Phase separation of competing memories along the human hippocampal theta rhythm.

Casper Kerrén[1,2*], Sander van Bree[1,3], Benjamin J. Griffiths[1,4], Maria Wimber[1,3*]

Affiliations

1. Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, United Kingdom
2. Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany.
3. Centre for Cognitive Neuroimaging, School of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom
4. Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

* Corresponding authors:

Casper Kerrén – casper.kerren@gmail.com

Maria Wimber - Maria.Wimber@glasgow.ac.uk

Abstract

Competition between overlapping memories is considered one of the major causes of forgetting and it is still unknown how the human brain resolves such mnemonic conflict. In the present MEG study, we empirically tested a computational model that leverages an oscillating inhibition algorithm to minimise overlap between memories. We used a proactive interference task, where a reminder word could be associated with either a single image (non-competitive condition) or two competing images, and participants were asked to always recall the most recently learned word-image association. Time-resolved pattern classifiers were trained to detect the reactivated content of target and competitor memories from MEG sensor patterns, and the timing of these neural reactivations was analysed relative to the phase of the dominant hippocampal 3Hz theta oscillation. In line with our preregistered hypotheses, target and competitor reactivations locked to different phases of the hippocampal theta rhythm after several repeated recalls. Participants who behaviourally experienced lower levels of interference also showed larger phase separation between the two overlapping memories. The findings provide evidence that the temporal segregation of memories, orchestrated by slow oscillations, plays a functional role in resolving mnemonic competition by separating and prioritising relevant memories under conditions of high interference.

Introduction

Each day is a flow of events that take place at different times but often in overlapping contexts. Unavoidably, many of the stored memories share similar features, and this overlap poses a major challenge for our memory system (McClelland et al., 1995; Norman & O'Reilly, 2003). The present MEG study investigates the possibility that the human brain uses a temporal phase code to adaptively separate overlapping memories, enabling the targeted retrieval of goal-relevant information.

Competition between similar memories is considered one of the major causes of forgetting (Anderson & Neely, 1996; Underwood, 1957). A prominent case is proactive interference, where access to a target memory is impaired when overlapping information has been stored prior to target learning (Tulving & Watkins, 1974). This impairment is typically ascribed to the conflict arising from the co-activation of competing memories (Kliegl & Bäuml, 2021). On a neurophysiological level, mid-frontal theta oscillations (3-8 Hz) have been identified as a reliable marker of cognitive conflict in general (Cavanagh & Frank, 2014), and mnemonic conflict specifically (Ferreira et al., 2014; Hanslmayr et al., 2010; Johansson et al., 2007; Staudigl et al., 2010). Increased theta amplitudes may facilitate the emergence of phase separation, where populations of neurons transmitting potentially interfering information are segregated in time along a theta cycle (Buzsáki & Draguhn, 2004).

On a theoretical level, computational models assign a central role to the phase of the hippocampal theta oscillation in the ordered representation of multiplexed information (Lakatos et al., 2005; Lisman, 2005; Lisman & Idiart, 1995; Lisman & Jensen, 2013; Nyhus & Curran, 2010; O'Keefe & Recce, 1993). Empirically, in rodents it has been shown that bursts of gamma power reflecting sequences of spatial information processing are represented at

different phases of the hippocampal theta oscillation (Bragin et al., 1995; Colgin et al., 2009). Several recent studies in humans have shown that the neural representations of items held in working memory (Bahramisharif et al., 2018; Canolty et al., 2006; Fuentemilla et al., 2010; ten Oever et al., 2020), or encoded into or retrieved from episodic memory (Heusser et al., 2016; Kerrén et al., 2018; Kunz et al., 2019), cluster at distinct phases of the (hippocampal) theta rhythm. These findings suggest that low-frequency oscillations provide time windows for the selective processing and readout of distinct units of information (Lisman & Jensen, 2013). The present study investigates phase separation as a potential mechanism to avoid interference between multiple competing memories that are simultaneously reactivated by a cue.

