Two kinds of memory signals in neurons of the human hippocampus

Zhisen Urgolites (✉ zurgolites@ucsd.edu)
UCSD

John Wixted (✉ jwixted@ucsd.edu)
UCSD

Stephen Goldinger
Arizona State University

Megan H. Papesh
New Mexico State University

David M. Treiman
Barrow Neurological Institute

Larry R. Squire
Veterans Affairs Medical Center, San Diego; UCSD

Peter N. Steinmetz
Neurtex Brain Research Institute

Research Article

Keywords: memory, hippocampus, sparse coding, pattern separation, single-unit activity

DOI: https://doi.org/10.21203/rs.3.rs-588780/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Abstract

Some studies of the neural representation of memory in the human hippocampus have identified memory signals reflecting the categorical status of test items (novel vs. repeated). Others have identified pattern-separated, episodic memory signals reflecting recognition of particular test items. Here, we report that both kinds of memory signals can be found in the hippocampus, and we consider their possible functions. We recorded single-unit activity from four brain regions (hippocampus, amygdala, anterior cingulate, and prefrontal cortex) of epilepsy patients as they performed a continuous recognition task. The generic signal was found in all four regions, whereas the sparse, pattern-separated signal was limited to the hippocampus, as predicted by longstanding computational models.

Background

Two kinds of memory signals in neurons of the human hippocampus

The hippocampus is known to be essential for the formation of declarative (conscious) memory (1, 2), including both episodic memory (memory for events) and semantic memory (factual knowledge). Episodic memories represent the “what, when, and where” information about remembered events (3). Here, we focus on the neural representation of episodic memory for events, specifically words presented and later repeated in a continuous recognition memory format (4).

Bilateral hippocampal lesions result in substantial anterograde amnesia for events, whether memory is tested by recall or recognition (5). By contrast, bilateral lesions to a more anterior medial temporal lobe structure—the amygdala—have no such effect (6). One might therefore expect to find single-unit activity associated with episodic memory in the hippocampus but not in the amygdala. Yet, the earliest single-neuron studies failed to detect hippocampal neurons that fired differentially to recently presented test items vs. novel items. This was true in studies with humans (7, 8) and monkeys (9-11). One early study with monkeys did identify a few such neurons (e.g., 12).

Subsequent studies have successfully detected such activity. These found that ~10% of hippocampal neurons exhibited differential firing rates for repeated vs. novel items, some firing more for repeated items and others firing more for novel items. Thus, these memory-selective neurons responded differentially to the generic, categorical status of test items (repeated vs. novel). Surprisingly, similar “memory-selective” neurons were also reliably detected in the amygdala at approximately the same frequency (13-15).

Yet, these are not episodic memory signals (i.e., signals representing memory for specific events). According to neurocomputational models dating back to Marr (16), episodic memory representations in the hippocampus are supported by sparse neural codes (17-19). If memories for individual items are sparsely coded in largely non-overlapping (pattern-separated) neural assemblies, then, one should be able to find, not neurons that simply report novelty vs. familiarity, but neurons that respond differentially to particular repeated items. Two recent single-unit studies with humans detected such neurons in the hippocampus but not in the amygdala (24, 25), apparently reflecting sparsely-coded episodic memories.
Why are two different kinds of memory signals being found in different studies? We now report that both memory signals are present in the hippocampus in the same dataset from the same experiment. Using a continuous recognition memory procedure, neurons were recorded from four brain structures: hippocampus, prefrontal cortex, anterior cingulate cortex, and amygdala. Altogether, 55 continuous recognition memory sessions were completed by 34 epilepsy patients who had implanted clinical depth electrodes with microwires measuring single-unit activity (SUA) and multi-unit activity (MUA) bilaterally (20). We limited our data analyses to SUA. Words were presented consecutively and repeated once after varying lags, and patients judged each word as either “novel” or “repeated.” Thus, repeated words differed from their earlier presentations as novel words only with respect to their combined “what, when, and where” episodic status (3).

Results

Behavioral Performance. On average, behavioral performance was well above chance and below ceiling: 83.5 ± 2.0% correct for the first presentation of words (i.e., correct rejection rate for novel words, 0.84; false alarm rate, 0.17) and 80.6 ± 2.8% correct for the second presentation of words (i.e., hit rate for repeated words, 0.81; miss rate, 0.19). The average $d'$ across sessions was 2.20 (min = 0.12, max = 3.69).

