

Electrophysiological signatures underlying variability in human memory consolidation

Corresponding Author: Dr Liang Wang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors present an interesting study that uses targeted memory reactivation in epilepsy patients to investigate electrophysiological activity related to memory consolidation. The authors aim at relating hippocampal ripples to connectivity between the hippocampus and neocortex, which promises to provide novel insights into the neural foundations of memory processes and is potentially interesting to a large audience. Overall, the paper is well-written and the results are described clearly. However, the authors need to address several critical issues that question the robustness of some of the major findings. These and other issues are detailed below:

1. Reactivation time windows:

1.1 Figure 3B: Are these two separate time windows? They are rather close in time and might just appear separate because of statistical thresholds. I am not convinced that the permutation test procedure is designed to evaluate more than one cluster. How does the procedure treat second clusters? Does it compare those clusters to the same distribution as first clusters? If so, is this justified, given that second clusters are not independent of first clusters?

More generally, the test procedure tests whether two conditions are different and not when (in time) these conditions differ. Inferences on what drives these differences would need to be inspected separately, for example on the basis of grand averages. Do grand-averages reflect two clusters? A similar criticism applies to the results reported Figure 4E.

1.2 In Figure S4, it does not look like there are two clusters per subject and the timing of the two largest clusters per subject does not correspond to the two time bins reported in Fig. 4E, in particular with respect to the timing on the x-axis. There seems to be a discrepancy between group level and single subject results which the authors would need to explain. This also questions the statement in the discussion (l. 362) that the time-frequency analysis found consistent results.

1.3 As far as I understand, the control analyses depicted in Fig. S4 and S5 were done by correlating the whole matrix coming from single round/subject data with the mean (excluding the round/subject in question). I do not think this strongly supports the stability of the relevant patterns (the temporal clusters), because the correlations might reflect overall similarities between the patterns across all time points and not specifically reflect the two clusters in question.

1.4 The reactivation windows are rather late, in particular the second time window which also displays differences in spindle power. Usually, these differences are found earlier (e.g., Cairney et al.). How do the authors explain this delayed pattern? The presented results seem to be somewhat at odds with some previous findings (e.g., Cairney et al.), which should be discussed in more detail. Cairney et al. find reactivation connected to spindles in a similar time window to what is reported here as first reactivation. However, there is no relation to spindles in the present study.

2. Ripples:

2.1 The authors filtered the LFP in a very broad band, between 70-180 Hz. This is very uncommon, usually more narrow bands are used, e.g., 80-120 Hz (see Vaz et al., 2019).

Why would the authors choose such a broad band?

Please provide a description of the distribution of ripples across frequencies, e.g. a histogram that shows ripple counts per 10 Hz. Alternatively, the authors could repeat their analyses with ripples filtered in a narrower, more commonly used frequency band, e.g., 80-120 Hz, to make results comparable to previous studies.

2.2 The authors should include a minimum of 3-4 cycles in the raw signal as a criterion for ripple detection. The duration criterion of min. 20 ms is insufficient to detect ripples that display sufficient numbers of cycles, which makes this method prone to include sharp spikes as they occur during epileptic activity.

2.3 Please provide a power spectrum of epochs around all detected ripple events (e.g., ripple $\pm$ 500 ms) that covers SO, spindle and ripple frequencies. Are there distinct peaks in the power spectrum for SO, spindles, and also in the ripple band?

2.4 Related to this issue, why does the ripple average (Fig. 2C, 4B) display such clear cycles (of a particular frequency?) if ripples were detected between 70 and 180 Hz? Averaging across distinct ripples with a particular frequency across a broad frequency range should result in a smoother, flatter curve.

2.5 For each participant, only one hippocampal contact was used for ripple detection. A lot of data gets lost with this approach. Please detect ripples on all hippocampal electrodes and repeat the main analyses.

The authors state that the hippocampal contact had the highest probability of being located in the hippocampus during the automatic brain segmentation. Is this based on co-registered anatomical data using an atlas? If so, this should be done in native patient space and verified by individual visual inspection. Fig. 1A is too small to understand this selection, it seems like some of the sphere are right at the edge of the hippocampal volume while other, unselected contacts are right in the center.

2.6 clipping of extreme values: how was this done (please provide criteria) and why was this necessary?

2.7 The authors should provide an epileptiform spike-locked average and plot this next to their ripple-locked average. Can one clearly distinguish pathological from non-pathological averages?

2.8 100 ms intervals between epileptiform spikes and other detected ripples might be too short, given that typical sharp wave after an epileptiform spike can be around 200ms long. Could the authors repeat the main analyses with a larger time interval between epileptiform spikes and ripples?

2.9 Do ripple rates differ between states? Previous studies have shown higher ripple rates during NREM than REM sleep (e.g., Liu et al., 2022, Nat Com). Does this pattern hold in the current data set? The authors should report ripple rates for REM sleep and directly compare those to NREM sleep. A comparison with wake ripple rates would also be beneficial?

Please report ripple rates in Hz (i.e., events / second), to make them comparable to previous studies (e.g., Vaz et al., 2019).

3. ERPs and spectral power locked to the TMR cue.

3.1 Basic information on how brain activity relates to TMR cues is missing. Please add figures for TMR cue-locked time-frequency power as well as ERPs of cortical contacts and hippocampus (without statistical comparisons, masking or computing differences). This will help to understand whether TMR cues elicited the typical patterns in the SO and spindle bands. Please add figures that display grand-average of: all experimental cues; all control cues; each condition (strengthening, decaying and control); please add separate plots for hippocampus and cortical contacts figures, respectively;

3.2 Fig. S8 is puzzling. The legend says that cluster-based permutation tests were done and gives one p-value per panel, but A and B show many outlines across various time-frequency ranges. Are all of these significant clusters?

3.3 In relation to this, I do not agree with the statement in the results section (l. 260 ff) that there are specific patterns related to the first vs. second time window of reactivation. In Fig. S8A, for example, there seems to be a large increase (outlined in black, thus significant?) from around 10 to 20Hz just after 2 s, which would be in the first cluster. How is that in line with the authors' statement that they found significant cortical spindle activity enhanced only during the second reactivation? This also questions the statement in the discussion (l. 362) that the time-frequency analysis found consistent results.

4. Slow oscillations, SO-spindle coupling

It is surprising that the paper lacks information on slow oscillations and their relationship to spindles and ripples, given that the study investigates neural activity during NREM sleep. Previous studies showed that slow oscillations and their coupling with sleep spindles is important for memory reactivation during sleep (e.g, Schreiner et al., 2021, Nat Com). Moreover, ripples also couple to SO-spindle complexes (e.g., Staresina et al., 2023). To relate to these findings, the authors could analyze SO-spindle coupling and SO-spindle-ripple coupling during NREM sleep in relation to TMR cues and investigate whether a potential coupling is related to memory reactivation.

5. Connectivity analysis

5.1 Coherence can be contaminated by volume conduction. To avoid this, please report a connectivity metric that excludes zero-lag connectivity (e.g., imaginary coherence).

5.2 Coherence (and other connectivity metrics) can be heavily biased by trial numbers, which is particularly relevant when comparing conditions. This is even more problematic when trial numbers are generally low, as it is the case here: the first two participants had a maximum of 27 / 36 trials per condition. Since I did not find information on this in the methods section,

I assume no correction for number of trials was performed. The authors should repeat the analyses using appropriate metrics.

5.3 Please perform the connectivity analyses based on all hippocampal contacts vs. all cortical contacts. Please report the number of trials that included into each of the three conditions for each of the displayed coherence analyses.

5.4 related to this issue: how many trials went into the RSS/RRS analyses? Are there enough trials to support stable estimations of these metrics?

6. Relation of neural signatures and memory performance

6.1 Was the ripple rate during learning related to successfully remembering the location of the item? One would expect more ripples during correctly remembered locations (smaller drop error) than forgotten locations (larger drop error), in each of the retrievals (pre-sleep, post-sleep). This would substantially strengthen the link between the reported ripples and memory processes and relate to previous findings (e.g., Vaz et al. (2019).

6.2 The authors should add a similar contrast to their reactivation metric during learning. Was the reported higher WI similarity related to memory performance? Does a higher WI similarity correspond smaller drop error in a given trial? Please add figures depicting these contrasts.

7. Multiple Comparisons

Are all reported p-values corrected for multiple comparisons? For example, when conducting multiple t-tests displayed in Fig. 4G? If not, please use appropriate corrections and report corrected p-values. For comparisons between 3 conditions, an ANOVA that tests the difference between the 3 conditions should be computed first and post-hoc t-tests in case the ANOVA shows a significant difference.

8. Sound cues:

Please provide mean duration (and SDs) for the of the sound cues. Did the duration of the sound cues differ between conditions? How would different durations affect the outcomes? Would locking to sound cue offset change the pattern of the results?

9. Title: The word 'mechanism in the title could be misleading to large parts of the audience. I suggest to remove this highly debated term.

10. Questions regarding the intro / discussion:

10.1 Abstract: 'produced nonuniform amounts of memory enhancement across a set of object-location associations that were reactivated'

What do the authors mean by 'nonuniform'? They show an overall benefit of TMR cuing. This effect varies across items, but that is a typical pattern seen in many TMR studies. The same questions apply to I. 317.

10.2 I. 95: 'In short, endogenous tags based on emotion or reward, or exogenous sensory cues, can systematically improve consolidation during sleep and thus strengthen the corresponding memories.'

Memory reactivation has been shown without emotional or reward tags, see e.g., Schreiner et al., 2022.

10.3 I. 334: 'Furthermore, our study revealed multiple TMR induced memory reactivations. This result suggests that memory consolidation during NREM sleep may require multiple reactivations.'

An interesting question related to this statement is whether it is the same item or different items that are reactivated multiple times. Can the authors address this?

10.4 I. 355 : 'However, considering the echoing effect of TMR-induced reactivation, these previous studies have not indicated whether there are hippocampal ripples during each cortical reactivation.'

Where do the authors show that there is a ripple for each cortical reactivation? I might have missed this, but if there is no evidence to support the claim, please remove it.

10.5 I. 388: 'indicating communication between hippocampus and cortex.'

The authors should acknowledge the ongoing debate on the relationship between coherence and neuronal communication (Schneider et al., 2021, Neuron; Dowdall & Vinck, 2023, Neurolmage; Fries, 2015, Neuron) and carefully evaluate their own interpretation.

10.6 I. 391: 'transmitted to cortex'

Where do the authors show information transfer to the cortex? ? I might have missed this, but if there is no evidence to support the claim, please remove it.

Minor:

-Paradigm: Some terms are misleading in my opinion. Why is the initial phase of the experiment called 'preview phase' and not 'study phase', which is much more common? Also, the "learning stage" is actually a retrieval task. Together, the reader may confuse these phases in the present study.

- Figure 6 indicates that multiple distinct cortical areas were investigated, which was not the case.

- Figure S7. Please add occurrence rate of retrieval ripples

- Line 218 and Figure 3: please specify what a (gamma) pattern is.

- Figure S2: What is encoding-sleep similarity (ESS) and why was it compared to Retrieval-Sleep Similarity (RSS)? ESS is not explained in the methods.

- Line 952: 'marginally significant': readers might object to this wording

- What is the rationale behind choosing 50-140 Hz as high gamma band?

-l. 106: TMR cue appears to occur in multiple consecutive sweeps and produce a temporal sequence of reactivation. Where do Cairney et al., 2018 show this?

Reviewer #2

(Remarks to the Author)

In the present study, the authors employ targeted memory reactivation during sleep to demonstrate that ripples and sleep spindles reflect two distinct signatures of memory reactivation. They recorded intracranial EEG from 11 participants and simultaneously recorded from neocortex and the hippocampus. The findings are interesting, since the authors describe two distinct time points after reactivation that are either characterized by ripples (ca. 2 s after reactivation) or spindle activity (ca. 3-4 s after reactivation) that differentiate items that were later remembered or forgotten. These observations are interesting since they extend the finding of spindle 'pulsing' every 3-6 s (Antony et al., 2018, *Curr Biol*; Sirota et al., 2003, *PNAS*) that may have differential effects on the underlying circuit (e.g., Niethard et al., 2018, *PNAS*) in support of memory consolidation.

While this constitutes an interesting finding, it is unfortunately based on a relatively small sample of only 11 participants, with an overall underpowered statistical design (25 reactivated patterns, 50 in total) that yielded a weak statistical behavioral effect ($p = 0.049$). Moreover, the study does not involve a control night, which would have been easy given that patients stay in the EMU for multiple nights. In addition, multiple analytical problems hamper the interpretation. If this work can be replicated in a larger cohort and with suitable controls and appropriate analyses and statistics, then it might open new avenues to studying memory reactivation and the role of a 2-step process, where ripples and spindles play distinct roles.

A major caveat is how the study is phrased with regard to the literature. First, it narrowly focuses on the systems memory consolidation theory, but then key aspect, such as the dependence of both spindles and ripples on slow wave activity are omitted. How do the reported effects relate to the orchestration of ripples and spindles by the underlying SO phase (for related intracranial work see Staresina 2015/2023 *Nature Neuro* or Helfrich et al., 2019 *Nature Commun.*). It should also be highlighted that systems memory consolidation is a theory, but it is written very much as a matter-of-fact

The authors include a lengthy discussion on the comparison of human vs. rodent ripples and cite evidence from Buzsaki et al., but omit more recent work, such as the consensus paper by Liu et al. (2022) in *Nature Commun* or direct comparative evidence by e.g., Schalkwijk et al. (2023) *Progress Neurobiol*.

The authors also describe a set of finding that is rather in accordance with non-oscillatory broadband effects (e.g., 16-45 Hz), which are reminiscent of the aperiodic changes observed during wakefulness (Donoghue et al., 2020 *NatureNeuro*) and sleep (Ameen et al., 2024, *bioRxiv*, Lendner et al., 2020, *eLife*) and might be distinct from band-limited activity, in particular during sleep (Bodizs et al. 2024 *Progr Neurobiol*).

Likewise, some of the similarity effects are likely confounded, given that ripple activity is correlated with high frequency activity that overlaps in its spectral content. In other words, similarity analyses encompass the broadband signal (50-140 Hz) that also contains the ripple itself. How would the results look like if the ripple-band is omitted from the subsequent population analyses, e.g., by only focusing on the 100-140 Hz range? The same concerns hold true for Fig 5 and the reported coherence effects, which are again broadband in nature, and cannot be interpreted if power is also changing.

Per the reported ripple intervals – what is the min. inter-ripple interval, in its present form Fig2E indicates that these follow in succession at very short delays? How do the authors explain the mismatch between the number of spindles and ripples, spindles exceed by a factor of ~10, while their key result in Fig. 4F/G describes a very similar 'normalized' rate? It'd be more helpful to see the absolute rate to judge the true behavioral relevance.

The writing is sub-optimal in several instances and not very precise. For example, in the abstract very general terms are used ('much remains to be learned') or 'to address this issue' (which is not precisely defined), or unclear ('TMR produces nonuniform amounts') or in the results 'Patients ... were employed'

Minor:

- It would be desirable to have more information about the precise experimental timing. Given the clinical realities, this was probably rather variable when training and testing took place
- Fig. 2F should include a baseline period prior 0 to fully evaluate the temporal evolution of the described effect
- Fig. 3C constitutes double-dipping, since the statistical effect has already been established in Fig 3B

Reviewer #3

(Remarks to the Author)

The topic of the paper is highly relevant and important. The authors can be congratulated for attempting to conduct a challenging cognitive paradigm that requires multiple testing sessions and a monitored sleep session in a bedside setting. Not only did they study and report the behavioral effects of the nap-intervention experiment, but also aimed to understand the electrophysiological signatures of these effects. These signatures are highly relevant to our understanding of how TMR influences sleep-associated declarative memory consolidation and how this intervention can uncover natural mechanisms of memory consolidation during sleep. However, I do have some concerns regarding the data analysis and interpretations that I detailed below, and I hope the authors would be able to answer to allow the reader to judge their work.

Major

Subjects learned to associate 50 unique items with specific locations on a grid, each paired with a characteristic sound. The authors report a weak memory-improvement effect for TMR, which is expected based on the small cohort that iEEG research enables and also the difficulty of controlling all the experiment's parameters in a bedside setting. That said, since the experiment is performed out of the lab, it is critical to report the details clearly and acknowledge the inherent variability (i.e. auditory cues delivered in NREM-1-3, the sleep phases on which they were delivered etc.).

TMR literature, performed based on non-invasive EEG now covers a great deal of options that vary the efficacy of the intervention, one of the most notable are the differences of the intervention when targeting up/down states of slow wave oscillation (for example Ngo, PNAS 2022) and different relationships regarding timing of cues relative to spindle occurrence (Antony, Curr Biol., 2018). I believe the manuscript will benefit from a breakdown of the timings of auditory cues during sleep relative to sleep oscillations, as these can make a great difference concerning the intervention's effect on memory, and also – regarding the electrophysiological analysis below – these will impact the frequency profile that is expected after auditory cueing that will determine the expected baseline of sleep oscillations (spindles and ripples) during the re-activation periods (Fig. 4 F-G). There is also variability in the sleep-period the cues are delivered on, it would make sense to focus on auditory cues that are delivered on the same sleep-phase when conducting the analysis regarding spindle and ripple rates (Fig. 4 F-G) as their baseline probability differs between different sleep-phases.

Strengthening and decaying sub-group are defined per subject, such that a fixed number of items are assigned to each sub-group. It is important to point out that there is a significant difference between the performance of these sub-groups before sleep and that without proper control, these differences can be the driving force of later effects. This point is laid out thoroughly in the discussion section and a control analysis was added as a supplementary figure.

Item-specific cortical pattern around hippocampal ripples during learning:

[row 200-210] The authors investigated whether item-specific neural representations can be detected around the occurrence of ripples during the retrieval phase. The natural question is whether these specific representations are unique to ripple-locked analysis or are these a limited window into a process of re-instantiation that is related to the time-period following the stimuli exposure but not unique to the selected ripple windows. It would be helpful to perform the same analysis on ripple-less epochs to understand whether this item-specific similarity arises specifically around ripple times. Since a wide-time window is used, and ripple-rate peaks around 0.75 after stimuli presentation, this analysis correlates the post-presentation epochs between items. Can the effect (Fig. 2E) can be alternatively attributed to differences in memory-processing, i.e. – whether the baseline pre-sleep error of the item is high or low? I think the item-specific title is misleading because the analysis lacks the effort to characterize whether a minority of items that have a similar neural signature (for example – those that were learned efficiently or those that patients found very interesting) dominate the effect, while most of the others do not. Without a proper control or deeper analysis of whether the average represents the distribution of WI-BI matrices (via a shuffling procedure), the results suggest weak evidence for enhanced similarity for same-item representations that are specific to ripple periods following stimuli presentation.

It would also be helpful to understand whether all learning rounds and all cortical contacts (gray/white matter) were included in this analysis? Some contacts seem to be outside the brain in the figure (Fig. 1G).

Two behavioral-relevant item reactivations during sleep:

The authors used an analysis approach that searches for similarity between neural activity that arises during ripple activity during learning to that that arises during auditory cueing during sleep and then asks whether this similarity is predictive of memory consolidation. While this is a fundamental question that is interesting and important to investigate, I am not convinced that the methods chosen can lead to the suggested conclusions. High-frequency ripples are brief oscillatory events, but here the analysis is chosen using a wide window – [-0.5 0.5]s around ripples, making the reader wonder whether the activity is related to ripple occurrence at all. One would wonder if an analysis that correlates the activity time-locked to image presentation during learning would reveal similar results or is this a property of ripple-occurrence? These time periods may be confounded as ripples seem to be occurring more at [0.5-1]s after onset during learning (Fig. 2F) so important to acknowledge this confounding factor. This concern is backed up by the subject-specific reactivation plots presented in Figure S-4 – the rightmost cluster (Fig. 3B) seems to be common to 4/11 patients and the left cluster to 3/11 so there's quite a large variability in the specificity of the time-window around ripples and after cue-presentation that isn't addressed in the main analysis presented. As Figure S-5 reveals – 7/11 patients' similarity matrices do not align with the suggested process in the main text – namely, they either have a single reactivation cluster, or very small clusters that may be due to spurious stats.

Is the double-reactivation unique to patients whose memory benefited from auditory-cueing (those with a larger difference between cued and un-cued items, Fig 2L)? For transparency, it is important to acknowledge these limitations in the main text where these analyses are mentioned but not described (line 225).

Fig 4 D-E – Here the authors test their hypothesis by dividing the analysis to two sub-groups of cued items – those that benefited from cuing (strengthened memory) and those in which memory decayed. They compare the double-reactivation correlation maps between these two groups and show a significant difference. My concern is that these groups' differences are predetermined before the TMR intervention. Based on Figure S1 – there is a significant difference between the pre-sleep performance of these two item groups so perhaps there are additional differences between them (visual attributes, emotional valence, etc.) that contribute to the reported effects. What are the differences in the timing of the auditory cues between these groups? Can the authors plot the frequency profiles time-locked to auditory cues resulting in enhanced/degraded cues and comment on how the differences between these profiles contribute to the presented results? These differences can account for the differences between ripple rate and spindle rates (Fig. 4 F-G) as their baseline rates vary between the up/down phases of the slow oscillations.

The definition of the control condition of using non-learning-ripple trial data (awake TMR) was unclear.

Dynamic hipp-cortical synchronization during two reactivations:

Continuing my comment for the previous section – as previous literature has identified dynamic changes in the level of coordinated activity between neocortical and hippocampal areas during sleep (Helfrich et al, Nat. Commun 2019, Sanda et al, Cereb Cortex, 2021, and others) it is very important to characterize the baseline coherence of cue-less sleep-periods which are matched in their distance from slow-oscillation (or separated based on sleep-phases in which they were delivered) to the cued-periods used in the coherence analysis to understand the baseline without cueing. The null hypothesis would be that the timing of the cue is the driving force for most of the results presented in Fig. 5.

Minor

Fig 1A – it is extremely difficult to understand the locations of contacts based on the current panel's size.

Reference to Fig. 4A-C appears in text after discussing Fig. 4D-E. Fig. 4D-E relates to analysis in Fig. 3 rather than in Fig. 4 so is a bit confusing for reader to find.

Fig S-6 legend – the authors refer to awake TMR – but write that sounds were cued during NREM sleep.

Fig. 4C – spindle activity is usually paired with slow-oscillations, and the fact that the slow wave activity is very low here makes the reader wonder maybe the majority of the spindles are detected in early NREM sleep or mixed with IED activity?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

NCOMMS-24-07019A

The authors responded to all of my comments, which is appreciated. However, the authors responded to many of my comments without including the information in the manuscript (or supplements) which I do not find helpful. The point of these comments is to provide more detailed information to potential readers, not only the reviewer. Therefore, please include Response Figures 1, 2, 4, 6, 12, 15, 16, 20, 22, 23, 24 with relevant details on analyses statistics as well as all of the Response Tables in the manuscript /supplements and refer to them in the main text.

The following responses by the authors need further clarification:

Response to 1.2:

Rather than general differences in laboratory vs. bedside setups, I was interested in how the inconsistent patterns on the single patient level can result in a consistent group-level effect which remains unclear to me.

Response to 2.1

It seems that ripples beyond 120 Hz contribute very little to the results, but the manuscript does not reflect this, since it features a ripple band up to 180 Hz. This could mislead readers into assuming that these high frequency ripples are important here. I suggest to change the main analyses to 70-120 Hz (the author's response to 2.4 supports this view).

Response to 2.2:

I do not agree with the authors line of arguments here.

A duration < 42 ms cannot be translated into a minimum number of cycles at different frequencies. Instead, a duration of < 42 ms only ensures that relatively low frequency ripples (e.g., around 70 Hz) will have less than 3 cycles but high frequency ripples will have more cycles: a ripple at 170 Hz will have approx. 7 cycles in 42 ms. Since these fast ripples with many cycles contribute to the left panels of Resp. Fig. 5, this might explain why there seem to be multiple oscillations.

Moreover, how do the authors explain that the left panel shows approx. equal amounts of oscillations in the grand average as the right panel, even though the right panel is supposed to include many more "long ripples"? These "long ripples" at relatively low frequencies should result in longer oscillations the grand average in the right panel. In the left panel, the oscillation in the grand average covers a time window (between 100 to 150 ms) that is way larger than +/- 21 ms. How can

this be the case, if this is a ripple-centered average and no ripple has a duration longer than 42 ms?

I think that including only ripples with more than 3 cycles would improve reliability of ripple detection and would thus ask the authors to report their main analyses using this criterion.

Response to 2.3:

I asked for a power spectrum (no time information), not a time-frequency plot, and whether there are peaks in this power spectrum, in particular in the ripple band, but also for spindles and SOs. My apologies if this was not clear in the previous round.

Response to 2.5:

The authors detected all ripples in all hippocampus contacts, as requested. However, they did not repeat the main analyses based on all ripples in all contacts. I maintain the opinion that this analysis would provide more robust results than basing their findings on one selected hippocampus contact.

Response to 2.7:

The authors show epileptiform spike-locked spectrograms. However, this is unfortunately not what I asked for. Instead, I asked for epileptiform spike-locked averages or "Grand average peri-ripple field potentials", equal to what the authors show in Fig. 2C or 4B, left.

Response to 2.9:

No REM ripple rates provided. I still think it would make the manuscript more robust if the authors would try their ripple detection in other states such as REM and see whether their detected rates match previous studies. In particular, since the ripple rates in wake and NREM are very similar, which is not what has been reported in previous studies.

Why would the authors provide ripple rates in events/minute and not events/second, which seems to be more common?

Please provide the average ripple/spindle rates for each column in Table S3.

The authors report similar ripple rates in wake and NREM. Previous studies found higher ripple rates in NREM than wake, e.g., Dickey et al 2022; Schreiner et al., 2024. How do the authors explain this discrepancy?

Response to 3.1:

What is the increase in power at time = 0 in the time-frequency plots in Resp. Fig. 13 that spans all frequencies from 10 Hz and higher? Please show the same plots including a pre-TMR interval of 1 second.

Response to 5.2:

I cannot find the "added detailed descriptions" about the balancing of trial numbers in the methods section.

In general, it would help a lot if the authors would indicate the line numbers when referring to changes in the updated manuscript.

-remove "obviously" from

"But it should be pointed out that the intensity of the difference was obviously lower than the subsequent second reactivation cluster."

-A recent paper in Nature Communications by Schreiner et al. seems highly relevant for this: Spindle-locked ripples mediate memory reactivation during human NREM sleep.

The authors could discuss overlap / differences in their findings.

Reviewer #2

(Remarks to the Author)

The authors addressed all queries in detail, which further strengthened the manuscript and its conclusions. In sum, this might be published as it stands.

Reviewer #3

(Remarks to the Author)

I appreciate the authors' time and effort in answering every question in detail and providing additional analysis and text revisions to support their answers.

Thank you for expanding Table S3 and adding the stats of the TMR cues timing regarding different sleep stages which will assist others who are attempting to reproduce the experimental procedure. I do have a few open questions that were not answered by the authors and I hope they can provide additional answers. Some critical answers were kept as answers to reviewers while they should be added to the text to provide a complete picture to the reader. I list these below and believe they can be clarified quite easily in the text -

1. Item-level cortical pattern around hippocampal ripples in learning stage

The provided analysis did not adequately answer my question regarding the existence of sub-groups of items dominating this analysis, rather than a unique signature per item. The adjusted title still states that it is the identity of the item, rather than its learned/unlearned status that determines the cortical pattern.

Some of the analysis may be misleading because it groups subgroups that have differing trends. For example - Figure 2F displays an increase in ripple rate after cue onset, while response-figure 30 very strongly demonstrates that this result is driven by the correctly encoded items (based on pre-testing) while the poorly learned ones do not reflect an increased rate of

ripples. This leads to an alternative hypothesis that would explain the results - efficiently learned items have a unique neural signature during learning and averaging this group with the unlearned group masks this pattern.

-- response figure 30 is important to add to the supplementary data to complement Figure S7a because it demonstrates that ripple rate during learning is indeed modulated by the subsequent memory effect.

-- A control to contest the hypothesis that the representation is item-specific, is to compare the WI and BI of all strengthened items (completely removing the decayed ones from analysis) to confirm that the pattern does not reflect a successful operation of the memory-network, but rather a specific pattern per item. Response Figure 31, is breaking down Fig. 2L to items and does not address my concern.

2. Figure S1B demonstrates what looks like a significant difference in the error during the pre-sleep stage between the strengthened and the decayed group (blue line). Please add the stats of this comparison to the paper because it is critical to assess whether there is a difference between items assigned to these two groups that are not related to the TMR intervention. The authors need to acknowledge this difference and discuss how it may impact their analysis.

3. In the rebuttal letter, I read this addition but could not find it in the main text -

In addition, in the revised manuscript, we rephrased the definition of control condition as follows: "To verify the relationship between the learning ripples and subsequent reactivations during sleep, we selected trials without ripple events during the learning (149 ± 62 no-ripple trials) and investigated the reactivation of their patterns during sleep. For example, there were two learning trials of one item. There was a ripple event at 0.75 second after the image onset in trial 1, while there was no ripple in trial 2. We mark a fake ripple event for trial 2 at 0.75 second after the image onset. Then we repeated our RSA analysis just using trials of these fake ripple events." The current description of the methods to perform the control plotted in Figure S6A legend is not sufficient for the reader to understand how the control analysis was performed, nor reproduce it. Also – because the result is so close to significance ($p=0.058$ based on the rebuttal letter), it's important to provide the full statistics, be transparent about the result, and state that it follows the same trend as ripple-triggered analysis. Would be recommended to add stats for Figure S6B as well.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Remaining questions and comments:

Authors' response to 2.1:

Please include details on how Figure S10 was made. For example, what method was used to remove fractal components? On what data segments, ect.?

I assume that all the code will be made publicly available, not only the code for the main figures?

Authors' response to 2.3:

Please include Response Figure 2 in the manuscript / supplements.