The computations by which the human brain achieves a separation of multiple overlapping memories are currently unknown. A recent computational model leverages different phases of a theta oscillation to iteratively differentiate target from competitor memories (Norman et al., 2006). In this model, which we will henceforth refer to as the *oscillating interference resolution model*, a retrieval cue will activate associated units, representing target and competitor features, in a phase-dependent manner. In the most desirable output state, at medium levels of inhibition, the cue only activates the target units and no or few competitor features. When inhibition is raised towards the peak of the oscillation, only strong features of the target memory will remain active, and the model learns (via Contrastive Hebbian Learning) to strengthen through Long-Term Potentiation (LTP) the weaker target nodes that did not survive the higher inhibition levels. Conversely, during the transition from a medium to a lower inhibition state towards the trough of the oscillation, activation spreads to more weakly associated units, including some competitor features. This opposite phase is used to identify and punish overly strong features of the competing memory through Long-Term Depression (LTD). The mechanism (see Figure 1D) is repeated across several cycles of an oscillation, which changes the similarity structure

of memories into a state in which they are less likely to interfere with each other. This model offers a mechanistic but not yet tested explanation for how the brain solves mnemonic competition.

A previous study (Kerrén et al., 2018) provided proof-of-principle evidence that the reactivation of a single visual memory is rhythmically modulated by the hippocampal theta rhythm. In the current study, we applied the same analytic approach in an associative interference paradigm (Figure 1A and 1B), to investigate the phase modulation of competing memories. Hypotheses for the experiment were pre-registered on OSF. Based on the oscillating interference resolution model (Norman et al., 2006) and related theta-gamma models (e.g., Lisman & Idiart, 1995), we expected that target and competitor memories become active, and can thus be decoded, at different phases of the hippocampal theta oscillation, with targets most active at the high inhibition phase, and competitors surfacing as inhibition ramps down (see Figure 1D). We also hypothesised that this phase separation would become more prominent across several repetitions of competitive retrieval, as overlapping memories become less interfering and hence less likely to activate during the high-inhibition, target phase. Lastly, across repetitions we expected target memories to become strengthened and competing memories to be weakened, which should be reflected in changes in their decodability.

Results

Participants (n = 24) completed an associative memory task including one proactive interference condition and two control conditions (Figure 1A-B). In each learning trial, participants were instructed to form a novel association between a unique verb and an image, which could be either an object or a scene. In the critical competitive condition (CC), the word was associated with two images, one object and one scene image, learned on different trials.

Participants were asked to continuously update their memory to remember the most recently encoded associate of each word, thus allowing us to investigate the extent to which previously encoded associations interfered with newly encoded information (proactive interference). In the single-exposure non-competitive condition (NC1), a word was associated with only one image, and never re-appeared during learning. This condition served as a baseline for behavioural interference effects. In the dual-exposure non-competitive condition (NC2), a word was also associated with one image only, but this association was learned twice. This condition was included as a control for neural (cue) repetition and order effects. Both non-competitive conditions were also used to train and validate our multivariate classifiers. CC trials were pseudo-randomly intermixed with NC1 and NC2 trials during learning and recall blocks.

In a subsequent memory test, participants were asked to remember the most recent image when prompted with the cue word, and cued recall was repeated 3 times per word-image association, in a spaced fashion, irrespective of condition. Once participants indicated that they had the associated picture back in mind, two follow-up questions were asked, one about the supra-ordinate (object/scene) and one about the sub-ordinate (animate/inanimate for objects, indoor/outdoor for scenes) category. This task design, with its supra- and sub-ordinate categories, is critical because it allowed us to disentangle, on a neural level, the representational evidence for target and competitor reactivation at any given time point, using two independently trained linear decoding algorithms: one to discriminate indoor and outdoor scenes, and one to discriminate animate and inanimate objects (see example in Figure 1C).