Identifying the Generic Memory Signal. A generic memory signal is evident in neurons with firing rates that differentiate novel items from repeated items (memory-selective neurons). In prior work, some memory-selective neurons exhibited higher firing rates for repeated items (familiarity detectors), whereas others exhibited higher firing rates for novel items (novelty detectors). That pattern was also observed in our data. Aggregated across sessions and patients, 38 hippocampal neurons (out of 397 recorded; 9.6%) exhibited statistically significant familiarity/novelty signals (Fig.1). Of these memory-selective neurons, 20 were familiarity detectors and 18 were novelty detectors (see Methods). Using an alpha level of .05, the expected number of memory-selective neurons detected by chance would be 397 * .05 = 19.9. The observed number of 38 was greater than chance ($p < .001$, Fig.1).

We found a similar pattern in the amygdala, where 30 neurons (out of 378 recorded neurons; 7.9%) exhibited significant familiarity/novelty signals (Fig.1). There were 10 familiarity detectors and 20 novelty detectors. The expected number of neurons detected by chance would be 378 * .05 = 18.9. The observed number of 30 was greater than chance ($p < .01$, Fig.1). We repeated this analysis in the other two brain regions (i.e., prefrontal cortex and anterior cingulate) and found that in each of the two regions there was also a significant number of familiarity/novelty neurons (Fig.1). In the prefrontal cortex, we found 22 memory-selective neurons (6 familiarity detectors + 16 novelty detectors) out of 276 recorded neurons (8.0%). In the anterior cingulate, we found 47 memory-selective neurons (18 familiarity detectors + 29 novelty detectors) out of 344 recorded neurons (13.7%). Each number (22 and 47) was higher than the number expected by chance (prefrontal cortex: 276 * .05 = 13.8, $p < .05$; anterior cingulate: 344 * .05 = 17.2, $p < .001$). Thus, the familiarity/novelty signal was found in every brain region that we examined.
**Identifying the Item-Specific, Sparsely Coded Memory Signal.** A sparsely coded signal can be identified by comparing the shapes of the full distributions of normalized spike counts (for all recorded neurons across all test stimuli, sessions, and patients) separately for novel items vs. repeated items. A visual method for making such comparisons is to generate empirical quantile-quantile (QQ) plots (27). Quantiles refer to rank-ordered data broken into subgroups containing equal percentages of the data. For our analyses, the relevant empirical QQ plots consist of the quantiles of the normalized spike counts for novel words (x-axis) vs. the quantiles of the normalized spike counts for repeated words (y-axis), constructed separately for each of the four brain regions (Fig. 2). If these two distributions of normalized spike counts have the same shape (e.g., two unskewed Gaussian distributions or two positively skewed lognormal distributions), the QQ plot will be linear, even if the distributions have different means and/or standard deviations. However, if the repeated-item distribution is more skewed to the right than the novel-item distribution, as predicted by neurocomputational models, the QQ plot will instead exhibit a characteristic, nonlinear deflection.

For the hippocampus, the points on the QQ plot fall mostly along the diagonal line but show a sharp upward deflection for the highest-ranking points (Fig. 2A). As indicated by the light shading of the data points at the rightmost portion of the QQ plot, the deflection reflects the fact that a very small proportion of recordings exhibits a markedly strong response to specific repeated words. Indeed, when only 0.25% of the top-ranking recordings from both distributions were removed, retaining 99.75% of the recordings, the upward deflection was no longer apparent (Fig. 2B). Similar patterns were not apparent in the amygdala, prefrontal cortex, or anterior cingulate (Fig. 2C-H).

We next asked whether the nonlinear QQ plot apparent in the hippocampus reflected a statistically significant departure from linearity. If the effect is real, the distribution of spikes for repeated words should be more positively skewed than the distribution for novel words. Indeed, although both distributions were positively skewed (repeated: skewness = 2.77; novel: skewness = 2.09), the repeated distribution was significantly more skewed ($p = .002$) according to a bootstrap test. The repeated distribution also had significantly higher kurtosis than the novel distribution (20.11 vs. 6.44, $p = .006$). By contrast, in the other three brain regions, neither skewness nor kurtosis differed significantly across the repeated vs. novel distributions (Table 1).

The evidence for sparse coding in the hippocampus was not based on only a few patients, sessions, neurons, or stimuli (Table 2). Instead, the top 0.25% of normalized spikes to repeated words came from 25 different patients (out of 30 patients with single-unit data from the hippocampus) across 38 different sessions (out of 51 sessions with single-unit data from the hippocampus). In addition, each of these 38 sessions contributed 1 to 11 unique single neurons to the top 0.25% (mean = 2.95 neurons per session). On average, each of these neurons responded strongly to 1.66 unique repeated words (range = 1 to 11).

