Human hippocampal ripples signal encoding of episodic memories

John J. Sakon[1], David J. Halpern[1], Daniel R. Schonhaut[2], Michael J. Kahana,[1*]
[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104, USA
[2]Department of Neuroscience, Perelman School of Medicine at the University of Pennsylvania
Philadelphia, PA, 19104, USA
*To whom correspondence should be addressed: kahana@psych.upenn.edu

October 3, 2022

Keywords: ripples, episodic memory, hippocampus, contextual reinstatement, medial temporal lobe, encoding, subsequent clustering effect

Abstract

Recent human electrophysiology work has uncovered the presence of high frequency oscillatory events, termed ripples, during awake behavior. This work focuses on ripples in the medial temporal lobe (MTL) during memory retrieval. Few studies, however, investigate ripples during item encoding. Many studies have found neural activity during encoding that predicts later recall, termed subsequent memory effects (SMEs), but it is unclear if encoding ripples also predict subsequent recall. Detecting ripples in 124 neurosurgical participants performing an episodic memory task, we do not find ripple SMEs in any MTL region, even as these regions exhibit robust high frequency activity (HFA) SMEs. Instead, hippocampal ripples increase during encoding of items leading to recall of temporally or

semantically associated items, a phenomenon known as clustering. This subsequent clustering effect (SCE) arises specifically when hippocampal ripples occur during both encoding and retrieval, suggesting that ripples mediate the encoding and future reinstatement of episodic memories.

Introduction

Decades of work in animal models have identified discrete, high frequency events in MTL, termed ripples (*Buzsáki, 2015*). This animal work has suggested a specific role for hippocampal ripples in memory formation during learning and offline replay (*Buzsáki, 2015*) and more recently memory retrieval (*Joo and Frank, 2018*). A series of recent studies have investigated ripples in human intracranial recordings (see Liu *et al.* (2022) for a review). Many of these investigations relate medial temporal lobe (MTL) ripples and memory retrieval, with ripple rate increasing just before participants vocalize recalls (*Sakon and Kahana, 2022, Norman et al., 2019, Norman et al., 2021, Henin et al., 2021, Chen et al., 2021, Vaz et al., 2019, Vaz et al., 2020, Dickey et al., 2022*). The few studies, however, that have measured ripples during memory encoding find conflicting evidence regarding their relation to subsequent recall (*Henin et al., 2021, Norman et al., 2019*).

A separate but related body of research has shown that $\geq$60 Hz spectral power (often termed high frequency activity [HFA] or fast-gamma oscillations) predicts subsequent recall (*Fell et al., 2001, Osipova et al., 2006, Paller and Wagner, 2002, Lachaux et al., 2012, Herweg et al., 2020*). Indeed, numerous intracranial studies using HFA detectors distinguish subsequently recalled and not-recalled items, termed a subsequent memory effect (SME) (*Burke et al., 2015, Griffiths et al., 2019, Henin et al., 2021*). The overlapping frequency ranges used to detect HFA and ripples raise questions about whether and how these signals may be related to one another (*Buzsáki and da Silva, 2012*).

Recent human intracranial studies find hippocampal ripples preferentially occur during recall of episodic memories (*Norman et al., 2021, Chen et al., 2021*). In particular, Sakon & Kahana (2022) demonstrate that hippocampal ripples signal reinstatement of context during memory retrieval, a mechanism considered crucial to the "jump back in time" phenomenology of episodic memory (*Howard and Kahana, 2002a, Polyn et al., 2009*). Meanwhile, theories of ripple function suggest memory formation and memory retrieval share mechanisms, as neural activity reinstated during memory retrieval overlaps with activity repeatedly reinstated during consolidation (*Joo and Frank, 2018, Vaz et al., 2020*). Considering these two ideas, do ripples also reflect context reinstatement during the formation of episodic memories? For example, if you attend a Philadelphia Phillies game and enjoy a cheesesteak, future cheesesteak orders may retrieve the context of the event, which will lead to reinstatement of memories clustered with the game (*Healey et al., 2019*). We hypothesize that during the formation of the memory, hippocampal ripples signal engagement of episodic memory mechanisms, which strengthen the association between the item (cheesesteak) and context (Phillies game). This association subsequently increases the likelihood of reinstating the context of the game when later cued by cheesesteak. Previous work has shown evidence of this phenomenon, termed a subsequent clustering effect (SCE), using HFA in hippocampus (*Long et al., 2014*), hinting that ripples may underlie subsequent clustering.

Analyzing intracranial EEG recordings of 124 participants performing a delayed free recall task, we ask if ripples show an SME or SCE in the hippocampus and surrounding cortical regions in MTL. We partitioned our data into two parts: an initial $\sim$35% of participants for developing initial hypotheses and analyses, and a second part held out so that we could confirm our findings with the whole dataset. We pre-registered the initial hypotheses and figures supporting them on the Open Science Framework (https://osf.io/e98qp). Therefore, we defined the analysis parameters for our figures in the pre-registration based on the first part of the dataset,

and here we present figures and statistics on the full dataset (*Methods*).

The analyses build the case that hippocampal ripples specifically signal the formation of episodic memories. First, we do not find a significant ripple SME throughout any MTL regions despite replicating an SME for HFA. However, when partitioning words into those that lead to subsequent clustering of recalls vs. those that do not, we find a significant ripple subsequent clustering effect (SCE) specifically in the hippocampus. Evidencing its role in task performance, participants with a stronger hippocampal ripple SCE exhibit increased clustering and superior memory. Finally, we show that hippocampal ripples during memory formation lead to subsequent clustering precisely when ripples also occur prior to word recall, implying the SCE incites ripple-mediated reinstatement.

Results

Hippocampal ripples do not exhibit a subsequent memory effect (SME).

To clarify the relation between ripples and memory encoding we align hippocampal recordings (**Fig. 1e**) to the onset of word presentation during the study phase of a categorized, delayed free recall task (**Fig. 1b**), in which participants view a list of words and subsequently recall as many as possible after a distractor period. We use an algorithm recently shown to isolate ripples in human hippocampus and surrounding MTL during both memory encoding and retrieval (*Norman et al., 2019, Sakon and Kahana, 2022*) (*Methods*). A raster of ripples from five sample participants illustrates an encoding-related rise in ripples occurring $\sim$0.5 seconds after word onset (each row in **Fig. 2a** represents a word presentation recorded on a single channel, and each dot represents the start time of a single ripple). These findings accord with prior work where hippocampal ripple onset occurs hundreds of ms after presentation of face or place stimuli (*Norman et al., 2019, Chen et al., 2021, Henin et al., 2021*).

First, we ask if ripples show an SME. For both hippocampal subfields CA1 (163 sessions

from 86 participants) and CA3/dentate gyrus (CA3/DG: 117 from 59) we average across participants to create peri-stimulus time histograms (PSTHs) for both subsequently recalled and not recalled words. We find only a modest difference in ripple rates between these groups beginning $\sim$0.5 seconds after word presentation in both regions (**Fig. 2b**). Although the dataset is adequately powered to find a ripple SME in each region (power$>$0.97 using effect sizes for SMEs reported using HFA (*Long et al., 2014*) and ripple (*Henin et al., 2021*) detectors, Methods), we find no significant difference in the ripple rates during encoding of subsequently recalled and not recalled words for CA1 and CA3/DG (**Fig. 2b; Eq. 2**). Meanwhile, amygdala (AMY; 104 sessions from 50 participants) and entorhinal/parahippocampal cortex (ENTPHC, 96 from 52) show lower ripple rates than hippocampus overall and also fail to show a ripple SME (**Fig. 2b**).

Considering that previous studies find strong HFA SMEs in the hippocampus and neighboring MTL subregions (*Burke et al., 2015*) we apply a high frequency activity (HFA) detector on the same trials as in the ripples analysis reported above. Measuring HFA in a frequency range almost completely overlapping that of our ripple detector, we find a clear HFA SME in all MTL subregions (**Fig. 2c**). Using the same linear mixed effects model as with ripples, the HFA SME is significant for CA1, CA3/DG, AMY, and ENTPHC (**Fig. 2c; Eq. 1**). In sum, HFA exhibits an SME across the MTL, while the ripple detector fails to detect an SME in any MTL region.

Hippocampal ripples exhibit a subsequent clustering effect (SCE).

The SME contrast fails to take advantage of the rich behavioral structure of the categorized free recall task (**Fig. 1b-c**). Specifically, the order in which people free recall recently studied items reveals information about the organization of memory. When participants strongly bind items to their encoding context, which includes both temporal and semantic information (*Howard and Kahana, 2002b, Polyn et al., 2009*), they tend to retrieve clusters of temporally and semantically similar items (*Weidemann et al., 2019, Solomon et al., 2019, Kragel et al., 2021*). Our previous work showed an increase in hippocampal ripples just prior to par-

ticipants recalling a cluster of related words, suggesting that ripples signal the reinstatement of context (*Sakon and Kahana, 2022*). Here, we hypothesize that ripples might also signal contextual reinstatement during encoding. If this is true, an increase in ripples during initial presentation of a word would predict that word will subsequently lead to clustering during retrieval (**Fig. 3a**). We refer to this phenomenon as a subsequent clustering effect (SCE) (*Long and Kahana, 2015*).

In categorized free recall, transitions between clustered recalls neatly divide into a handful of groups (**Fig 1c**). Referring to the example words in **Fig 1c**: **adjacent semantic** indicates two words from the same categorical pair recalled consecutively, e.g. dolphin and octopus (20% of recalls); **remote semantic** indicates two words from the same category but not the same pair recalled consecutively, e.g. dolphin and fish (20% of recalls); **remote unclustered** indicates two words from different categories that are not presented back-to-back recalled consecutively, e.g. dolphin and pliers (17% of recalls); and **dead end** indicates the last recall from each list, which therefore does not transition to another recall (26% of recalls). Adjacent unclustered, in which participants recall words that appear back-to-back from different categories, is rare (3%) so we do not analyze this type further, while the remaining recalls are incorrect or repeats (14%). We then measure ripples during the presentation of the first word in each subsequent transition (except for dead ends, where there is no transition) and average within each group to find the ripple rate.