The "peak" with the highest frequency seems not to be in the ripple band, but below 70 Hz. Unfortunately, the authors do not provide the frequency of the "peaks", but looking at the plot tells me the "peak" is at approx. 65 Hz. This is outside the ripple band in their main analyses (70 to 180 Hz), as well as outside of ripple band in the additional analyses (80-120 Hz). How do the authors explain this? Does the "peak" at 65 Hz still reflect activity in the ripples band and, if so, why would the frequency be different? Alternatively, there is no peak in the power spectrum that reflects the activity of the ripples, and if so, why would that be?

Authors' response to 2.5:

How do the authors explain the apparent discrepancies between Response Figure 4 and main Figure 3B regarding the y-axes? In Response Figure 4, activity before the ripples seems to be most relevant but not in Fig. 3B.

Authors' response to 2.9:

- "Leaving the hospital" is not a sufficient argument for not analyzing REM data. Was REM data not recorded? This would be very unusual, since patients are in the hospital to record their intracranial EEG during epileptic seizures.

- Schreiner et al. (2024) show ripple rates during NREM / TMR that are clearly higher than 0.19 (Figure 4d) and also clearly higher than the ripple rate during wake.

Reviewer #3

(Remarks to the Author)

The authors addressed all queries in detail, added details to the main text and the requested explanations to the supplementary material which allow the reader to better understand the analysis and results. They are clearly stating which hypotheses are met and which aren't which I appreciate. In sum, I support publishing the manuscript in its current state.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response letter to comments

Reviewer #1

The authors present an interesting study that uses targeted memory reactivation in epilepsy patients to investigate electrophysiological activity related to memory consolidation. The authors aim at relating hippocampal ripples to connectivity between the hippocampus and neocortex, which promises to provide novel insights into the neural foundations of memory processes and is potentially interesting to a large audience. Overall, the paper is well-written and the results are described clearly. However, the authors need to address several critical issues that question the robustness of some of the major findings. These and other issues are detailed below:

We appreciate the reviewer's interests in our study and the valuable suggestions on control analyses to strengthen our conclusions. We have diligently addressed all concerns, and the supplementary results further support our main findings. Below, please find our detailed responses to each question.

Major comments

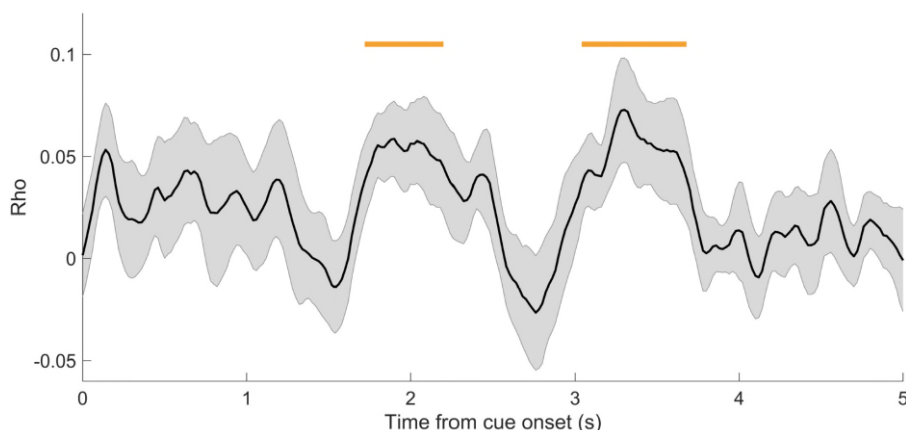
1. Reactivation time windows:

1.1 Figure 3B: Are these two separate time windows? They are rather close in time and might just appear separate because of statistical thresholds. I am not convinced that the permutation test procedure is designed to evaluate more than one cluster. How does the procedure treat second clusters? Does it compare those clusters to the same distribution as first clusters? If so, is this justified, given that second clusters are not independent of first clusters? More generally, the test procedure tests whether two conditions are different and not when (in time) these conditions differ. Inferences on what drives these differences would need to be inspected separately, for example on the basis of grand averages. Do grand-averages reflect two clusters? A similar criticism applies to the results reported Figure 4E.

A: In the cluster-based permutation test, we calculate the summed T-values of the largest cluster after each shuffle, and then compare the real data clusters size with the shuffled clusters distribution (all real data clusters were sorted in the same distribution). This procedure also can correct multiple comparisons (Zhang et al., 2018). Previous studies have used the same (Zhang et al., 2018) or similar (Schreiner et al., 2018) methods for statistical analysis and affirmed the multiple occurrences of reactivation. In our study, we are searching for clusters in a two-

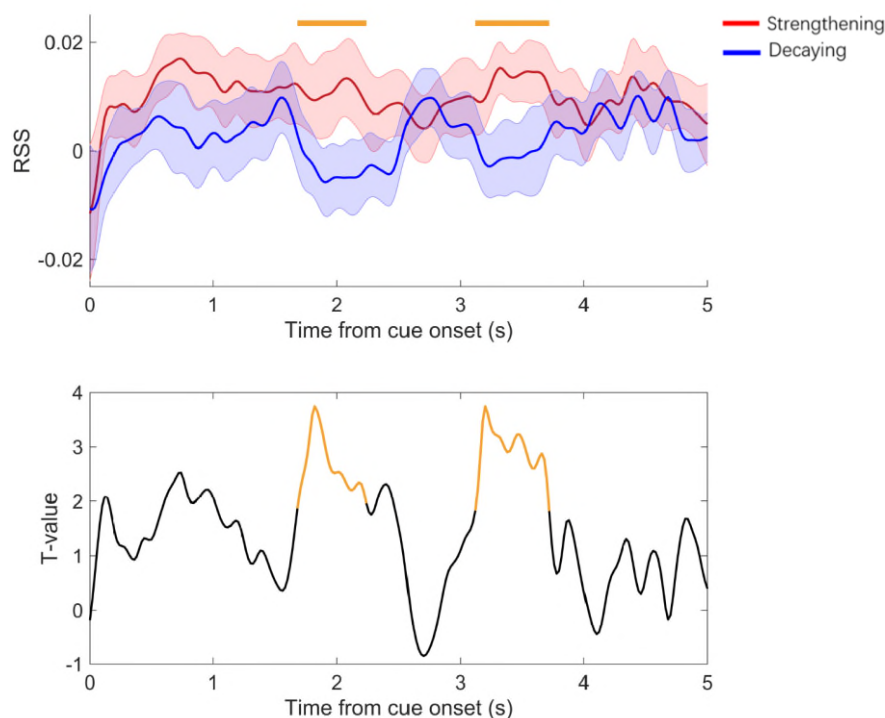
dimensional matrix. Since the x-axis represents the time after TMR cues onset, we believe that it is possible to determine when they are significant.

In Figure 3, we calculated the Spearman correlation between reactivation matrix and behavioral performance, and took the average values across rounds and subjects. It means that the analysis is actually based on the grand average value. To demonstrate the dynamic changes more clearly, here we compressed the matrix in Figure 3B into a vector (averaged over y-axis [-0.3s to +0.3s]) and performed permutation tests. Then we can obviously see the temporal dynamic of the grand average value in the Response Figure 1. As shown in the figure, the grand-average value also reflects two clusters (cluster1: 1740ms to 2220ms; cluster2: 3060ms to 3700ms) in the similar time windows seen as Figure 3B.



Response Figure 1. The Spearman correlation (Rho) between RSS and memory change scores. Two significant clusters emerged after the cluster-based permutation test ($p = 0.043$, $p = 0.024$, from left to right, yellow lines marked).

Figure 4E (Figure 4B in the revised manuscript) shows the difference of reactivation between strengthening and decaying items averaged across participants. If we plot the grand-average of the two conditions separately (strengthening and decaying), we can see obviously the temporal dynamic of the grand average value. As shown in the Response Figure 2, the grand average value of the two conditions is significant difference in two separate time periods (cluster 1: 1700ms to 2260ms; cluster 2: 3140ms to 3740ms). It is worth noting that our first cluster is very similar to the cluster shown in Cairney et al.'s study (Cairney et al., 2018). However, as Cairney et al.'s data was only analyzed for 2.5 seconds, the second reactivation cluster cannot be observed in their data.



Response Figure 2. Upper: Retrieval-Sleep Similarity (RSS) in strengthening and decaying conditions. Yellow lines mark the significant clusters after cluster-based permutation test ($p = 0.030$, $p = 0.019$, from left to right). Below: the test statistics of the RRS between strengthening and decaying conditions.

References:

- Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat Commun*, 9(1), 4103. <https://doi.org/10.1038/s41467-018-06553-y>
- Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Rep*, 25(2), 296-301. <https://doi.org/10.1016/j.celrep.2018.09.037>
- Cairney, S. A., Guttesen, A. A. V., El Marj, N., & Staresina, B. P. (2018). Memory Consolidation Is Linked to Spindle-Mediated Information Processing during Sleep. *Curr Biol*, 28(6), 948-954 e944. <https://doi.org/10.1016/j.cub.2018.01.087>

1.2 In Figure S4, it does not look like there are two clusters per subject and the timing of the two largest clusters per subject does not correspond to the two time bins reported in Fig. 4E, in particular with respect to the timing on the x-axis. There seems to be a discrepancy between group level and single subject results which the authors would need to explain. This also questions the statement in the discussion (l. 362) that the time-frequency analysis found consistent results.

A: Thank you for the careful inspection. For intracranial EEG recordings, there are large differences in the number and position of electrodes across patients, which certainly limits the stability of RSA analysis. In addition, the uncontrolled non-laboratory environment also amplified the variation across patients. When conducting experiments in the hospital bedside in this study, we did not control factors such as the phases (up/down states) of the SO when the TMR cues were delivered. The above factors could affect the stability of RSA analysis at the participant-level in intracranial EEG research. Therefore, previous iEEG studies generally only showed group-level RSA results. It is worth noting that the present study found a stable reactivation pattern at the round level. This also supports the stability of the reactivation pattern (two clusters). In response to the above limitations, we have added a “limitations” section in the discussion of the revised manuscript. In future research, we will attempt to implement stricter controls, such as the closed-loop TMR intervention method.

In addition, the description of this sentence “the time-frequency analysis found consistent results.” It was indeed unclear as the reviewer pointed out. If only looking at the time-frequency analysis results (see Figure S8), our study found difference in spindle activity during the first reactivation period. However, it should be noted that the intensity of the difference is obviously lower than the subsequent second reactivation cluster. This result indicated that the changing trend of spindle activity differences (strengthening>decaying) was consistent with the results of spindle events detection. We have made the following modifications (Page 7, line 271 - 275; Page 10, line 375 - 378):

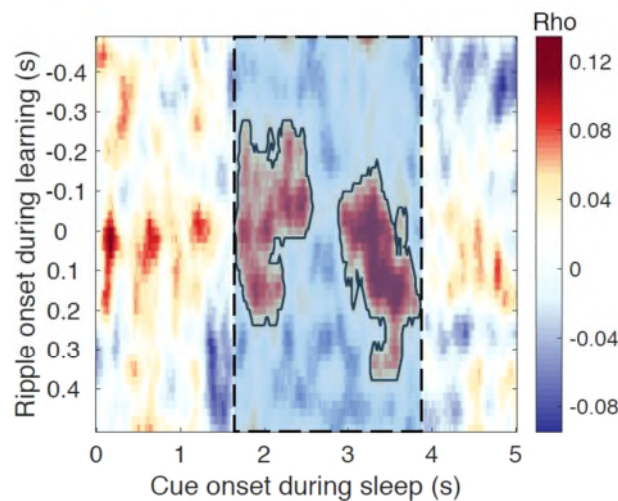
“When comparing cortical spectrograms between the two conditions, we found difference in spindle activity during the first reactivation period. But it should be pointed out that the intensity of the difference was obviously lower than the subsequent second reactivation cluster. This result suggests that the changing trend of spindle activity differences (strengthening>decaying) is consistent with the result of spindle events detection.”

“In addition, the time-frequency analysis found similar trends with the spindle detection analysis. The spindle activity differences (strengthening>decaying) during the first reactivation was obviously lower than the subsequent second reactivation.”

1.3 As far as I understand, the control analyses depicted in Fig. S4 and S5 were done by correlating the whole matrix coming from single round/subject data with the mean (excluding the round/subject in question). I do not think this strongly supports the stability of the relevant patterns (the temporal clusters), because the correlations might reflect overall similarities between the patterns across all time points and not specifically reflect the two clusters in question.

A: Thanks for this good suggestion. We completely agree with the reviewer's viewpoint that using the whole matrix to calculate correlation is indeed inappropriate. We apologize for the unclear description of the analysis method. Instead of using the whole matrix, actually we used the data from the beginning of the first cluster (1.6s after cue) to the end of the second cluster (3.8s after cue) to calculate the similarity between the matrices, as shown in Response Figure 3. We have added this information in the supplementary materials (Page 33, lines 1099 - 1100) as follows:

“The red circles in each panel denote the actual correlation between the Rho matrix (from 1.6s to 3.8s after TMR cues) of single subject and the matrix of main result (excluding the corresponding subject).”



Response Figure 3. The black box marks the time window selected for calculating similarity between single round/subject and the matrix of main result.

1.4 The reactivation windows are rather late, in particular the second time window which also displays differences in spindle power. Usually, these differences are found earlier (e.g., Cairney et al.). How do the authors explain this delayed pattern? The presented results seem to be somewhat at odds with some previous findings (e.g., Cairney et al.), which should be discussed in more detail. Cairney et al. find reactivation connected to spindles in a similar time window to what is reported here as first reactivation. However, there is no relation to spindles in the present study.

A: Firstly, previous studies have shown differences in the number of reactivations. Cairney et al. found one reactivation around 2 seconds (Cairney et al., 2018). It should be noted that they only analyzed 2.5s after the TMR cue. On the other hand, Schreiner et al. analyzed a wider time window after TMR cues and found three reactivation clusters: the first peak at 390ms, the second peak at 1990ms and the third peak at 3310ms (Schreiner et al., 2018). In a recent review, Lewis and Bendor speculated that the first reactivation directly after TMR may be driven by sensory stimuli, while subsequent reactivations (i.e., echoes) may be hippocampal driven (Lewis & Bendor, 2019). In the present study, we found two reactivated clusters, the first cluster located in 1.6-2.6s, and the second cluster located in 2.8-3.8s. They are very similar to the second and third clusters shown in Schreiner et al.'s study. In addition, we also observed a trend of early reactivation (before 1s, see Figure 4E), but it did not reach statistical significance. Overall, the second cluster observed in the present study may correspond to the third cluster discovered in Schreiner et al.'s study, which indeed was a late reactivation.

Secondly, we would clarify the spindle activity from two aspects during the reactivation period induced by TMR. On the one hand, actually the findings of previous studies are inconsistent. Although Cairney et al. (Cairney et al., 2018) found significant differences in spindle activity during the second reactivation (around 2s), they did not analyze spindle activity over a larger time window (up to 2.5s). Therefore, it was impossible to determine whether the spindle activity occurred during a later reactivation process. However, Liu et al. discovered later differences in spindle activity related to memory consolidation effects (Liu et al., 2023). On the other hand, if only looking at the time-frequency analysis results, our study also found differences in spindle activity during the first reactivation period (see Figure S8, similar to the

cluster discovered by Cairney et al., 2018), although the intensity of spindle power is lower than the subsequent second reactivation cluster. Since the description of time-frequency analysis results in the original manuscript was unclear, we have made the corresponding modifications in the revised manuscript (Page 7, line 271 - 275; Page 10, line 375 - 378):

“When comparing cortical spectrograms between the two conditions, we found difference in spindle activity during the first reactivation period. But it should be pointed out that the intensity of the difference was obviously lower than the subsequent second reactivation cluster. This result suggests that the changing trend of spindle activity differences (strengthening>decaying) is consistent with the result of spindle events detection.”

“In addition, the time-frequency analysis found similar trends with the spindle detection analysis. The spindle activity differences (strengthening>decaying) during the first reactivation is obviously lower than the subsequent second reactivation.”

References:

- Cairney, S. A., Guttesen, A. A. V., El Marj, N., & Staresina, B. P. (2018). Memory Consolidation Is Linked to Spindle-Mediated Information Processing during Sleep. *Curr Biol*, 28(6), 948-954 e944. <https://doi.org/10.1016/j.cub.2018.01.087>
- Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Rep*, 25(2), 296-301. <https://doi.org/10.1016/j.celrep.2018.09.037>
- Lewis, P. A., & Bendor, D. (2019). How Targeted Memory Reactivation Promotes the Selective Strengthening of Memories in Sleep. *Curr Biol*, 29(18), R906-R912. <https://doi.org/10.1016/j.cub.2019.08.019>
- Liu, J., Xia, T., Chen, D., Yao, Z., Zhu, M., Antony, J. W., Lee, T. M. C., & Hu, X. (2023). Item-specific neural representations during human sleep support long-term memory. *PLoS Biol*, 21(11), e3002399. <https://doi.org/10.1371/journal.pbio.3002399>

Ripples: 2.1 The authors filtered the LFP in a very broad band, between 70-180 Hz. This is very uncommon, usually more narrow bands are used, e.g., 80-120 Hz (see Vaz et al., 2019). Why would the authors choose such a broad band? Please provide a description of the distribution of ripples across frequencies, e.g. a histogram that shows ripple counts per 10 Hz. Alternatively,

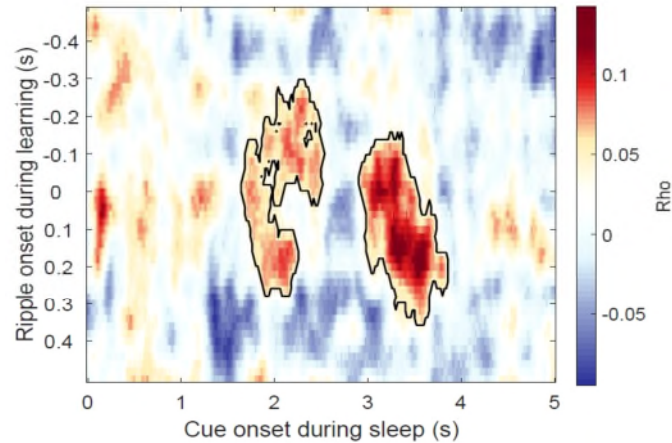
the authors could repeat their analyses with ripples filtered in a narrower, more commonly used frequency band, e.g., 80-120 Hz, to make results comparable to previous studies.

A: According to the reviewer's great suggestion, here we provided a description of the distribution of the detected ripples across frequencies (a histogram that shows ripple counts, see Response Table 1). The results showed that the majority of ripples were in the narrow bands (< 120Hz). In addition, when we repeated the analyses only using those low frequency ripples (< 120 Hz), the RSA result was very consistent with the previous result (see Response Figure 4).

Furthermore, it should be clarified that there is currently no unified standard for detecting ripples. Recently Liu et al. reviewed the studies on detecting ripples using primate intracranial recordings (Liu et al., 2022). Among the nine studies that they reviewed, six of them used a broad band (e.g., 95-180 Hz), while the other three articles used narrow band (i.e., 80-120 Hz).

Response Table 1. *The distribution of ripples across frequencies.*

Ripple peak frequency	Proportion (TMR)	Proportion (Learning)
70-80Hz	40.6%	43.4%
80-90 Hz	27.8%	28.4%
90-100 Hz	14.3%	12.6%
100-110 Hz	5.4%	4.3%
110-120 Hz	3.8%	4.1%
>120 Hz	8.1%	7.2%



Response Figure 4. The average Rho across subjects and rounds (exclude those ripples with greater than 120 Hz).

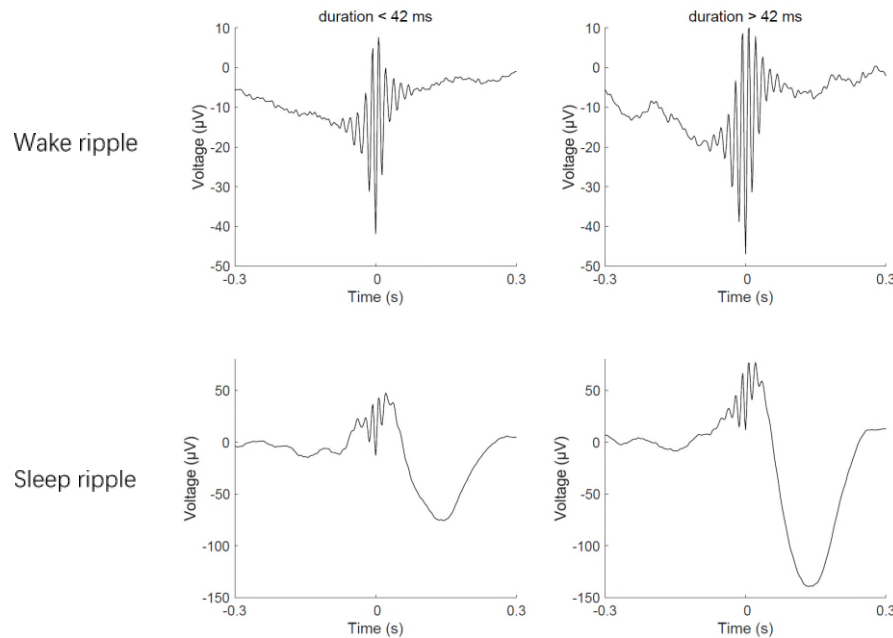
References:

Liu, A. A., Henin, S., Abbaspoor, S., Bragin, A., Buffalo, E. A., Farrell, J. S., Foster, D. J., Frank, L. M., Gedankien, T., Gotman, J., Guidera, J. A., Hoffman, K. L., Jacobs, J., Kahana, M. J., Li, L., Liao, Z., Lin, J. J., Losonczy, A., Malach, R., van der Meer, M. A., McClain, K., McNaughton, B. L., Norman, Y., Navas-Olive, A., de la Prida, L. M., Rueckemann, J. W., Sakon, J. J., Skelin, I., Soltesz, I., Staresina, B. P., Weiss, S. A., Wilson, M. A., Zaghoul, K. A., Zugaro, M., & Buzsáki, G. (2022). A consensus statement on detection of hippocampal sharp wave ripples and differentiation from other fast oscillations. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-33536-x>

2.2 The authors should include a minimum of 3-4 cycles in the raw signal as a criterion for ripple detection. The duration criterion of min. 20 ms is insufficient to detect ripples that display sufficient numbers of cycles, which makes this method prone to include sharp spikes as they occur during epileptic activity.

A: Thanks for your careful inspection. What we need to clarify is that there is currently no unified standards for detecting ripples. Liu et al. reviewed the researches on detecting ripples using primate intracranial EEG (Liu et al., 2022). Based on that review, we can find that the minimum duration is generally 20 or 25ms (6/9 studies in Liu's review paper). In addition, to prevent the influence of epileptic activity, we have preprocessed the iEEG data using the ISARG (Intracranial Signal Analysis Research Group) method (Janca et al., 2015), and inter-ictal epileptic discharges have been detected and removed before further analysis. In addition,

based on the frequency band used to detect ripple (70-180 Hz), we showed the grand average of ripples that less than 3 cycles (duration < 42 milliseconds) and those more than 3 cycles (duration > 42 milliseconds), respectively in Response Figure 5. It can be seen that those shorter ripples still have obvious multiple oscillations, rather than a sharp spike.



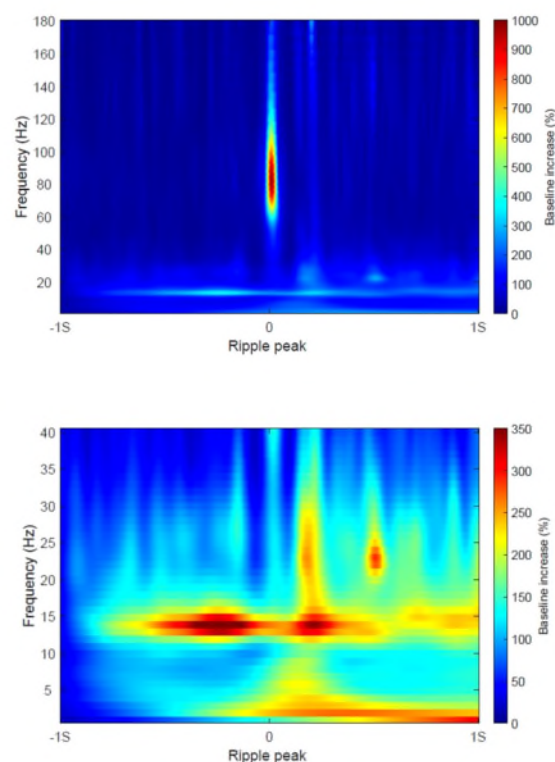
Response Figure 5. Grand average peri-ripple field potential for the shorter ripples (left) and longer ripples (right) during wake and sleep. 65.2% of ripples are shorter than 42 milliseconds during sleep and 81.4% of ripples are shorter than 42 milliseconds during wake.

References:

- Janca, R., Jezdik, P., Cmejla, R., Tomasek, M., Worrell, G. A., Stead, M., Wagenaar, J., Jefferys, J. G., Krsek, P., Komarek, V., Jiruska, P., & Marusic, P. (2015). Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. *Brain Topogr*, 28(1), 172-183. <https://doi.org/10.1007/s10548-014-0379-1>
- Liu, A. A., Henin, S., Abbaspoor, S., Bragin, A., Buffalo, E. A., Farrell, J. S., Foster, D. J., Frank, L. M., Gedankien, T., Gotman, J., Guidera, J. A., Hoffman, K. L., Jacobs, J., Kahana, M. J., Li, L., Liao, Z., Lin, J. J., Losonczy, A., Malach, R., van der Meer, M. A., McClain, K., McNaughton, B. L., Norman, Y., Navas-Olive, A., de la Prida, L. M., Rueckemann, J. W., Sakon, J. J., Skelin, I., Soltesz, I., Staresina, B. P., Weiss, S. A., Wilson, M. A., Zaghoul, K. A., Zugaro, M., & Buzsáki, G. (2022). A consensus statement on detection of hippocampal sharp wave ripples and differentiation from other fast oscillations. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-33536-x>

2.3 Please provide a power spectrum of epochs around all detected ripple events (e.g., ripple +/- 500 ms) that covers SO, spindle and ripple frequencies. Are there distinct peaks in the power spectrum for SO, spindles, and also in the ripple band?

A: According to such constructive suggestion, we show the spectrograms of epochs around all detected ripple events that covers SO, spindle and ripple frequencies. Similar to previous research (Staresina et al., 2015), time-locked TFRs of all epochs were normalized as percent change from pre-event baseline (−1.5s to −1 s for ripples) and averaged over participants. As shown in the Response Figure 6, there are clearly distinct peaks in the power spectrum for SO, spindles, and ripple band.



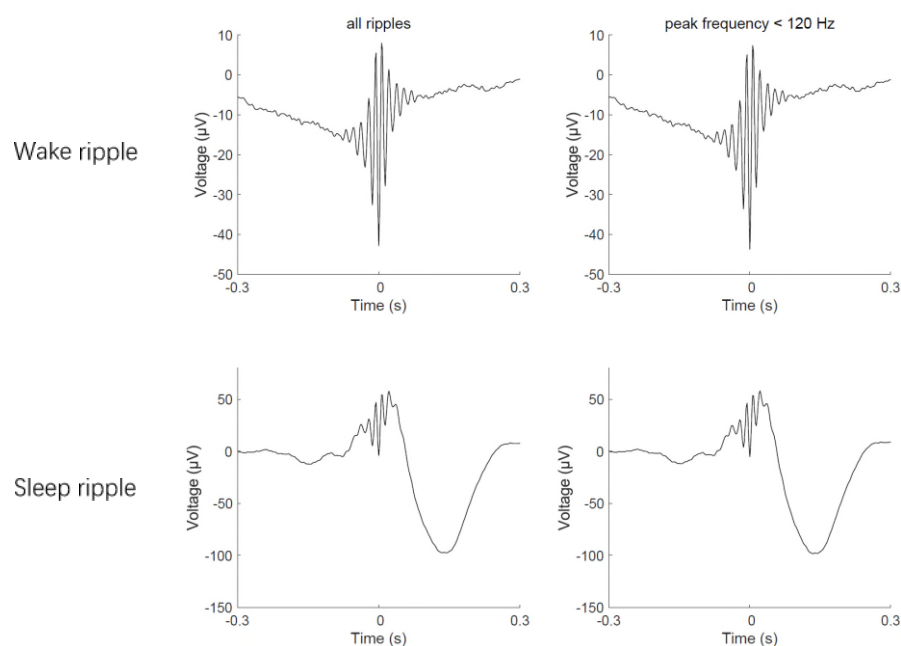
Response Figure 6. Average of ripple peak–locked TFR (percentage change from pre-event baseline). The bottom figure narrows the range of the colorbar and focuses on 1-40 Hz to highlight the nesting of spindle/SO power around the ripple peak.

References:

Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Rep*, 25(2), 296-301. <https://doi.org/10.1016/j.celrep.2018.09.037>

2.4 Related to this issue, why does the ripple average (Fig. 2C, 4B) display such clear cycles (of a particular frequency?) if ripples were detected between 70 and 180 Hz? Averaging across distinct ripples with a particular frequency across a broad frequency range should result in a smoother, flatter curve.

A: We provide a description of the distribution of ripples across frequencies (see Response Table 1 above). The results showed that the majority of ripples was in the relative narrow bands ($< 120\text{Hz}$). Response Figure 7 shows the grand averages of ripples which peak frequency is below 120 Hz (right panel). It is very similar to the result from all detected ripples (left panel).