**Discussion**
In this study, we found that single neurons of the human hippocampus exhibit two distinct memory signals. One signal is generic: The recorded neurons respond differentially depending on the categorical status of the test item (novel vs. repeated). This signal was found in all four brain regions that we examined (hippocampus, amygdala, prefrontal cortex, and anterior cingulate). The other memory signal is item-specific: Each neuron responded strongly to a small fraction of repeated words, and each repeated word elicited strong responding in a small fraction of neurons. This signal was found only in the hippocampus.

The method for identifying the item-specific signal in the hippocampus involved comparing the shapes of neuron-by-item normalized spiking distributions for novel vs. repeated items. This method of analysis is not intuitive and likely would not have been pursued absent predictions made by neurocomputational models (16-19). Those models predict that pattern-separated, item-specific memory representations are sparsely coded, in which case the memory signals they generate should be hard to detect. The models further predict that, for test items equated in every respect except their episodic occurrence in the experimental context (novel vs. repeated), these signals should be selectively detected for repeated items. The QQ plots of data from the hippocampus confirmed this prediction (Fig. 2A). This elusive memory signal was predicted to exist only in the hippocampus and, indeed, was observed only in the hippocampus.

Previous single-unit studies with humans typically searched only for the generic memory signal, finding it in the amygdala and the hippocampus (13-15), and later in parietal cortex (22). Similarly, we found the generic memory signal in all four brain regions that were examined. Thus, this signal is not only generic (responding to the categorical status of test items), it is also widespread.

What role might the widespread generic memory signal play? Hippocampal neurons that sparsely code item-specific, episodic memories may distribute generic information to other brain regions (and to other neurons in the hippocampus) to perform memory-related functions not requiring item-specific information. For example, the prefrontal cortex may determine whether the memory signal associated with a test item is strong enough to exceed a decision criterion. Because that decision would depend on memory strength, represented by the generic signal, item-specific content would not be required. Similarly, parietal cortex may play a role in assessing confidence (23, 24), which again requires only information about the strength of the memory signal. On this view, the item-specific memory signal in the hippocampus is fundamental, whereas the generic memory signal – the focus of much prior research – is secondary and derivative.

Our findings contradict a recent claim (25) that episodic memories in the human hippocampus are not stored as sparsely-coded, pattern-separated representations. The proposal is that concept neurons code episodic memories. Concept neurons are neurons that fire selectively when specific concepts (e.g., “Eiffel Tower”) are evoked, whether the evoking stimulus is a picture representing the concept, its printed name, or its spoken name (26). Critically, concept neurons can expand their tuning, responding to unrelated stimuli that have been paired with the relevant concept (27). For example, if the Eiffel Tower and the actor
Jackie Chan are shown together during an experiment, the “Eiffel Tower neuron” will subsequently increase its firing rate to a Jackie Chan image presented alone.

In our view, these neural signals do not reflect episodic memories. Concept neurons respond to any stimulus that evokes their preferred concepts (e.g., the Eiffel Tower), which suggests that they code factual knowledge (i.e., semantic memory). Neurons that code episodic memory code the episodic aspects of a memory, namely, what, when, and where (3). As noted by Suthana et al. (28), the tasks used in concept cell studies do not require the patient to appreciate what, when, and where information about test stimuli. By contrast, our study required the patient to appreciate precisely that kind of information, and it yielded clear evidence of sparsely-coded, pattern-separated episodic memory signals in the hippocampus. Studies using other methodologies have also identified pattern-separated memory signals in the hippocampus (29-31).

We measured differences in single-unit activity for novel vs. repeated words: Each word was presented twice differing only in its status as being novel or repeated. Concept neurons that respond to semantic meaning would respond to both the first and second presentation of a word. By contrast, we described neurons that fired differentially for particular repeated words and that therefore integrated stimulus information about the attributes of “what” (i.e., this particular word), “when” (presented a few minutes ago), and “where” (in this experimental context). This item-specific, neural code was sparse, pattern-separated, and limited to the hippocampus.