Measuring the average ripple rates between these behaviorally-defined groups reveals clear evidence that hippocampal ripples exhibit an SCE. In particular, testing the ripple rates of words that lead to subsequent clustering (adjacent semantic and remote semantic) vs. those that are subsequently recalled but do not lead to clustering (remote unclustered and dead ends) yields a significant difference in both CA1 and CA3/DG (**Fig 3b, Eq. 2**). When grouping non-hippocampal MTL (non-HPC MTL) regions (including AMY, ENT, PHC, and perirhinal

cortex) we do not find a significant difference. When making comparisons between the individual categories in the clustering group (i.e. adjacent semantic and remote semantic), each of these also show significantly more ripples during their presentation compared to unclustered recalls for both CA1 and CA3/DG ($p<0.02$, FDR-corrected, **Eq. 2**) but not non-HPC MTL ($p>0.19$, FDR-corrected, **Eq. 2**). Using a linear mixed effects model to directly compare each hippocampal subfield with the non-HPC MTL group to account for differences both within and between participants, we find a significant interaction between region and clustering that predicts higher ripple rates in each hippocampal region compared to non-HPC MTL **Fig. 3b; Eq. 3**.

The previous contrasts isolate clustering as we compare words that subsequently lead to clustering vs. those words that are still recalled but do not lead to the subsequent semantic or temporal transitions that are hallmarks of context reinstatement. In a similar manner, we can isolate the SME by contrasting those words that are recalled and do not lead to clustering vs. words that are not recalled. Using this contrast, we find no evidence of an SME in CA1, CA3/DG, or non-HPC MTL (**Fig. 3c, Eq. 2**). Further, when making comparisons between the individual categories in the remembered but not clustered group (i.e. remote semantic and dead ends) vs. words not recalled, each of these also show no difference in ripples rates for CA1, CA3/DG, and non-HPC MTL ($p>0.33$, FDR-corrected, **Eq. 2**).

Finally, we can directly compare the SCE and SME in the same model, thereby assessing if the significance of the ripple rise for subsequently clustered words exists after taking into account ripple rates during the encoding of words that are recalled but not clustered. For both CA1 and CA3/DG, we find a significant factor for the SCE ($p<0.004$, **Eq. 4**, FDR-corrected across two tests), but not for the SME ($p>0.08$, **Eq. 4**, FDR-corrected across two tests), suggesting the presence of hippocampal ripples during encoding more strongly predicts subsequent clustering than subsequent memory.

The hippocampal ripple SCE is associated with better memory and increased clustering.

Next, we ask if the ripple SCE correlates with participant behavior. First, measuring the SCE for each individual participant as the difference in ripples during words that lead to subsequent clustering vs. words that are remembered but do not lead to clustering, we compare this change to the average number of recalls for that person per list. Participants that recall more words display a significantly larger ripple SCE in CA1 and CA3/DG but not in non-HPC MTL (**Fig. 4a, Eq. 5**). When directly comparing proportion of words recalled vs. SCE between hippocampus and non-HPC MTL in a single model, both CA1 and CA3/DG show a significant interaction between region and recall that predicts SCE (**Fig. 4a, Eq. 6**). Therefore, the hippocampal ripple SCE in particular predicts superior memory across participants.

We next ask whether participants who show a larger ripple SCE also exhibit greater clustering. Indeed, we find a significant correlation between SCE and proportion of recalls that lead to subsequent clustering for CA1 (**Fig. 4b, Eq. 7**). CA3/DG and non-HPC MTL do not show a significant difference. When comparing hippocampus directly with non-HPC MTL in a single model, CA1 (but not CA3/DG) shows a significantly greater interaction between region and clustering that predicts SCE (**Fig. 4b, Eq. 8**), indicating the hippocampal SCE also predicts participant clustering.

Hippocampal ripples during both the first and second pair of words from a category lead to improved recall of the second pair.

On each list in the free recall task two pairs of words from the same semantic category appear: one in the first half of the list and another in the second half (with the constraint that pairs from the same category are never shown back-to-back) (**Fig. 1b**). This task structure allows us to investigate if ripples occurring as participants encode the first pair of words from a given category (X_{1-2}) influence memory for the second pair (X_{3-4}) despite the intervening word presentations. Considering the SCE results (**Fig. 3**), in which increased ripples during

word presentation predict the word will subsequently lead to context reinstatement (and therefore clustering) during the retrieval period, we hypothesize that ripples during X_{1-2} might also lead to context reinstatement during the presentation of words X_{3-4}. And if such context reinstatement manifests during the presentation of words X_{3-4}, we anticipate likelier subsequent recall of these words (**Fig. 5a**).

To test this hypothesis, we measure the accuracy of X_{3-4} words from each category on each list after assigning the category to one of four pools: 1) those where $\geq$1 ripple occurs during the presentation of X_{1-2} (but not X_{3-4}), 2) those where $\geq$1 ripple occurs during X_{3-4} (but not X_{1-2}), 3) those where $\geq$1 ripple occurs during both X_{1-2} and X_{3-4}, and 4) those where no ripple occurs during either. Averaging within each pool, we find words with a ripple during both X_{1-2} and X_{3-4} exhibit the highest recall accuracy, followed by lists with ripples only during X_{1-2} (**Fig. 5b**).

To evaluate differences in the accuracy of X_{3-4} recall among the pools, we create a linear mixed model that takes into account $\geq$1 ripple during presentation of X_{1-2}, $\geq$1 ripple during presentation of X_{3-4}, the interaction of a ripple occurring for both pairs, and also the ripple rate for the remaining (eight) words on the list to remove possible list-level ripple rate effects (**Eq. 9**). This model reveals that CA1 ripples during X_{1-2} predict X_{3-4} recall, but *only* if a ripple also occurs during X_{3-4} (**Fig. 5b**). Thus, if a ripple occurs during both pairs of words from a category, the likelihood of recalling the 2nd pair (X_{3-4}) increases. However, if a CA1 ripple occurs only during X_{1-2}, we find no significant difference in recall accuracy of X_{3-4}. CA3/DG does not show a significant difference for either comparison, even though the effect is in the same direction for higher X_{3-4} recall when a ripple occurs during both pairs (**Fig. 5b**). If the increase in X_{3-4} recalls comes from X_{1-2} ripples leading to context reinstatement and therefore ripples during X_{3-4}, as opposed to an additive effect where increased ripples during same category words leads to more recalls from that category, we anticipate that ripples

during both pairs of words will not improve recall of X_{1-2} recalls. Indeed, when ripples occur during both X_{1-2} and X_{3-4}, recall of X_{1-2} words does not increase when measuring either CA1 or CA3/DG ripples (p>0.37, **Eq. 9**). These findings suggest ripples during early list words promote category reinstatement later in the list, as reflected by ripples occurring for words from the same category later in the list.

Hippocampal ripples during encoding and retrieval of the same word predict clustering. The previous analysis suggests that ripples can reflect context reinstatement during encoding, where ripples during early list words promote ripples during late list words when the words carry strong semantic relations. Our previous work finds ripples reflect context reinstatement during retrieval, as ripples occur just prior to vocalization of clustered recalls (the pre-retrieval effect (PRE), see Discussion). Here we ask whether clustering emerges specifically when ripples occur during both encoding and retrieval of the same words (**Fig. 6a**).

To answer this question, for every recalled word we determine if $\geq$1 ripple occurs during its presentation and/or during the PRE window. Assigning each recall to one of four conditions—encoding $\pm$ ripple crossed with retrieval $\pm$ ripple—we assess the proportion of recalls within each condition that lead to clustering. As predicted, recalls with ripples during both encoding and retrieval exhibit the highest clustering rates **Fig. 6b**. Using a linear mixed effects model to assess if ripples during encoding, retrieval, or both lead to clustering, only when CA1 ripples occur in both conditions do we find a significant increase in clustering **Fig. 6b, left**. Ripples measured in CA3/DG, however, do not significantly predict clustering regardless of their presence during encoding, retrieval, or both periods **Fig. 6b, right**.

Discussion

Measuring medial temporal lobe (MTL) ripples as participants encode and then free recall lists of words, we find that clustering of recalls significantly increases during memory

retrieval specifically when hippocampal ripples occur during word presentation. This ripple subsequent clustering effect (SCE) appears more prominently than a ripple subsequent memory effect (SME), specifying a role for ripples in binding items to their semantic and/or temporal associates when forming memories. The magnitude of the hippocampal ripple SCE also aligns with task behavior, as participants with a larger rise in SCE exhibit better clustering of recalls and superior memory. Finally, two analyses provide evidence that ripples signal context reinstatement. First, ripples during words shown early in the list lead to ripples during presentation of semantically-related words many seconds later in the list and, combined, predict increased recall of these later words. Second, when ripples occur during encoding of a word, that word leads to clustering significantly more often when a ripple also occurs prior to its retrieval. These findings, in which hippocampal ripples during memory formation predict subsequent ripple-mediated reinstatement during both later list items and retrieval, suggest ripples specifically signal encoding and reinstatement of episodic memories.