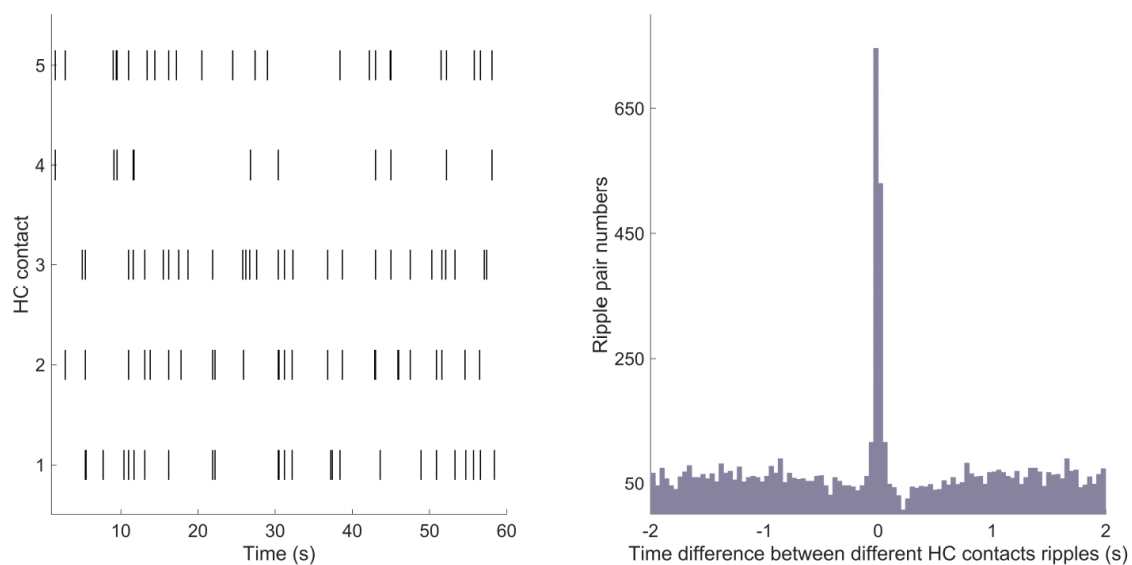


Response Figure 7. Grand ripple averages during wake and sleep.

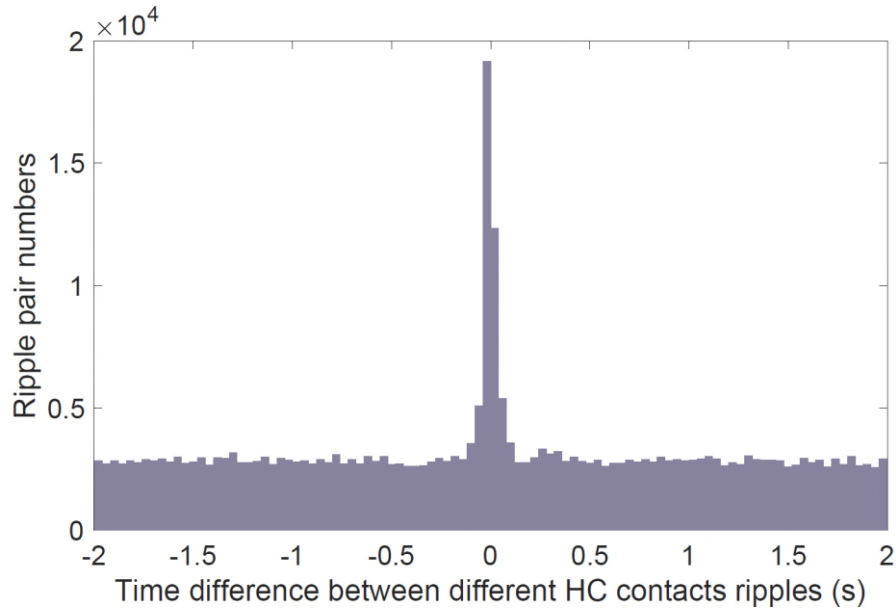
2.5 For each participant, only one hippocampal contact was used for ripple detection. A lot of data gets lost with this approach. Please detect ripples on all hippocampal electrodes and repeat the main analyses. The authors state that the hippocampal contact had the highest probability of being located in the hippocampus during the automatic brain segmentation. Is this based on co-registered anatomical data using an atlas? If so, this should be done in native patient space and verified by individual visual inspection. Fig. 1A is too small to understand this selection, it seems like some of the sphere are right at the edge of the hippocampal volume while other, unselected contacts are right in the center.

A: Firstly, we apologize for the unclear descriptions. Actually, we chose one hippocampal contact only for the purpose of ripple detection during learning and TMR stages. Then a time-window (-0.5s to 0.5s) around ripples was used for RSA analysis. This method referred to Norman's study (Norman et al., 2019). In order to test whether the ripples can be reliably detected in the hippocampus, we detected ripples using all hippocampal contacts and calculated the time interval between the ripples over different hippocampal contacts for each subject. The results showed high overlap in the detected ripple time over hippocampal contacts (see Response Figure 8 and 9).

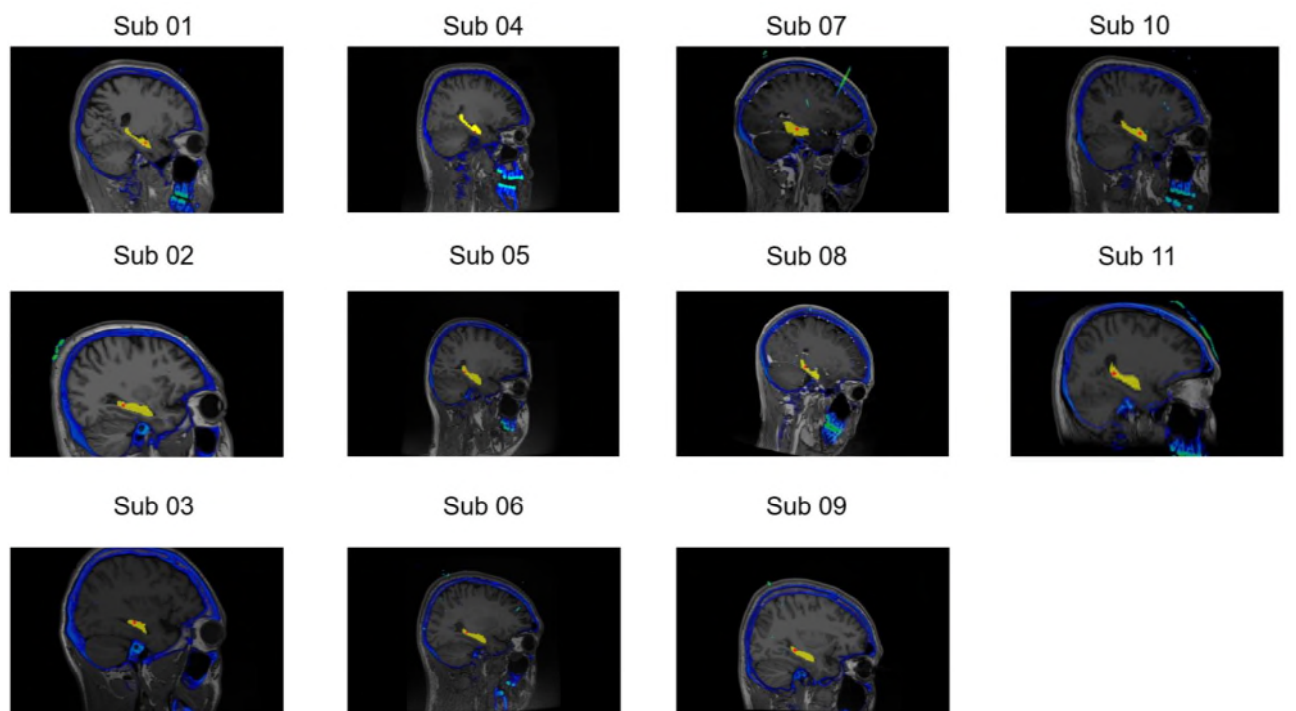
In addition, we need to clarify that in the present study, the electrode reconstruction and localization were done in native patient space. Response Figure 10 shows that the representative hippocampal electrode contact of each participant (used for ripple detection) is indeed located in the hippocampus area verified by individual visual inspection.



Response Figure 8. *The consistency of the ripples detected from different hippocampal contacts of one representative subject. The left part shows the examples of the occurrence of ripples across different hippocampal contacts. The right part shows the time intervals between different contact's ripples.*



Response Figure 9. The consistency of the ripples detected from all hippocampal contacts of all 11 subjects.



Response Figure 10. The hippocampal contact marked in the red dots (used for ripples detection) in native patient space. The yellow area represents the hippocampus region.

References:

Norman, Y., Raccach, O., Liu, S., Parvizi, J., & Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection.

2.6 clipping of extreme values: how was this done (please provide criteria) and why was this necessary?

A: According to the previous studies (Norman et al., 2019; Norman et al., 2021), to minimize ripple-rate induced biasing, the extreme values were robustly estimated using Least-Median-Squares and clipped to 4 SD above the mean value. We have added this part in the method section as follows (Page 14, line 568 - 569):

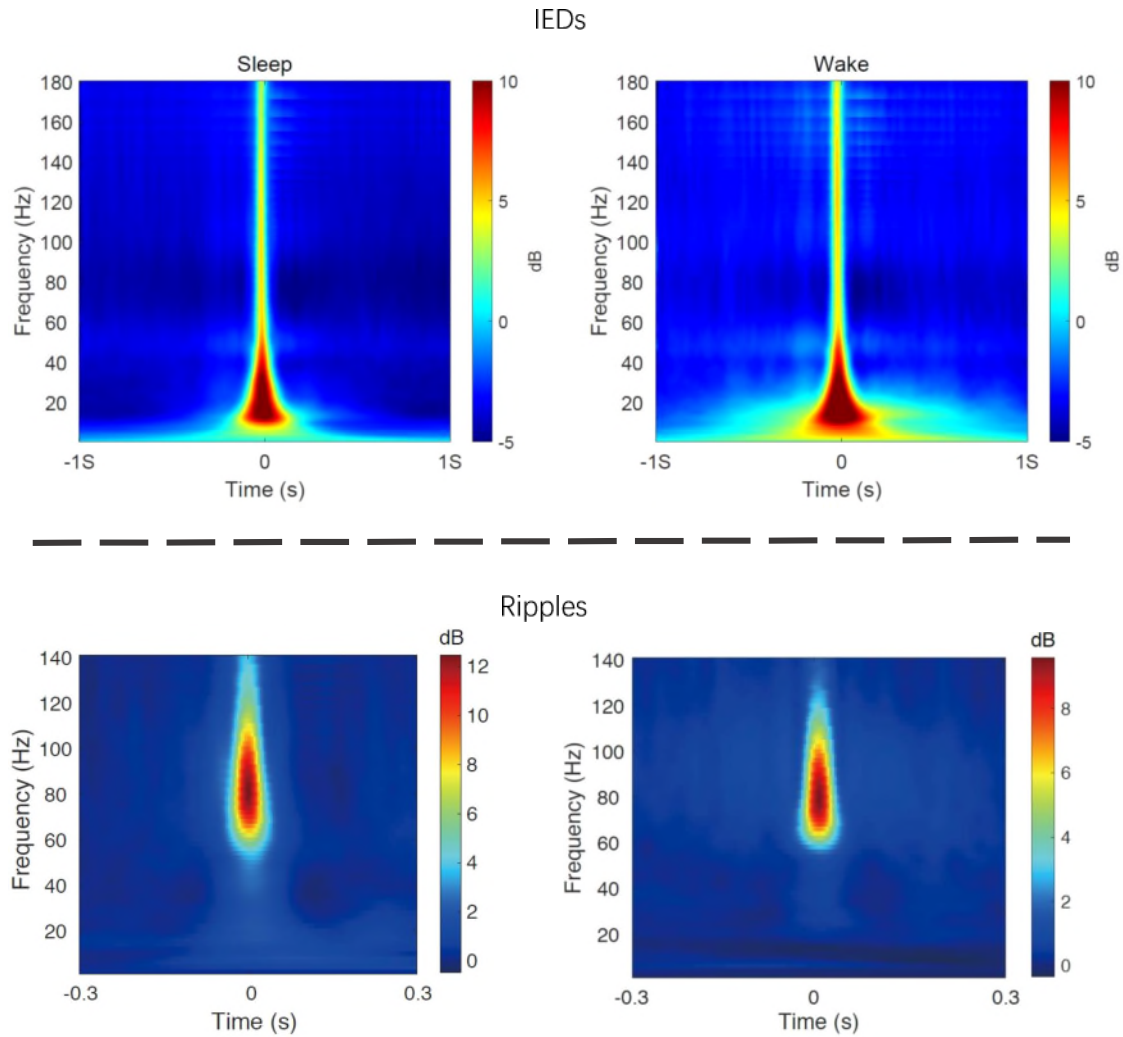
“To minimize ripple-rate induced biasing, the extreme values were robustly estimated using Least-Median-Squares and clipped to 4 SD above the mean value.”

References:

- Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454). <https://doi.org/10.1126/science.aax1030>
- Norman, Y., Raccach, O., Liu, S., Parvizi, J., & Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron*, 109(17), 2767-2780 e2765. <https://doi.org/10.1016/j.neuron.2021.06.020>

2.7 The authors should provide an epileptiform spike-locked average and plot this next to their ripple-locked average. Can one clearly distinguish pathological from non-pathological averages?

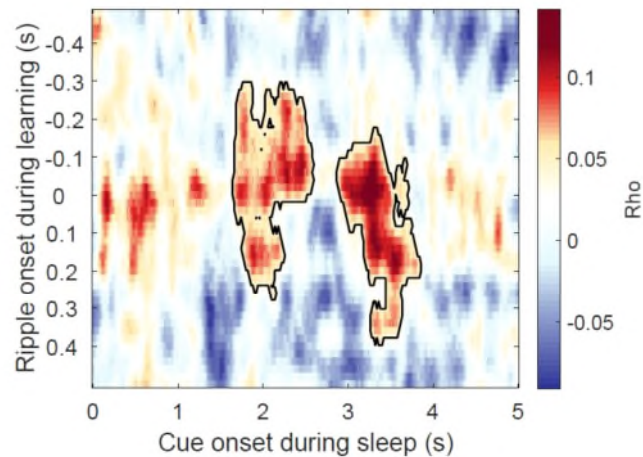
A: Thanks for this constructive suggestion. We have provided the epileptiform spike-locked average for wake and NREM sleep stages. As shown in the Response Figure 11, it is easy to distinguish between pathological and non-pathological conditions.



Response Figure 11. Grand average epileptiform spike-locked spectrogram (Top left: TMR stage, right: wake learning stage) and ripple-locked spectrogram (Bottom left: TMR stage, right: wake learning stage).

2.8 100 ms intervals between epileptiform spikes and other detected ripples might be too short, given that typical sharp wave after an epileptiform spike can be around 200ms long. Could the authors repeat the main analyses with a larger time interval between epileptiform spikes and ripples?

A: Thanks for the careful inspection. We calculated the interval between epileptiform spikes and other detected ripples and found that 93.3% of ripples have an over 200 milliseconds interval with an epileptiform spike. In addition, we repeated the main analyses using those ripples over 200 milliseconds interval with IEDs and found that the RSA results were consistent with previous results (see Response Figure 12).



Response Figure 12. The average Rho across subjects and rounds (excluding those ripples near IEDs).

2.9 Do ripple rates differ between states? Previous studies have shown higher ripple rates during NREM than REM sleep (e.g., Liu et al., 2022, Nat Com). Does this pattern hold in the current data set? The authors should report ripple rates for REM sleep and directly compare those to NREM sleep. A comparison with wake ripple rates would also be beneficial? Please report ripple rates in Hz (i.e., events / second), to make them comparable to previous studies (e.g., Vaz et al., 2019).

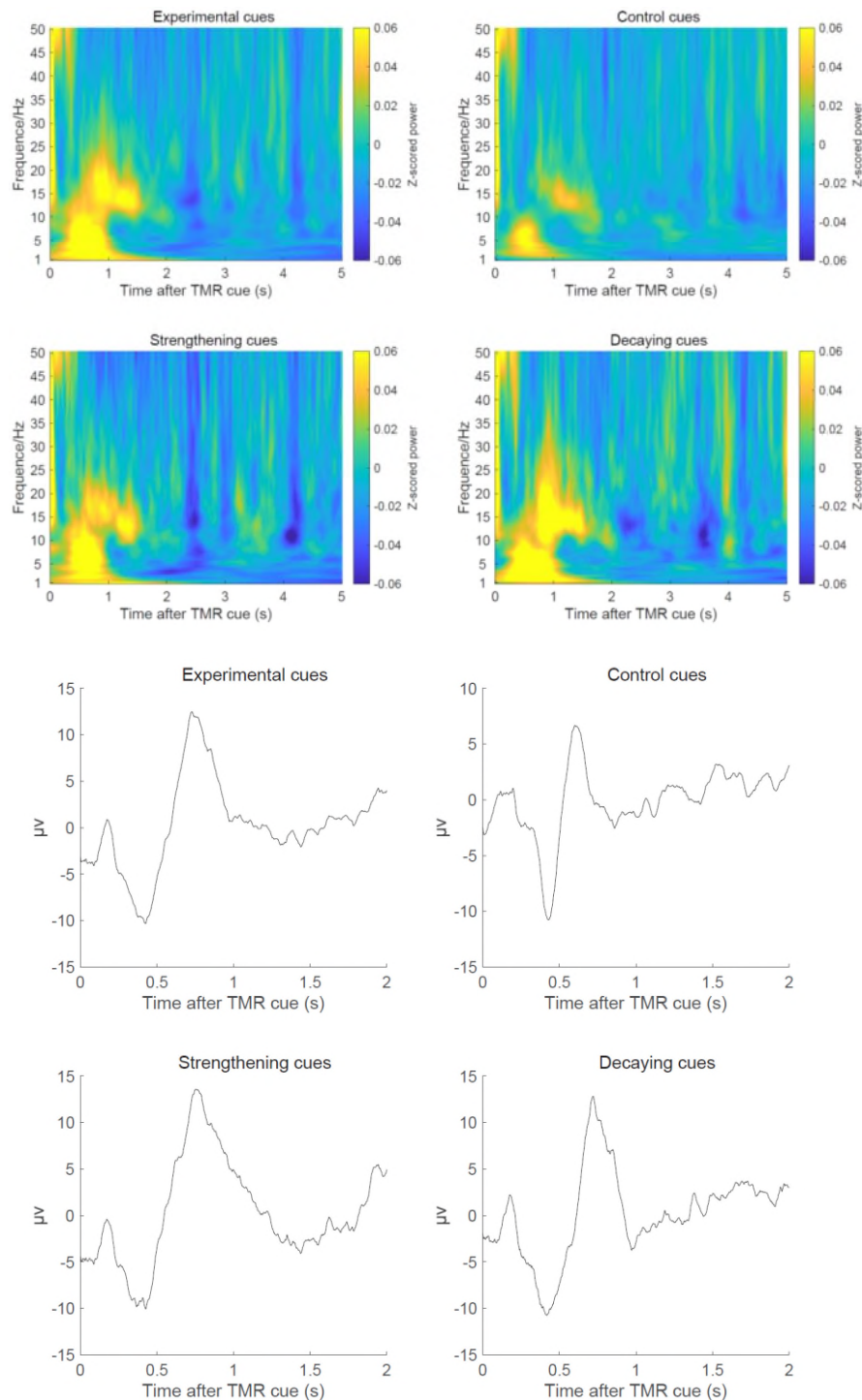
A: We have reported the ripple rates during the wake learning stage and the NREM sleep period (see Table S3 in the revised manuscript). The ripple rates during wake and NREM sleep stages in the present study was 0.40 ± 0.14 Hz and 0.37 ± 0.10 Hz, respectively. Given this study merely focused on memory consolidation during NREM sleep, we did not analyze the data during REM stage. In future study, it's worth performing this comparison between NREM and REM sleep.

3. ERPs and spectral power locked to the TMR cue.

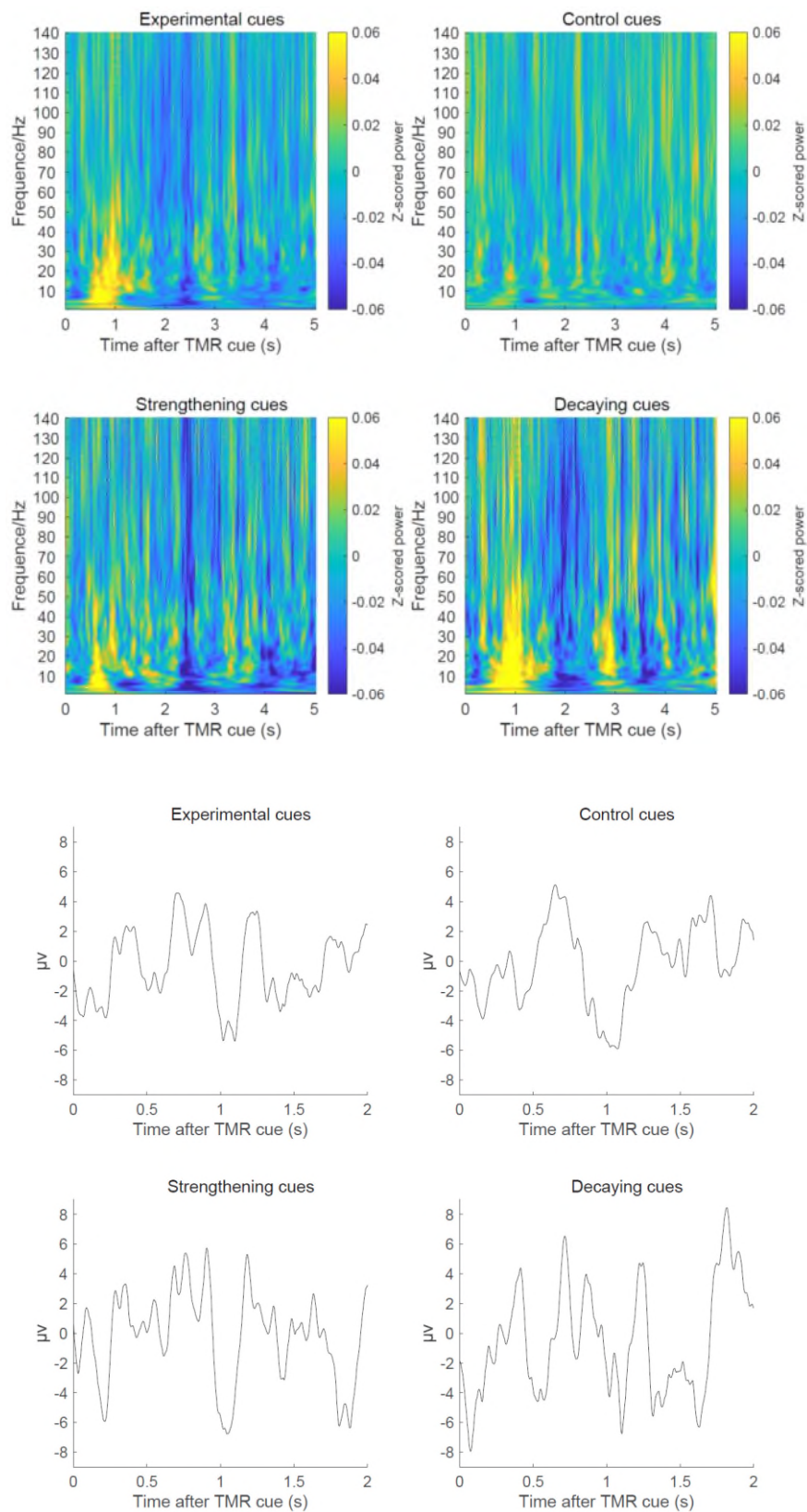
3.1 Basic information on how brain activity relates to TMR cues is missing. Please add figures for TMR cue-locked time-frequency power as well as ERPs of cortical contacts and hippocampus (without statistical comparisons, masking or computing differences). This will help to understand whether TMR cues elicited the typical patterns in the SO and spindle bands. Please add figures that display grand average of: all experimental cues; all control cues; each

condition (strengthening, decaying and control); please add separate plots for hippocampus and cortical contacts figures, respectively;

A: Thanks for such great suggestions. We showed figures for TMR cue-locked time-frequency power as well as ERPs of auditory cortex and hippocampal contacts (see Response Figure 13 and 14). Time-frequency power and ERP responses can be clearly seen in the auditory cortex contacts.



Response Figure 13. TMR cue-locked time-frequency power and ERPs in the auditory cortex contacts.



Response Figure 14. TMR cue-locked time-frequency power and ERPs in the hippocampal contacts.

3.2 Fig. S8 is puzzling. The legend says that cluster-based permutation tests were done and gives one p-value per panel, but A and B show many outlines across various time-frequency ranges. Are all of these significant clusters?

A: These clusters outlined are indeed significant. All of those empirical cluster values were compared to a distribution of the surrogate data based on the permutation test. The p value per panel was the threshold for detecting significant clusters.

3.3 In relation to this, I do not agree with the statement in the results section (l. 260 ff) that there are specific patterns related to the first vs. second time window of reactivation. In Fig. S8A, for example, there seems to be a larger increase (outlined in black, thus significant?) from around 10 to 20Hz just after 2 s, which would be in the first cluster. How is that in line with the authors' statement that they found significant cortical spindle activity enhanced only during the second reactivation? This also questions the statement in the discussion (l. 362) that the time-frequency analysis found consistent results.

A: Thanks for this careful inspection. Ripples and spindles are generally described as specific, discrete events. Therefore, in the present study, we specifically detected the ripple and spindle events during two reactivation clusters, while the time-frequency analysis of the ripple and spindle bands were only used as supplementary validation (included in the supplementary materials, Fig. S8). The detection results showed that cortical spindle rate in the strengthening condition was significantly higher than that in the decaying and control conditions during the second reactivation, but not the first reactivation. On the other hand, to further validate this finding, we conducted time-frequency analysis (as shown in Fig.S8). The time-frequency analysis found that although significant difference (strengthening>decaying) in spindle band power were also observed during the first reactivation period as the reviewer mentioned, it is worth noting that compared to the first reactivation, the difference (strengthening>deactivating) in spindle band power during the second reactivation was much stronger. Hence this result indicates that the changing trend of spindle activity differences (strengthening>decaying) is consistent with the results of spindle events detection. We completely agreed that the statement

in the results section of the original manuscript (l. 260) was unclear. Therefore, according to the reviewer's suggestions, we have made the following modifications for the statement in the results and discussion section (Page 7, line 271 - 275; Page 10, line 375 - 378):

“When comparing cortical spectrograms between the two conditions, we found difference in spindle activity during the first reactivation period. But it should be pointed out that the intensity of the difference is obviously lower than the subsequent second reactivation cluster. This result suggests that the changing trend of spindle activity differences (strengthening>decaying) is consistent with the result of spindle events detection.”

“In addition, the time-frequency analysis found similar trends with the spindle detection analysis. The spindle activity differences (strengthening>decaying) during the first reactivation was obviously lower than the subsequent second reactivation.”

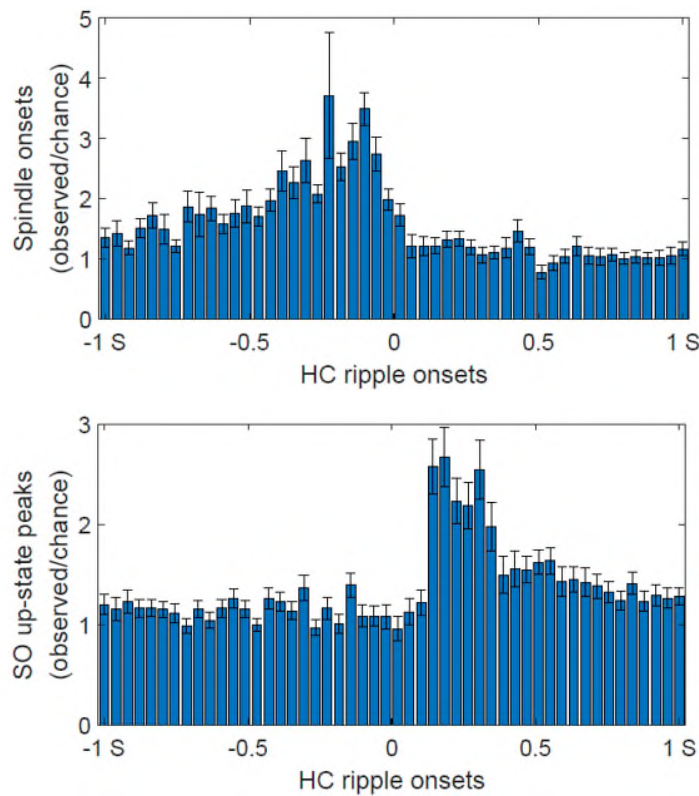
4. Slow oscillations, SO-spindle coupling

It is surprising that the paper lacks information on slow oscillations and their relationship to spindles and ripples, given that the study investigates neural activity during NREM sleep. Previous studies showed that slow oscillations and their coupling with sleep spindles is important for memory reactivation during sleep (e.g., Schreiner et al., 2021, Nat Com). Moreover, ripples also couple to SO-spindle complexes (e.g., Staresina et al., 2023). To relate to these findings, the authors could analyze SO–spindle coupling and SO–spindle–ripple coupling during NREM sleep in relation to TMR cues and investigate whether a potential coupling is related to memory reactivation.

A: Thanks so much for such constructive suggestion. As the reviewer pointed out, slow oscillations (SO) and their coupling with spindles are important for memory reactivation during sleep. According to the reviewer's suggestion, here SO was detected based on the algorithm used by Kim and colleagues (Kim et al., 2019). We indeed found the coupling relationship among SO, spindle and ripple (see Response Figure 15, consistent to previous research; Dickey et al., 2022). Then, SO–ripple events (ripple centers within ± 1 s of SO down-state), SO–spindle events (spindle centers within ± 1 s of SO down-state) and spindle-ripple events (ripple centers

within ± 1 s of spindle centers) were also detected based on previous methods (Staresina et al., 2023).

The literatures mentioned by the reviewer are all studies on spontaneous sleep, which included sufficient data for the above analysis. However, the iEEG data available for testing the coupling relationship among SO, spindle and ripple is very limited (only about 15 minutes) due to the use of the TMR paradigm in our study (25 TMR cues, each of which was played 6 times during NREM sleep). Considering the occurrence rate of coupling events (see Response Table 3), the present study was not suitable for testing whether the coupling was related to memory reactivation.