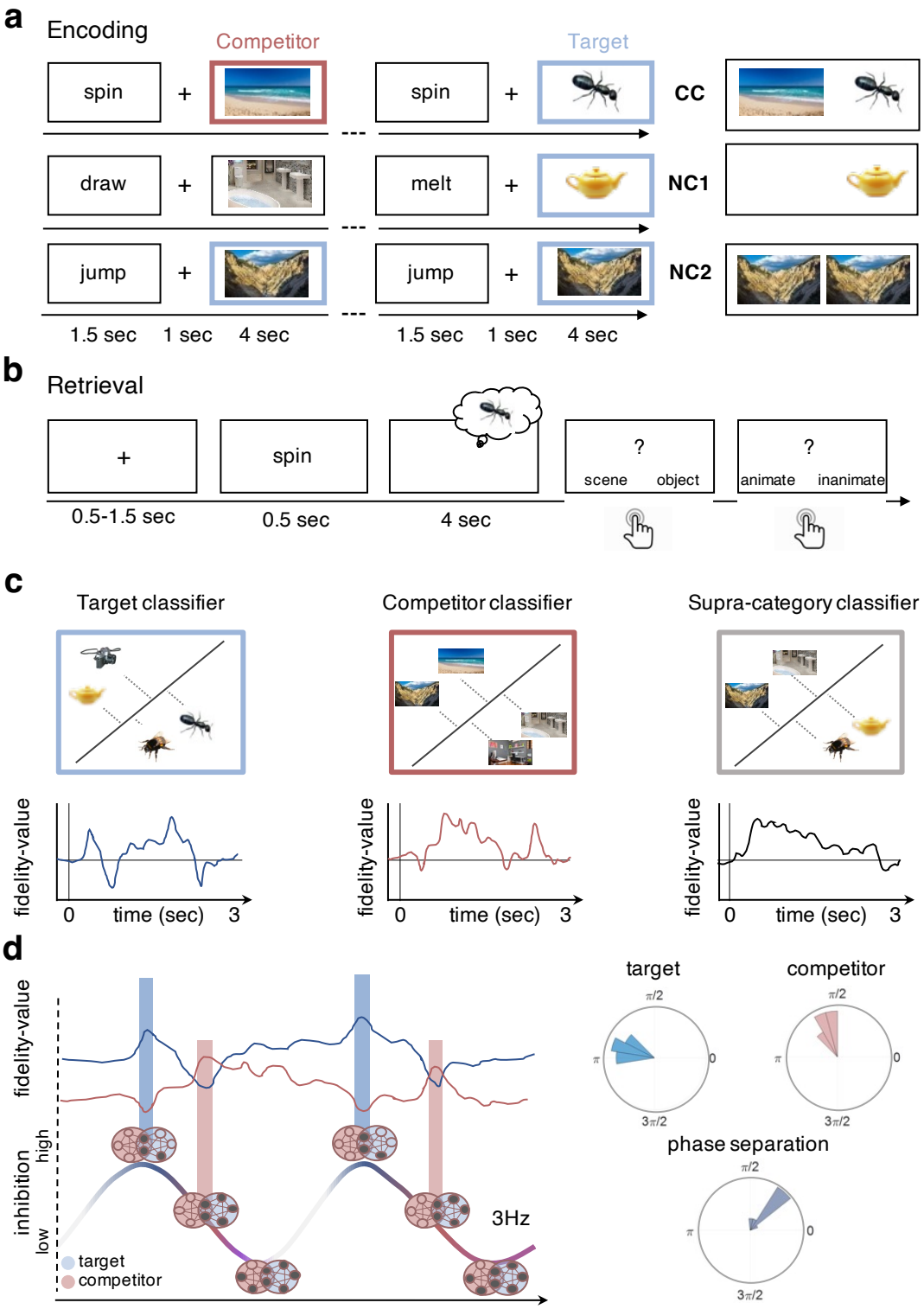


Figure 1. **Paradigm and rationale for decoding analyses. A)** At encoding, subjects were instructed to memorize the word-image associations using an imagery strategy, and to constantly update their memory with the most recent associate to each word. The experiment consisted of 3 different conditions. In NC1, subjects encoded the word together with one association; in NC2, subjects encoded the word with the same associate twice; and in the competitive condition (CC), subjects encoded the word together with two different associates (one scene and one object). **B)** At retrieval, participants were instructed to remember the most recently encoded associate when prompted with a word cue. **C)** Subordinate-category classifiers (animate/inanimate for objects, and indoor/outdoor for scenes) were used to obtain independent evidence for target and competitor reactivation at each sample point. Note that the super-ordinate (object/scene) classifier cannot discriminate between evidence for the target and

against the competitor, and vice versa. **D)** We hypothesised that target (blue) and competitor (red) fidelity time courses would rhythmically fluctuate, and that their reactivation peaks would be locked to different phases of the hippocampal theta oscillation, while both locking to the retrieval part of the cycle (ramp-down). Along the 3Hz oscillation, darker colours indicate the predicted phase for target (blue) and competitor (red) reactivation, with purple indicating a mixture of target and competitor reactivation. Grey represents the encoding part of the cycle, where we expected no retrieval-related reactivation. The bottom left shows the hypothesis derived from the oscillating interference resolution model (adjusted from Norman et al., 2006). Blue circles represent the target memory, whereas the pink circles represent the competitor memory. The black circles inside show the strength of their feature reactivation at a given phase of the oscillation. When inhibition is high, only few features of the target memory are active. In contrast, during the low-inhibition phase, features of both target and competitor memories are simultaneously active. The Phase of maximum target and competitor reactivation was identified and subtracted per participant, allowing us to test for a consistent phase segregation across participants, as shown in the example on the bottom right.