During free recall, hippocampal ripples occur just prior to the retrieval of a previously studied item, termed the pre-retrieval effect (PRE) (*Sakon and Kahana, 2022*). The strongest PRE occurs prior to pairs of recalls bearing strong temporal and/or semantic relations, suggesting that hippocampal ripples reflect an item-to-context reinstatement process (*Kahana, 2020*). A recent review hypothesizes that sharp-wave ripples perform a dual function by mediating both memory formation and retrieval, as repetition in support of consolidation (*Vaz et al., 2020*) may share mechanisms with reinstatement during retrieval (*Joo and Frank, 2018*). In light of this hypothesis and the ripple SCE results (**Fig. 3**), we asked if the SCE relates to the PRE. Our final analysis substantiates the hypothesis: recalls with ripples during both the initial word presentation and in the PRE window lead to clustering significantly more than recalls without ripples in both periods (**Fig. 6**). In other words, both the SCE and the PRE appear to reflect a related process (*Kahana, 2020*), where items bind to context during encoding and subsequently

reinstate context from items during retrieval. Further, considering that participants have prior knowledge of the semantics of the common nouns used in this study and that 40% of recalls lead to clustering (**Fig. 1c**), the SCE also may reflect reinstatement of categorical context during word presentation (e.g. sea animals in **Fig. 1b-c**). That participants with larger ripple SCEs show more subsequent clustering of recalls (**Fig. 4b**) supports this interpretation. **Figure 5** also supports categorical reinstatement during encoding, as ripples during a semantic category early in a list X_{1-2} predict better recall of words from that same category shown later in the list (X_{3-4}). This effect only happens when ripples also occur during X_{3-4}, and not when ripples occur solely during X_{1-2}, supporting the idea that ripples mediate reinstatement during the X_{3-4} words and this process leads to their increased recall.

A more conservative interpretation of our data is that the SCE simply reflects engagement of the hippocampal memory system. Memory tasks with larger demands are more likely to necessitate hippocampal involvement. For example, studies of hippocampal amnesics on delayed memory tasks found that deficits only occur if task demand is sufficient (e.g. relatively large set size or retention delays (*Jeneson et al., 2011*)). And when MTL amnesics performed a delayed free recall task similar to ours they specifically showed deficits in reinstating context compared to healthy controls, but no difference recalling the most recently-shown items, suggesting deficits occur due to defects specific to the episodic system (*Palombo et al., 2019*). Similarly, single unit recordings support the idea that recruitment of the hippocampus only occurs with sufficient task demands, as hippocampal neurons fail to fire above baseline levels until memory demands are relatively large (*Kamiński et al., 2017, Boran et al., 2019*). Therefore, when participants engage their hippocampal memory system, whether through increased attention or by forming associations between words from semantic categories, the ripple SCE may reflect the increase in hippocampal activity. Indeed, the SCE is larger for participants with higher recall rates (**Fig. 4a**), which is compatible with the idea that participants successfully

recruiting their episodic system show improved memory. And in the case of **Fig. 6**, where the same words reflect ripple-mediated episodic encoding and retrieval, in both cases we expect the hippocampus to be engaged as participants encoding semantically-associated items during encoding and subsequently retrieve semantically-associated items during retrieval.

We replicate previous work (*Sederberg et al., 2007, Burke et al., 2014a, Long et al., 2014, Henin et al., 2021*) showing high frequency activity (HFA) SMEs throughout the MTL, as each region we test has significantly stronger signal for subsequently recalled vs. not recalled words. The ubiquity of the HFA SME throughout the MTL is possibly of physiological relevance, as high gamma, which largely overlaps with HFA, is thought to synchronize regions during cognitive tasks (*Jensen et al., 2007*). Surprisingly, and contrary to a hypothesis from our pre-registration (https://osf.io/e98qp), we do not find a significant ripple SME in either of the hippocampal subfields we test **Fig. 2b**. And while subsequently recalled words show a peak $\sim$0.6 s above not recalled words when looking at the PSTHs for CA1 or CA3/DG, even when we use a narrower 400 to 1100 ms window, we still do not find a significant ripple SME in either (P>0.22, each FDR-corrected across 4 tests of Eq. 2). These results suggest the algorithms designed to detect ripples in rodents (*Stark et al., 2014*) achieve a level of specificity that separates ripples from more ubiquitous high frequency signals. What differences in the detector for ripples vs. HFA account for this specificity? Two components are likely to be responsible. First, the ripple detector only considers events with power exceeding a high threshold (3 SD). Second, the detector requires these candidate events to stay above a lower threshold (2 SD) for a minimum duration (20 ms) to be considered a ripple. Therefore, we speculate that high frequency activity that does not reach sufficiently high powers or arises only transiently accounts for the HFA SME. Future work splitting individual events into ripple vs. HFA groups will be necessary to test these hypotheses.

The present report argues that hippocampal ripples signal the encoding of episodic memo-

ries, as the presence of ripples during item encoding predicts the subsequent, ripple-mediated reinstatement of context during retrieval. Considering the specificity in which hippocampal ripples signal this subsequent clustering effect **Fig. 3b**, as opposed to the more ubiquitous HFA subsequent memory effect found throughout MTL **Fig. 2c** and other regions (*Burke et al., 2014a*), future work might take advantage of ripples as a biomarker of episodic memory formation. In particular, considering that classification of brain states that predict memory encoding can be used to time stimulation for the purpose of improving memory (*Ezzyat et al., 2018*), future work might incorporate ripple detection to specifically target episodic memory formation for use in translational work.

Acknowledgements

Data were collected as part of the DARPA RAM program (Cooperative Agreement N66001-14-2-4032). This work is supported by NIH Grant R01NS106611-02 and US Army Medical Research and Development Command Medical Technology Enterprise Consortium Grant MTEC-20-06-MOM-013. The views, opinions, and/or findings contained in this material are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government. We thank the Kahana lab and the Joshua Jacobs lab for providing valuable feedback on this work.

References

Avants et al., 2011. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–2044.

Boran et al., 2019. Boran, E., Fedele, T., Klaver, P., Hilfiker, P., Stieglitz, L., Grunwald, T., and Sarnthein, J. (2019). Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. *Science Advances*, 5(3).

Burke et al., 2014a. Burke, J. F., Long, N. M., Zaghloul, K. A., Sharan, A. D., Sperling, M. R., and Kahana, M. J. (2014a). Human intracranial high-frequency activity maps episodic memory formation in space and time. *NeuroImage*, 85:834–843.

Burke et al., 2015. Burke, J. F., Ramayya, A. G., and Kahana, M. J. (2015). Human intracranial high-frequency activity during memory processing: Neural oscillations or stochastic volatility? *Current Opinion in Neurobiology*, 31:104–110.

Burke et al., 2014b. Burke, J. F., Sharan, A. D., Sperling, M. R., Ramayya, A. G., Evans, J. J., Healey, M. K., Beck, E. N., Davis, K. A., Lucas, T. H., and Kahana, M. J. (2014b). Theta and high–frequency activity mark spontaneous recall of episodic memories. *Journal of Neuroscience*, 34(34):11355–11365.

Buzsáki, 2015. Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, 25(10):1073–1188.

Buzsáki and da Silva, 2012. Buzsáki, G. and da Silva, F. L. (2012). High frequency oscillations in the intact brain. *Progress in Neurobiology*, 98(3):241–249.

Chen et al., 2021. Chen, Y. Y., Aponik-Gremillion, L., Bartoli, E., Yoshor, D., Sheth, S. A., and Foster, B. L. (2021). Stability of ripple events during task engagement in human hippocampus. *Cell Reports*, 35(13):109304.

Desikan et al., 2006. Desikan, R., Segonne, B., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R. L., Dale, A., Maguire, A., Hyman, B., Albert, M., and Killiany, N. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–80.

Dickey et al., 2022. Dickey, C. W., Verzhbinsky, I. A., Jiang, X., Rosen, B. Q., Kajfez, S., Stedelin, B., Shih, J. J., Ben-Haim, S., Raslan, A. M., Eskandar, E. N., Gonzalez-Martinez, J., Cash, S. S., and Halgren, E. (2022). Widespread ripples synchronize human cortical activity during sleep, waking, and memory recall. *Proceedings of the National Academy of Sciences*, 119(28).

Dykstra et al., 2012. Dykstra, A. R., Chan, A. M., Quinn, B. T., Zepeda, R., Keller, C. J., Cormier, J., Madsen, J. R., Eskandar, E. N., and Cash, S. S. (2012). Individualized localization and cortical surface-based registration of intracranial electrodes. *NeuroImage*, 59(4):3563–3570.

Ezzyat et al., 2018. Ezzyat, Y., Wanda, P., Levy, D. F., Kadel, A., Aka, A., Pedisich, I., Sperling, M. R., Sharan, A. D., Lega, B. C., Burks, A., Gross, R., Inman, C. S., Jobst, B. C., Gorenstein, M., Davis, K. A., A., W. G., Kucewicz, M. T., Stein, J. M., Gorniak, R. J., Das, S. R., Rizzuto, D. S., and Kahana, M. J. (2018). Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nature Communications*, 9(1):365.

Fell et al., 2001. Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., and Fernandez, G. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience*, 4(12):1259–1264.

Fischl et al., 2004. Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., et al. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22.

Gelinas et al., 2016. Gelinas, J. N., Khodagholy, D., Thesen, T., Devinsky, O., and Buzsáki, G. (2016). Interictal epileptiform discharges induce hippocampal–cortical coupling in temporal lobe epilepsy. *Nature medicine*, 22(6):641.

Griffiths et al., 2019. Griffiths, B. J., Parish, G., Roux, F., Michelmann, S., van der Plas, M., Kolibius, L. D., Chelvarajah, R., Rollings, D. T., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., Staresina, B., Wimber, M., and Hanslmayr, S. (2019). Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *PNAS*, 116(43):21834–21842.

Healey et al., 2019. Healey, M. K., Long, N. M., and Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 26(3):699–720.