Response Figure 15. *Upper: times of spindle onsets plotted relative to ripple onsets during NREM. Below: times of SO up-state peaks plotted relative to ripple onsets during NREM.*

Response Table 3

# Subjects	Occurrence rate (events/min)			Trials with coupling events		
	Ripple-SO	Ripple-spindle	Spindle-SO	Ripple-SO	Ripple-spindle	Spindle-SO
1	2.01	0.98	0.95	14%	10%	8%
2	0.88	0.74	1.09	5%	5%	6%

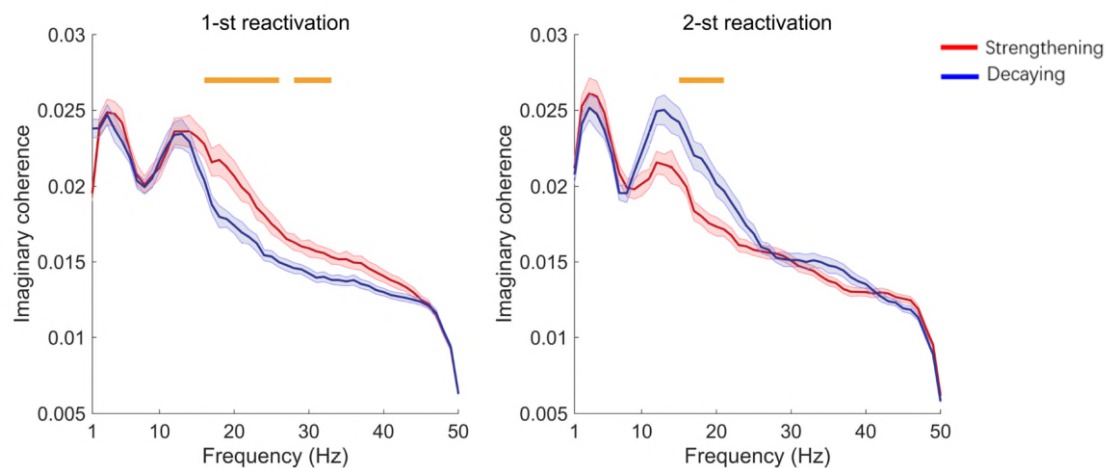
3	2.12	1.14	1.24	16%	8%	10%
4	3.08	1.08	0.96	25%	10%	9%
5	3.33	1.76	1.49	28%	16%	14%
6	4.41	1.84	2.08	37%	15%	19%
7	2.76	1.61	2.45	23%	13%	20%
8	2.00	1.53	0.86	18%	14%	8%
9	3.10	1.14	1.62	26%	10%	13%
10	4.29	2.17	2.10	32%	16%	16%
11	1.96	0.49	0.60	18%	5%	6%
Average	2.72	1.32	1.40	22%	11%	12%

References:

- Kim, J., Gulati, T., & Ganguly, K. (2019). Competing Roles of Slow Oscillations and Delta Waves in Memory Consolidation versus Forgetting. *Cell*, 179(2), 514-526 e513. <https://doi.org/10.1016/j.cell.2019.08.040>
- Dickey, C. W., Verzhbinsky, I. A., Jiang, X., Rosen, B. Q., Kajfez, S., Eskandar, E. N., Gonzalez-Martinez, J., Cash, S. S., & Halgren, E. (2022). Cortical Ripples during NREM Sleep and Waking in Humans. *The Journal of Neuroscience*, 42(42), 7931-7946. <https://doi.org/10.1523/jneurosci.0742-22.2022>
- Staresina, B. P., Niediek, J., Borger, V., Surges, R., & Mormann, F. (2023). How coupled slow oscillations, spindles and ripples coordinate neuronal processing and communication during human sleep. *Nat Neurosci*. <https://doi.org/10.1038/s41593-023-01381-w>

5. Connectivity analysis 5.1 Coherence can be contaminated by volume conduction. To avoid this, please report a connectivity metric that excludes zero-lag connectivity (e.g., imaginary coherence).

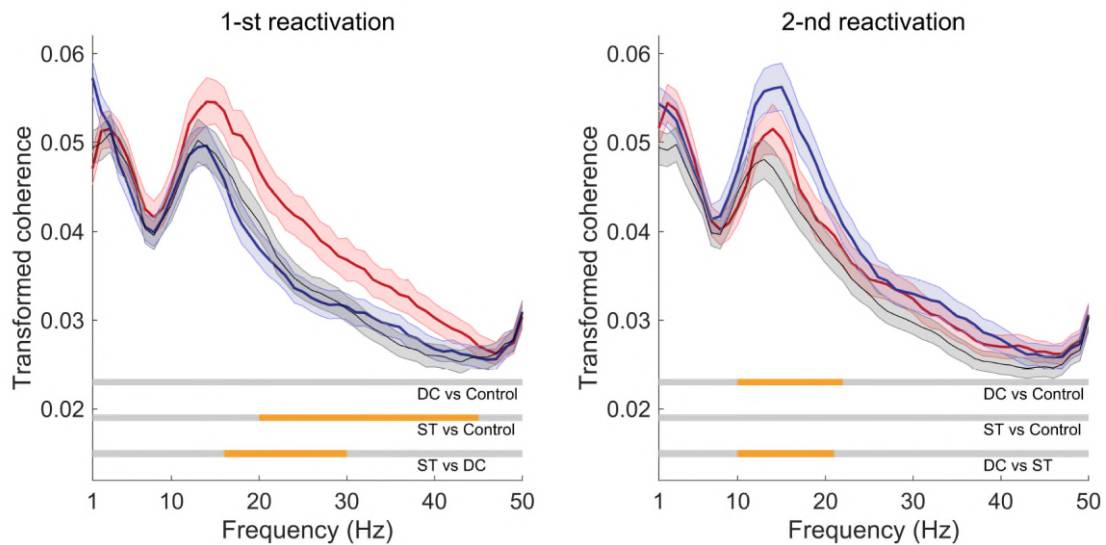
A: Thanks for this valuable suggestion. We conducted a connectivity analysis using imaginary coherence and found similar result with the main result (see Response Figure 16).



Response Figure 16. Imaginary coherence between cortex and hippocampus during the first reactivation (left) and second reactivation (right). Yellow lines marked the significant differences based on the permutation test.

5.2 Coherence (and other connectivity metrics) can be heavily biased by trial numbers, which is particularly relevant when comparing conditions. This is even more problematic when trial numbers are generally low, as it is the case here: the first two participants had a maximum of 27 / 36 trials per condition. Since I did not find information on this in the methods section, I assume no correction for number of trials was performed. The authors should repeat the analyses using appropriate metrics.

A: Thank you for your careful reading. Considering that connectivity analysis can be affected by the number of trials, so to avoid this impact, we balanced the trial number between the conditions (see Response Table 4). We have added detailed descriptions in the methods section (1. The trial number of different conditions is balanced; 2. The reasons for the fewer trials in the first two participants). In addition, we did correct the coherence value by taking the number of trials into account (Bokil et al., 2007) and found similar results with the main result (see Response Figure 17).



Response Figure 17. Transformed coherence between cortex and hippocampus in 1-st reactivation (left) and 2-nd reactivation (right).

Reference:

Bokil, H., Purpura, K., Schoffelen, J. M., Thomson, D., & Mitra, P. (2007). Comparing spectra and coherences for groups of unequal size. *J Neurosci Methods*, 159(2), 337-345. <https://doi.org/10.1016/j.jneumeth.2006.07.011>

5.3 Please perform the connectivity analyses based on all hippocampal contacts vs. all cortical contacts. Please report the number of trials that included into each of the three conditions for each of the displayed coherence analyses.

A: The connectivity analysis in our study indeed used all hippocampal and cortical electrodes (See the description of connectivity analysis in the Methods section). According to the reviewer's suggestions, we reported the number of trials that were included into each of the three conditions for each of the coherence analyses (see Response Table 4).

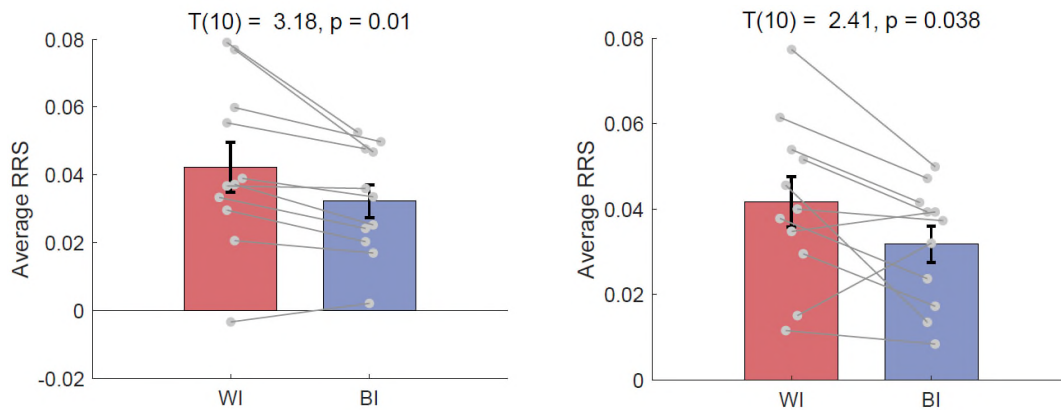
Response Table 4

	Coherence			RRS	RSS	
# Subjects	Strengthening	Decaying	Control		Learning	TMR
1	27	27	27	97	83	63
2	36	36	36	127	85	100
3	54	54	54	66	48	132

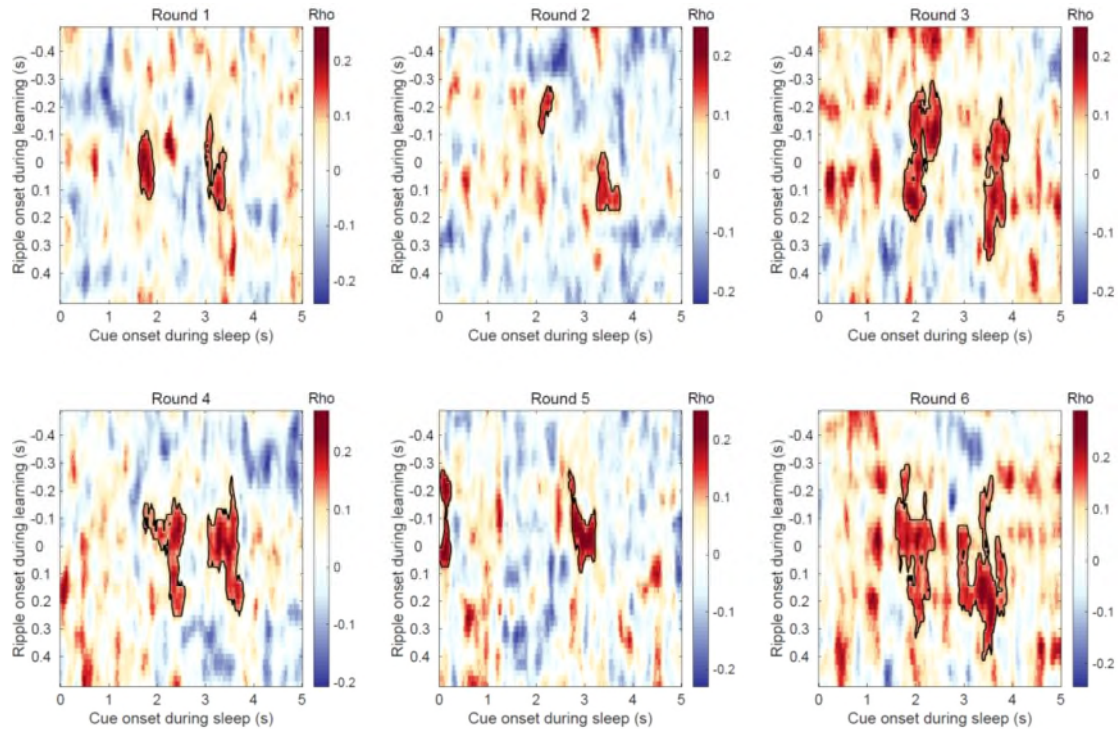
4	54	54	54	206	167	150
5	54	54	54	165	132	144
6	54	54	54	118	102	138
7	54	54	54	37	23	78
8	54	54	54	77	53	114
9	54	54	54	103	80	126
10	54	54	54	135	107	126
11	54	54	54	122	84	138

5.4 related to this issue: how many trials went into the RSS/RRS analyses? Are there enough trials to support stable estimations of these metrics?

A: The number of trials for RSA analysis is as follows: RRS: 114 ± 46.6 ; RSS learning: 88 ± 39.8 ; RSS TMR: 119 ± 27.8 (see Response Table 4). Compared to previous studies, this trial number is at a similar level to them. For example, the average trial numbers used for RSA analysis was 80 in Zhang et al.'s study (Zhang et al., 2018) and 112 trials in Norman et al.'s study (Norman et al., 2019). To further demonstrate whether the trial number is sufficient, we conducted a new analysis by dividing all trials into two parts in RRS analysis and calculating RSA results, separately. We found that the results were consistent with the main result (Response Figure 18). In addition, for RSS analysis, you can refer to the results of each round in our supplementary materials. This result indicates that even partial trials can produce stable results, indirectly supporting that the trial numbers for RSA analysis are sufficient (Response Figure 19).



Response Figure 18. Item-specific representation is calculated using the first half (left) and the second half of the trials (right).



Response Figure 19 (identical to Figure S2). Behavioral-relevant item reactivations at each TMR round. Spearman correlation is calculated between memory change scores and RSS of each TMR round. The permutation test is performed within each round, and significant areas are extracted as clusters. The two largest clusters are shown for each round.

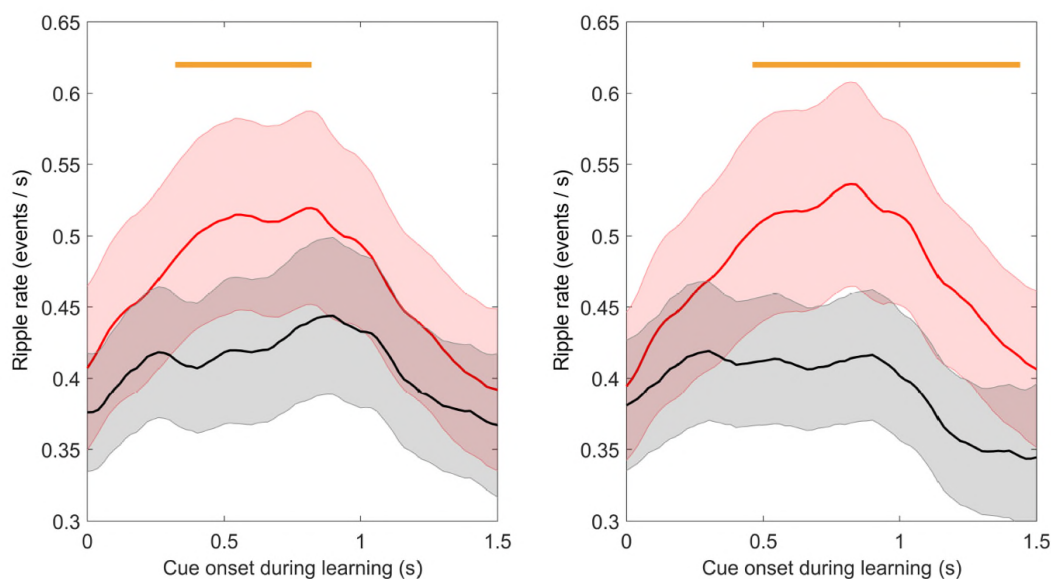
References:

- Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat Commun*, 9(1), 4103. <https://doi.org/10.1038/s41467-018-06553-y>
- Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454). <https://doi.org/10.1126/science.aax1030>

6. Relation of neural signatures and memory performance

6.1 Was the ripple rate during learning related to successfully remembering the location of the item? One would expect more ripples during correctly remembered locations (smaller drop error) than forgotten locations (larger drop error), in each of the retrievals (pre-sleep, post-sleep). This would substantially strengthen the link between the reported ripples and memory processes and relate to previous findings (e.g., Vaz et al. (2019).

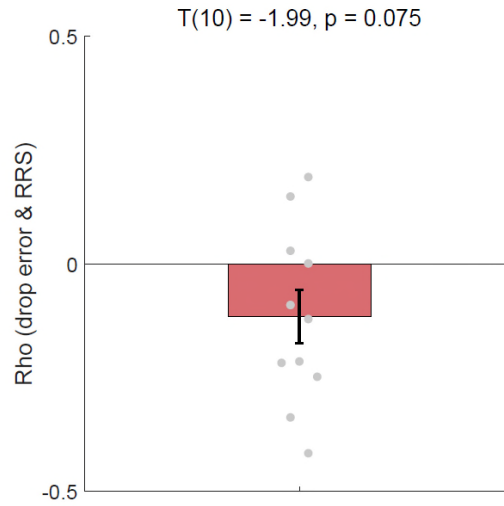
A: Thanks for this constructive suggestion. To test whether the ripple rate during learning is related to successfully remembering the location of the item, we divided the learning trials into two groups based on the drop error in pre-sleep test (the left part of Response Figure 20) or post-sleep test (the right part of Response Figure 20): smaller error group (red line) and larger error group (black line). As the reviewer expected, the result indeed showed that ripple rate of the smaller error group was significantly higher than those of the larger error group.



Response Figure 20. Ripple rate during the learning stage. We divided the learning trials into two groups based on the drop error in pre-sleep test (left part) or post-sleep test (right part). Yellow bar indicates the time period of significant difference between two groups.

6.2 The authors should add a similar contrast to their reactivation metric during learning. Was the reported higher WI similarity related to memory performance? Does a higher WI similarity correspond smaller drop error in a given trial? Please add figures depicting these contrasts.

A: Thanks for the valuable comment. To test whether WI similarity during learning was related to memory performance, we calculated the correlation coefficient between the WI similarity and drop error in pre-sleep test. Negative correlation indicates that the higher the WI similarity is, the smaller drop error in pre-sleep test occurs (see Response Figure 21).



Response Figure 21. The Rho between WI similarity and drop error in pre-sleep test for each subject. Negative correlation indicates that the higher the WI similarity is, the smaller drop error in pre-sleep test occurs.

7. Multiple Comparisons. Are all reported p-values corrected for multiple comparisons? For example, when conducting multiple t-tests displayed in Fig. 4G? If not, please use appropriate corrections and report corrected p-values. For comparisons between 3 conditions, an ANOVA that tests the difference between the 3 conditions should be computed first and post-hoc t-tests in case the ANOVA shows a significant difference.

A: Thank you for your careful reading. For these results with cluster-based permutation test (Fig 3b, 4b, 5a, 5e, s8, s10), we have considered multiple comparisons in the permutation test method according to previous literature (Zhang et al., 2018). Specifically, we compared these empirical cluster values with a distribution of maximum surrogate cluster of each shuffle to correct for multiple comparisons. For those results that involve multiple comparisons (Fig 3c, s1b), we updated the p-values after Bonferroni correction.

For the results involving comparisons between three conditions (Fig 4f, 4g, 5b, 5f), we conducted the repeated measure ANOVA:

Fig. 4F (left): A repeated measure ANOVA revealed a significant difference between the ripple rate for the three conditions ($F_{(2,20)} = 3.7, p = 0.043$) during the first reactivation. Post-hoc t-tests revealed that strengthening condition was higher than control conditions ($t(10) =$

2.31, $p = 0.043$). Fig 4f (right): A repeated measure ANOVA revealed no significant condition effects ($F_{(2,20)} = 0.21$, $p = 0.81$).

Fig. 4G (left) during the second reactivation: A repeated measure ANOVA revealed no significant condition effects ($F_{(2,20)} = 0$, $p = 0.998$). Fig 4g (right): there was a significant difference between the spindle rate for the three conditions ($F_{(2,20)} = 5.58$, $p = 0.012$). Post-hoc t-tests revealed that the spindle rate in the strengthening condition was significantly higher as compared to the decaying and control condition ($t(10) = 2.637$, $p = 0.025$; $t(10) = 3.226$, $p = 0.009$, respectively).

Fig. 5B: A repeated measure ANOVA (using a LME model) indicated a significant difference between three conditions ($F_{(2,3276)} = 47.858$, $p < 0.001$) during the first reactivation. Post-hoc t-tests revealed that differences between two conditions were significant in the three conditions (LME strengthening vs. decaying, $t(1092) = 9.064$, $p < 0.001$; LME strengthening vs. control $t(1092) = 5.943$, $p < 0.001$; LME decaying vs. control $t(1092) = -4.909$, $p < 0.001$).

Fig. 5F: A repeated measure ANOVA (using a LME model) indicated a significant difference between three conditions ($F_{(2,3276)} = 25.498$, $p < 0.001$). Post-hoc t-tests revealed that the HFB power in the control condition was significantly higher as compared to the strengthening and decaying condition (LME control vs. strengthening, $t(1092) = 5.394$, $p < 0.001$; LME control vs. decaying, $t(1092) = 7.771$, $p < 0.001$). There was no significant difference between strengthening condition and decaying condition (LME strengthening vs. decaying, $t(1092) = 1.536$, $p = 0.125$).

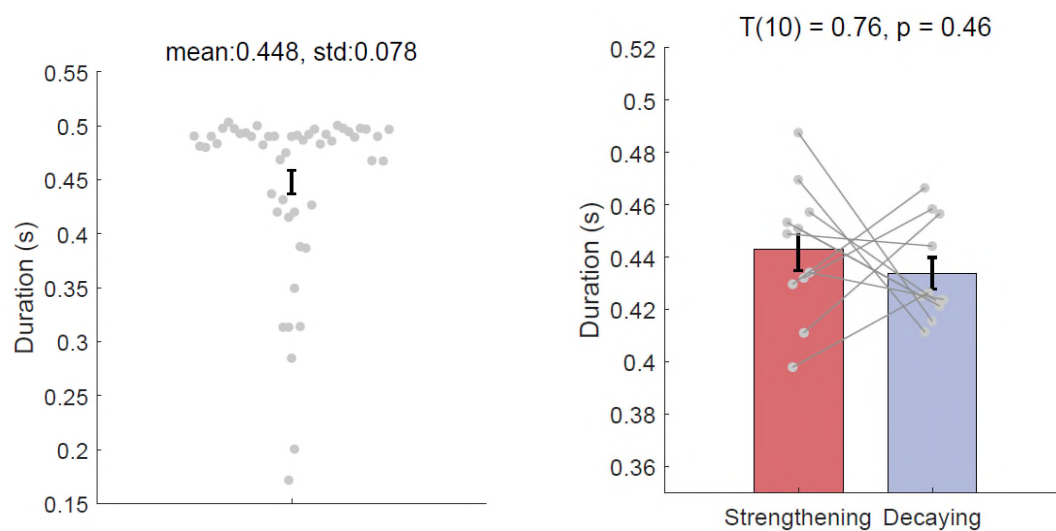
We have updated description of the Results about ANOVA and post-hoc t-tests in the revised manuscript.

Reference:

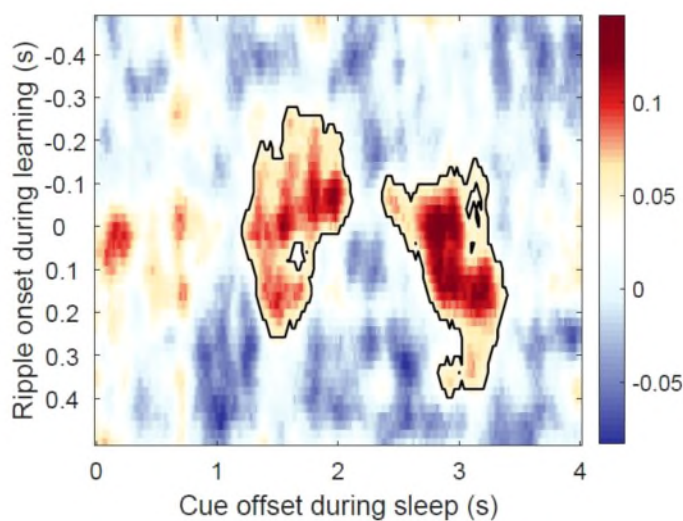
Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat Commun*, 9(1), 4103. <https://doi.org/10.1038/s41467-018-06553-y>

8. Sound cues: Please provide mean duration (and SDs) for the of the sound cues. Did the duration of the sound cues differ between conditions? How would different durations affect the outcomes? Would locking to sound cue offset change the pattern of the results?

A: According to the reviewer's suggestions, we reported mean duration (and SDs) of the sound cues. There is no significant difference in the duration of the two conditions (strengthening vs. decaying, see Response Figure 22). In addition, we used the sound cue offset as the zero point for RSA analysis, and the result was consistent with the main result (see Response Figure 23, please note that the zero point in this figure represents the end of the sound cue).



Response Figure 22. Left: the duration of 50 sound cues. Right: The difference of sound duration between strengthening and decaying items.



Response Figure 23. *The average Rho locking to sound cue offset across subjects and rounds. The result is very similar to the main result.*

9. Title: The word ‘mechanism in the title could be misleading to large parts of the audience. I suggest to remove this highly debated term.

A: We have removed the word “mechanism” from the title. The new title is “Electrophysiological signatures underlying variability in human memory consolidation”.

10. Questions regarding the intro / discussion:

10.1 Abstract: ‘produced nonuniform amounts of memory enhancement across a set of object-location associations that were reactivated’ What do the authors mean by ‘nonuniform’? They show an overall benefit of TMR cuing. This effect varies across items, but that is a typical pattern seen in many TMR studies. The same questions apply to l. 317.

A: The “nonuniform” refers to the fact that although TMR cues were given to 25 items, their benefits are different. Some items benefit more than others (strengthening group vs. decaying group). In the present study, we firstly showed the overall effect of TMR only to validate the effectiveness of the TMR paradigm. However, it needs to be clarified that our core research objective is to explore why some TMR-cued items are preferentially strengthened during sleep than others. Therefore, we emphasized and described the differences of TMR effect across items in the manuscript. Given that the word 'nonuniform' may mislead readers, we have replaced it with the word 'variable' in the revised manuscript.

10.2 l. 95: ‘In short, endogenous tags based on emotion or reward, or exogenous sensory cues, can systematically improve consolidation during sleep and thus strengthen the corresponding memories.’ Memory reactivation has been shown without emotional or reward tags, see e.g., Schreiner et al., 2022.

A: In the Introduction section, we focused on discussing the tagged activation, in order to introduce the usefulness of the TMR paradigm in memory consolidation. We have not concluded that memory reactivation can only occur under these endogenous and exogenous

tags. As the reviewer pointed out, memory reactivation can also occur in the absence of emotional or reward tags. To avoid such misunderstandings, we have made the following modifications in the revised manuscript (Page 3, lines 93 - 97):

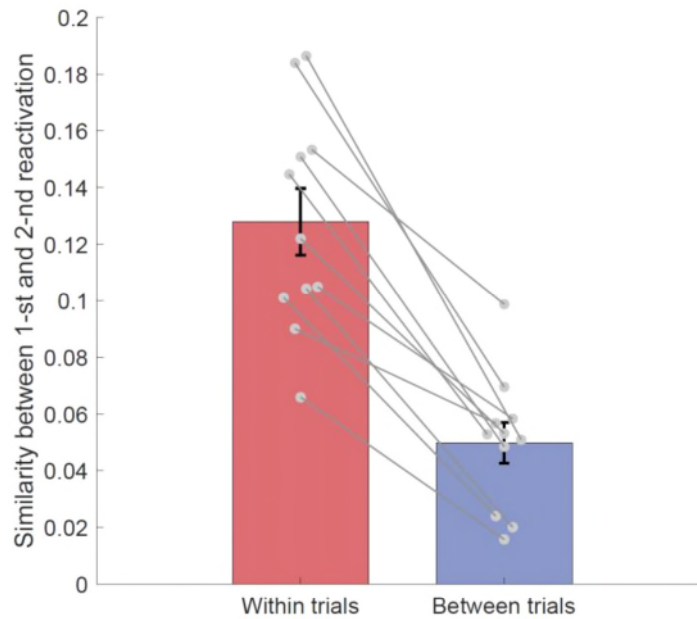
“In short, endogenous tags based on emotion or reward, or exogenous sensory cues, can systematically improve consolidation during sleep and thus strengthen the corresponding memories. On the other hand, it is worth noting that endogenous and exogenous tags are not necessary for memory reactivation. For example, Schreiner et al. found that memory reactivation can occur without external cues (Schreiner et al., 2021).”

References:

Schreiner, T., Petzka, M., Staudigl, T., & Staresina, B. P. (2021). Endogenous memory reactivation during sleep in humans is clocked by slow oscillation-spindle complexes. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-23520-2>

10.3 l. 334: ‘Furthermore, our study revealed multiple TMR induced memory reactivations. This result suggests that memory consolidation during NREM sleep may require multiple reactivations.’ An interesting question related to this statement is whether it is the same item or different items that are reactivated multiple times. Can the authors address this?

A: Thanks for this interesting comment. We calculated the pattern similarity between the first and second reactivations and found that it was significantly higher within trials than that between trials (see Response Figure 24). Therefore, we believe that the two reactivations are probably for the same item. Given that the present study was limited by the numbers of trials during learning, it was difficult to train a high-precision decoder for reactivation pattern during sleep. In the future, we may need to use more sophisticated techniques, such as single-cell recording, and more trials to train classifiers with higher decoding accuracy to clearly address this question.



Response Figure 24. Pattern similarity between the first and second reactivations.

10.4 l. 355: ‘However, considering the echoing effect of TMR-induced reactivation, these previous studies have not indicated whether there are hippocampal ripples during each cortical reactivation.’ Where do the authors show that there is a ripple for each cortical reactivation? I might have missed this, but if there is no evidence to support the claim, please remove it.

