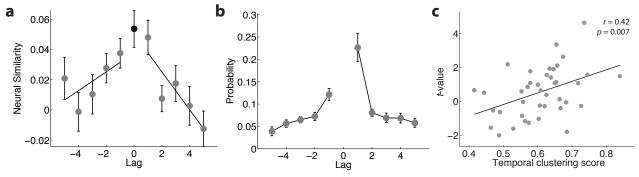


Fig. 4. A neural signature of temporal context reinstatement. a. Neural similarity between the feature vector corresponding to recall of a word from serial position i and study of a word from serial position $i + \log$ (black dot denotes study and recall of the same word, i.e., $\log = 0$). **b.** Participants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

The decrease in neural similarity with absolute lag elegantly mirrors the contiguity effect - people's striking tendency to make transitions to neighboring items rather than remote ones, as seen in behavioral data for the same participants (Fig. 4b). Consistent with the hypothesis that the contiguity effect arises due to reinstatement of context [23, 6, 5], participants with stronger neural signatures of context reinstatement exhibited more pronounced contiguity effects than did participants with weaker reinstatement effects (r = 0.42, p = 0.007, Fig. 4c). In recalling a list item, people not only reinstate that item's representation, as has been recently documented [24, 25] 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 [26, 22, 27, 28, 29] and prefrontal cortex [28, 21, 30] 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: temporal lobe (including the hippocampus and medial temporal lobe), frontal lobe (including prefrontal cortex), parietal lobe, and occipital lobe (Fig. S2a). We found that neural activity recorded from temporal lobe electrodes exhibited a decrease in similarity with increasing absolute lag in both the positive and negative directions (Fig. S2b; positive: t(20) = -2.20, p = 0.04; negative: t(20) = -2.82, p = 0.01). As in the whole brain analysis, the neural signature of context reinstatement in the temporal lobe was significantly correlated with the temporal clustering of participants' recalls (Fig. S2c; r = 0.48, p = 0.03). The frontal lobe exhibited a weak neural signature of context reinstatement that trended towards significance (positive: t(20) = -2.85, p = 0.01; negative: t(20) = -1.54, p = 0.14). However, this frontal signature of context reinstatement was not correlated with temporal clustering of participants' recalls (r = -0.08, p = 0.73). Our findings in the parietal and occipital lobes were inconclusive due to insufficient data.

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 similarity 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, 23, 5, 6]. These theories posit that a gradually changing contextual state becomes associated with each experienced event, and that recalling an event revives the contextual state associated with the original experience. This retrieved context, in turn, activates other memories that were associated with similar contexts, producing the contiguity effect seen in recall tasks (Fig. 4b). The present findings provide critical neurobiological evidence in support of context reinstatement by showing that remembering an item reinstates the patterns of distributed oscillatory activity associated with surrounding (contextual) items from the original study episode. This neural signature of context reinstatement was observed both for the whole brain analysis and for recordings taken only from the temporal lobe.

What we have called a context representation might also be attributed to a rehearsal-based model where participants engage in short bouts of covert recall during the study period. By a rehearsal model, words are rehearsed after they are presented, and more recently presented items are more likely to be rehearsed than distant items. Rehearsal models have been shown to be difficult to distinguish from contextbased models [31, 32], likely because a context-based mechanism is necessary to explain the pattern of rehearsals made in a free-recall task. To distinguish between these two classes of models 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 [33, 34, 35, 36, 37]). Thus if our basic findings were driven by rehearsal, one would expect that participants exhibiting strong neural context reinstatement should also show a strong primacy effect. However, we observed no significant correlation between primacy and neural context reinstatement (r = 0.12, p = 0.45; see Supplemental analysis methods), indicating that rehearsal during study is unlikely to account for our findings.