Henin et al., 2021. Henin, S., Shankar, A., Borges, H., Flinker, A., Doyle, W., Friedman, D., Devinsky, O., Buzsáki, G., and Liu, A. (2021). Spatiotemporal dynamics between interictal epileptiform discharges and ripples during associative memory processing. *Brain*, 5.

Herweg et al., 2020. Herweg, N. A., Solomon, E. A., and Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Science*, 24(3):208–227.

Howard and Kahana, 2002a. Howard, M. W. and Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299.

Howard and Kahana, 2002b. Howard, M. W. and Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, 46:85–98.

Jeneson et al., 2011. Jeneson, A., Mauldin, K. N., Hopkins, R. O., and Squire, L. R. (2011). The role of the hippocampus in retaining relational information across short delays: The importance of memory load. *Learning & Memory*, 18(5):301–305.

Jensen et al., 2007. Jensen, O., Kaiser, J., and Lachaux, J. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, 30(7):317–324.

Joo and Frank, 2018. Joo, H. R. and Frank, L. M. (2018). The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. *Nature Reviews Neuroscience*, 19(12):744–757.

Kahana, 2020. Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, 71(1):107–138.

Kamiński et al., 2017. Kamiński, J., Sullivan, S., Chung, J. M., Ross, I. B., Mamelak, A. N., and Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat Neurosci*, 20(4):590–601.

Klein and Tourville, 2012. Klein, A. and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6.

Kragel et al., 2021. Kragel, J. E., Ezzyat, Y., Worrell, G. A., Sperling, M. R., Gross, R. E., Lega, B. C., Jobst, B. C., Sheth, S. A., Zaghloul, K. A., Stein, J. M., and Kahana, M. J. (2021). Distinct cortical systems reinstate content and context information during memory search. *Nature Communications*, 12.

Lachaux et al., 2012. Lachaux, J. P., Axmacher, N., Mormann, F., Halgren, E., and Crone, N. E. (2012). High-frequency neural activity and human cognition: Past, present, and possible future of intracranial EEG research. *Progress in Neurobiology*, 98(3):279–301.

Long et al., 2014. Long, N. M., Burke, J. F., and Kahana, M. J. (2014). Subsequent memory effect in intracranial and scalp EEG. *NeuroImage*, 84:488–494.

Long and Kahana, 2015. Long, N. M. and Kahana, M. J. (2015). Successful memory formation is driven by contextual encoding in the core memory network. *NeuroImage*, 119:332–337.

Norman et al., 2021. Norman, Y., Raccah, O., Liu, S., Parvizi, J., and Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron*, 109(17):2767–2780.e5.

Norman et al., 2019. Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., and Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454):eaax1030.

Osipova et al., 2006. Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., and Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience*, 26(28):7523–7531.

Paller and Wagner, 2002. Paller, K. A. and Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2):93–102.

Palombo et al., 2019. Palombo, D. J., Lascio, J. M. D., Howard, M. W., and Verfaellie, M. (2019). Medial temporal lobe amnesia is associated with a deficit in recovering temporal context. *Journal of Cognitive Neuroscience*, 31(2):236–248.

Polyn et al., 2009. Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1):129–156.

Rey et al., 2014. Rey, H. G., Fried, I., and Quiroga, R. Q. (2014). Timing of single-neuron and local field potential responses in the human medial temporal lobe. *Current Biology*, 24(3):299–304.

Sakon and Kahana, 2022. Sakon, J. J. and Kahana, M. J. (2022). Hippocampal ripples signal contextually mediated episodic recall. *Proceedings of the National Academy of Sciences*, 119(40):e2201657119.

Sederberg et al., 2007. Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., Tully, M. S., and Kahana, M. J. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, 17(5):1190–1196.

Solomon et al., 2019. Solomon, E. A., Lega, B. C., Sperling, M. R., and Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences*, 116(48):24343–24352.

Stark et al., 2014. Stark, E., Roux, L., Eichler, R., Senzai, Y., Royer, S., and Buzsáki, G. (2014). Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron*, 83(2):467–480.

Sullivan et al., 2011. Sullivan, D., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., and Buzsaki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: Influence of dentate and entorhinal cortical activity. *The Journal of Neuroscience*, 31(23):8605–8616.

Vaz et al., 2019. Vaz, A. P., Inati, S. K., Brunel, N., and Zaghloul, K. A. (2019). Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science*, 363(6430):975–978.

Vaz et al., 2020. Vaz, A. P., Wittig, J. H., Inati, S. K., and Zaghloul, K. A. (2020). Replay of cortical spiking sequences during human memory retrieval. *Science*, 367(6482):1131–1134.

Weidemann et al., 2019. Weidemann, C. T., Kragel, J. E., Lega, B. C., Worrell, G. A., Sperling, M. R., Sharan, A. D., Jobst, B. C., Khadjevand, F., Davis, K. A., Wanda, P. A., Kadel, A., Rizzuto, D. S., and Kahana, M. J. (2019). Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *Journal of Experimental Psychology: General*, 148(1):1–12.

Yushkevich et al., 2015. Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L., Kliot, D., Das, S. R., and Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36(1):258–287.

Figures

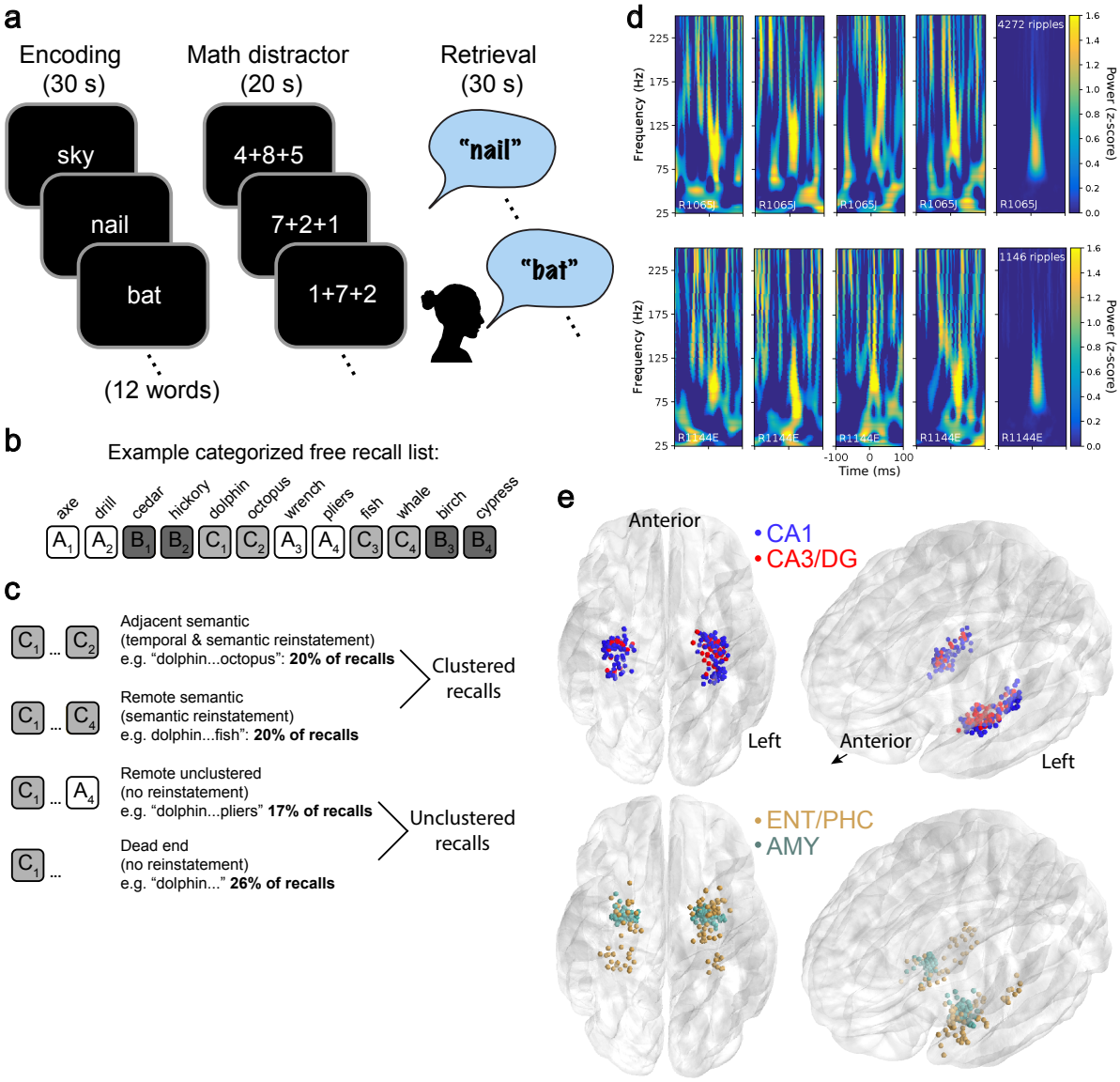

Figure 1. Free recall task design and ripple detection details. (a) Task diagram of delayed free recall, in which participants perform a math distractor in between word presentations and a retrieval period. **(b)** Structure of categorized word lists used in this task variant. A, B, and C are each semantic categories (tools, trees, and sea animals in this case). The two pairs of words from the same category are never shown back-to-back **(c)** Types of recall transitions in the categorized free recall task and percentage of recalls that lead to each. Note that adjacent, non-semantic transitions are only 3% of recalls due to the semantic nature of the task so are not analyzed. **(d)** Each row displays EEG spectrograms aligned to the start of ripples occurring during word presentation for two participants with hippocampal CA1

electrodes. The first four columns show single trial examples while the fifth column shows the average across all ripples during word presentation for all CA1 electrodes in all sessions for each participant. **(e)** Electrode bipolar pair midpoint localizations for all participants performing catFR. Shown are hippocampal subfields CA1 and CA3/dentate gyrus (CA3/DG), entorhinal (ENT) and parahippocampal (PHC) cortex, and amygdala (AMY).