Behavioural indices of proactive interference

We first evaluated behavioural evidence for proactive interference, that is, the extent to which cued recall performance suffered from having encoded two different pictures with one word (CC) compared to only one picture (NC1). Only trials where participants correctly responded to both follow-up questions (i.e., they were able to provide the supra- and sub-ordinate category of the target image) were considered a successful recall. A one-factorial repeated-measures ANOVA, with memory accuracy as the dependent variable, and condition (CC, NC1, NC2) as within-subject factor, showed a significant main effect of condition, $F(1.453, 46) = 57.99$, $p < .05$. The main planned comparison of interest showed that memory performance was indeed significantly reduced in the competitive (M=53.18%, SD = 19.20%) compared to the single-exposure (M = 57.71%, SD = 13.98%) condition, $t(23) = 1.80$, $p < .05$ (Figure 2A). This confirms the first behavioural hypothesis (OSF) that retrieving a memory is more difficult when a previously encoded, now irrelevant associate competes for recall, indicative of proactive interference. For later analyses relating neural and behavioural indices, an interference score for each participant was calculated by subtracting average memory performance in the CC from the NC1 condition. A second posthoc test showed that a significant difference was also obtained between NC1 and NC2 (M = 78.88%, SD = 10.06%), $t(23) = -11.91$, $p < .05$, demonstrating that, unsurprisingly, a second encoding exposure improved performance compared to a single exposure.

Another behavioural index of competition is the intrusion score; a participant's tendency to choose the competitor instead of the target. Our design allowed us to quantify intrusions based on how often participants selected the specific sub-category of the competitor (i.e., "outdoor scene" in the example in Figure 1) during the staged retrieval. On trials where participants made an error (i.e., selected "object" though the target was a scene and vice versa), they showed a significantly above-chance (i.e., above 50%) probability of selecting the sub-category of the competitor (on average 88.72%, SD = 10.62%; $t(23) = 10.48$, $p < .05$), suggesting that their errors were not random but biased towards the category of the previously learned but now irrelevant associate (Figure 2B). The proportion of intrusions did not differ between first (88.08%, SD = 12.04%) and third (88.85%, SD = 14.72%) repetition, $t(23) = -.19$, $p = .85$, which might be explained by the absence of feedback in our recall task. Across participants, the average number of intrusions correlated negatively with memory performance in the non-competitive conditions, $r = -.44$, $p < .05$ (see Supplemental Figure 1), indicating that better memory performance was generally associated with fewer intrusion errors.

Mnemonic competition is associated with a frontal theta power increase

Having established that the paradigm elicits competition on a behavioural level, we next checked if the competitive condition was associated with the neural indices of competition typically reported in the literature. A well-documented effect is an early increase in theta power (3-8 Hz) over frontal sensors, which has been related to cognitive conflict processing in tasks with and without memory demands (Cavanagh & Frank, 2014; Ferreira et al., 2014; Hanslmayr et al., 2010). In the present study the competitive condition (CC) and the non-competitive neural baseline (NC2) were equated for repetitions of the cue. When contrasting these two conditions we expected to see a similar power increase in the theta range in the present

experiment. A non-parametric cluster-based permutation test, focused on a time window of interest 0 to 1 second following cue onset and on the previously reported 3-8 Hz theta range, revealed a significant power increase when retrieving the target association in the CC compared to the NC2 condition ($p_{cluster} < .05$), ranging from approximately 500 to 700 ms post-cue (Figure 2C). In addition to behavioural evidence for proactive interference, we thus also replicate previous work that established frontal theta power increases as a neural signature of competition. Note, however, that in terms of its topography, this frontal effect was more right-lateralized than the typically observed mid-frontal theta (Cavanagh & Frank, 2014).