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

Materials and Methods

Participants. We tested 69 patients with drug-resistant epilepsy who had arrays of subdural and/or depth electrodes surgically implanted for one to four weeks to localize the sites of seizure onset (see Tab. S1). The clinical team determined the placement of these electrodes with the goal of localizing suspected epileptogenic foci and identifying functional regions to be avoided in surgery (see Supplemental recording methods for details on recording methods). Our research protocol was approved by the appropriate institutional review boards and informed consent was obtained from the participants and their guardians. Data were collected as part of a long-term multicenter study with previously published articles describing separate analyses conducted on subsets of these data [38, 39, 19].

Behavioral methods. Participants studied lists of 15 or 20 high-frequency nouns for a delayed free-recall task. Following a fixation cue, the computer displayed each word for 1600 ms followed by a 800-1200 ms blank inter-stimulus interval. Each word was displayed at most once within a single testing session. For 18 s following list presentation, participants solved a series of single-digit addition problems of the form A+B+C=X. Participants were then given 45 s to recall list items in any order. Vocal responses, digitally recorded during the trial, were scored for subsequent analysis. Participants recalled $22.5\pm1.0\%$ (mean \pm SEM) of the studied words. Repetitions and incorrect recalls ($32.4\pm2.6\%$ of all responses) were excluded from our analyses, as were responses that occurred within 1 s of a prior vocalization.

Data analysis. We measured oscillatory power in the ECoG recordings by applying a Hilbert transform to the Butterworth-bandpassed signal in each of five frequency bands (see Results). To reduce edge artifacts, we computed power at each frequency for the entire recording session before parsing the recordings into experimental events. Before applying PCA to the frequency \times electrode matrices (Fig. 1b) we z-transformed power values relative to the distribution of all events in the recording session (the ztransformation was performed independently for each frequency-electrode pair). We used the Kaiser criterion to choose, for each participant, the principal components that explained a substantial proportion of the variance [40]. We next sought to identify principal components that changed gradually during the study period. Defining the value of component k for events at serial positions i and j as $v_k(i)$ and $v_k(j)$, we computed the correlation, r, between the absolute difference in the components $(|v_k(i)-v_k(j)|)$ and the serial position lag (|i-j|) for $lag=1\dots(l-1)$, where l was the length of the presented list. Components with r>0 and p<0.1were selected for further analysis (see Supplemental analysis methods). Features that met these criteria were identified in 132 (out of 144) recording sessions. We further excluded an additional 68 sessions in which fewer than five candidate context

features were identified. This threshold was chosen to balance two factors: first, we wanted to ensure that the feature vectors were of high enough dimensionality that it would be possible to observe neural signatures of context reinstatement (Fig. 4a) for each participant; second, we wanted to maximize the amount of data included in our analysis. We report the mean number of features selected for each participant in Table S1. To prevent selection bias, recall events were not used in the feature selection process.

- Bower, G. H. (1972) in Coding Processes in Human Memory, eds. Melton, A. W & Martin, E. (John Wiley and Sons, New York), pp. 85–121.
- 2. Tulving, E. (1983) Elements of Episodic Memory. (Oxford, New York).
- 3. Estes, W. K. (1955) Psychological Review 62, 145-154.
- 4. Anderson, J. R & Bower, G. H. (1972) Psychological Review 79, 97-123.
- Polyn, S. M, Norman, K. A, & Kahana, M. J. (2009) Psychological Review 116, 129–156.
- Sederberg, P. B, Howard, M. W, & Kahana, M. J. (2008) Psychological Review 115, 893–912.
- 7. Kahana, M. J. (1996) Memory & Cognition 24, 103-109.
- Howard, M. W & Kahana, M. J. (1999) Journal of Experimental Psychology: Learning, Memory, and Cognition 25, 923–941.
- Howard, M. W, Youker, T. E, & Venkatadass, V. (2008) Psychonomic Bulletin & Review 15, 58–63.
- 10. Raaijmakers, J. G. W & Shiffrin, R. M. (1981) Psychological Review 88, 93-134.
- 11. Jensen, O & Lisman, J. E. (1998) J. Neuroscience 18, 10688-10699.
- 12. Lisman, J. (1999) Neuron 22, 233-242.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005) Psychological Review 112, 3–42.
- 14. Rolls, E. T & Kesner, R. P. (2006) Progress in Neurobiology 79, 1 48.
- Fell, J, Klaver, P, Lehnertz, K, Grunwald, T, Schaller, C, Elger, C. E, & Fernandez, G. (2001) Nature Neuroscience 4, 1259–1264.
- 16. Paller, K. A & Wagner, A. D. (2002) Trends in Cognitive Sciences 6, 93-102.
- 17. Buzsáki, G. (2006) Rhythms of the Brain. (Oxford University Press, New York).
- Osipova, D, Takashima, A, Oostenveld, R, Fernndez, G, Maris, E, & Jensen, O. (2006)
 J Neurosci 26, 7523–7531.
- Sederberg, P. B, Schulze-Bonhage, A, Madsen, J. R, Bromfield, E. B, Litt, B, Brandt, A, & Kahana, M. J. (2007) Psychological Science 18, 927–932.
- 20. Jacobs, J & Kahana, M. J. (2009) Journal of Neuroscience 29, 10203-10214.
- 21. Polyn, S. M & Kahana, M. J. (2008) Trends in Cognitive Sciences 12, 24-30.
- 22. Manns, J. R, Howard, M. W, & Eichenbaum, H. (2007) Neuron 56, 530-540.