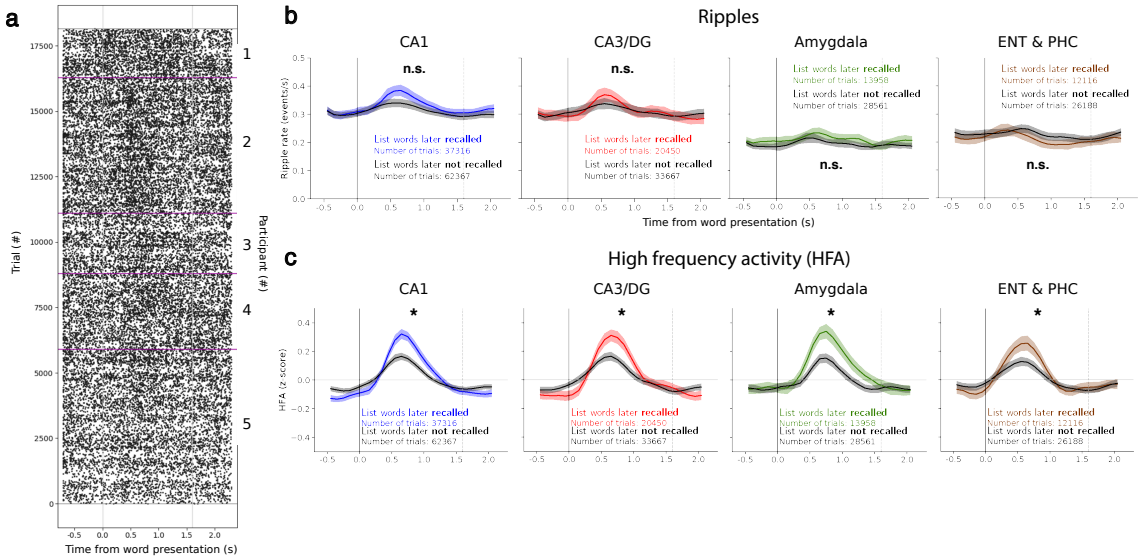

Figure 2. High frequency activity (but not ripples) shows a subsequent memory effect (SME) in the medial temporal lobe (MTL). (a) Raster plot for 5 example participants with EEG from hippocampal electrode pairs aligned to time of word presentation. Same participants as the first 5 shown in Sakon & Kahana 2021, Fig. 4b. Each dot represents the start time of a single detected ripples. Vertical gray lines denote the 1.6 s onscreen period for each word, and purple horizontal lines divide participants. We define a trial as a recording from a single bipolar pair during the presentation of a single word. **(b)** Ripple peri-stimulus time histograms (PSTH) averaged across all participants with bipolar electrode pairs localized to hippocampal subfields CA1 or CA3/DG, AMY, or ENT/PHC. We label the number of trials for words that are later recalled or not recalled during the subsequent retrieval period within each subpanel. Error bands are SE from a separate mixed model calculated at each time bin (Eq. 1). Significance of mixed model assessing ripple rates between words subsequently recalled vs. not recalled (Eq. 2): CA1, β = -0.0020±0.0071, P = 0.78; CA3/DG, β = -0.0052±0.010, P = 0.78; AMY, β = 0.0030±0.0066, P = 0.78; ENT/PHC, β = -0.017±0.0084, P = 0.19 (each FDR-corrected across 4 tests of Eq. 2). **(c)** PSTH for high frequency activity (HFA) using the frequency range 64-178 Hz. HFA is z-scored for each session by averaging across trials and time bins and normalizing with the standard deviation across trials. Error bands are SE from a separate mixed model calculated at each time bin (Eq. 1). Significance of mixed model assessing ripple rates between words subsequently recalled vs. not recalled (Eq. 2): CA1, β = 0.10±0.022, P = 1.3 X 10^{-5}; CA3/DG, β = 0.11±0.028, P = 7.9 X 10^{-5}; AMY, β = 0.16±0.032, P = 2.1 X 10^{-6}; ENT/PHC, β = 0.10±0.030, P = 6.1 X 10^{-4} (each FDR-corrected across 4 tests of Eq. 2).

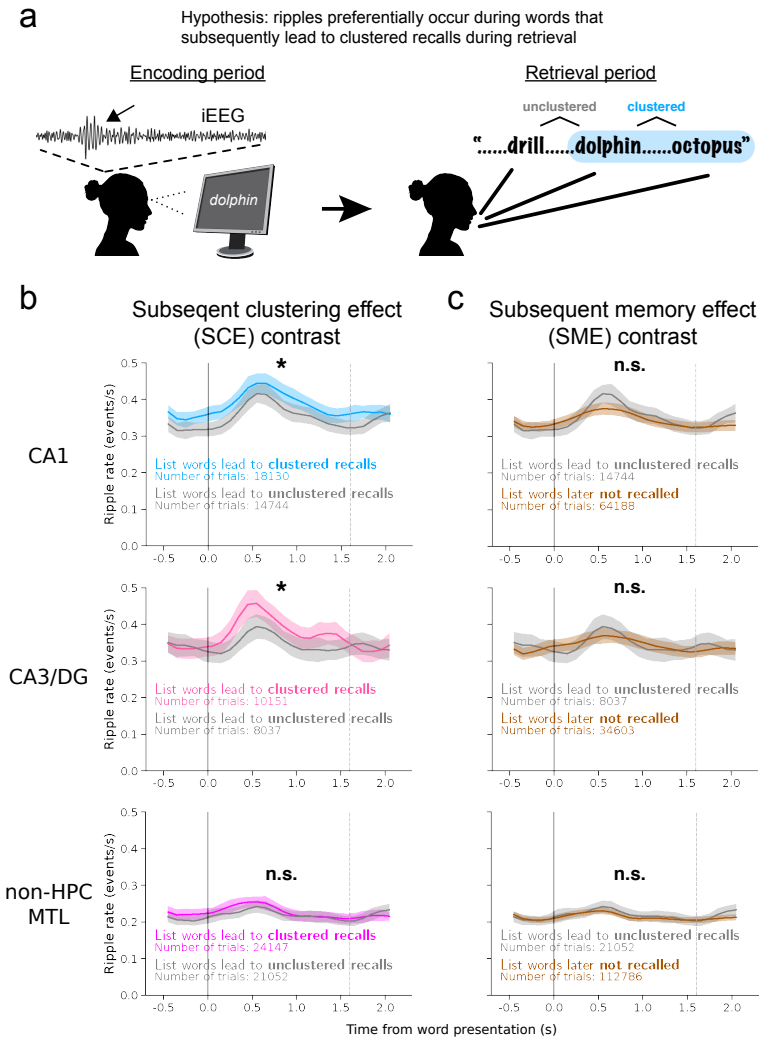

Figure 3. Hippocampal ripples signal a subsequent clustering effect (SCE). (a) Diagram explaining the subsequent clustering effect. When words are recalled during the retrieval period (right) we examine the relationships between the recall order to identify semantic or temporal relationships. Using the example list shown throughout the manuscript (see **Fig. 1b**), dolphin and octopus are adjacent semantic as they were a pair shown back-to-back and are from the same semantic category. We then measure ripples during the encoding period (left) when dolphin was presented as this was the word that led to the subsequent transition (or clustering) between recalls during retrieval. **(b)** Ripples rates grouped by clustering category for CA1, CA3/DG, and all other non-HPC MTL sites (including AMY, ENT, PHC, and perirhinal cortex). Each plot shows words that lead to subsequent clustering (adjacent semantic represents temporal and semantic clustering) vs. those that do not (remote unclustered combined with dead ends). Significance of mixed model comparing clustered vs. unclustered groups for each region: CA1, $\beta = 0.024 \pm 0.013$, $P = 6.2 \times 10^{-3}$; CA3/DG, $\beta = 0.043 \pm 0.010$, $P = 7.8 \times 10^{-5}$; non-HPC MTL, $\beta = 0.0069 \pm 0.0053$, $P = 0.19$ (each FDR-corrected across 3 tests of

Eq. 2). To directly assess if hippocampal regions show a stronger SCE than non-HPC MTL, we compare using a mixed model measuring the interaction between clustering groups and region to predict SCE: CA1 vs. non-HPC MTL, β = 0.026±0.0074, P = 4.1 X 10^{-4}; CA3/DG vs. non-HPC MTL, β = 0.046±0.0086, P = 2.6 X 10^{-7} (each FDR-corrected across 2 tests of Eq. 3). **(c)** Each plot shows a breakdown of words that are subsequently recalled but without clustering (remote unclustered and dead ends) vs. those that are not recalled. Significance of mixed model comparing these groups for each region: CA1, β = -0.0095±0.0076, P = 0.32; CA3/DG, β = -0.015±0.0089, P = 0.25; non-HPC MTL, β = -0.0039±0.0042, P = 0.35 (each FDR-corrected across 3 tests of Eq. 2). For all plots vertical black and gray lines denote word presentation onset and offset and error bands are SE from a separate mixed model calculated at each time bin (Eq. 1).