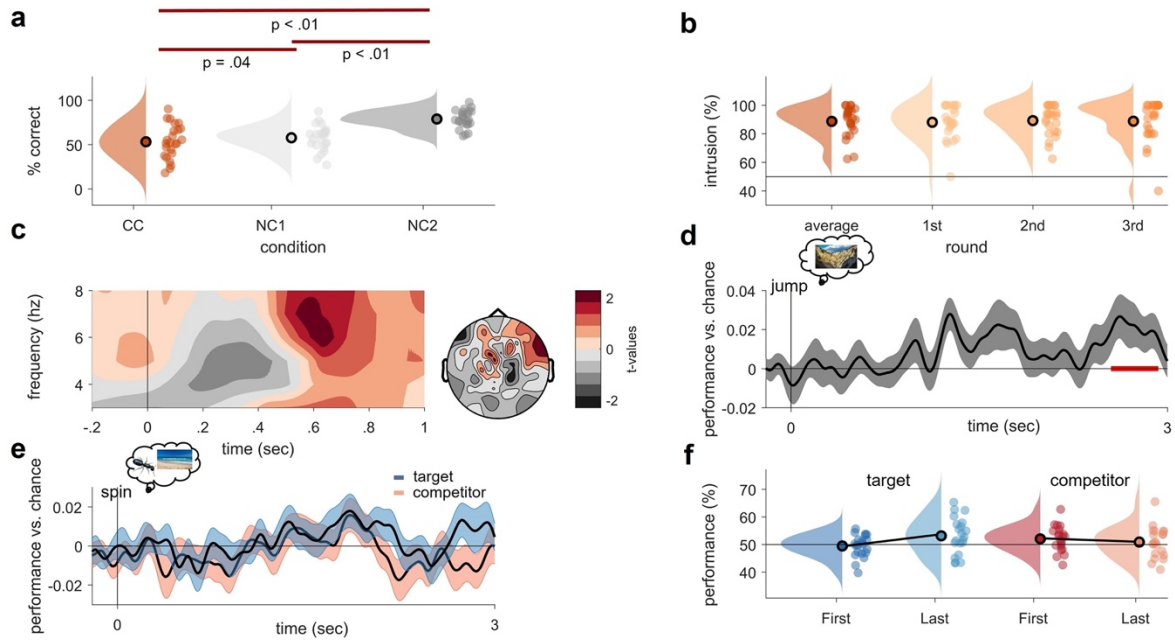


Figure 2. **Behavioural results, time-frequency analysis of theta power, and decoding accuracies in the non-competitive and competitive conditions. A**) As expected, we found that memory accuracy (on the two follow-up questions combined), averaged across the three recall repetitions, was significantly impaired when encoding a given word cue with two different images (CC) compared to just one image (NC1), indicative of proactive interference (red line indicating significant difference at $p < .05$). Recall performance also benefited from learning a cue word together with the same image twice (NC2) compared to once (NC1). **B)** The average intrusion score shows that errors were not random (50% black line), but instead were significantly biased towards the competitor's sub-category. The proportion of intrusions did not decrease significantly across repetitions. **C)** Contrasting oscillatory power elicited by the cue in the CC and the NC2 conditions resulted in a significant cluster (500-700ms, darkest red) in the theta frequency range (3-8 Hz), most prominently over right frontal electrodes (see right inlay). **D)** Results of an LDA-based classifier trained and tested on the non-competitive conditions (NC1 and NC2) at retrieval, showing a cluster of significantly above-chance decoding accuracy approximately 2500-3000ms post cue-onset, with an earlier decoding peak around 1-2 seconds not surviving cluster correction. **E)** Results from a classifier trained on the non-competitive conditions (NC1 and NC2) and tested on the competitive condition. Separate classifiers were used to detect target (blue) and competitor (red) evidence at the level of sub-categories. No significant cluster emerged when averaging across all repetitions. **F)** Contrasting decodability in the first and third retrieval repetition for target and competitor memories, respectively, yielded a significant interaction such that evidence for target memories increased as a function of repetition, whereas evidence for competitor memories

decreased as a function of repetition. Red line in panel D indicates time windows of significant clusters at $p_{cluster} < .05$ corrected for multiple comparisons across time.