A: We do apologize for the unclear description. The "each cortical reactivation" refers to the time window during which multiple reactivations (for example, three reactivation clusters in Schreiner’s study, and two clusters in our study) occur. We attempted to detect the occurrence of hippocampal ripples corresponding to the multiple reactivation windows. We have made the modifications to the text as follows (Page 9, lines 368 - 370):

“However, considering the echoing effect of TMR-induced reactivation, these previous studies have not indicated whether there are hippocampal ripples during the multiple cortical reactivation periods after TMR cue”.

10.5 l. 388: ‘indicating communication between hippocampus and cortex.’ The authors should acknowledge the ongoing debate on the relationship between coherence and neuronal communication (Schneider et al., 2021, Neuron; Dowdall & Vinck, 2023, NeuroImage; Fries, 2015, Neuron) and carefully evaluate their own interpretation.

A: Thanks for this suggestion. We have made the following modifications to this sentence (Page 10, lines 414 - 417):

“During the first reactivation, (i) strengthening memories showed more pronounced hippocampal-cortical coupling and cortical HFB activities coupled to hippocampal ripples, may imply associations between hippocampus and cortex during memory reactivation”.

10.6 1. 391: ‘transmitted to cortex’ Where do the authors show information transfer to the cortex? ? I might have missed this, but if there is no evidence to support the claim, please remove it.

A: Thanks for pointing out this mistake. We have removed this content in the revised manuscript.

Minor:

-Paradigm: Some terms are misleading in my opinion. Why is the initial phase of the experiment called ‘preview phase’ and not ‘study phase’, which is much more common? Also, the “learning stage” is actually a retrieval task. Together, the reader may confuse these phases in the present study.

A: The initial stage of the experiment is called the "preview stage" for the following reasons: First, the task requires participants to remember 50 object-location associations. Since the large number of stimuli may make the experiment challenge for patients, the preview stage was added to familiarize them with the experimental materials, spatial background, and requirements before the formal learning stage. In sum, the purpose of this part was to facilitate learning of many object-location associations during the main learning session. Secondly, the experimental paradigm and stimulus materials used in this study were derived from our collaborator’s previous studies (Rudoy et al., 2009; van Dongen et al., 2012, Creery et al., 2022). Therefore, the naming of the experimental stages was consistent with their researches.

The “learning stage” used several retrieve (test)-feedback (study) cycles. This pattern is common in previous memory research. For example, in Hsieh's research, their learning session

included study and test phase (Hsieh et al., 2014). We made the modification in the Methods section in the revised manuscript (Page 12, line 477-478):

“In the learning stage (including study and retrieval phase), participants attempted to place each object in the correct location (i.e., retrieval phase).”

References:

- Creery, J. D., Brang, D. J., Arndt, J. D., Bassard, A., Towle, V. L., Tao, J. X., Wu, S., Rose, S., Warnke, P. C., Issa, N. P., & Paller, K. A. (2022). Electrophysiological markers of memory consolidation in the human brain when memories are reactivated during sleep. *119*(44), e2123430119. <https://doi.org/doi:10.1073/pnas.2123430119>
- Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, *326*(5956), 1079. <https://doi.org/10.1126/science.1179013>
- Hsieh, L.-T., Gruber, Matthias J., Jenkins, Lucas J., & Ranganath, C. (2014). Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron*, *81*(5), 1165-1178. <https://doi.org/10.1016/j.neuron.2014.01.015>
- van Dongen, E. V., Takashima, A., Barth, M., Zapp, J., Schad, L. R., Paller, K. A., & Fernández, G. (2012). Memory stabilization with targeted reactivation during human slow-wave sleep. *Proceedings of the National Academy of Sciences*, *109*(26), 10575-10580. <https://doi.org/doi:10.1073/pnas.1201072109>

- Figure 6 indicates that multiple distinct cortical areas were investigated, which was not the case.

A: To avoid reader misunderstandings, we added the following sentence in the Figure 6 (page 29, line 1055-1056):

“Note that the cortical positions shown in Figure 6 are only for illustration purposes.”

- Figure S7. Please add occurrence rate of retrieval ripples

A: Figure S7a shows the occurrence rate of retrieval ripples both for strengthening and decaying conditions. The “learning stage” includes retrieval and feedback phases. We detected the incidence of ripples during the retrieval phase. We have modified the description in the legend of Figure S7 (page 35, line 1135-1136):

“Hippocampal ripple occurrence rate following cues during retrieval (test) phase of the learning stage.”

- Line 218 and Figure 3: please specify what a (gamma) pattern is.

A: The gamma pattern is a vector ($1 \times \text{contacts}$) composed of gamma power (averaged from 50 to 140 Hz) of every cortical contact. We added the description of “gamma pattern” in revised manuscript (page 6, line 222 - 223) as follows:

“The gamma pattern is a vector ($1 \times \text{contacts}$) composed of gamma power (averaged from 50 to 140 Hz) of every cortical contact.”

- Figure S2: What is encoding-sleep similarity (ESS) and why was it compared to Retrieval-Sleep Similarity (RSS)? ESS is not explained in the methods.

A: Sorry about this typo. It should be read as RSS. We have corrected it .

- Line 952: ‘marginally significant’: readers might object to this wording

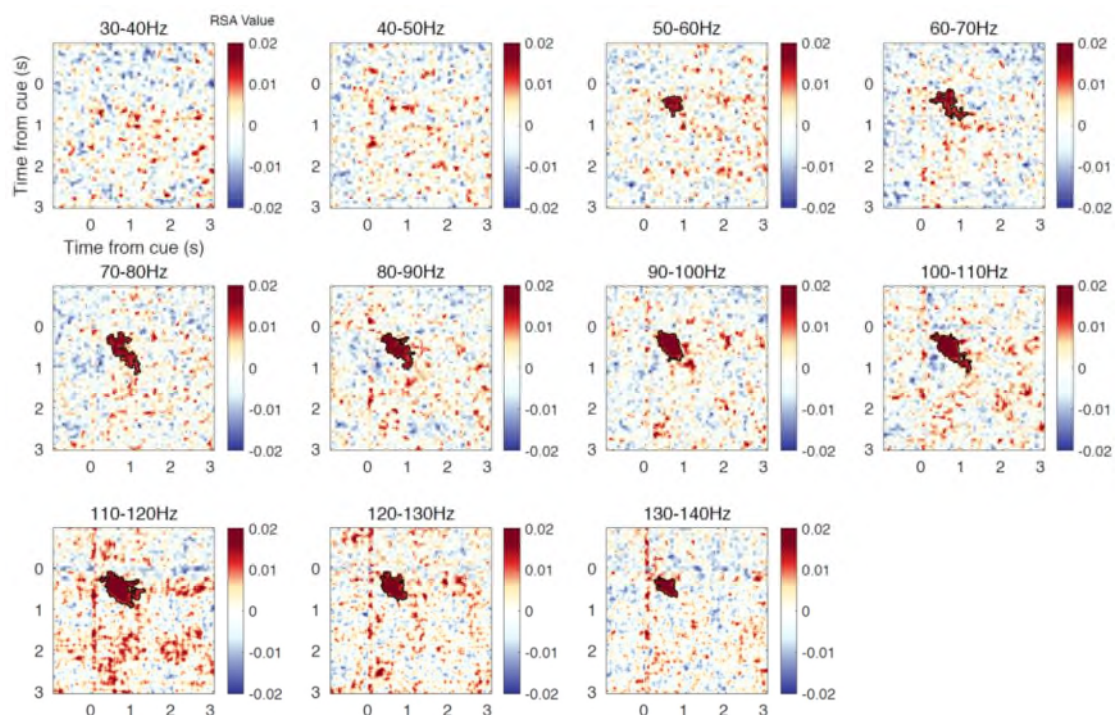
A: According to the reviewer's suggestions, we have made the following modifications (page 27, line 1027 - 1030):

“Hippocampal ripple rate of the strengthening condition is significantly higher than that of the control conditions and showed an increasing trend than that of the decaying conditions, but did not reach statistical significance ($p = 0.054$)”.

- What is the rationale behind choosing 50-140 Hz as high gamma band?

A: In previous studies, different frequency bands were used as high gamma band in calculating brain patterns (30-90 Hz (Zhang et al., 2018) or 60-140 Hz (Norman et al., 2019)), for the RSA analysis. In this study, we first determined the optimal frequency band by detecting item-level activity patterns during learning stage. We extracted epochs from 1,000ms before to 3,000ms after the onset of each item for each of 11 frequency bands with 10 Hz bandwidth (30-40Hz, 40-50Hz, ..., 130-140Hz). Since the learning stage had several rounds, we calculated the RSA

of either the same item (RSA_WI, selected the same object across multiple rounds) or the different item (RSA_BI, randomly selected one from the other 49 items). Then we used cluster-based permutation test to detect significant clusters between RSA-WI and RSA-BI for each frequency band (*Response Figure 25*). The results indicate that a broad frequency band (50-140Hz) contained item-level representations. Therefore, we chose 50-140 Hz as high gamma band for subsequent representation similarity analysis. We have added this part in the revised manuscript.



Response Figure 25 (identical to Figure S13 in the revised manuscript). RSA values for step-wise analysis of frequency bands

RSA values of $[RSA_WI - RSA_BI]$ on cortical patterns during the learning stage over 11 frequency bands. The black frame outlining the cluster shows the time range of significant difference between same items and different items ($p < 0.05$, cluster-based permutation test).

References:

- Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454). <https://doi.org/10.1126/science.aax1030>
- Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat Commun*, 9(1), 4103. <https://doi.org/10.1038/s41467-018-06553-y>

-l. 106: TMR cue appears to occur in multiple consecutive sweeps and produce a temporal sequence of reactivation. Where do Cairney et al., 2018 show this?

A: This point was originated from a review paper (see Figure 1B in Lewis & Bendor's paper). We have corrected the citation (Cairney et al., 2018).

References:

Lewis, P. A., & Bendor, D. (2019). How Targeted Memory Reactivation Promotes the Selective Strengthening of Memories in Sleep. *Curr Biol*, 29(18), R906-R912. [https://doi.org/ 10.1016/j.cub.2019.08.019](https://doi.org/10.1016/j.cub.2019.08.019)

Reviewer #2

In the present study, the authors employ targeted memory reactivation during sleep to demonstrate that ripples and sleep spindles reflect two distinct signatures of memory reactivation. They recorded intracranial EEG from 11 participants and simultaneously recorded from neocortex and the hippocampus. The findings are interesting, since the authors describe two distinct time points after reactivation that are either characterized by ripples (ca. 2 s after reactivation) or spindle activity (ca. 3-4 s after reactivation) that differentiate items that were later remembered or forgotten. These observations are interesting since they extend the finding of spindle ‘pulsing’ every 3-6 s (Antony et al., 2018, Curr Biol; Sirota et al., 2003, PNAS) that may have differential effects on the underlying circuit (e.g., Niethard et al., 2018, PNAS) in support of memory consolidation.

We appreciate the reviewer's interests in our study and many valuable suggestions. We have diligently addressed all concerns, and the supplementary results further support our main findings. Please see below for detailed responses to each question.

While this constitutes an interesting finding, it is unfortunately based on a relatively small sample of only 11 participants, with an overall underpowered statistical design (25 reactivated patterns, 50 in total) that yielded a weak statistical behavioral effect ($p = 0.049$). Moreover, the study does not involve a control night, which would have been easy given that patients stay in the EMU for multiple nights. In addition, multiple analytical problems hamper the interpretation. If this work can be replicated in a larger cohort and with suitable controls and appropriate analyses and statistics, then it might open new avenues to studying memory reactivation and the role of a 2-step process, where ripples and spindles play distinct roles.

A: Thanks for the constructive suggestions. We completely agree that the sample size finally used in this iEEG TMR study was relatively small compared with the scalp EEG TMR study. There are many reasons that limited our efforts to collect a large cohort: 1) it was difficult to successfully perform the TMR task for patients. In the present study, we require participants to meet the following conditions: 1) They were able to successfully complete the learning task; 2) There must be no epileptic seizures during the experimental period; 3) The subjects need to

have relatively normal sleep. Due to the difficulty in collecting TMR iEEG data during sleep, the number of participants in intracranial EEG research is generally small. For example, Creery et al.'s intracranial TMR study used 5 epilepsy participants (Creery et al., 2022). Some iEEG studies that focus on the memory also used about 11 epileptic patients (Pacheco Estefan et al., 2019; Treder et al., 2021). Although multiple factors limit the sample size, we will further increase sample size in the future to enhance statistical power.

Comparing TMR sleep with natural sleep (control night) is an important and interesting question. However, in the present study, the hospital attached scalp EEG to the subjects only on the day of the TMR experiment. Therefore, we did not record scalp EEG data during control night. According to the reviewer's good suggestion, in future we plan to conduct scalp EEG and iEEG during TMR sleep and natural sleep.

References:

- Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H., Axmacher, N., & Verschure, P. F. M. J. (2019). Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-09569-0>
- Treder, M. S., Charest, I., Michelmann, S., Martin-Buro, M. C., Roux, F., Carceller-Benito, F., Ugalde-Canitrot, A., Rollings, D. T., Sawlani, V., Chelvarajah, R., Wimber, M., Hanslmayr, S., & Staresina, B. P. (2021). The hippocampus as the switchboard between perception and memory. *Proc Natl Acad Sci U S A*, 118(50). <https://doi.org/10.1073/pnas.2114171118>

A major caveat is how the study is phrased with regard to the literature. First, it narrowly focuses on the systems memory consolidation theory, but then key aspect, such as the dependence of both spindles and ripples on slow wave activity are omitted. How do the reported effects relate to the orchestration of ripples and spindles by the underlying SO phase (for related intracranial work see Staresina 2015/2023 *Nature Neuro* or Helfrich et al., 2019 *Nature Commun.*). It should also be highlighted that systems memory consolidation is a theory, but it is written very much as a matter-of-fact.

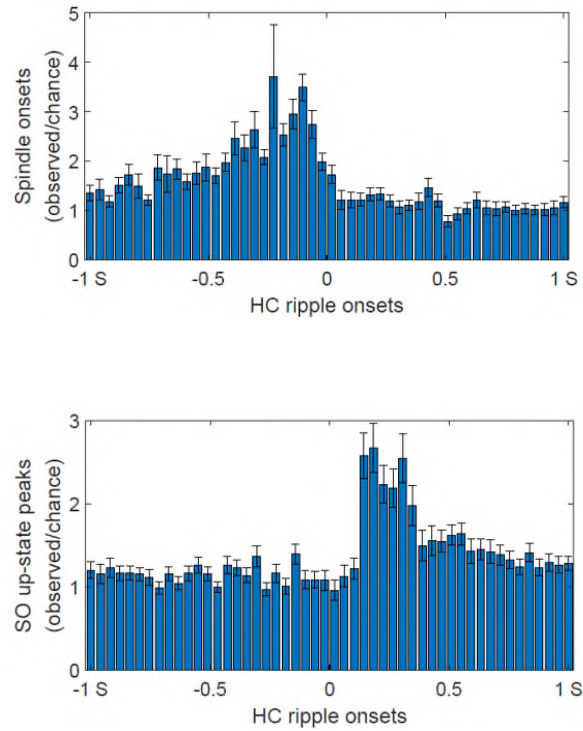
A: Thanks so much for this great comment. As the system consolidation theory proposes, the triple coupling of SO-spindle-ripple is crucial. According to the reviewer's suggestion, here SO was detected based on the algorithm used by Kim and colleagues (Kim et al., 2019). We indeed

found the coupling relationship among SO, spindle and ripple (see Response Figure 15, consistent to previous research; Dickey et al., 2022). Then, SO–ripple events (ripple centers within ± 1 s of SO down-state), SO–spindle events (spindle centers within ± 1 s of SO down-state) and spindle-ripple events (ripple centers within ± 1 s of spindle centers) were also detected based on previous methods (Staresina et al., 2023). However, due to the use of the TMR paradigm (25 TMR cues, each played 6 times during NREM sleep), the iEEG data available for accurately testing this relationship is very limited (only about 15 minutes). Therefore, considering the occurrence rate of coupling events (see Response Table 3), the present study is not suitable for conducting such analysis (the literature mentioned by the reviewer are all studies on spontaneous sleep, and there is sufficient data for the above analysis).

In addition, we have modified the description on systems memory consolidation in the revised manuscript (Page 4, line 126 – 129; Page 10, line 398 - 404).

“The systems consolidation theory hypothesizes that newly encoded memory traces are interactively processed in neocortical and hippocampal networks in order to become stable long-lasting memories (Brodt et al., 2023; Klinzing et al., 2019; Rasch & Born, 2013; Tonegawa et al. 2018).”

“This discovery may be consistent with some assumptions of the system consolidation theory. This theory hypothesized that newly encoded memories are initially distributed in the hippocampus and cortex (Klinzing et al., 2019; Squire, 1992; Squire et al., 1984). Subsequently, in NREM sleep, the new memory needs to be transformed into a stable long-lasting memory through hippocampal-cortical interactions. The hippocampus may play a key role in integrating initial memory information, and transmitting the integrated memory to the cortex (Klinzing et al., 2019; Lewis & Bendor, 2019; Rasch & Born, 2013; Rothschild et al., 2017).”



Response Figure 15 (shown again). Upper: times of spindle onsets plotted relative to ripple onsets during NREM. Below: times of SO up-state peaks plotted relative to ripple onsets during NREM.

Response Table 3 (shown again)

# Subjects	Occurrence rate (events/min)			Trials with coupling events		
	Ripple-SO	Ripple-spindle	Spindle-SO	Ripple-SO	Ripple-spindle	Spindle-SO
1	2.01	0.98	0.95	14%	10%	8%
2	0.88	0.74	1.09	5%	5%	6%
3	2.12	1.14	1.24	16%	8%	10%
4	3.08	1.08	0.96	25%	10%	9%
5	3.33	1.76	1.49	28%	16%	14%
6	4.41	1.84	2.08	37%	15%	19%
7	2.76	1.61	2.45	23%	13%	20%
8	2.00	1.53	0.86	18%	14%	8%
9	3.10	1.14	1.62	26%	10%	13%
10	4.29	2.17	2.10	32%	16%	16%
11	1.96	0.49	0.60	18%	5%	6%
Average	2.72	1.32	1.40	22%	11%	12%

References:

Kim, J., Gulati, T., & Ganguly, K. (2019). Competing Roles of Slow Oscillations and Delta Waves in

- Memory Consolidation versus Forgetting. *Cell*, 179(2), 514-526 e513.
<https://doi.org/10.1016/j.cell.2019.08.040>
- Dickey, C. W., Verzhbinsky, I. A., Jiang, X., Rosen, B. Q., Kajfez, S., Eskandar, E. N., Gonzalez-Martinez, J., Cash, S. S., & Halgren, E. (2022). Cortical Ripples during NREM Sleep and Waking in Humans. *The Journal of Neuroscience*, 42(42), 7931-7946.
<https://doi.org/10.1523/jneurosci.0742-22.2022>
- Staresina, B. P., Niediek, J., Borger, V., Surges, R., & Mormann, F. (2023). How coupled slow oscillations, spindles and ripples coordinate neuronal processing and communication during human sleep. *Nat Neurosci*. <https://doi.org/10.1038/s41593-023-01381-w>

The authors include a lengthy discussion on the comparison of human vs. rodent ripples and cite evidence from Buzsaki et al., but omit more recent work, such as the consensus paper by Liu et al. (2022) in *Nature Commun* or direct comparative evidence by e.g., Schalkwijk et al. (2023) *Progress Neurobiol*.

A: Thanks for sharing the relevant studies. We have cited them in the revised manuscript.

The authors also describe a set of finding that is rather in accordance with non-oscillatory broadband effects (e.g., 16-45 Hz), which are reminiscent of the aperiodic changes observed during wakefulness (Donoghue et al., 2020 *Nature Neuro*) and sleep (Ameen et al., 2024, *bioRxiv*, Lendner et al., 2020, *eLife*) and might be distinct from band-limited activity, in particular during sleep (Bodizs et al. 2024 *Progr Neurobiol*).

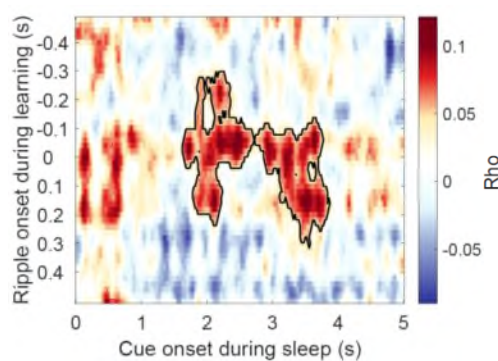
A: Thanks for your valuable comments. We carefully read these literatures and found them to be relevant to our study. As the reviewer pointed out, our results suggest some trends in aperiodic activity beyond classical sleep oscillations (e.g., spindle). Previous studies have demonstrated that non-oscillatory aperiodic activities have different characteristics from periodic oscillations during both wake and sleep periods. According to reviewer's suggestion, we have discussed this point (page 10, line 404 - 411) in the revised manuscript.

“It is worth noting that coupling between hippocampus and cortex not only exists in periodic oscillations (e.g., spindle), but also covers broadband activities (e.g., 16-45Hz). This may suggest the potential value of aperiodic activities in memory consolidation during sleep. Previous studies have suggested that aperiodic activities have different characteristics from

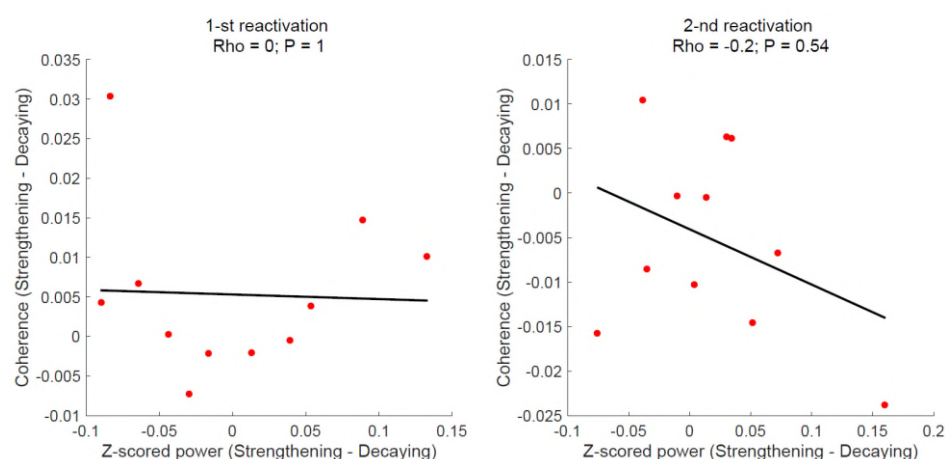
periodic oscillations during both wake and sleep periods (Ameen et al., 2024; Bódizs et al., 2024; Donoghue et al., 2020; Lendner et al., 2020). Therefore, in future research, we can attempt to investigate the characteristics of aperiodic neural activities during sleep memory consolidation process.”

Likewise, some of the similarity effects are likely confounded, given that ripple activity is correlated with high frequency activity that overlaps in its spectral content. In other words, similarity analyses encompass the broadband signal (50-140 Hz) that also contains the ripple itself. How would the results look like if the ripple-band is omitted from the subsequent population analyses, e.g., by only focusing on the 100-140 Hz range? The same concerns hold true for Fig 5 and the reported coherence effects, which are again broadband in nature, and cannot be interpreted if power is also changing.

A: According to the reviewer’s suggestion, we conduct RSA analysis using narrowband signal (100-140 Hz). The analysis shows a very similar RSA pattern as the main result (see Response Figure 26). To exclude the influence of power in connection analysis, we calculated the correlation analysis between connectivity difference and power difference (see Response Figure 27), indicating that the difference in coherence between two conditions was not related to spectral power.



Response Figure 26: Rho between similar patterns (using narrowband signals :100-140 Hz) and behavioral performance.



Response Figure 27. Spearman correlation between the difference of coherence and difference of power.

Per the reported ripple intervals – what is the min. inter-ripple interval, in its present form Fig 2E indicates that these follow in succession at very short delays? How do the authors explain the mismatch between the number of spindles and ripples, spindles exceed by a factor of ~ 10 , while their key result in Fig. 4F/G describes a very similar ‘normalized’ rate? It’d be more helpful to see the absolute rate to judge the true behavioral relevance.

A: The min. inter-ripple interval was 31ms. The present study used steps of Norman et al.’s study for the detection of ripple events (Norman et al., 2019; Norman et al., 2021). Ripples with an interval of more than 30ms are defined as two independent ripple events. The distribution of inter-ripple intervals shown in Fig. 2E was consistent with previous studies (Norman et al., 2019; Norman et al., 2021).

We have reported absolute rate of ripples and spindles (Fig. S7, Table S3). The occurrence rate of ripples is significantly higher than that of spindle. The reason for using normalized rate was mainly due to individual differences in occurrence rate of spindles and ripples. Specifically, we normalized the occurrence rate after the cue onset divided by the average occurrence rate at each contact for each subject.

References:

Norman, Y., Raccach, O., Liu, S., Parvizi, J., & Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron*, 109(17), 2767-2780 e2765. <https://doi.org/10.1016/j.neuron.2021.06.020>

Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454). <https://doi.org/10.1126/science.aax1030>

The writing is sub-optimal in several instances and not very precise. For example, in the abstract very general terms are used ('much remains to be learned') or 'to address this issue' (which is not precisely defined), or unclear ('TMR produces nonuniform amounts') or in the results 'Patients ... were employed'

A: Thanks for pointing out this. We have carefully checked the revised manuscript.

Minor:

- It would be desirable to have more information about the precise experimental timing. Given the clinical realities, this was probably rather variable when training and testing took place

A: According to the reviewer's suggestion, we have added precise information about the experimental timing in the "Method" part of the revised manuscript (page 12, line 497 - 498) as follows:

"The average time spent by participants in the experiment as follows: learning, 29.9±9.4 minutes; pre-test, 5.9 ± 0.6 minutes; post-test, 5.8 ± 1.1 minutes."

- Fig. 2F should include a baseline period prior 0 to fully evaluate the temporal evolution of the described effect

A: Fig. 2F showed the occurrence of ripples during the learning stage. There was no clear interval between learning trials (i.e., the next picture is displayed once the feedback disappeared). Even if a baseline period prior 0 is selected, its activity is closely related to the previous trial and cannot be used as an ideal baseline. So, we use the average rate throughout the entire learning period as the baseline.

- Fig. 3C constitutes double-dipping, since the statistical effect has already been established in Fig 3B

A: Fig. 3C is mainly used to support Fig. 3B. Fig. 3B only displayed the results at the group level, while Fig. 3C further displayed the results of each subject.

Reviewer #3:

The topic of the paper is highly relevant and important. The authors can be congratulated for attempting to conduct a challenging cognitive paradigm that requires multiple testing sessions and a monitored sleep session in a bedside setting. Not only did they study and report the behavioral effects of the nap-intervention experiment, but also aimed to understand the electrophysiological signatures of these effects. These signatures are highly relevant to our understanding of how TMR influences sleep-associated declarative memory consolidation and how this intervention can uncover natural mechanisms of memory consolidation during sleep. However, I do have some concerns regarding the data analysis and interpretations that I detailed below, and I hope the authors would be able to answer to allow the reader to judge their work.

We appreciate the reviewer's positive feedback on our study and the valuable suggestions. We have carefully addressed all concerns, and performed additional analysis to confirm our findings. Please find below our point-to-point responses to all the queries and suggestions.

Major

Subjects learned to associate 50 unique items with specific locations on a grid, each paired with a characteristic sound. The authors report a weak memory-improvement effect for TMR, which is expected based on the small cohort that iEEG research enables and also the difficulty of controlling all the experiment's parameters in a bedside setting. That said, since the experiment is performed out of the lab, it is critical to report the details clearly and acknowledge the inherent variability (i.e. auditory cues delivered in NREM-1-3, the sleep phases on which they were delivered etc.).

A: Thanks for the constructive comments. We fully agree with the reviewer's view that the non-laboratory environment limited the experimental effect. When conducting experiments in the hospital, it is indeed difficult to control factors such as the sleep stages (NREM 1-3) and the phases (up/down states) of the SO in which the TMR cues were delivered. In response to the above limitations, we have added a limitation section in the revised manuscript. In future, it's worth implementing strict controls, such as closed-loop TMR.

TMR literature, performed based on non-invasive EEG now covers a great deal of options that vary the efficacy of the intervention, one of the most notable are the differences of the intervention when targeting up/down states of slow wave oscillation (for example Ngo, PNAS 2022) and different relationships regarding timing of cues relative to spindle occurrence (Antony, Curr Biol., 2018). I believe the manuscript will benefit from a breakdown of the timings of auditory cues during sleep relative to sleep oscillations, as these can make a great difference concerning the intervention's effect on memory, and also – regarding the electrophysiological analysis below – these will impact the frequency profile that is expected after auditory cueing that will determine the expected baseline of sleep oscillations (spindles and ripples) during the re-activation periods (Fig. 4 F-G).

A: Thanks for these great suggestions. We reported the sleep phases in which the TMR cues were delivered in Table S3. In addition, we detected the SO events based on the algorithms used by Kim and colleagues (Kim et al., 2019). According to the reviewer's suggestion, we picked out up-state TMR cues which fell around SO up-state peaks ($\pm 45^\circ$) and down-state TMR cues which fell around SO down-state peaks ($\pm 45^\circ$). Only a small portion ($\sim 14\%$) of the sound cues fell in a specific phase (up/down) of the SO (see Response Table 4). So, the present study was not suitable for conducting such analysis (e.g., breakdown of the timings of auditory cues during sleep relative to the specific phase of SO). Since the two references mentioned by the reviewer are the studies using a closed-loop stimulation, they are very suitable for this type of analysis.