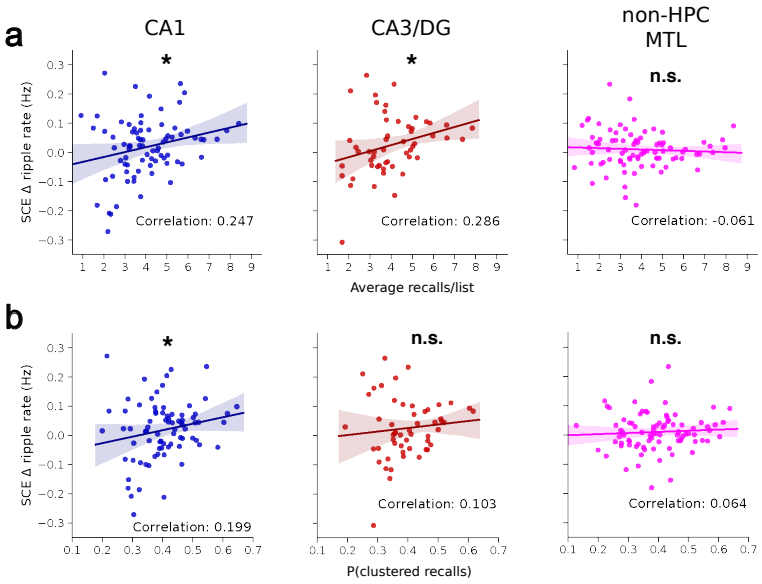

Figure 4. The ripple subsequent clustering effect (SCE) relates to memory performance. (a) For each participant, we relate the change in ripple rate during word presentation between words that lead to subsequent clustering (adjacent semantic and remote semantic) vs. no clustering (remote unclustered and dead ends) to the average number of words recalled on each list by that participant. Significance of mixed model comparing this change in ripple rate SCE vs. the average recalls per list: CA1, β = 0.015±0.0056, P = 0.017; CA3/DG, β = 0.022±0.0.0089, P = 0.017; non-HPC MTL, β = -0.0014±0.0033, P = 0.68 (each FDR-corrected across 3 tests of Eq. 5). To directly assess if hippocampal regions show a stronger relationship between SCE and recalls than non-HPC MTL, we compare using a mixed model measuring the interaction between recalls and region to predict SCE: CA1 vs. non-HPC MTL, β = 0.017±0.0061, P = 6.5 X 10^{-3}; CA3/DG vs. non-HPC MTL, β = 0.023±0.0079, P = 6.5 X 10^{-3} (FDR-corrected across 2 tests of Eq. 6). **(b)** For each participant, we relate the same change in ripple rate from **a** to the average proportion of clustered recalls (i.e. recalls that are adjacent semantic and remote semantic). Significance of mixed model comparing the SCE vs. the proportion of clustered recalls: CA1, β = 0.25±0.099, P = 0.040; CA3/DG, β = 0.14±0.14, P = 0.49; non-HPC MTL, β = -0.0095±0.053, P = 0.86 (each FDR-corrected across 3 tests of Eq. 7). To directly assess if hippocampal regions show a stronger relationship between SCE and clustering than non-HPC MTL, we compare using a mixed model measuring the interaction between proportion of clustered recalls and region to predict SCE: CA1 vs. non-HPC MTL, β = 0.25±0.095, P = 0.017; CA3/DG vs. non-HPC MTL, β = 0.11±0.12, P = 0.34 (FDR-corrected across 2 tests of Eq. 8).

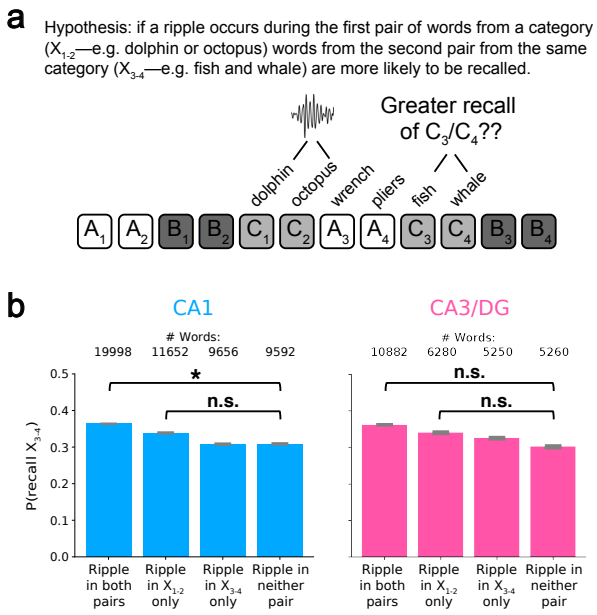

Figure 5. Presence of hippocampal ripples during initial category presentation leads to better recall of words from the same category. (a) Diagram of hypothesis that ripples during presentation of words from a category will increase likelihood of recalling subsequently presented words from same category. **(b)** Accuracy of recall for the second pair of words from a category (X_{3-4}) when a ripple occurs during either of the first pair of words from a category (X_{1-2}), either of the second pair of words from a category (X_{1-2}), both, or neither. The number of total words for each of these pools is indicated above the bars. Error bars are SE of proportions. Significance of mixed model term assessing the impact on accuracy for X_{3-4} based on the presence of ripples during X_{1-2}: CA1, β = -0.012±0.0063, P = 0.12; CA3/DG, β = -0.0059±0.0087, P = 0.50 (each FDR-corrected across two tests of Eq. 9). Significance of mixed model term assessing the impact on accuracy for X_{3-4} based on the presence of ripples during both X_{1-2} and X_{3-4}: CA1, β = 0.023±0.0083, P = 0.010; CA3/DG, β = 0.012±0.011, P = 0.30 (each FDR-corrected across two tests of Eq. 9).

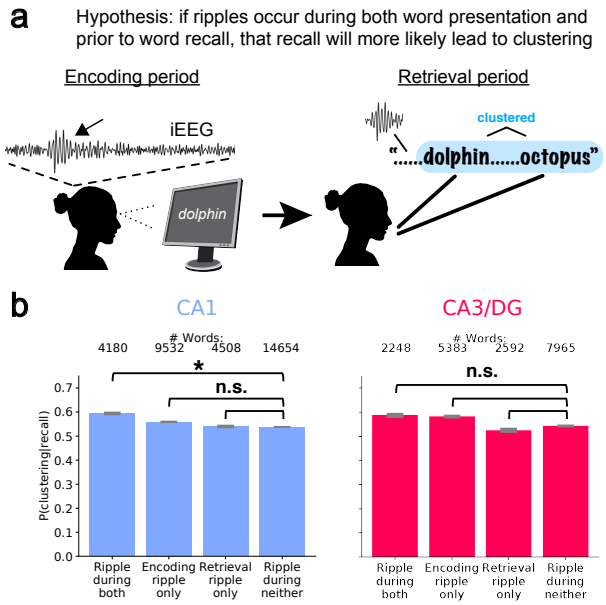

Figure 6. Words with ripples during both word presentation and prior to recall lead to clustering. (a) Diagram of hypothesis that clustering arises when ripples occur during both the presentation of and prior to the recall of words. **(b)** Proportion of recalls that lead to clustering conditioned on whether the recalled word has $\geq$1 ripple occur during its initial presentation and/or prior to its vocalization. The number of total recalls for each condition is indicated above the bars. Error bars are SE of proportions. Significance of mixed model terms assessing the impact on clustering of the presence of ripples during both word encoding or retrieval: CA1, $\beta = 0.037 \pm 0.0074$, $P = 9.5 \text{ X } 10^{-7}$; CA3/DG, $\beta = 0.023 \pm 0.017$, $P = 0.17$ (each FDR-corrected across 2 tests of **Eq. 10**). P-values for remaining terms are not significant ($P \geq 0.065$, each FDR-corrected).

Materials and Methods

Human participants. Comprising the dataset are intracranial recordings from 129 adult participants in the hospital for drug-resistant epilepsy surgery with subdural electrodes placed on the cortical surface or within the brain to localize epileptic activity. Collaborating hospitals include Thomas Jefferson University Hospital (Philadelphia, PA), University of Texas Southwestern Medical Center (Dallas, TX), Emory University Hospital (Atlanta, GA), Dartmouth-Hitchcock Medical Center (Lebanon, NH), Hospital of the University of Pennsylvania (Philadelphia, PA), Mayo Clinic (Rochester, MN), and Columbia University Hospital (New York, NY). All participants consented to research under a protocol approved by the Institutional Review Board at the University of Pennsylvania via a reliance agreement with each hospital.

Free recall task. Participants were tested on a delayed free recall task in which each ”list” comprised viewing a sequence of common nouns with the intention to commit them to memory. The task was run at bedside on a laptop and participants were tasked to finish up to 25 lists for a whole session or 12 lists for a half-session. The free recall task consisted of four phases per list: countdown, encoding, distractor, and retrieval (**Fig. 1a**). Each list began with a 10-second countdown period with numbers displayed from 10 to 1. For encoding, participants were sequentially presented 12 words centered on the screen that were selected at random–without replacement in each whole session or two consecutive half sessions–from a pool of 300 high frequency, intermediate-memorable English or Spanish nouns (http://memory.psych.upenn.edu/WordPools (*Weidemann et al., 2019*)). Each word was presented for 1600 ms with a jittered 750-1200 ms (randomly sampled uniform distribution) blank screen shown after each word. After encoding was a distractor period where participants performed 20 seconds of arithmetic math problems to disrupt their memory for recently-shown items. Math problems were of the form A+B+C=??, where each letter corresponds to a random integer and participants typed their responses into the laptop keyboard. The final phase is retrieval, in which participants had 30 seconds to recall as many words–in any order–from the most recent list as possible. Retrieval began with a series of asterisks accompanied by a 300 ms, 60 Hz beep to signal for the participants to begin vocalizing recalled words. Vocalizations were recorded and later annotated offline using Penn TotalRecall (http://memory.psych.upenn.edu/TotalRecall) to determine correct and incorrect recalls. For each session the participant began with a practice list of the same words that we do not include in the analysis.