**Methods**

**Participants.** The participants were 34 patients (mean age, 41 ± 2.02 y; 21 females and 14 males; all but 2 were right-handed) who had temporal lobe, drug-resistant epilepsy that required implantation of depth electrodes (Ad-Tech Medical) for clinical evaluation and consideration of possible surgical resection of their seizure foci. The patients participated in a total of 55 sessions, with each patient completing one, two, three, or four sessions. An additional 10 sessions from 6 of the 34 patients and all of the data from 1 additional patient were excluded from analysis because of low recognition memory performance (d’ no greater than 0). All patients provided informed consent to participate in the research, using a protocol approved by the Institutional Review Board of St. Joseph’s Hospital and Medical Center.

**Materials and Procedure.** The patients were tested using a continuous recognition task for words. Words were presented one after another, and most of the words were presented a second time after 0, 1, 3, 7, 15, or 31 intervening words. The task was to judge whether each word was novel (i.e., presented the first time) or repeated (i.e., presented a second time). The words were presented either visually on a computer screen or auditorily through headphones. Among 55 sessions (from 34 patients), there were 35 visual sessions and 20 auditory sessions. QQ plot data from the visual sessions only (focusing on recordings from the hippocampus and amygdala only) were analyzed in a previous report (32). Single-unit recordings from each structure were available for most of these sessions (hippocampus = 51 sessions, amygdala = 45 sessions, prefrontal cortex = 38 sessions, and anterior cingulate = 45 sessions). As is
typically done, we treated these 55 sessions as if they were independent, although some patients completed more than one session. Different sessions for any given patient were conducted on different days, and it is typically assumed that different neurons were recorded during each session because the depth electrodes shift slightly as patients move around.

For the visual sessions, 360 words (120 one-syllable, 120 two-syllable, and 120 three-syllable words) were used, each of which was repeated once. The words were taken from the Medical Research Council Psycholinguistic database (33). Another set of 45 one-syllable words were used as fillers that were never repeated. There were three separate sets of words that could be presented, and these were used for patients who volunteered for multiple sessions. Each session consisted of a sequence of 255 trials (i.e., 240 trials where each of the 120 words were presented twice and 15 filler trials using words that never repeated). One patient completed more than four sessions, and thus saw one repeated set of words, but the repetition of the stimuli set did not affect memory performance. In each trial, a word was visually displayed on a computer screen for 1,500 ms, followed by a question mark as a prompt for response. Patients had up to 2,000 ms to indicate whether the word was novel or repeated. The trial ended when a response was made. There was a jittered intertrial interval (mean = 888 ms; SD = 552 ms). (On some trials of the visual task, the pre-stimulus time period included the time when patients made their response on the previous trial. However, excluding those trials from the analysis had a negligible effect on the results.) The procedure for the auditory sessions was similar to that for the visual sessions. Each auditory session consisted of a sequence of 615 trials. In these sessions, 300 different words were repeated once and 15 filler words were never repeated. The prompt for response appeared at the end of the audio file for the trial. There was a jittered intertrial interval that lasted for an average of 1,055 ms with a SD of 53 ms. Data are available at the Open Science Foundation repository at https://osf.io/9tgmx/.