Response Table 4. The sleep phases in which the TMR cues were delivered

# Subjects	Total cue numbers	TMR cues		Control cues	
		SO up-state	SO down-state	SO up-state	SO down-state
1	150	7	6	7	7
2	200	6	4	7	6
3	300	11	13	15	10
4	300	8	7	12	8
5	300	4	6	7	3
6	300	15	11	12	13
7	300	13	3	4	13
8	300	19	10	17	7
9	300	14	13	10	12
10	300	4	6	6	3

11	300	11	9	10	5
Average	277	10.2	8.0	9.7	7.9

There is also variability in the sleep-period the cues are delivered on, it would make sense to focus on auditory cues that are delivered on the same sleep-phase when conducting the analysis regarding spindle and ripple rates (Fig. 4 F-G) as their baseline probability differs between different sleep-phases.

A: Firstly, during NREM sleep, a randomized stimulus sequence including 25 TMR cued sounds and 25 control sounds (all guitar strum) was presented. All 50 sounds are randomly played one round before the next round. Therefore, this presentation mode can ensure that TMR cues of two conditions (strengthening vs. decaying) can be delivered in the same sleep-period as much as possible (see Response Table 5).

Secondly, because our analysis mainly focused on comparing the differences between two conditions (strengthening vs. decaying). Then, we counted the trial number of strengthening and decaying conditions during each NREM sleep period (Response Table 5), and did not find significant difference between the two conditions (all $p > 0.2$, permutation test). This result indicated that the baseline levels of spindle and ripples rates during different sleep-phases cannot affect our results. In the future, we would ensure that TMR cues are delivered in the same sleep period through a closed-loop method.

Response Table 5 (identical to Table S3). Ripple and spindle occurrence rate and TMR information

Subject ID	HC ripple	HC ripple	Cortical spindle	# The number of TMR cues: total (Strengthening, decaying)				
	events/min (learning)	events/min (TMR)	events/min (TMR)	All	N3	N2	N1	Wake
1	23.8	25.9	2.86	75 (27,27)	75 (27,27)	0	0	0
2	27.9	23.9	4.73	100 (36,36)	88 (33,31)	12 (3,5)	0	0
3	22.7	21.9	3.38	150 (54,54)	134 (48,49)	12 (4,4)	4 (2,1)	0
4	31.6	28.7	5.20	150 (54,54)	106 (37,38)	44 (17,16)	0	0

5	24.1	27.2	4.67	150 (54,54)	150 (54,54)	0	0	0
6	30.3	25.7	3.64	150 (54,54)	109 (41,39)	41 (13,15)	0	0
7	5.6	8.7	4.63	150 (54,54)	44 (17,15)	91 (30,36)	7 (3,1)	8 (4,2)
8	12.4	14.2	4.94	150 (54,54)	83 (32, 32)	61 (20,22)	0	6 (2,0)
9	32.0	26.1	4.31	150 (54,54)	92 (31,35)	58 (23,19)	0	0
10	28.6	19.8	3.50	150 (54,54)	0	140 (50,51)	10 (4,3)	0
11	25.0	22.9	3.96	150 (54,54)	0	78 (27,30)	69 (25,24)	3 (2,0)

HC: hippocampus; N1: NREM stage 1; N2: NREM stage 2; N3: NREM stage 3.

Strengthening and decaying sub-group are defined per subject, such that a fixed number of items are assigned to each sub-group. It is important to point out that there is a significant difference between the performance of these sub-groups before sleep and that without proper control, these differences can be the driving force of later effects. This point is laid out thoroughly in the discussion section and a control analysis was added as a supplementary figure.

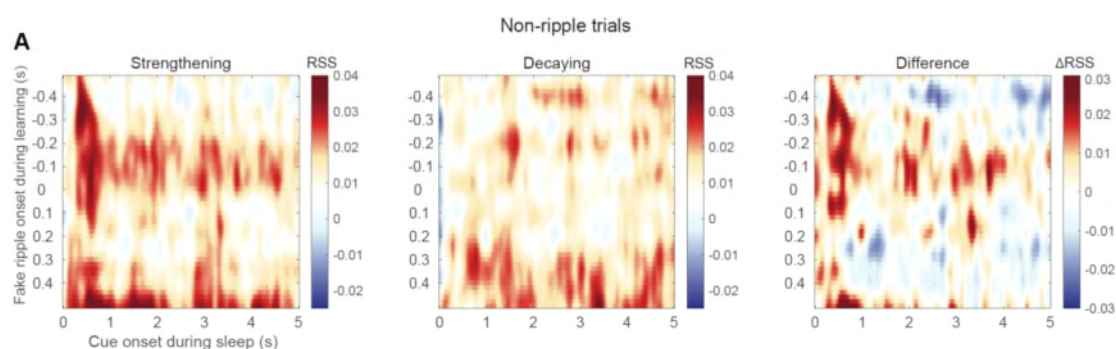
A: Thanks for this valuable suggestion. As the reviewer mentioned, the memory performance of the two sub-groups was different during pre-sleep test. Hence we indeed conducted a control analysis on the difference in pre-sleep memory accuracy (see Fig. S12), and discussed this point in the discussion section of the revised manuscript (page 7, line 242 - 247).

“In addition, we tested whether the reactivation results were driven by differences in pre-sleep accuracy of strengthening vs. decaying items, and compared reactivation of two groups of items determined entirely by pre-sleep accuracy but matched for memory change score. RSS results showed no significant reactivation differences between such two groups (Fig. S12), indicating that higher reactivation of strengthening items cannot be explained by differences in pre-sleep accuracy.”

Item-specific cortical pattern around hippocampal ripples during learning: [row 200-210] The authors investigated whether item-specific neural representations can be detected around the occurrence of ripples during the retrieval phase. The natural question is whether these specific

representations are unique to ripple-locked analysis or are these a limited window into a process of re-instantiation that is related to the time-period following the stimuli exposure but not unique to the selected ripple windows. It would be helpful to perform the same analysis on ripple-less epochs to understand whether this item-specific similarity arises specifically around ripple times. Since a wide-time window is used, and ripple-rate peaks around 0.75 after stimuli presentation, this analysis correlates the post-presentation epochs between items.

A: Thanks for your constructive suggestions. We perform the same RSA analysis using those non-ripple trials (using the time window 0.75 ± 0.5 s after picture onset) during the learning stage, and did not find significant difference between the two conditions ($p = 0.058$). In addition, in order to clarify the question of whether these specific representations are unique to ripple-locked analysis, we further tested the relationship between the learning ripples and subsequent reactivations during sleep. We selected trials without ripple events during the learning (149 ± 62 no-ripple trials) and investigated the reactivation of their patterns during sleep. The difference between strengthening and decaying items is consistent with the trend of ripple trials, but it has not reached statistical significance (see Response Figure 28). The results above suggest that there still exists item-level neural representations in non-ripple trials, but the intensity is lower than in ripple trials.

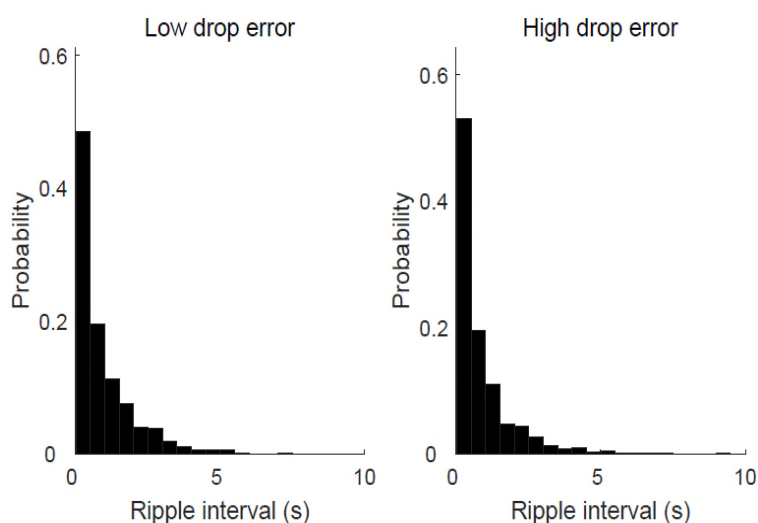


Response Figure 28 (identical to Figure S6). *Control RSA analyses using non-ripple learning trials.*

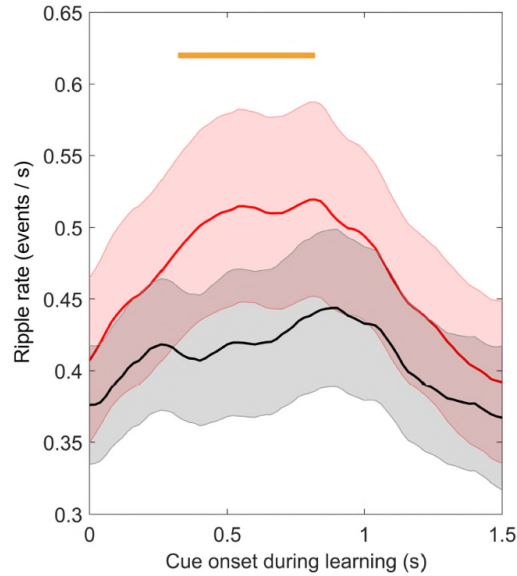
Can the effect (Fig. 2E) can be alternatively attributed to differences in memory-processing, i.e.
– whether the baseline pre-sleep error of the item is high or low?

A: If all items during learning stage are divided into high and low drop-error groups according to the average pre-sleep error, we could compare their inter-ripple interval. However, very few learning trials have more than one ripple (65% of trials have only one ripple) due to the low occurrence rate of ripples. According to the reviewer's suggestion, we selected trials with two or more ripples (35% of all trials). Then, we divided them into high and low drop-error groups, and calculated their inter-ripple intervals (see Response Figure 29). No difference was found between the two groups on each participant (all $p > 0.1$, permutation test).

In addition, we speculate that the reviewer's comment actually refers to Fig 2F. Therefore, we conducted another control analysis. To test whether the ripple rate during learning is related to successfully remembering the location of the item, we divided the learning trials into two groups based on the drop error in pre-sleep test (Response Figure 30): low error group (red line) and high error group (black line). As the reviewer expected, the result indeed showed that ripple rate of the low error group was significantly higher than those of the high error group.



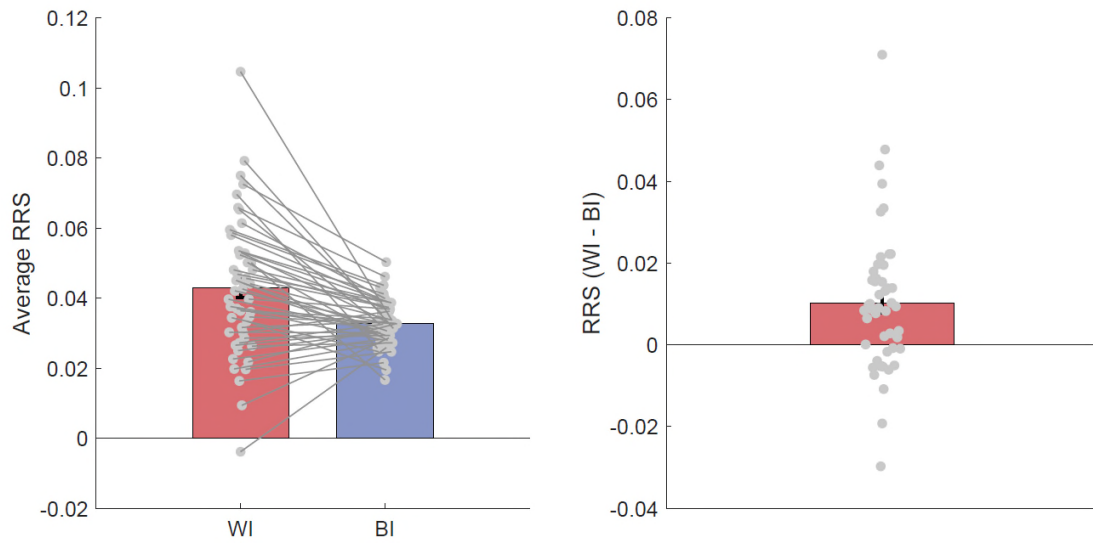
Response Figure 29. Distribution of inter-ripple intervals in high and low drop-error trials.



Response Figure 30. *Ripple rate during the learning stage. We divided the learning trials into two groups based on the drop error in pre-sleep test. Yellow bar indicates the time period of significant difference between two groups.*

I think the item-specific title is misleading because the analysis lacks the effort to characterize whether a minority of items that have a similar neural signature (for example – those that were learned efficiently or those that patients found very interesting) dominate the effect, while most of the others do not. Without a proper control or deeper analysis of whether the average represents the distribution of WI-BI matrices (via a shuffling procedure), the results suggest weak evidence for enhanced similarity for same-item representations that are specific to ripple periods following stimuli presentation.

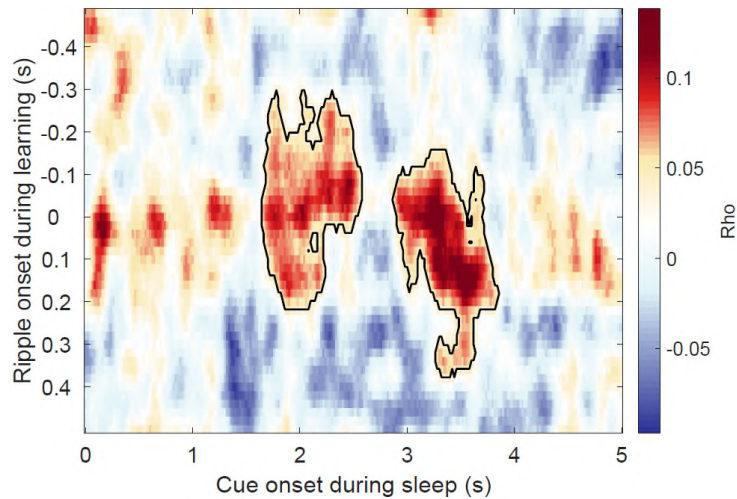
A: In the present study, “item specific” refers to the comparison of within-item (WI) and between-item (BI) representation similarity, which did not indeed characterize whether a minority of items dominated the effect. In addition, we conducted a control analysis to test whether a minority of items that have a similar neural pattern dominate the effect. We compared the WI similarity with BI similarity for each item, and the results showed that 72% of items showed greater WI than BI (Binomial distribution test, $p < 0.001$; see Response Figure 31). In general, it means that this effect did not arise from a minority of items. To avoid reader misunderstandings, we have changed “item specific” to “item level” in the revised manuscript.



Response Figure 31. RRS averaged across subjects (each point represents an item). Left: the WI similarity and BI similarity for each item; right: 72% of items shows greater WI than BI.

It would also be helpful to understand whether all learning rounds and all cortical contacts (gray/white matter) were included in this analysis? Some contacts seem to be outside the brain in the figure (Fig. 1G).

A: In the present study, we included all learning rounds and all cortical contacts (gray/white matter) in this analysis. We also carefully examined the position of each contact in the native brain space and found that few contacts were located at the edge of the cortex (4.9% of the contacts). Therefore, when normalized to the standard MNI space, the electrode contacts appeared outside the brain. In order to eliminate the influence of these contacts, we removed them and calculated our main results again, and found that it had no impact on the RSA result (see Response Figure 32).



Response Figure 32. Average ρ between RSS matrix (exclude the contacts outside the edge of cortex) and behavioral performance. After removing the contacts outside the brain, we found very similar result.

Two behavioral-relevant item reactivations during sleep:

The authors used an analysis approach that searches for similarity between neural activity that arises during ripple activity during learning to that that arises during auditory cueing during sleep and then asks whether this similarity is predictive of memory consolidation. While this is a fundamental question that is interesting and important to investigate, I am not convinced that the methods chosen can lead to the suggested conclusions. High-frequency ripples are brief oscillatory events, but here the analysis is chosen using a wide window $[-0.5, 0.5]$ s around ripples, making the reader wonder whether the activity is related to ripple occurrence at all. One would wonder if an analysis that correlates the activity time-locked to image presentation during learning would reveal similar results or is this a property of ripple-occurrence? These time periods may be confounded as ripples seem to be occurring more at $[0.5-1]$ s after onset during learning (Fig. 2F) so important to acknowledge this confounding factor. This concern is backed up by the subject-specific reactivation plots presented in Figure S-4 – the rightmost cluster (Fig. 3B) seems to be common to 4/11 patients and the left cluster to 3/11 so there's quite a large variability in the specificity of the time-window around ripples and after cue-presentation that isn't addressed in the main analysis presented. As Figure s-5 reveals – 7/11 patients' similarity matrices do not align with the suggested process in the main text – namely,

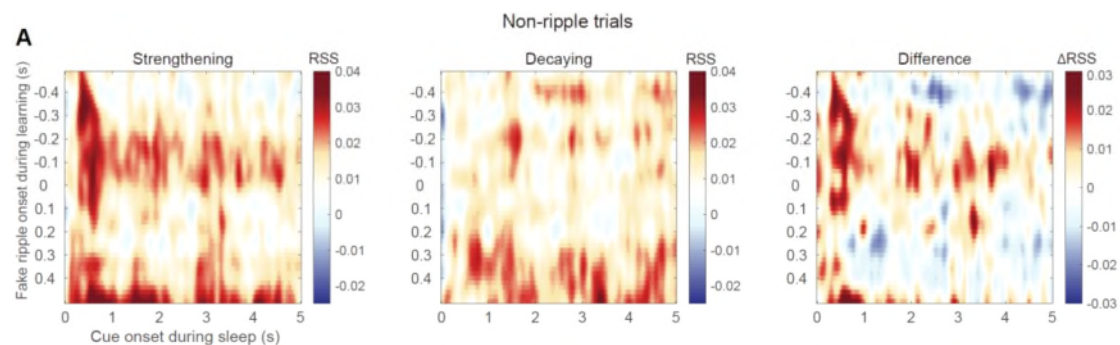
they either have a single reactivation cluster, or very small clusters that may be due to spurious stats. Is the double-reactivation unique to patients whose memory benefited from auditory-cueing (those with a larger difference between cued and un-cued items, Fig 2L)? For transparency, it is important to acknowledge these limitations in the main text where these analyses are mentioned but not described (line 225).

A: Thanks so much for such great suggestions. To verify the relationship between the learning ripples and subsequent reactivations during sleep, we selected trials that have no ripple events during the learning (149 ± 62 no-ripple trials) and investigated the reactivation of their patterns during sleep. For example, there were two learning trials of one item. There was a ripple event at 0.75 second after the image onset in trial 1, while there was no ripple in trial 2. We created a fake ripple event for trial 2 at 0.75 second after the image onset. Then we repeated our RSA analysis just using trials of these fake ripple events. The difference between strengthening and decaying items is consistent with the trend of ripple trials, but it has not reached statistical significance (see Response Figure 28). In addition, we analyzed the relationship between the double-reactivation pattern and patients' TMR effect (memory benefited more from auditory-cueing), and the result showed no significant correlation between the them. That is to say, patients with more pronounced TMR effects did not show more stable double-reactivation patterns.

For intracranial EEG recordings, there are large differences in the number and position of electrodes across patients, which certainly limits the stability of RSA analysis. In addition, the uncontrolled non-laboratory environment also amplified the variation across patients. When conducting experiments in the hospital bedside in this study, we did not control factors such as the phases (up/down states) of the SO when the TMR cues were delivered. The above factors could affect the stability of RSA analysis at the participant-level in intracranial EEG research. For the limitations of reactivation pattern stability (two reactivations or single), we added a “limitations” section in the Discussion of the revised manuscript (page 11, line 445 - 451) as follows:

“For intracranial EEG recording, there are differences in the number and position of electrodes for each subject, which limits the stability of RSA analysis (especially for item-level

similarity analysis). In addition, when conducting experiments in the bedside, it is difficult to control factors such as the sleep stages (NREM 1-3) on which the TMR cues were delivered and the phases (up/down states) of the SO on which the TMR cues were delivered. Considering the limitations, in future, we will use more sophisticated recording method (such as single-neuron recording) and close-loop TMR intervention to test the stability of multiple reactivations phenomena.”



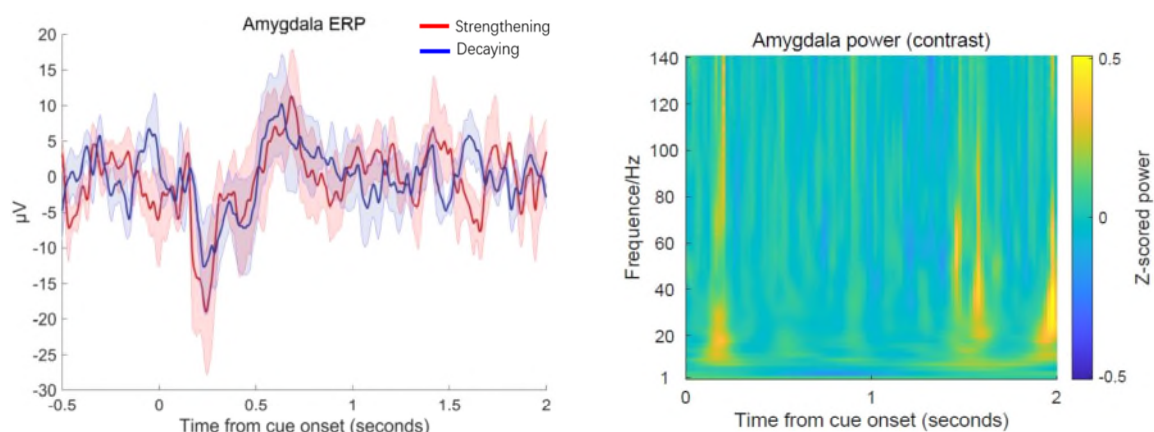
Response Figure 28 (shown again, identical to Figure S6). Control RSA analyses using non-ripple learning trials.

Fig 4 D-E – Here the authors test their hypothesis by dividing the analysis to two sub-groups of cued items – those that benefited from cuing (strengthened memory) and those in which memory decayed. They compare the double-reactivation correlation maps between these two groups and show a significant difference. My concern is that these groups' differences are predetermined before the TMR intervention. Based on Figure S1 – there is a significant difference between the pre-sleep performance of these two item groups so perhaps there are additional differences between them (visual attributes, emotional valence, etc.) that contribute to the reported effects.

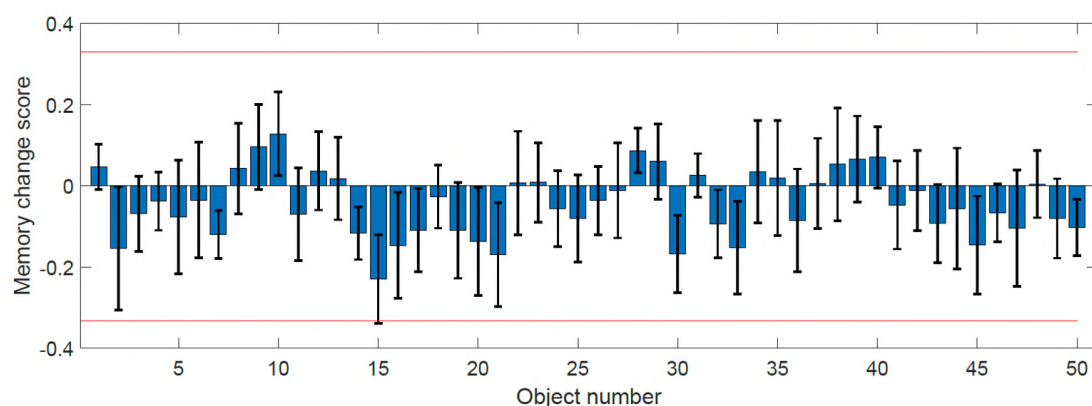
A: Thanks for your careful inspection. In this study, we did not explore the potential impact of emotional valence. We attempt to eliminate the potential impact from the following aspects. Firstly, the experimental paradigm and stimulus materials used in this study were derived from previous studies (Rudoy et al., 2009; van Dongen et al., 2012, Creery et al., 2022). In van Dongen et al.'s 2012 MRI study (covering the whole brain), no activity was found in emotion-related brain regions such as the amygdala (van Dongen et al., 2012). To further eliminate the

potential impact of emotional valence, we conducted a new control analysis. We calculated the ERP and power response after stimulus presentation during the learning phase on 5 subjects including amygdala contacts (see Response Figure 33). The permutation test results did not show significant differences between the two conditions (strengthening and decaying).

In addition, to further control for the influence of additional variables (e.g., visual attributes, emotional valence), we calculated memory change score for each stimulus. We hypothesize that if these additional variables contribute to the reported effects, then a minority of items would dominate the experimental effect. The results showed that memory change scores were not specific for some items across participants (see Response Figure 34).



Response Figure 33. ERP/power responses of Amygdala during learning phase.



Response Figure 34 (identical to Figure S1c). Memory change score of each item. The red lines generated by randomized items positions (1000 times shuffles) are the threshold value of

$p < 0.05$. This result shows that the variability in memory change is not consistent for specific objects across patients. *** $p < 0.001$.

References:

- Creery, J. D., Brang, D. J., Arndt, J. D., Bassard, A., Towle, V. L., Tao, J. X., Wu, S., Rose, S., Warnke, P. C., Issa, N. P., & Paller, K. A. (2022). Electrophysiological markers of memory consolidation in the human brain when memories are reactivated during sleep. *119*(44), e2123430119. <https://doi.org/doi:10.1073/pnas.2123430119>
- Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, *326*(5956), 1079. <https://doi.org/10.1126/science.1179013>
- van Dongen, E. V., Takashima, A., Barth, M., Zapp, J., Schad, L. R., Paller, K. A., & Fernández, G. (2012). Memory stabilization with targeted reactivation during human slow-wave sleep. *Proceedings of the National Academy of Sciences*, *109*(26), 10575-10580. <https://doi.org/doi:10.1073/pnas.1201072109>

What are the differences in the timing of the auditory cues between these groups? Can the authors plot the frequency profiles time-locked to auditory cues resulting in enhanced/degraded cues and comment on how the differences between these profiles contribute to the presented results? These differences can account for the differences between ripple rate and spindle rates (Fig. 4 F-G) as their baseline rates vary between the up/down phases of the slow oscillations. The definition of the control condition of using non-learning-ripple trial data (awake TMR) was unclear.

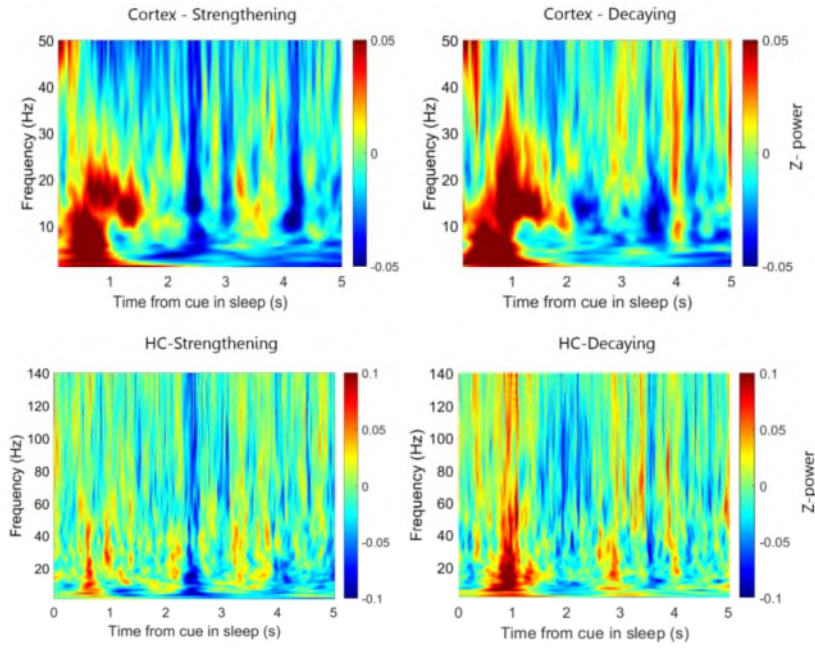
A: Thank you for this constructive suggestion. Firstly, we analyzed the timing of auditory cues and found that a very small number of cues fell on the up/down state of SO (see Response Table 4). Therefore, due to the limited number of trials located on up/down state of SO in the present study, it is difficult to explore the impact of this difference on the results. Secondly, we plot the spectrograms time-locked to auditory cues resulting in strengthening/decaying cues (see Response Figure 35). The differences between these profiles were shown in Figure S8 in the manuscript (see Response Figure 36). The spectrogram differences showed similar trends with the ripples/spindle detection analysis. The ripple band activity differences (strengthening>decaying) during the first reactivation were obviously higher than the subsequent second reactivation. On the contrary, the spindle activity differences

(strengthening>decaying) during the first reactivation is obviously lower than the subsequent second reactivation.

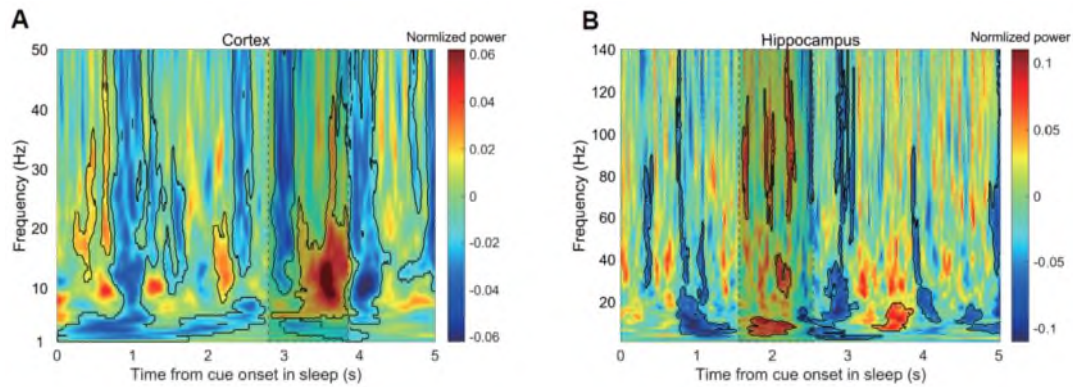
In addition, in the revised manuscript, we rephrased the definition of control condition as follows: “In order to verify the relationship between the learning ripples and subsequent reactivations during sleep, we selected trials without ripple events during the learning (149 ± 62 no-ripple trials) and investigated the reactivation of their patterns during sleep. For example, there were two learning trials of one item. There was a ripple event at 0.75 second after the image onset in trial 1, while there was no ripple in trial 2. We mark a fake ripple event for trial 2 at 0.75 second after the image onset. Then we repeated our RSA analysis just using trials of these fake ripple events.”.