All analyses in this manuscript are done on a variant called categorized free recall, in which each list is comprised of words with semantic relationships. For every whole session (or consecutive half sessions), words were drawn from a pool of 300 that included 12 words each from 25 categories created using Amazon Mechanical Turk to crowdsource typical exemplars for each category (*Weidemann et al., 2019*). For each list, three semantic categories were randomly chosen, and the four words from each category were presented sequentially in pairs. Pairs from the same category were never shown back-to-back (in other words, the four words from the same category were never shown in a row). This setup allowed us to study both adjacently (same pair) and remotely presented words from the same category.

Intracranial electroencephalogram (iEEG) recordings. iEEG was recorded from macroelectrodes on subdural grids and strips (intercontact spacing 10.0 mm) or depth electrodes (intercontact spacing 3-6

mm) using DeltaMed XlTek (Natus), Grass Telefactor, Nihon-Kohden, Blackrock, or custom Medtronic EEG systems. Signals were sampled at 500, 512, 1000, 1024, 1600, 2000 or 2048 Hz and downsampled using a Fourier transformation to 500 Hz for all analyses. Initial recordings were referenced to a common contact, the scalp, or the mastoid process, but to eliminate possible system-wide artifacts and to better isolate localized high frequency signals we applied bipolar rereferencing between pairs of neighboring contacts. Bipolar referencing is ideal as the spatial scale of ripples is unlikely to exceed intercontact spacing of our recordings (3-10 mm) (*Vaz et al., 2019*). Line removal is performed between 58-62 using a 4th order Butterworth filter (120 Hz is in our sensitive ripple range and we did not find artifacts in these frequencies).

Ripple detection. Detection of ripples is identical to our previous work, where we performed numerous control analyses to ensure the detector is robust to vocalization artifacts, frequency window selection, correlations across channels, and seizurogenic activity (*Sakon and Kahana, 2022*), and is based on prior human work (*Norman et al., 2019, Gelinas et al., 2016*). Briefly, local field potential from bipolar iEEG channels is bandpass Hamming filtered from 70-178 Hz, rectified, squared, smoothed, and normalized to find candidate events exceeding 3 standard deviations (SD) that are expanded to find their duration above 2 SDs. Events are considered ripples if the expanded duration is between 20 and 200 ms and not within 30 ms of another expanded event (in which case the events are merged). To avoid pathological interictal epileptiform discharges (IEDs), LFP is bandpass Hamming filtered from 25-58 Hz rectified, squared, smoothed and normalized to detect events 4 SD above the mean. Ripples within 50 ms of an IED event are removed.

We treat ripples as discrete events (**Fig. 2a**) with the timestamp set to the beginning of each ripple (**Fig. 1d**). The average power of events is $\sim$90 Hz, although individual events peak throughout the 70-178 Hz range (**Fig. 1d** shows 8 single ripple examples). Most participants had multiple MTL contacts within their montage, thereby providing iEEG recordings from multiple channels for every word presentation. As with previous work (*Norman et al., 2019, Vaz et al., 2019, Sakon and Kahana, 2022*), since the spacing of clinical electrodes (3-10 mm) is much farther than ripples are expected to travel in the brain ($<$0.2mm, (*Sullivan et al., 2011*)), we consider each presented word for each channel as a separate "trial". To ensure ripples are not double-counted across neighboring channels we use a combination of automated channel and session removal (by measuring correlations across trials and channels, respectively) and manual inspection of raster plots (**Fig. 2a**) as detailed in previous work (*Sakon and Kahana, 2022*).

High frequency activity (HFA). We calculate HFA by averaging oscillatory power extracted using Morlet wavelets at 10 logarithmically-spaced frequencies from 64-178 Hz, with the lower bound as in previous HFA work (*Burke et al., 2014a, Burke et al., 2014b, Henin et al., 2021*) and the upper bound the same as for the ripple detector. To measure powers, we use the following procedure using the bipolar-referenced iEEG from each trial from 1000 ms before word presentation until 2600 ms after word presentation. This window includes a 300 ms buffer on both sides to avoid edge effects during Morlet transform and 700 ms (comprised of the inter-trial interval) both before and after word presentation to incorporate as part of the normalization procedure. The signal is then Butterworth filtered from 118-122 Hz and high-pass filtered from 0.5 Hz. A Morlet wavelet transform (using PTSA, see notebooks 5 and 6 on

https://github.com/pennmem/CMLWorkshop) is done for each of the 10 frequencies (64.0 , 71.7, 80.3, 90.0, 100.8, 113.0, 126.6, 141.8, 158.9, and 178), the buffers are removed, and the log of each value is taken. Next, we resample to 100 ms bins, which leaves us with a FREQUENCY X WORD X CHANNEL X 30 BIN array. We then z-score this array by subtracting the average across words and bins, and dividing by the standard deviation across words after averaging across bins. Finally, we average across the 10 frequencies to arrive at a final HFA value for each WORD X CHANNEL X BIN.

To make the fairest comparison between HFA and ripples, we use the exact same set of trials as selected by our criteria for the ripple detection algorithm. That is, the same word presentations recorded in the same channels (note the identical trial counts in **Fig 2b-c**).

Anatomical localization. Localization of contacts is identical to previous work (*Sakon and Kahana, 2022*). Briefly, pre-implant structural T1- and T2-weighted MRI scans were used to define the anatomical regions for each participant in addition to a post-implant CT scan to localize electrodes in the participant brain, which were coregistered using Advanced Normalization Tools (*Avants et al., 2011*). The point source of iEEG for bipolar electrode pairs is considered to be the midpoint between adjacent electrode contacts. Center to center electrode spacing was between 3-10 mm as chosen by the neurosurgical teams for medical reasons.

Similar to our previous work (*Sakon and Kahana, 2022*), we split channels localized to hippocampus into two groups, CA1 and CA3/DG, since we have sufficient sample size to test our hypotheses in each. However, since we use the midpoint of bipolar electrode pairs for signal localization (hippocampal pairs are 3-6 mm apart as only stereo-EEG depth electrodes reach hippocampus), and considering an estimated 350,000 neurons contribute to macroelectrode LFP (*Sakon and Kahana, 2022*), many of the channels are likely to reflect ripples crossing subfields.

Bipolar electrode pairs in hippocampal subfields CA1 and dentate gyrus (DG) were localized using a combination of neuroradiologist labels (Joel M. Stein and Sandhitsu Das, Penn Medicine) and the automated segmentation of hippocampal subfields (ASHS) technique utilizing the T2 scan (*Yushkevich et al., 2015*). However, we label the DG pairs as CA3/DG due to the difficulty in delineating these regions. Sites localized to CA3 are not included in this group as ASHS achieves poor classification of this subfield compared to CA1 and DG (*Yushkevich et al., 2015*)), and because of its relatively small volume $\sim$15x fewer channels are localized to CA3 than DG.

We also analyze electrode pairs in non-hippocampal cortical regions, which include entorhinal (ENT), parahippocampal (PHC), and perirhinal cortex and amygdala. We used a combination of neuroradiologist labels and an automated segmentation pipeline combining whole-brain cortical reconstructions from the T1 scan in Freesurfer (*Fischl et al., 2004*), an energy minimization algorithm to snap electrodes to the cortical surface (*Dykstra et al., 2012*), and boundaries and labels from the Desikan-Killiany-Tourville cortical parcellation protocol (*Klein and Tourville, 2012, Desikan et al., 2006*)).

Plots and binning. Raster plots are formed by aligning the iEEG to the time of word presentation and plotting the time of the beginning of each detected ripple. Peri-stimulus time histograms (PSTHs) are formed by binning ripples (100 ms bins) and averaging the raster plots across participants after sep-

arating words into groups (e.g. subsequently recalled vs. not recalled words). For visualization only, these PSTHs are triangle smoothed using a 5-bin window (*Norman et al., 2019, Sakon and Kahana, 2022*) and a separate linear mixed model with sessions nested in participants is run at each bin to calculate the mean and error bars (SE) (**Eq. 1**). Ripple rates are the frequency in Hz. within each bin.

The default analysis window used to assess the ripple subsequent memory effect (SCE) and the subsequent clustering effect (SCE) throughout the paper is from 100 to 1700 ms from beginning of word presentation. We offset 100 ms from time on screen to account for latencies from the time of presentation until signals reach MTL circuits (*Rey et al., 2014*). The analysis window for HFA is from 400 to 1100 ms after word presentation. These windows are based on pilot analyses done on the first half of the data and pre-registered on the Open Science Framework (OSF, https://osf.io/e98qp). To measure pre-retrieval effect (PRE) ripples during the retrieval period we use the window from -1100 to -100 ms prior to recall vocalization (*Sakon and Kahana, 2022*).

Clustering. When participants correctly recall a series of words during the retrieval period, the order of word recall provides a window into the organization of their memory. For categorized free recall, as participants transition from one recall to the next, we expect them to cluster recalls based on semantic and/or temporal relationships between words on the list. As explained in the Free Recall task section above, each 12-word list in this task had words drawn from 3 categories, with the 4 words from the same category presented in non-contiguous pairs. This setup provides three distinct forms of clustering between consecutive recalls: adjacent semantic (20% of recalls lead to this transition), remote semantic (20%), and adjacent non-semantic (3%) (examples given in **Fig. 1b-c**). Adjacent semantic are two words from the same category shown as a consecutive pair during encoding while remote semantic are two words from the same category from pairs separated by other words. Adjacent, non-semantic transitions were not analyzed due to their small sample size. Recalls that do not lead to clustering include remote unclustered (17%), where consecutive words were neither from the same category or shown back-to-back, and dead ends (26%), which are the last recall that do not lead to a subsequent recall. The remaining recalls were those that led to intrusions or repeats (14%).

For the SCE contrast we pool clustered and unclustered recalls in **Fig. 3b**, and for the SME contrast we pool unclustered recalls in **Fig. 3c**. However, in the caption for the SCE contrast, we also provide statistics for pairwise models between each of the clustering types (adjacent semantic and remote semantic) vs. unclustered recalls. And in the caption for the SME contrast, we also provide statistics for pairwise models between each of the unclustered types (remote unclustered and dead ends) vs. not recalled words.