ACKNOWLEDGMENTS. We thank D. Brainard, M. Howard, J. Jacobs, L. Lohnas, L. Ungar, and C. Weidemann for useful discussions. We also thank E. Rosenberg for help testing patients and J. Miller for helping to organize patient information. This work was supported by National Institutes of Mental Health grants MH61975, MH55687, and F31MH088118, and a grant from the Dana Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of our supporting organizations.

- Howard, M. W & Kahana, M. J. (2002) Journal of Mathematical Psychology 46, 269–299.
- Polyn, S. M, Natu, V. S, Cohen, J. D, & Norman, K. A. (2005) Science 310, 1963– 1966.
- Gelbard-Sagiv, H, Mukamel, R, Harel, M, Malach, R, & Fried, I. (2008) Science 3, 96–101.
- Skaggs, W. E, McNaughton, B. L, Wilson, M. A, & Barnes, C. A. (1996) Hippocampus 6. 149–172.
- Pastalkova, E, Itskov, V, Amarasingham, A, & Buzsáki, G. (2008) Science 321, 1322

 1327.
- 28. Jenkings, L. J & Ranganath, C. (2010) Journal of Neuroscience In press.
- 29. Howard, M. W, Viskontas, I. V, Shankar, K. H, & Fried, I. (Submitted) Submitted.
- 30. Schacter, D. L. (1987) Psychobiology 15, 21-36.
- Davelaar, E. J, Usher, M, Haarmann, H. J, & Goshen-Gottstein, Y. (2008) Psychological Review 115, 1116–1118.
- Laming, D. (2006) Journal of Experimental Psychology: Learning, Memory, and Cognition 32, 1146–1163.
- 33. Welch, G. B & Burnett, C. T. (1924) American Journal of Psychology 35, 396-401.
- Postman, L & Phillips, L. W. (1965) Quarterly Journal of Experimental Psychology 17, 132–138.
- Glanzer, M & Cunitz, A. R. (1966) Journal of Verbal Learning and Verbal Behavior 5, 351–360.
- Atkinson, R. C & Shiffrin, R. M. (1968) in The psychology of learning and motivation, eds. Spence, K. W & Spence, J. T. (Academic Press, New York) Vol. 2, pp. 89–105.
- 37. Rundus, D. (1971) Journal of Experimental Psychology 89, 63-77.
- Sederberg, P. B, Kahana, M. J, Howard, M. W, Donner, E. J, & Madsen, J. R. (2003) Journal of Neuroscience 23, 10809–10814.
- Sederberg, P. B, Schulze-Bonhage, A, Madsen, J. R, Bromfield, E. B, McCarthy, D. C, Brandt, A, Tully, M. S, & Kahana, M. J. (2007) Cerebral Cortex 17, 1190–1196.
- 40. Kaiser, H. F. (1960) Educational and Psychological Measurement 20, 141 151.