Response Table 4 (shown again)

# Subjects	Total cue numbers	TMR cues		Control cues	
		SO up-state	SO down-state	SO up-state	SO down-state
1	150	7	6	7	7
2	200	6	4	7	6
3	300	11	13	15	10
4	300	8	7	12	8
5	300	4	6	7	3
6	300	15	11	12	13
7	300	13	3	4	13
8	300	19	10	17	7
9	300	14	13	10	12
10	300	4	6	6	3
11	300	11	9	10	5
Average	277	10.2	8.0	9.7	7.9



Response Figure 35. Spectrograms depicting the mean activity for the strengthening and decaying conditions. The upper part shows cortical spectrograms, while the lower part shows hippocampal spectrograms.



Response Figure 36 (identical to Figure S8). Difference of power spectrograms of cortex and hippocampus during TMR.

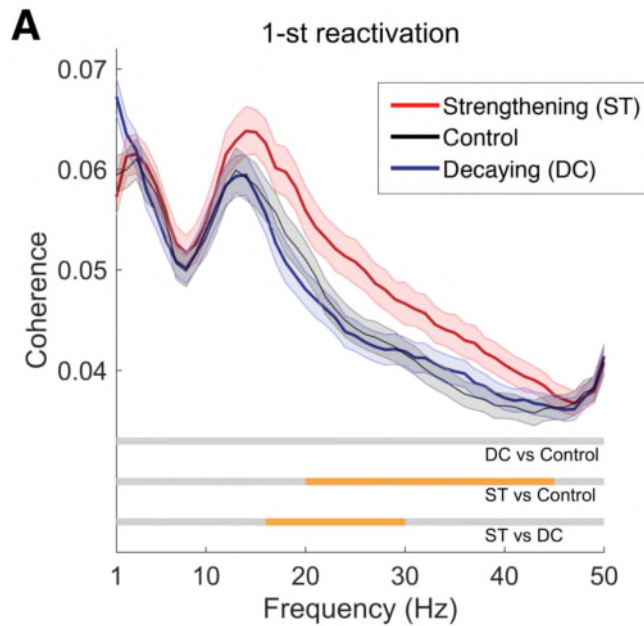
(A) Difference of cortical power ($p < 0.01$, cluster-based permutation test) between memory strengthening and decaying condition. The gray shaded area marks the time range of the second reactivation (2.8-3.8 seconds). (B) Difference of hippocampal power ($p < 0.01$, cluster-based permutation test) between memory strengthening and decaying conditions. The gray shaded area marks the time range of the first reactivation (1.6-2.6 seconds).

Dynamic hipp-cortical synchronization during two reactivations:

Continuing my comment for the previous section – as previous literature has identified dynamic changes in the level of coordinated activity between neocortical and hippocampal areas during sleep (Helfrich et al, Nat. Commun 2019, Sanda et al, Cereb Cortex, 2021, and others) it is very important to characterize the baseline coherence of cue-less sleep-periods which are matched in their distance from slow-oscillation (or separated based on sleep-phases in which they were delivered) to the cued-periods used in the coherence analysis to understand the baseline without cueing. The null hypothesis would be that the timing of the cue is the driving force for most of the results presented in Fig. 5.

A: In order to test the null hypothesis (the timing of the cue is the driving force for most of the results), we calculated and compared the distribution of TMR cues and control cues on the up/down state of slow-waves, as shown in Response Table 4, and there was no significant difference between them (all $p > 0.7$). This indicates that there is no significant difference in the distance from slow-oscillation between the TMR cued period and control period in the present study.

In addition, in the present study, the reason why we used control sound (guitar strum) periods to calculate the baseline coherence rather than the cue-less sleep-periods, is mainly because previous studies have found that sound stimulation (pink noise) can also significantly affect slow oscillation and spindle activity, but does not affect behavioral performance (Henin et al., 2019). This result suggests that simple meaningless sounds may also affect brain oscillation, but do not have the memory enhancement effect shown by TMR cues. Therefore, in the present study, we used meaningless sound (guitar strum) as a high-level control condition. If cue-less sleep-periods are used as a control condition, it may potentially cause confounding effect among factors.



Response Figure 37 (identical to Figure 5a). Coherence between cortex and hippocampus under strengthening, decaying and control condition (guitar strum).

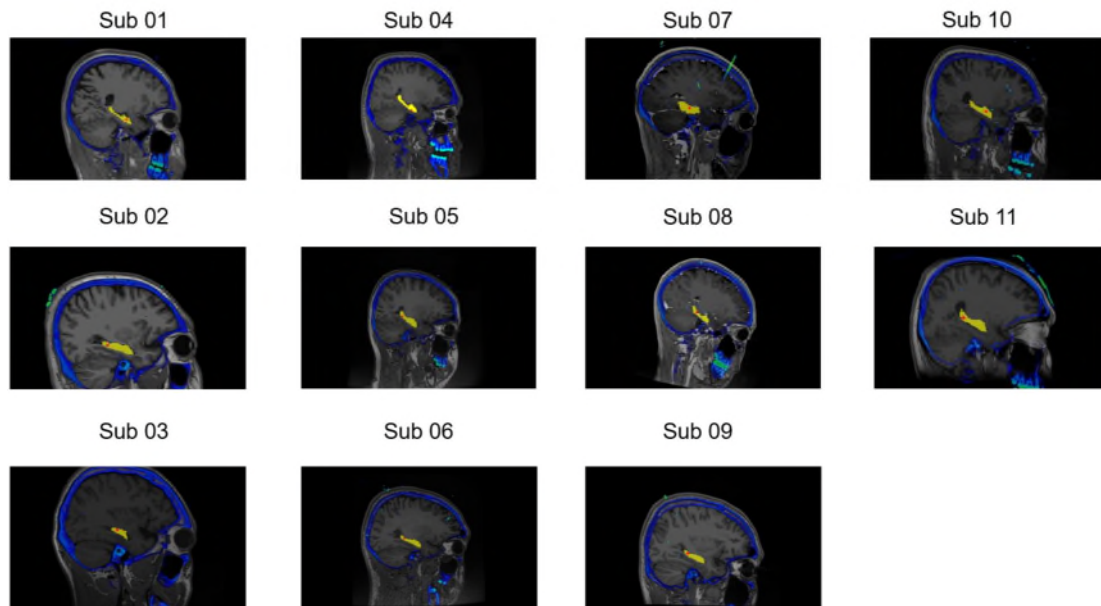
References:

Henin, S., Borges, H., Shankar, A., Sarac, C., Melloni, L., Friedman, D., Flinker, A., Parra, L. C., Buzsaki, G., Devinsky, O., & Liu, A. (2019). Closed-Loop Acoustic Stimulation Enhances Sleep Oscillations But Not Memory Performance. *eneuro*, 6(6). <https://doi.org/10.1523/eneuro.0306-19.2019>

Minor

Fig 1A – it is extremely difficult to understand the locations of contacts based on the current panel's size.

A: Thanks for the careful inspection. In order to show the locations of hippocampal contacts clearly, we increased the figure size. In addition, we drew brain maps for each participant to clearly display the location of the representative hippocampal contacts (see Response Figure 10).



Response Figure 10 (show again). The hippocampal contact marked in the red dots (used for ripples detection) in native patient space. The yellow area represents the hippocampus region.

Reference to Fig. 4A-C appears in text after discussing Fig. 4D-E. Fig. 4D-E relates to analysis in Fig. 3 rather than in Fig. 4 so is a bit confusing for reader to find.

A: Thanks for pointing out this. We have redrawn Figure 4 in a correct order.

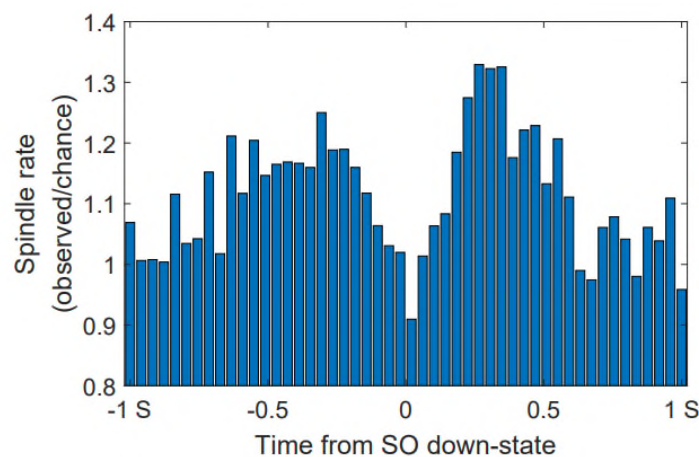
Fig S-6 legend – the authors refer to awake TMR – but write that sounds were cued during NREM sleep.

A: We do apologize that our description may have caused potential confusion for the readers. We have modified the description in the legend of Figure S6 in the revised manuscript (Page 35, line 1122 – 1124).

“We employ learning trials (i.e., learning trials of items that were cued during sleep) without ripples to calculate the RSS of the strengthening and decaying conditions (i.e., pattern similarity between non-ripple learning trials and corresponding NREM sleep cued trials).”

Fig. 4C – spindle activity is usually paired with slow-oscillations, and the fact that the slow wave activity is very low here makes the reader wonder maybe the majority of the spindles are detected in early NREM sleep or mixed with IED activity?

A: Thanks for this great comment. Fig. 4B, C (Fig. 4D, E in the revised manuscript) are mainly used to demonstrate typical ripples and spindle activities. Thus, the color bar used in the figures is to highlight these two as obvious as possible. To verify the coupling between SO and spindle, we detected SO and calculated the number of spindles relative to SO down-state (see Response Figure 38, consistent to the previous research; Staresina et al., 2023). This shows the coupling relationship between SO and spindle in our study.



Response Figure 38. The spindle rate relative to SO down-state during NREM sleep.

References:

Staresina, B. P., Niediek, J., Borger, V., Surges, R., & Mormann, F. (2023). How coupled slow oscillations, spindles and ripples coordinate neuronal processing and communication during human sleep. *Nat Neurosci.* <https://doi.org/10.1038/s41593-023-01381-w>

Response letter to comments

Reviewer #1:

The authors responded to all of my comments, which is appreciated. However, the authors responded to many of my comments without including the information in the manuscript (or supplements) which I do not find helpful. The point of these comments is to provide more detailed information to potential readers, not only the reviewer. Therefore, please include Response Figures 1, 2, 4, 6, 12, 15, 16, 20, 22, 23, 24 with relevant details on analyses statistics as well as all of the Response Tables in the manuscript /supplements and refer to them in the main text. The following responses by the authors need further clarification:

We thank the Reviewer for the positive feedback on our first-round response. In this round, we have made the following key revisions:

1. In accordance with your suggestion, we have included the referenced Response Figures and Tables in the main text and supplementary materials (Figure S2, S3, S4, S5, S10, S11, S16, S21, S22; Tables S4, S5, S6).
2. In response to Comment 1.2, we analyzed the periodic characteristics of reactivation similarity time course following the sound played during NREM and observed strong low-frequency oscillations in each subject. This further supports the existence of multiple reactivations.
3. In response to Comments 2.1-2.5, we reanalyzed the hippocampal ripples using the reviewer's suggested criteria (filter to 80–120 Hz, at least 3 cycles, all hippocampus contacts) and replicated the reactivation results.

Response to 1.2:

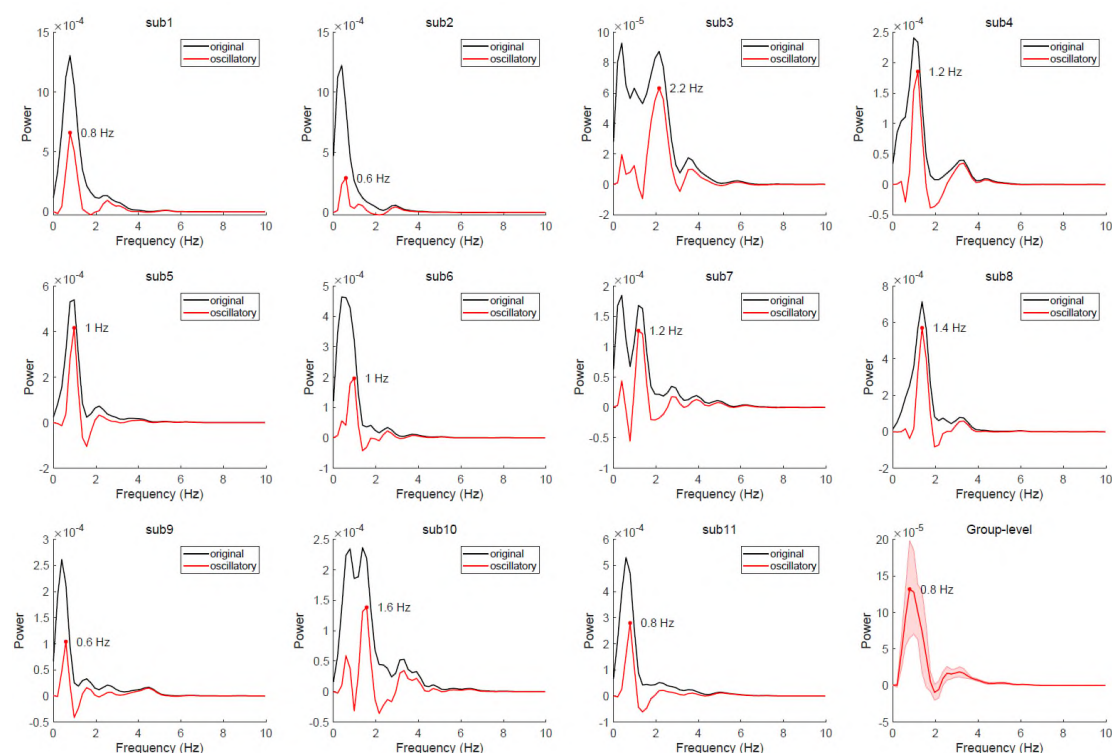
Rather than general differences in laboratory vs. bedside setups, I was interested in how the inconsistent patterns on the single patient level can result in a consistent group-level effect which remains unclear to me.

A: Firstly, we would like to sincerely thank you for recognizing the general differences between laboratory and bedside setups, which primarily stem from two key factors. One is the inherent variability of iEEG data, where the electrode contacts across brain regions can largely differ between patients. The second factor is the relatively small sample size for iEEG studies. We believe that these two aspects are the primary reasons for the inconsistent patterns observed at the single-patient level in the current dataset. We also addressed these limitations in the discussion.

However, we are pleased to report that, despite these challenges, we still observed the consistent patient-level effect in the dataset. We apologize for not clarifying this in the manuscript and in the first round response. In this revision, we provide a more detailed explanation. The hypothesis of our study is grounded in the memory consolidation echo model (Lewis & Bendor, 2019). This model posits that the hippocampal-cortical loop forms a dynamic system that relies on iterative information exchange between the hippocampus and cortex, which is thought to enhance memory traces through hippocampal-cortical replay. The theory predicts that, in a TMR paradigm, multiple reactivations could occur after cue presentation. Several empirical studies support this theory. For example, in a rat experiment, the researchers trained rats on an auditory-spatial association task while recording neuronal activity in the hippocampus. They found that a cue-evoked increase in hippocampal replay lasted for 10 seconds, until another acoustic cue was presented (Bendor & Wilson, 2012). A human scalp EEG study also found that

neural activity patterns associated with learned vocabulary word pairs were reinstated when the cue words were played back during NREM sleep. Interestingly, the reactivation occurred up to three times after the TMR cue, echoing back roughly once per second (Schreiner et al., 2018). Furthermore, we know that slow waves play a role in modulating memory consolidation. Given the intervals between reactivations reported in the studies above, we speculate that the echo model may involve repeated hippocampal-cortical dynamics mediated by slow waves. More interestingly, similar phenomena to the echo model have also been observed during the awake task period in previous studies. For example, several studies have found that memory representations are repeatedly activated during the maintenance phase and modulated by theta rhythms (Staresina & Wimber, 2019).

Based on these new findings and theories, we hypothesize that multiple reactivations may be observed in our dataset. We also note that some recent studies have used spectral decomposition methods to detect periodicities in representation similarity time course (e.g., Liu et al., 2020). We applied a similar approach: building on Supplementary Figure S8 from the manuscript, we averaged the columns of the Rho matrix for each subject to obtain a one-dimensional Rho curve, and then performed spectral decomposition. Indeed, we observed a peak around 1 Hz for each patient. The results are as follows:



Response Figure 1 (identical to Figure S10). Rhythmic activity in behavioral-relevant reactivation during TMR. We averaged the Y-axis of the Rho matrix for each subject to obtain a time series representing similarity, and performed spectral decomposition on it (shown in the black lines). Additionally, we extracted the oscillatory components from the time series (by subtracting the fractal components from the original spectrum, as shown in the red lines). The bottom-right panel shows the group-level average of the oscillatory components.

We observed strong low-frequency oscillations in the behavioral-relevant reactivations for each subject, with an average frequency peak of 0.8 Hz. This suggests that multiple reactivations indeed occurred at the subject level. In an ideal scenario with perfect subjects and data, we would expect to observe approximately 4 reactivations within 5 seconds. However, due to the characteristics of iEEG

mentioned earlier, results may vary across subjects. Therefore, we observed consistent results at the group level around the 2- and 3-second intervals. We speculate that by increasing the sample size and ensuring greater homogeneity of electrode placements, researchers could potentially observe reactivations occurring roughly once per second.

A key focus of this paper is to discuss the hippocampal-cortical dynamic differences underlying these reactivations, which is a unique advantage of iEEG. We have included the relevant figures in the supplementary materials and have elaborated on these points in the manuscript (page 7, line 249-253). We hope these results will help readers better understand the scope of our study.

“In addition, to further verify whether multiple reactivations occurred individually, we extracted the oscillatory components (Wen & Liu, 2016) from the behavioral-relevant reactivation matrix and observed strong low-frequency oscillations for each subject, with an average frequency peak of 0.8 Hz (Fig. S10). This indicated that multiple reactivations indeed occurred at the subject level.”

References:

- Liu, J., Zhang, H., Yu, T., Ni, D., Ren, L., Yang, Q., Lu, B., Wang, D., Heinen, R., Axmacher, N., & Xue, G. (2020). Stable maintenance of multiple representational formats in human visual short-term memory. *Proc Natl Acad Sci U S A*, 117(51), 32329-32339.
- Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. *Trends Cogn Sci*, 23(12), 1071-1085.
- Wen, H., & Liu, Z. (2016). Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topogr*, 29(1), 13-26.
- Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Rep*, 25(2), 296-301.
- Lewis, P. A., & Bendor, D. (2019). How Targeted Memory Reactivation Promotes the Selective Strengthening of Memories in Sleep. *Curr Biol*, 29(18), R906-R912.
- Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nat Neurosci*, 15(10), 1439-1444.

Response to 2.1

It seems that ripples beyond 120 Hz contribute very little to the results, but the manuscript does not reflect this, since it features a ripple band up to 180 Hz. This could mislead readers into assuming that these high frequency ripples are important here. I suggest to change the main analyses to 70-120 Hz (the author's response to 2.4 supports this view)

A: Thank you for your suggestion. In light of this additional comment regarding ripples, we have re-detected ripples and repeated our main results. Please refer to our response following Response to 2.5 for further details.

Response to 2.2:

I do not agree with the authors line of arguments here.

A duration < 42 ms cannot be translated into a minimum number of cycles at different frequencies. Instead, a duration of < 42 ms only ensures that relatively low frequency ripples (e.g., around 70 Hz) will have less than 3 cycles but high frequency ripples will have more cycles: a ripple at 170 Hz will

have approx. 7 cycles in 42 ms. Since these fast ripples with many cycles contribute to the left panels of Resp. Fig. 5, this might explain why there seem to be multiple oscillations.

Moreover, how do the authors explain that the left panel shows approx. equal amounts of oscillations in the grand average as the right panel, even though the right panel is supposed to include many more “long ripples”? These “long ripples” at relatively low frequencies should result in longer oscillations the grand average in the right panel. In the left panel, the oscillation in the grand average covers a time window (between 100 to 150 ms) that is way larger than ± 21 ms. How can this be the case, if this is a ripple-centered average and no ripple has a duration longer than 42 ms?

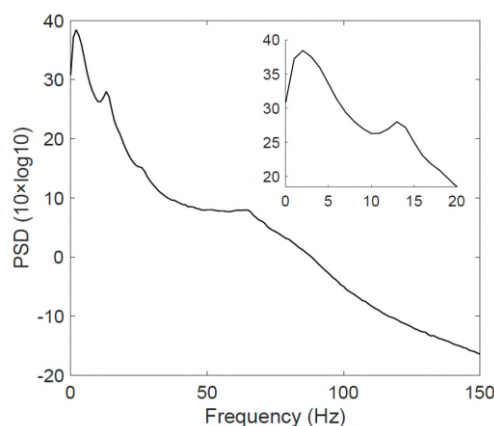
I think that including only ripples with more than 3 cycles would improve reliability of ripple detection and would thus ask the authors to report their main analyses using this criterion.

A: Thank you for your suggestion. We agree the view that only using duration to determine the cycles of ripple was not an appropriate approach. Based on your recommendation, we have now defined the number of cycles by identifying discrete peaks or troughs of neural oscillations in the filtered raw signal. Please refer to our response following Response to 2.5 for further details.

Response to 2.3:

I asked for a power spectrum (no time information), not a time-frequency plot, and whether there are peaks in this power spectrum, in particular in the ripple band, but also for spindles and SOs. My apologies if this was not clear in the previous round.

A: We apologize for the misunderstanding of your suggestion. We present the spectrogram of ripple events ± 500 ms during sleep (**Response Figure 2**). Three distinct peaks are clearly visible, corresponding to slow waves, spindle waves, and ripples, respectively.



Response Figure 2. Power spectral density (PSD) averaged across all detected ripples (± 500 ms).

Response to 2.5:

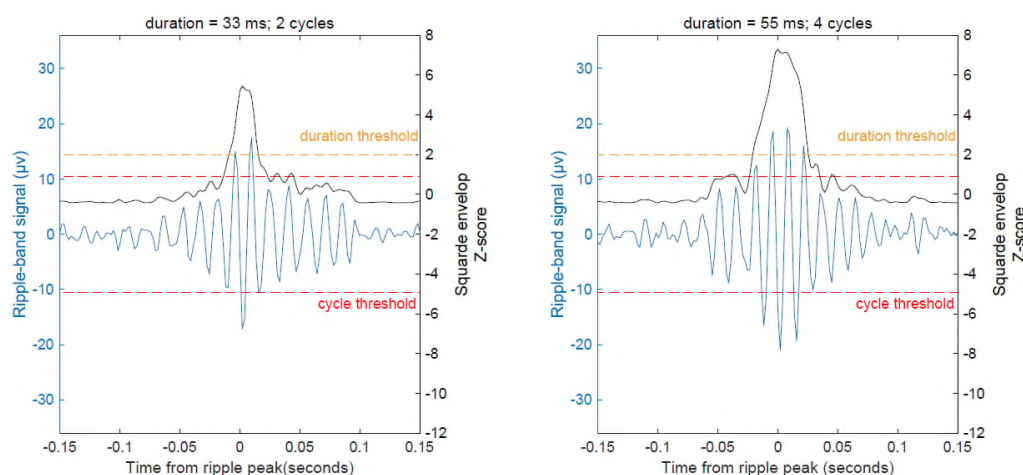
The authors detected all ripples in all hippocampus contacts, as requested. However, they did not repeat the main analyses based on all ripples in all contacts. I maintain the opinion that this analysis would provide more robust results than basing their findings on one selected hippocampus contact.

A: Thank you for your valuable suggestions. We now address Responses to 2.1, 2.2, 2.5 together, as they all relate to the ripple analysis.

We have integrated the suggestions from Responses to 2.1, 2.2, and 2.5 and revisited the selection of the ripple frequency band (80-120 Hz) and the criteria for ripple identification (at least three cycles).

Based on these revised criteria, we re-detected the ripples across all hippocampal electrodes. The figure below illustrates both a short ripple (2 cycles) and a long ripple (4 cycles) to demonstrate the distinction.

The duration of a ripple is defined as the period during which the amplitude (right axis) exceeds two standard deviations from the baseline. We now require that within this duration window, there must be at least three cycles. Specifically, we defined the threshold for identifying a ripple as requiring at least three discrete peaks or three discrete troughs, corresponding to the 99th percentile of the RMS values (as indicated by the red dashed line in the **Response Figure 3**) according to Staerens et al. (2015).

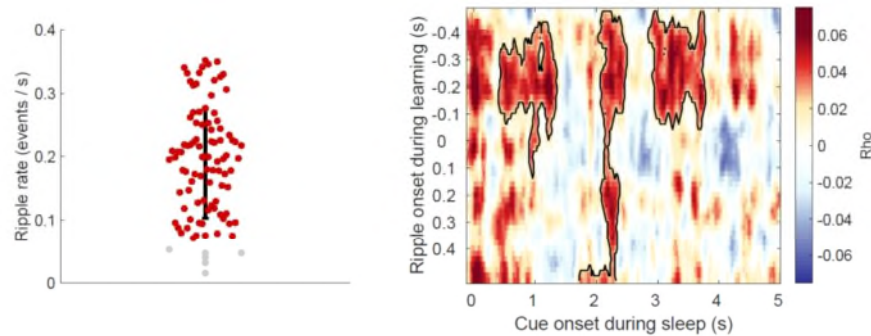


Response Figure 3 (identical to Figure S11A). Example of ripple detection approach. We demonstrated a short ripple (left) and a long ripple (right). Now we not only require ripple duration to be greater than 20 milliseconds, but it must also include at least 3 cycles within its duration. The blue line corresponds to the filtered signal (left axis), and the black line is the normalized envelope (right axis). The yellow and red dashed lines represent the thresholds for ripple detection and cycle detection, respectively.

Based on this criterion, we detected a total of 29,955 ripples from 96 hippocampal contacts during the learning phase, with a mean $\pm$ standard deviation of 312 ± 146 . The ripple occurrence rate was 0.19 ± 0.09 Hz. Due to the influence of ripple occurrence rates on the number of items included in the reactivation analysis, we excluded 6 contacts with particularly low ripple rates (ripple rate < 0.06 Hz, indicated by the gray circles in the **Response Figure 4**). Using these detected ripples, we re-conducted our pattern reactivation analysis, as shown in the Response Figure 4 (right). We observed three significant reactivation clusters (cluster1: from 400 ms to 1320 ms, $p = 0.006$; cluster2: from 2010 ms to 2460 ms, $p = 0.008$; cluster3: from 2840 ms to 3740 ms, $p = 0.004$; from left to right), with the second and third reactivation events aligning well with our previous main findings. In addition, a significant reactivation cluster emerged around one second post-stimulus, which was consistent with previous research suggesting that reactivation directly after TMR may be driven by sensory stimuli (Lewis & Bendor, 2019). We greatly appreciate the reviewer's valuable suggestion, which provides a more robust validation of the hypothesis of multiple reactivations. We have added this result as Figure S11 in the supplementary materials and included the following description in the Results section (page 7, line 254-262):

“To further validate the robustness of behavioral-relevant reactivation, we applied a different ripple detection criterion (filtering in 80-120 Hz and at least 3 cycles) across all hippocampal contacts. We then recalculated the reactivation, as shown in Fig. S11. There were three significant reactivation clusters (cluster1: from 400 ms to 1320 ms, $p = 0.006$; cluster2: from 2010 ms to 2460 ms, $p = 0.008$;

cluster3: from 2840 ms to 3740 ms, $p = 0.004$; from left to right), with the second and third reactivation clusters aligning with our main results. In addition, a significant reactivation cluster emerged around one second post-cue, which was consistent with previous research suggesting that reactivation directly after TMR may be driven by sensory stimuli (Lewis & Bendor, 2019). ”



Response Figure 4 (identical to Figure S11B/C). Behavioral-relevant reactivation using the ripple detected by the updated approaches. Six contacts with low ripple occurrence rates (< 0.06 Hz) were excluded (indicated by the gray circles), and pattern reactivation analysis was conducted using the ripple from remaining 90 contacts. Three significant clusters were identified following the cluster-based permutation test.

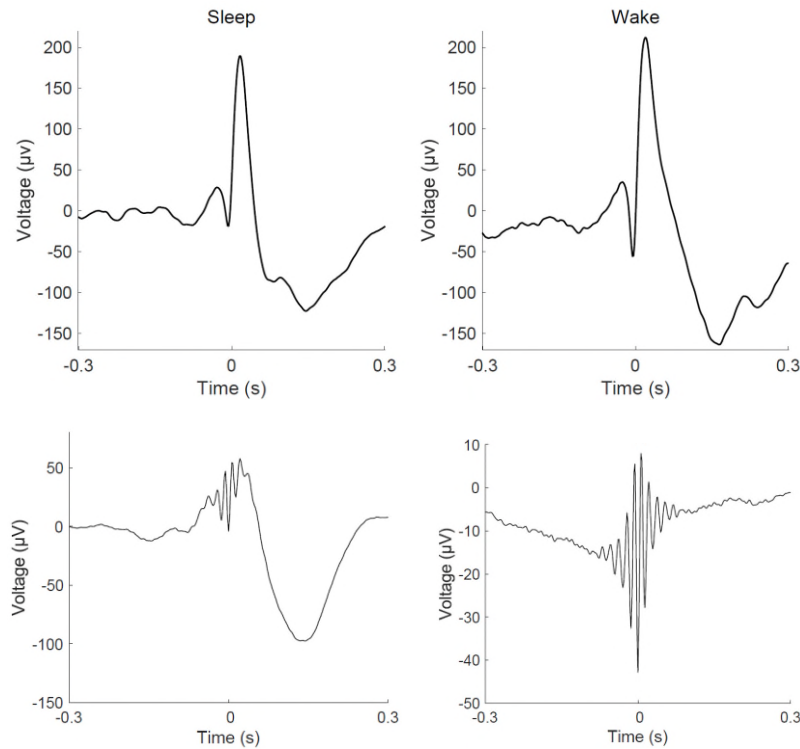
References:

- Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C. E., Axmacher, N., & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, 18(11), 1679-1686.
- Lewis, P. A., & Bendor, D. (2019). How Targeted Memory Reactivation Promotes the Selective Strengthening of Memories in Sleep. *Curr Biol*, 29(18), R906-R912.