Pattern reactivation of target and competitor memories

The main preregistered hypothesis of this study concerned the timing of neural reactivation of target and competitor memories relative to the phase of the theta rhythm. As a first step, we thus checked whether and when the categorical content of the associated memories could be decoded from the MEG sensor patterns in the recall phase of the experiment. Two separate Linear Discriminant Analysis (LDA) based classifiers were trained to discriminate the sub-ordinate picture class, i.e., animate vs inanimate for objects, and indoor vs outdoor for scenes. As noted above, the two sub-ordinate categories were an important feature of the task design that allowed us to obtain independent indices of target and competitor reactivation. Both classifiers were trained on distinguishing sub-category membership (animate/inanimate and indoor/outdoor for objects and scenes, respectively) in the non-competitive ("pure") conditions during recall.

As a sanity check, we then first tested these classifiers' ability to detect the retrieval of objects and scenes in the non-competitive conditions (NC1 and NC2). Training and testing were conducted time point per time point, in steps of 8 msec, each time point centred in the middle of a Gaussian smoothing window (FWHM = 40 msec). MEG gradiometer sensor patterns (i.e., signal amplitude on each of the 204 channels) were used as classifier features, independently per participant and per time bin (from -500 to 3000 around cue onset). Since the same data was used for training and testing in this case, a 10-fold cross-validation was used and repeated 5 times. Results are shown in Figure 2D, where highlighted in red are significant time points corrected for multiple comparisons over time using a non-parametric cluster-based permutation test (as implemented in Fieldtrip toolbox version 2019, Oostenveld et al., 2011). This first

decoding analysis indicated that reliable ($p_{cluster} < .05$) above-chance decoding performance was found in a time-window from 2.55-2.92 sec post cue onset, somewhat later than the typical time window for reinstatement of associative memories in cued recall tests (for a review, see Staresina & Wimber, 2019). Earlier clusters did not survive stringent cluster correction. For decoding during visual presentation of the stimuli at encoding, see Supplemental Figure 3A.

To test for reactivation of target and competitor memories in the competitive condition, trials were split into those where the target was an object and the competitor was a scene, and vice versa. The corresponding object (animate vs inanimate) and scene (indoor vs outdoor) classifiers, again trained on non-competitive recall trials, were then used to indicate evidence for target and competitor reactivation, restricted to only correct trials. Figure 2E shows the average decoding accuracy in the competitive condition. Classification performance peaked in time-windows similar to the non-competitive condition. However, no significant cluster could be identified when correcting for multiple comparisons across time points, indicating that overall, reinstatement was less robust in the condition where two associates competed for recall.

In addition to general decodability when averaging across all retrieval repetitions, we also predicted in our pre-registration that in the competitive condition, the neural representation of the target memories would become stronger and the competing associate would become weaker across repetitions. A contrast of the decoding timelines between the first and third repetition revealed a significant increase ($p_{cluster} < .05$, 1.79 to 2.11 sec) in decodability of the target memory approximately 2 seconds after cue onset (Supplemental Figure 3C). No corresponding cluster emerged when comparing the entire timeline of competitor reactivation between the first and third recall (Supplemental Figure 3D). Zooming in on the time window where targets

and competitors were maximally decodable (1.77 to 1.93 sec after cue onset, $p_{uncorr} < .05$, see Supplemental Figure 3B), a significant repetition-by-associate interaction was found ($F(1,20) = 5.187$, $p = .034$), with a significant increase of target reactivation from first to third recall (first: 49.43%, SD = 4.08%, third: 53.14%, SD = 6.43%; $t(20) = -2.46$, $p = .02$), and a numerical decrease of competitor reactivation (first: 52.05%, SD = 4.55%, third: 50.82%, SD = 5.54%; $t(20) = 1.1$, $p = .29$) (Figure 2F). Thus, we observed both a strengthening of the relevant target memory, and a weakening of competing memory across repetitions in this select time window.

Together, the decoding results provide evidence that target memories can be decoded from MEG signals when no competition is present, while under conditions of interference, target reactivation is less robust in early trials but gradually increases over repetitions. This pattern could emerge due to the gradual strengthening of the relevant target memory, or due to reduced interference from the competing memory.