Held out data and pre-registration. The large size of our dataset allowed us to set aside $\sim$35% of trials in order to come up with initial figures and hypotheses that can then be confirmed with the entire dataset. That is, after creating a raster plot to ensure all data is in usable form after the data-cleaning steps outlined in Ripple Detection above, we used a random kernel to select a subset of participants comprising 35% of hippocampal trials. Once we set our initial analysis parameters and figures based on this exploratory 35% of data, we registered them along with hypotheses based on these figures on the Open Science Framework (https://osf.io/e98qp), which also contains specific details on the randomization and sampling plan. Here we present the statistics and figures for the entire dataset based on the

analysis parameters defined in this pre-registration.

Equations. Linear mixed effects models are run using the function MixedLM in the python package statsmodels with restricted maximum likelihood and Nelder-Mead optimization with a maximum of 2000 iterations. The following equations are written in pseudocode of the inputs to statsmodels. Statistics are presented as: $\beta \pm SE, P-value$, where β is the coefficient being tested and SE is the standard error of the coefficient being fit. For all comparisons the first group takes the indicator value 1 and the second takes 0 in the model. For example, clustered vs. unclustered trials are assigned 1 and 0, meaning if clustered is greater the coefficient will be positive.

We use mixed effects models to plot the mean and standard error of ripple rates for all peri-stimulus time histograms (PSTHs). For a given group of trials, a separate mixed effect model is run at each 100 ms bin:

$$ripple_rate \sim 1 + (1|participant) + (1|participant : session) \tag{1}$$

where $1|participant$ is a random intercept and slope for each participant, $(1|participant : session)$ is a random intercept and slope for each session nested in each participant, and $ripple_rate$ is the average ripple rate in that bin for a given trial. The solved coefficient and its standard error (SE) are used to plot the mean $\pm$ SE at each bin (after a 5-point triangle smooth of the means). Plotting the average ripple rates across trials looks similar, but plotting using the mixed effects model fits gives a closer approximation to the statistical model used to compare groups of trials in **Eq. 2**.

To test the hypothesis that ripples rates are higher during words that are subsequently recalled vs. subsequently not recalled (**Fig. 2b & 3b**), we use the linear mixed effects model:

$$\begin{aligned} ripple_rate \sim & recall_indicator + (recall_indicator|participant) + \\ & (recall_indicator|participant : session) \end{aligned} \tag{2}$$

where $recall_indicator$ is an indicator variable with value 1 for words subsequently recalled and 0 for those that are not, $(recall_indicator|participant)$ are random intercepts and slopes for each participant, $(recall_indicator|participant : session)$ are random intercepts and slopes for sessions nested in each participant, and $ripple_rate$ is the average ripple rate for each trial from 0.1 to 1.7 s following word presentation. The null hypothesis is no difference between ripple rates on words that are subsequently remembered v. subsequently not recalled.

We use the same model for comparisons between groups, such as words that subsequently lead to clustered recalls vs. unclustered recalls **Fig. 3**). In this case, instead of recall indicator, the predictor indicates if a recalled word subsequently leads to clustering or not (e.g. subsequently clustered vs. unclustered recalls, **Fig. 3b**). We also use this model to compare SMEs for HFA, although for this comparison use 0.4-1.1 s following word presentation.

To compare SCE between regions we use the linear mixed-effects model:

$$\begin{aligned} ripple_rate \sim & clustering_indicator * region_indicator + \\ & (clustering_indicator * region_indicator | participant) + \\ & (clustering_indicator * region_indicator | participant : session) \end{aligned} \tag{3}$$

where $clustering_indicator$ is 1 if a word subsequently is recalled and leads to clustering (i.e. adjacent semantic or remote semantic) or 0 if a word is recalled and does not lead to clustering (i.e. remote unclustered or dead end), and region_indicator is 0 or 1 for two given regions. The * indicates separate coefficients are calculated for each term and the interaction. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is no difference in the interaction between SCE and region.

We also compare the SCE and the SME directly in the same model:

$$\begin{aligned} ripple_rate \sim & recall_indicator + clustering_indicator + \\ & (recall_indicator + clustering_indicator | participant) + \\ & (recall_indicator + clustering_indicator | participant : session) \end{aligned} \tag{4}$$

where $recall_indicator$ and $clustering_indicator$ are the same as in **Eq. 2** and **Eq. 3**, respectively. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is no difference between clustered and unclustered words after taking into account recalled vs. not recalled words.

We hypothesize that participants that recall more words will show a bigger ripple subsequent clustering effect (SCE), in which words that subsequently are recalled and lead to clustering will have more ripples than words that are subsequently recalled and do not lead to clustering. To test this relationship we use the linear mixed effects model:

$$\begin{aligned} \Delta ripple_rate \sim & average_recalls + (average_recalls | participant) + \\ & (average_recalls | participant : session) \end{aligned} \tag{5}$$

where $average_recalls$ is the average number of recalls per 12-word list for the participant and $\Delta ripple_rate$ is the average difference in ripple rate from 0.1 to 1.7 s following word presentation for subsequently clustered (i.e. adjacent semantic and remote semantic trials) vs. unclustered (i.e. remote unclustered and dead ends) words. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is that SCE does not relate to memory performance.

To directly compare the SCE vs. memory interaction between regions, we use the linear mixed effects

model:

$$\begin{aligned}\Delta ripple_rate \sim & average_recalls * region_indicator + \\ & (average_recalls * region_indicator | participant) + \\ & (average_recalls * region_indicator | participant : session)\end{aligned} \tag{6}$$

The * indicates separate coefficients are calculated for each term and the interaction. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is no difference in the interaction between the average number of recalls and region ($average_recalls : region_in$

We also compare the SCE $\Delta ripple_rate$ with the amount of clustering at the participant-level using a similar linear mixed-effects model:

$$\begin{aligned}\Delta ripple_rate \sim & proportion_clustered + (proportion_clustered | participant) + \\ & (proportion_clustered | participant : session)\end{aligned} \tag{7}$$

where $proportion_clustered$ is the combined number of words that lead to adjacent semantic and remote semantic trials divided by the total number of words recalled for each participant. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is that SCE does not relate to the amount participants recall words via clustering.

To directly compare the SCE vs. clustering interaction between regions, we use the linear mixed effects model:

$$\begin{aligned}\Delta ripple_rate \sim & proportion_clustered * region_indicator + \\ & (proportion_clustered * region_indicator | participant) + \\ & (proportion_clustered * region_indicator | participant : session)\end{aligned} \tag{8}$$

The * indicates separate coefficients are calculated for each term and the interaction. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is no difference in the interaction between the proportion of clustered recalls and region ($proportion_clustered :$

Next we investigate the hypothesis that a ripple during the first pair of words from a category (X_{1-2}) will make it more likely to see reinstatement–and therefore a ripple (*Sakon and Kahana, 2022*)–during the second pair of words from a category (X_{3-4}). As a result, we expect likelier recall of X_{3-4} if a ripple occurs during X_{1-2}, and even likelier recall if a ripple occurs during both pairs. To test this hypothesis we use the linear mixed-effect model:

$$\begin{aligned}recall_X_{3-4} \sim & ripple_X_{1-2} * ripple_X_{3-4} + ripple_other_words + \\ & (ripple_X_{1-2} * ripple_X_{3-4} + ripple_other_words | participant) + \\ & (ripple_X_{1-2} * ripple_X_{3-4} + ripple_other_words | participant : session)\end{aligned} \tag{9}$$

where $recall_X_{3-4}$ indicates if a participant recalled a word from X_{3-4}, $ripple_X_{1-2}$ indicates a ripple occurred during X_{1-2}, $ripple_X_{3-4}$ indicates a ripple occurred during X_{3-4}, and $ripple_other_words$ is the ripple rate for the remaining (eight) words on the list not from that category. The * indicates

separate coefficients are calculated for each term and the interaction. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypotheses are that 1) recall of a word from X_{3-4} is not more likely if a ripple occurs during X_{1-2} (the coefficient for $ripple_X_{1-2}$) and 2) recall of a word from X_{3-4} is not more likely if a ripple occurs during both X_{1-2} and X_{3-4} (the coefficient for the interaction $ripple_X_{1-2} : ripple_X_{3-4}$).

As a control, we use the same model as above to predict X_{1-2} recalls (instead of X_{3-4} recalls). The null hypothesis is recall of words from X_{1-2} is not more likely if a ripple occurs during both X_{1-2} and X_{3-4}.

Finally, we test the hypothesis that a ripple during encoding of a word combined with a ripple in the PRE window during its subsequent recall will increase the likelihood that word leads to clustering. To test this hypothesis we use the linear mixed effects model:

$$\begin{aligned} clustering_indicator \sim & encoding_ripple * retrieval_ripple + \\ & (encoding_ripple * retrieval_ripple | participant) + \\ & (encoding_ripple * retrieval_ripple | participant : session) \end{aligned} \tag{10}$$

where $clustering_indicator$ is 1 if a recalled word leads to clustering and 0 if not (i.e. remote unclustered or dead end), $encoding_ripple$ is an indicator variable with the value 1 if $\geq$1 ripple occurred in the window from 0.1 to 1.7 s after word presentation, and $retrieval_ripple$ is an indicator variable with the value 1 if $\geq$1 ripple occurred in the window from -1.1 to -0.1 s aligned to vocalization of the word during retrieval. The * indicates separate coefficients are calculated for each term and the interaction. Random intercepts and slopes for sessions nested in participants follows the same structure as **Eq. 2**. The null hypothesis is no increase in clustering when a ripple occurs during encoding of a word and prior to its subsequent recall.