Response to 2.7:

The authors show epileptiform spike-locked spectrograms. However, this is unfortunately not what I asked for. Instead, I asked for epileptiform spike-locked averages or “Grand average peri-ripple field potentials”, equal to what the authors show in Fig. 2C or 4B, left.

A: We apologize for the misunderstanding of your suggestion. We have now plotted the epileptiform spike-locked averages below (**Response Figure 5**), which clearly differ from the typical ripples. Together with the previously presented time-frequency plot, we have combined these results into supplementary Figure S22.



Response Figure 5. IED (upper) and ripple-triggered (lower) grand average of all detected IEDs and ripples during learning (left) and TMR sleep (right) stages.

Response to 2.9: No REM ripple rates provided. I still think it would make the manuscript more robust if the authors would try their ripple detection in other states such as REM and see whether their detected rates match previous studies. In particular, since the ripple rates in wake and NREM are very similar, which is not what has been reported in previous studies. Why would the authors provide ripple rates in events/minute and not events/second, which seems to be more common? Please provide the average ripple/spindle rates for each column in Table S3. The authors report similar ripple rates in wake and NREM. Previous studies found higher ripple rates in NREM than wake, e.g. Dickey et al 2022; Schreiner et al., 2024. How do the authors explain this discrepancy?

A: We apologize for not clearly explaining the experimental procedure regarding REM sleep in our first round revision. In fact, after the subjects entered sleep, we observed the occurrence of slow waves via EEG and then began to deliver auditory stimuli, while simultaneously recording the corresponding iEEG data. Our TMR phase lasted for a continuous 30 minutes, during which none of the eleven subjects exhibited clear REM sleep characteristics. After TMR task finished, we left the hospital. In this study project, as we mainly focused on iEEG data during TMR stage, we did not get REM sleep data. We fully acknowledge the importance of the reviewer's suggestion, and in future sleep experiments, we will collect full sleep data to examine the differences between REM and NREM sleep. Additionally, in this version of the manuscript and table (Table S3), we provide the ripple rates in events per second (Hz).

The question of whether there is a difference in ripple rates between wakefulness and NREM sleep is indeed intriguing. We reviewed the two excellent papers you provided, as well as other literature, and found that the situation is complex. In the human intracranial electrophysiology study by Dickey et al.

(2022), the authors reported that "hippocampal ripples during NREM had a density of 7.0 ± 4.8 per minute, while during waking, the density was 7.4 ± 5.5 per minute." In Schreiner et al. (2024), Figure 3e shows ripple rates during waking with good memory (approximately 0.2-0.3 Hz), bad memory (approximately 0.1-0.2 Hz), and during NREM with a rate of 0.19 Hz. In addition, a human iEEG study found the number of ripple events did not differ between waking state and NREM sleep (Zhang et al., 2018). These results suggest that hippocampal ripple rates do not differ significantly between wakefulness and NREM. However, as you pointed out, some studies do report differences between these states, such as Staba et al. (2004) and Clemens et al. (2007), both of which found that ripple rates were higher during NREM sleep compared to wakefulness.

We agree that this point raised by the reviewer is valuable and warrants further investigation. We have addressed this in the Discussion section (page 10, line 403-408):

"Our study found no difference in ripple rate between wakefulness and NREM sleep, consistent with the findings in recent studies (Dickey et al., 2022; Zhang et al., 2018). However, other studies also reported significant differences in ripple rate between two states (Clemens et al., 2007; Staba et al., 2004). This suggests that the difference in ripple rate between wakefulness and NREM sleep remains inconclusive and requires further clarification in future studies."

References:

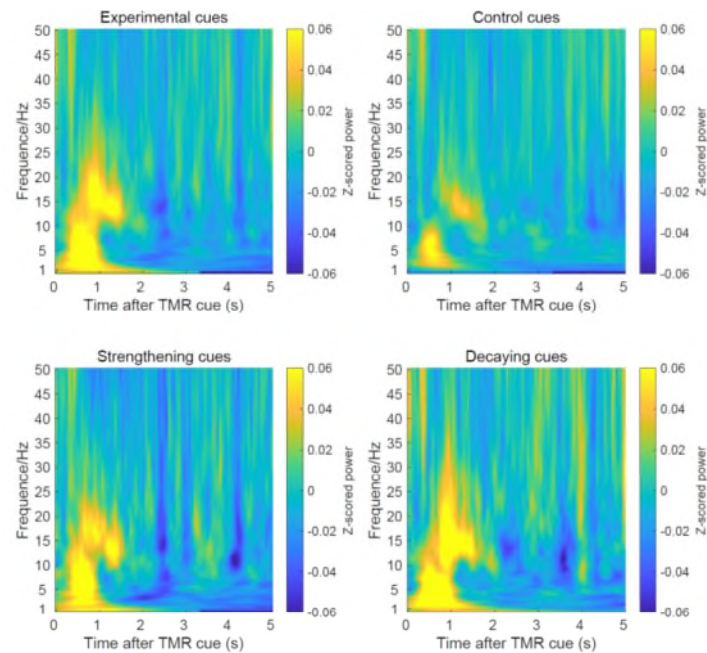
- Clemens, Z., Molle, M., Eross, L., Barsi, P., Halasz, P., & Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain*, 130(Pt 11), 2868-2878.
- Dickey, C. W., Verzhbinsky, I. A., Jiang, X., Rosen, B. Q., Kajfez, S., Stedelin, B., Shih, J. J., Ben-Haim, S., Raslan, A. M., Eskandar, E. N., Gonzalez-Martinez, J., Cash, S. S., & Halgren, E. (2022). Widespread ripples synchronize human cortical activity during sleep, waking, and memory recall. *Proc Natl Acad Sci U S A*, 119(28), e2107797119.
- Schreiner, T., Griffiths, B. J., Kutlu, M., Vollmar, C., Kaufmann, E., Quach, S., Remi, J., Noachtar, S., & Staudigl, T. (2024). Spindle-locked ripples mediate memory reactivation during human NREM sleep. *Nat Commun*, 15(1), 5249.
- Staba, R. J., Wilson, C. L., Bragin, A., Jhung, D., Fried, I., & Engel, J., Jr. (2004). High-frequency oscillations recorded in human medial temporal lobe during sleep. *Ann Neurol*, 56(1), 108-115.
- Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nature Communications*, 9(1), 4103. <https://doi.org/10.1038/s41467-018-06553-y>

Response to 3.1:

What is the increase in power at time = 0 in the time-frequency plots in Resp. Fig. 13 that spans all frequencies from 10 Hz and higher? Please show the same plots including a pre-TMR interval of 1 second.

A: We re-examined the data and found that, among the 11 participants, the iEEG signals from four cortical contacts in three individuals (sub 01 with one contact, sub 05 with one contact, sub 10 with two contacts) were clearly contaminated by the sound trigger sent to iEEG recording system via a muscle electric channel.

After excluding these contacts, the time-frequency results of the cortex did not show the phenomenon of spanning all frequencies (**Response Figure 6**).



Response Figure 6. TMR cue-locked time-frequency power in the cortex contacts.

Response to 5.2:

I cannot find the “added detailed descriptions” about the balancing of trial numbers in the methods section.

In general, it would help a lot if the authors would indicate the line numbers when referring to changes in the updated manuscript.

A: We apologize for the mistake in our previous revision. We have included the number of trials in the supplementary materials (Table S6). In addition, we have added this information in the Methods section (page 18, line 730-731).

“For each participant and each condition (strengthening, decaying and control), the number of trials between the conditions was balanced (Table S6).”

-remove “obviously” from “But it should be pointed out that the intensity of the difference was obviously lower than the subsequent second reactivation cluster.”

A: Thanks for the suggestion. We have removed that word. (page 8, line 307-309)

“But it should be pointed out that the intensity of the difference was lower than the subsequent second reactivation cluster.”

-A recent paper in Nature Communications by Schreiner et al. seems highly relevant for this: Spindle-locked ripples mediate memory reactivation during human NREM sleep.

The authors could discuss overlap / differences in their findings.

A: Thanks for pointing out this highly relevant finding shown in that elegant study. We have carefully read it and added descriptions in the Discussion section. (page 10, line 391-395)

“A recent study found that spindle-locked ripples in MTL were related to memory reactivation during human NREM sleep (Schreiner et al. 2024). However, they mainly focused on the early

reactivation after TMR cue (within 1.5 seconds), and did not investigate the information interaction between the hippocampus and cortex during memory reactivation.”

Reference:

Schreiner, T., Griffiths, B. J., Kutlu, M., Vollmar, C., Kaufmann, E., Quach, S., Remi, J., Noachtar, S., & Staudigl, T. (2024). Spindle-locked ripples mediate memory reactivation during human NREM sleep. *Nat Commun*, 15(1), 5249.

Reviewer #2:

The authors addressed all queries in detail, which further strengthened the manuscript and its conclusions. In sum, this might be published as it stands.

A: We sincerely appreciate your positive feedback on our first-round response. Your constructive comments have been invaluable in helping us improve the clarity and depth of our work.

Reviewer #3:

I appreciate the authors' time and effort in answering every question in detail and providing additional analysis and text revisions to support their answers.

Thank you for expanding Table S3 and adding the stats of the TMR cues timing regarding different sleep stages which will assist others who are attempting to reproduce the experimental procedure. I do have a few open questions that were not answered by the authors and I hope they can provide additional answers. Some critical answers were kept as answers to reviewers while they should be added to the text to provide a complete picture to the reader. I list these below and believe they can be clarified quite easily in the text -

A: We greatly appreciate your positive feedback on our work. Following the suggestions from you and Reviewer 1, we have incorporated some important results into the manuscript. Additionally, we have performed control analyses related to the hypothesis that the representation was indeed item-specific during the learning stage.

1. Item-level cortical pattern around hippocampal ripples in learning stage

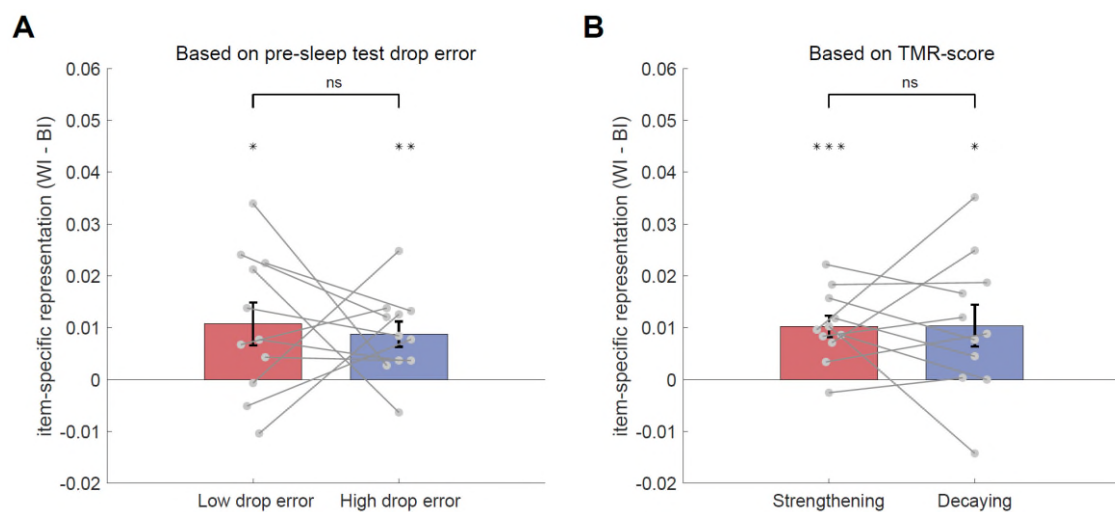
The provided analysis did not adequately answer my question regarding the existence of sub-groups of items dominating this analysis, rather than a unique signature per item. The adjusted title still states that it is the identity of the item, rather than its learned/unlearned status that determines the cortical pattern. Some of the analysis may be misleading because it groups subgroups that have differing trends. For example - Figure 2F displays an increase in ripple rate after cue onset, while response-figure 30 very strongly demonstrates that this result is driven by the correctly encoded items (based on pre-testing) while the poorly learned ones do not reflect an increased rate of ripples. This leads to an alternative hypothesis that would explain the results - efficiently learned items have a unique neural signature during learning and averaging this group with the unlearned group masks this pattern.

-- response figure 30 is important to add to the supplementary data to complement Figure S7a because it demonstrates that ripple rate during learning is indeed modulated by the subsequent memory effect.

-- A control to contest the hypothesis that the representation is item-specific, is to compare the WI and BI of all strengthened items (completely removing the decayed ones from analysis) to confirm that the pattern does not reflect a successful operation of the memory-network, but rather a specific pattern per item. Response Figure 31, is breaking down Fig. 2L to items and does not address my concern.

A: We greatly appreciate your suggestion on additional analyses. Based on the drop error from the pre-sleep test, we divided the items into two groups (low drop error group and high drop error group) and calculated the respective item-specific representations (WI - BI) for each group (**Response Figure 7A**). We found that both groups showed a significant difference from zero ($t(10) = 2.61, p = 0.026$, $t(10) = 3.62, p = 0.005$). However, there was no significant difference between the two groups ($t(10) = 0.36, p = 0.73$). As you mentioned, we observed a difference in hippocampal ripple occurrence rate following stimulus presentation between learned and not-learned items (which we have included in Supplementary Figure S2). However, the item-specific representation in the cortex may not be related to memory performance.

Similarly, we divided the items into strengthening and decaying groups based on their TMR scores (**Response Figure 7B**). Again, both groups showed significant differences from zero ($t(10) = 5.00, p < 0.001$, $t(10) = 2.58, p = 0.028$), and there was no significant difference between the groups ($t(10) = -0.03, p = 0.97$). This further suggests that the item-specific representation is not strongly linked to the effects of sleep memory consolidation.



Response Figure 7 (identical to Figure S3). Item-specific representations in subgroups during learning stage.

2. Figure S1B demonstrates what looks like a significant difference in the error during the pre-sleep stage between the strengthened and the decayed group (blue line). Please add the stats of this comparison to the paper because it is critical to assess whether there is a difference between items assigned to these two groups that are not related to the TMR intervention. The authors need to acknowledge this difference and discuss how it may impact their analysis.

A: Thank you for your valuable suggestion. We have added the statistics for this comparison. We have been closely monitoring this effect and have conducted control analyses to ensure that our neural results were not dominated by this effect (see Figure S13). We describe this in details in the Results section

(page 7, line 276-282). Additionally, we have expanded our discussion of this effect (page 12, line 469-482).

“In addition, we tested whether the reactivation results were driven by differences in pre-sleep accuracy of strengthening vs. decaying items ($t(10) = 7.3$, $p < 0.001$; Fig. S1B), and compared reactivation of two groups of items determined entirely by pre-sleep accuracy but matched for memory change score. RSS results showed no significant reactivation differences between the two groups (Fig. S13), indicating that higher reactivation of strengthening items cannot be purely explained by differences in pre-sleep accuracy.”

“Importantly, behavioral analysis showed that TMR benefits were more apparent for items with low pre-sleep accuracy. The performance difference before sleep could be one driving force for later effects during sleep. Previous TMR studies have produced similar results (Creery et al., 2022; Creery et al., 2015), and this effect may have multiple causes. One factor likely is that enhanced accuracy after sleep in those trials may reflect regression to the mean. Correspondingly, reduced accuracy after sleep for trials with high pre-sleep accuracy could also reflect regression to the mean, indicating that the memory differences between the two critical conditions could entirely reflect measurement variability. However, this factor cannot account for the present results because neural correlates of memory change would not be expected if memory change was due merely to such unsystematic fluctuations in performance. Another factor described in prior studies is that gains from TMR tend to be more apparent for weak memories than for strong memories (Creery et al., 2015; Oudiette & Paller, 2013). Notably, the control analysis showed that though two groups of trials showed significant differences in test accuracy at the pre-sleep stage, we did not find a significant reactivation difference if the two groups did not differ in memory change (Fig. S13).”

3. In the rebuttal letter, I read this addition but could not find it in the main text -

In addition, in the revised manuscript, we rephrased the definition of control condition as follows: “To verify the relationship between the learning ripples and subsequent reactivations during sleep, we selected trials without ripple events during the learning (149 ± 62 no-ripple trials) and investigated the reactivation of their patterns during sleep. For example, there were two learning trials of one item. There was a ripple event at 0.75 second after the image onset in trial 1, while there was no ripple in trial 2. We mark a fake ripple event for trial 2 at 0.75 second after the image onset. Then we repeated our RSA analysis just using trials of these fake ripple events.”. The current description of the methods to perform the control plotted in Figure S6A legend is not sufficient for the reader to understand how the control analysis was performed, nor reproduce it. Also – because the result is so close to significance ($p=0.058$ based on the rebuttal letter), it’s important to provide the full statistics, be transparent about the result, and state that it follows the same trend as ripple-triggered analysis. Would be recommended to add stats for Figure S6B as well.

A: Thanks a lot for your suggestion. We have added a description of the methods used in Figure S6A (Figure S12 in revised manuscript; page 40, line 1251-1256). Additionally, we have reported the p-values for the cluster tests in the figure legend, and in the Results section, we state that the “no ripple” trials follow the same trend as the ripple-triggered analysis (page 7, line 274-276).

“To verify the relationship between the learning ripples and subsequent reactivations during sleep, we selected trials without ripple events during learning (149 ± 62 no-ripple trials) and investigated the

reactivation of their patterns during sleep. For example, there were two learning trials of one item. There was a ripple event at 0.75 second after the image onset in trial 1, while there was no ripple in trial 2. We created a fake ripple event for trial 2 at 0.75 second after the image onset. Then we repeated our RSA analysis only using trials with these fake ripple events.”

“The difference between strengthening and decaying items of non-ripple trials showed a similar trend as ripple trials did (Fig. S12A, minimum P -value = 0.057, cluster-based permutation test).”

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Remaining questions and comments:

Authors' response to 2.1:

Please include details on how Figure S10 was made. For example, what method was used to remove fractal components? On what data segments, ect.?

I assume that all the code will be made publicly available, not only the code for the main figures?

Thank you for your suggestion. We have added the detailed descriptions of the calculation process for Figure S10 in the supplementary materials (page:39, line: 1230 to 1239).

“Extracting rhythmic activity from the behaviorally relevant reactivation matrix

To verify whether multiple reactivations occurred individually, we extracted the oscillatory components (Wen & Liu, 2016) from the behaviorally relevant reactivation matrix. The procedure was as follows: First, we averaged the columns of the Rho matrix for each subject to obtain a one-dimensional Rho curve (from 0 to 5 seconds following cue presentation). Next, similar to previous studies (Liu et al., 2020), we performed spectral decomposition to extract the original spectral and fractal components using the IRASA method from FieldTrip (frequency range: 0 to 10 Hz, frequency resolution: 0.5 Hz, overlap: 90%). Finally, the neural oscillatory components were obtained by subtracting the fractal component from the original spectrum.”

In addition, we have made the code for all the figures (including those in the supplementary materials) available. Please refer to the following link: <https://doi.org/10.5281/zenodo.14583770>.

References:

Liu, J., Zhang, H., Yu, T., Ni, D., Ren, L., Yang, Q., Lu, B., Wang, D., Heinen, R., Axmacher, N., & Xue, G. (2020). Stable maintenance of multiple representational formats in human visual short-term memory. *Proc Natl Acad Sci U S A*, 117(51), 32329-32339.

Wen, H., & Liu, Z. (2016). Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topogr*, 29(1), 13-26. <https://doi.org/10.1007/s10548-015-0448-0>

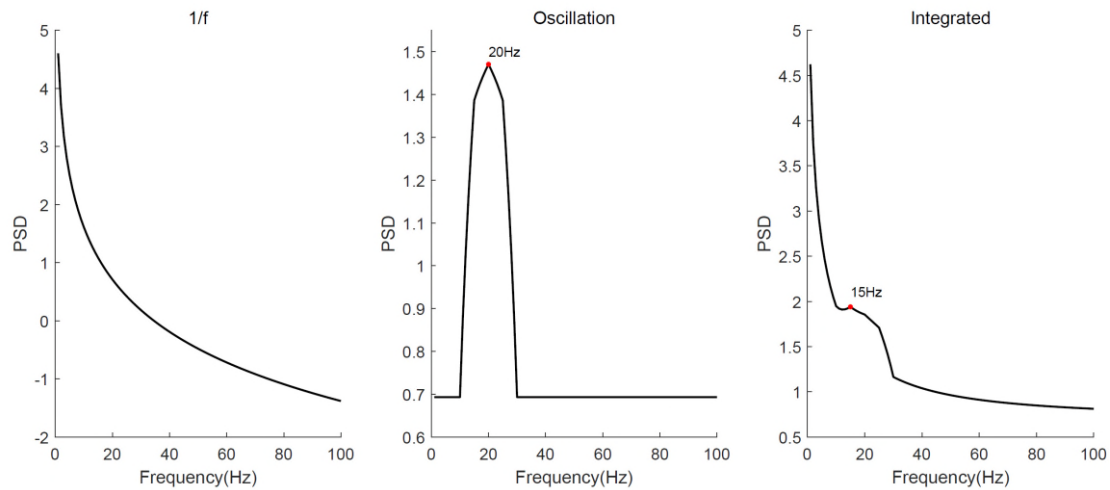
Authors' response to 2.3:

Please include Response Figure 2 in the manuscript / supplements.

The “peak” with the highest frequency seems not to be in the ripple band, but below 70 Hz. Unfortunately, the authors do not provide the frequency of the “peaks”, but looking at the plot tells me the “peak” is at approx. 65 Hz. This is outside the ripple band in their main analyses (70 to 180 Hz), as well as outside of ripple band in the additional analyses (80-120 Hz). How do the authors explain this? Does the “peak” at 65 Hz still reflect activity in the ripples band and, if so, why would the frequency be different? Alternatively, there is no peak in the power spectrum that reflects the

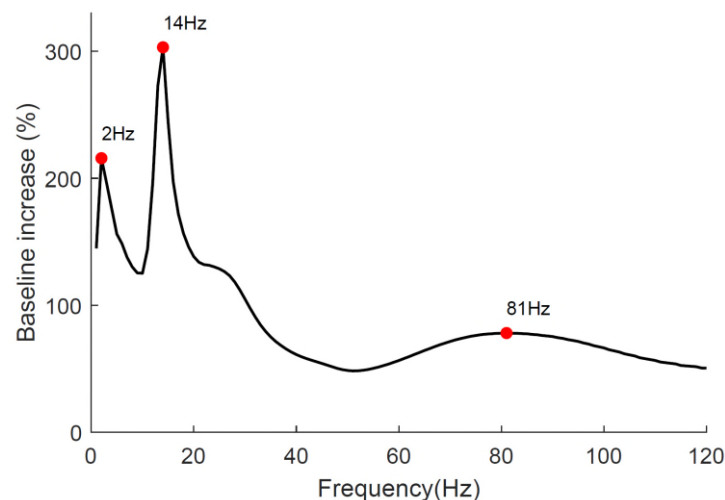
activity of the ripples, and if so, why would that be?

Thanks for this comment. We believe that this is due to the influence of $1/f$ noise, which may shift the peak frequency of neural oscillations to lower values. To better illustrate this, we simulated the spectral characteristics of pink noise (shown in the left part of the figure below) and the spectral characteristics of neural oscillatory activity with a peak at 20Hz (shown in the middle part of the figure). As shown, when these data are merged, the peak in the resulting spectrum slightly shifts to around 15Hz (shown in the right part of the figure).



Response Figure 1. The effect of $1/f$ noise on oscillatory peak frequency.

After carefully checking the previous studies, we found that when presenting the spectral characteristics of ripples, the influence of $1/f$ noise was often mitigated by subtracting the baseline activity (Staresina et al., 2015; Norman et al., 2019; Norman et al., 2021). We now adopted this approach and updated the spectral analysis around the ripple events, as shown in Response Figure 2. We selected a baseline period from 1.5 seconds to 1 second before the ripple event and calculated the ratio of spectral activity around the ripple event to the baseline period. We found a peak around 81Hz, which falls within the detection range of our ripple analysis (70 to 180 Hz). This figure is included in the supplementary materials (Fig. S21A).



Response Figure 2 (Identical to Figure S21A). Power spectral density (PSD) averaged across all detected ripples (± 500 ms), showing percentage change from the pre-event baseline (-1.5 s to -1 s).

References:

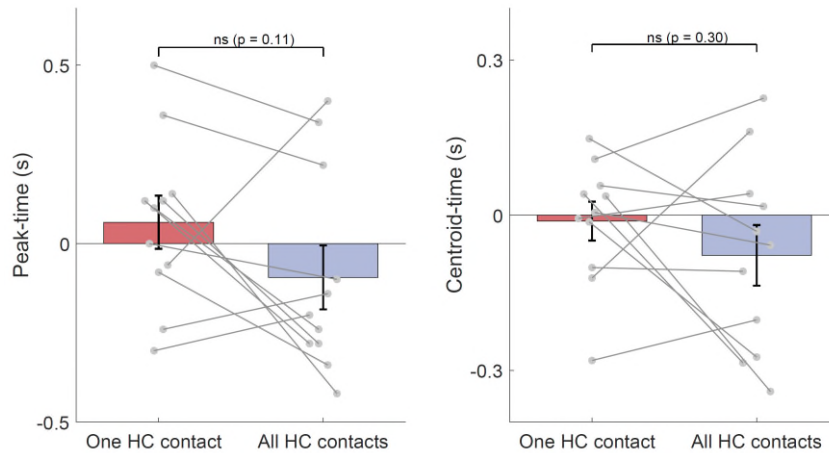
- Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C. E., Axmacher, N., & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, 18(11), 1679-1686.
- Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science*, 365(6454). <https://doi.org/10.1126/science.aax1030>
- Norman, Y., Raccach, O., Liu, S., Parvizi, J., & Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron*, 109(17), 2767-2780 e2765. <https://doi.org/10.1016/j.neuron.2021.06.020>

Authors' response to 2.5:

How do the authors explain the apparent discrepancies between Response Figure 4 and main Figure 3B regarding the y-axes? In Response Figure 4, activity before the ripples seems to be most relevant but not in Fig. 3B.

Thank you for this valuable suggestion. To address the concern regarding the timing of reactivation relative to the learning-related ripple, we performed the following analysis (Response Figure 3). First, we averaged the Rho matrices during two reactivation periods (1,660-2,560 ms and 2,880-3,840 ms, as in Fig. 3B) for each subject to derive a one-dimensional Rho curve, spanning from -0.5 to 0.5 seconds around the learning ripple. We then identified the peak time point at which the Rho value reached its maximum value (as shown in the left figure) or determined the centroid time point of the Rho vector (as shown in the right figure). The figure below illustrates the peak and centroid time points of reactivation based on the detected ripples from one single hippocampal electrode (as in Fig. 3B) and from all hippocampal electrodes (Fig. S11C, i.e., Response Figure 4 in last response letter). As shown, there was no significant difference between the two results ($t(10) = 1.77, p = 0.11$, left figure; $t(10) = 1.08, p = 0.30$, right figure), and no significant difference compared to the 0 s reference point ($t(10) = 0.84, p = 0.42$; $t(10) = -1.11, p = 0.29$, respectively, left figure; $t(10) = 0.32, p = 0.75$; $t(10) = -1.39, p = 0.19$, respectively, right figure). We completely agree that clarifying this point is crucial, and have included this figure in the supplementary materials (Fig. S11D).

We also noticed that there were some variations within the subjects between the two analyses (one hippocampal contact, all hippocampal contacts; see Response Figure 3), which may be related to changes in the ripple detection criteria and the number of hippocampal contacts included.



Response Figure 3 (identical to Figure S11D&E). The peak/centroid moment of reactivation.

Authors' response to 2.9:

-“Leaving the hospital” is not a sufficient argument for not analyzing REM data. Was REM data not recorded? This would be very unusual, since patients are in the hospital to record their intracranial EEG during epileptic seizures.

The main goal of this study was to explore the relationship between neural activity during NREM sleep and memory consolidation. As the reviewer pointed out, patients are indeed regularly in the hospital for intracranial EEG recording for epileptogenic zone estimation, while the scalp electrodes for sleep stage monitor are placed one night only for this study. Therefore, after completing the experiment for each patient, we only took the intracranial EEG and scalp EEG data during the TMR experiment from the hospital.

- Schreiner et al. (2024) show ripple rates during NREM / TMR that are clearly higher than 0.19 (Figure 4d) and also clearly higher than the ripple rate during wake.

Actually, we saw the ripple occurrence rate during NREM shown in the supplementary materials of Schreiner et al. (2024) (Fig. S16): ‘NREM: 0.19 ± 0.017 ripples per second’. According to your suggestion, we re-checked that study and found that the ripple occurrence rate reported in Figure 4d was high (around 0.5-0.6 Hz) after TMR cues which was indeed higher than the ripple rate during wake. Thanks for this kind remind.

Reviewer #3 (Remarks to the Author):

The authors addressed all queries in detail, added details to the main text and the requested explanations to the supplementary material which allow the reader to better understand the analysis and results. They are clearly stating which hypotheses are met and which aren't which I appreciate. In sum, I support publishing the manuscript in its current state.

We are truly grateful for your positive feedback on our response. Your thoughtful and constructive suggestions were vital in improving both the clarity and depth of our work.