References

1. L. R. Squire. Psychol Rev99, 195-231 (1992).
2. H. Eichenbaum, N. J. Cohen, From Conditioning to Conscious Recollection: Memory Systems of the Brain. (Oxford University Press, 2004).
3. E. Tulving, in Organization of memory, E. Tulving, W. Donaldson, Eds. (Academic Press, New York, 1972), pp. 381-403.
4. R. N. Shepard, M. Teghtsoonian. J Exp Psychol62, 302-309 (1961).
5. L. R. Squire, J. T. Wixted, R. E. Clark. Nat Rev Neurosci8, 872-883 (2007).
6. R. Adolphs, L. Cahill, R. Schul, R. Babinsky. Learn Mem4, 291-300 (1997).
7. G. Heit, M. E. Smith, E. Halgren. Nature333, 773-775 (1988).
8. G. Heit, M. E. Smith, E. Halgren. Brain113 (Pt 4), 1093-1112 (1990).
9. M. W. Brown, F. A. W. Wilson, I. P. Riches. Brain Research409, 158-162 (1987).
10. I. P. Riches, F. A. Wilson, M. W. Brown. Journal of Neuroscience11, 1763-1779 (1991).
11. J.-Z. Xiang, M. W. Brown. Neuropharmacology37, 757-767 (1998).
12. E. T. Rolls, P. M. Cahusac, J. D. Feigenbaum, Y. Miyashita. *Exp Brain Res* **93**, 299-306 (1993).
13. U. Rutishauser, A. N. Mamelak, E. M. Schuman. *Neuron* **49**, 805-813 (2006).
14. U. Rutishauser, E. M. Schuman, A. N. Mamelak. *Proc Natl Acad Sci U S A* **105**, 329-334 (2008).
15. U. Rutishauser *et al.* *Nat Neurosci* **18**, 1041-1050 (2015).
16. D. Marr. *Philos Trans R Soc Lond B Biol Sci* **262**, 23-81 (1971).
17. A. Treves, E. T. Rolls. *Hippocampus* **4**, 374-391 (1994).
18. K. A. Norman, R. C. O’Reilly. *Psychol Rev* **110**, 611-646 (2003).
19. J. L. McClelland, B. L. McNaughton, R. C. O’Reilly. *Psychol Rev* **102**, 419-457 (1995).
20. A. B. Valdez, E. N. Hickman, D. M. Treiman, K. A. Smith, P. N. Steinmetz. *J Neural Eng* **10**, 016001 (2013).
21. J. M. Chambers, W. S. Cleveland, B. Kleiner, P. Tukey, *Graphical methods of data analysis*. (Duxbury Press, Boston, MA, 1983).
22. U. Rutishauser, J. J. Slotine, R. J. Douglas. *Neural Comput* **30**, 1359-1393 (2018).
23. K. H. Hower, J. Wixted, M. E. Berryhill, I. R. Olson. *Neuropsychologia* **56**, 409-417 (2014).
24. U. Rutishauser, T. Aflalo, E. R. Rosario, N. Pouratian, R. A. Andersen. *Neuron* **97**, 209-220 e203 (2018).
25. R. Q. Quiroga. *Trends in cognitive Sciences* **24**, 994-1007 (2020).
26. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried. *Nature* **435**, 1102-1107 (2005).
27. M. J. Ison, R. Quian Quiroga, I. Fried. *Neuron* **87**, 220-230 (2015).
28. N. Suthana, A. D. Ekstrom, M. A. Yassa, C. Stark. *Trends Cogn Sci*, (2021).
29. J. K. Leutgeb, S. Leutgeb, M. B. Moser, E. I. Moser. *Science* **315**, 961-966 (2007).
30. M. A. Yassa, C. E. Stark. *Trends Neurosci* **34**, 515-525 (2011).
31. E. T. Rolls, T. Edmund, A. Treves, *Neural Network and Brain Function*. (Oxford University Press, 1998).
32. J. T. Wixted *et al.* *Proc Natl Acad Sci U S A* **115**, 1093-1098 (2018).
33. M. Coltheart. *Quarterly Journal of Experimental Psychology Section A* **33**, 497-505 (1981).

**Declarations**

**ACKNOWLEDGMENTS.** This is work is supported by Neurtex Brain Research Institute Grant 19-02, National Institute of Child Health and Human Development Grant HD075800-05, National Institute of Neurological and Communicative Disorders and Stroke Grant DC009781, Medical Research Service of the Department of Veterans Affairs (51K6CX001644), Award CX000359, and National Institute of Mental Health Grant 24600.

Competing interests: The authors declare no competing interests.

**Tables**
Due to technical limitations, the tables can be found as a download in the supplementary files section.

**Figures**

Number of neurons that exhibited generic memory signals (familiar vs. novel) in the hippocampus, amygdala, prefrontal cortex, and anterior cingulate. In each of the four regions, the total number of neurons exhibiting either a familiarity or novelty signal exceeded the number expected by chance with an alpha level of .05 (the dotted line) based on the total number of neurons per region.

**Figure 1**

![Bar graph showing the number of neurons in different brain regions](image)
Fig. 2

QQ plots of the normalized spike counts in response to novel words (x-axis) and repeated words (y-axis) for the hippocampus (A and B), amygdala (C and D), prefrontal cortex (E and F), and anterior cingulate (G and H). The top panels plot 100% of the data for each region (hippocampus: 77,431 recordings; amygdala: 65,219 recordings; prefrontal: 47,399 recordings; anterior cingulate: 62,205 recordings). The bottom panels show the same data after removing the 0.25% recordings with the highest spike counts from both the novel-word and repeated-word distributions. For the hippocampus, the points fell mostly along the diagonal line but exhibited a sharp upward deflection (A), indicating that some neurons responded strongly to some repeated words. After removing the top 0.25% of the data, the upward deflection for the hippocampus disappeared (B), indicating that only a very small percentage of neurons spiked more in response to repeated words than novel words. By contrast, the plots for the other three regions did not exhibit similar upward deflections. Thus, the activity pattern signaling episodic memory was identified only in the hippocampus.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialsNC.docx
- Tables.docx