Hippocampal theta oscillation clocks the reactivation of target and competitor memories

The next step was to assess if the fidelity with which target and competitor memories can be decoded fluctuates in a theta rhythm. The analysis was conducted to determine the dominant frequency in decodability, to replicate our previous findings for target memories (Kerrén et al., 2018), and to extend the principle to competitor memories. We estimated the oscillatory component of the fidelity values, following our pre-registration and the rationale of our previous study (Kerrén et al., 2018). We used IRASA (as implemented in the Fieldtrip toolbox version 2021), a method that robustly detects and separates the oscillatory from the fractal signal component in electrophysiological data (Wen & Liu, 2016). IRASA was applied to the single-trial fidelity time courses in the CC condition, in a time window from 500 to 3000ms

post-cue onset, and separately for the target and competitor memories. The first 500ms were excluded to attenuate the influence of early, cue-elicited event-related potentials, and because no neural reinstatement is expected during this early cuing period (Staresina & Wimber, 2019). The analysis identified 3Hz as the frequency with the highest power for both target and competitor memories, and power at 3Hz exceeded the 95th percentile of the empirical chance distribution generated from surrogate-label classifiers (Figure 3A-B, red line indicating significant deviation). We therefore used 3Hz as the modulating frequency in all subsequent phase-amplitude coupling analyses that related the hippocampal theta phase to the strength of memory reinstatement.

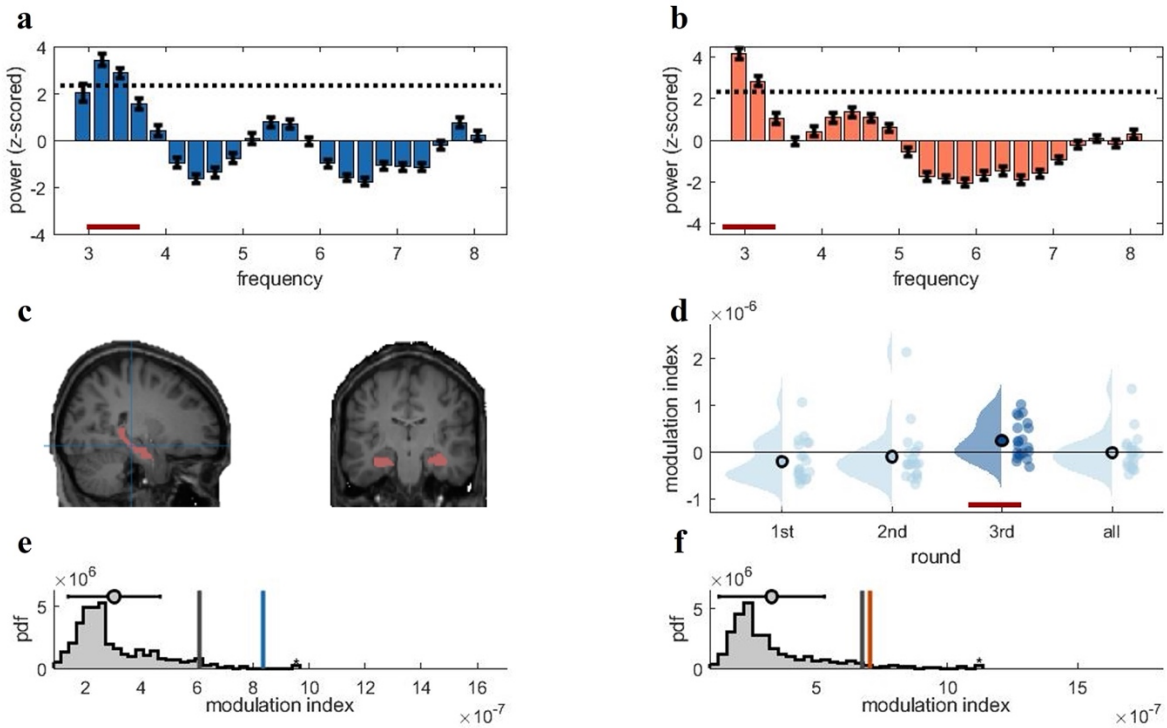


Figure 3. **Rhythmic fluctuations in classifier fidelity and phase-locking to hippocampal theta oscillation. A) and B)** An estimation of oscillatory components in the classifier fidelity values, using IRASA, revealed significant rhythmicity at 3Hz, compared with the same estimate from a label-shuffled baseline classifier (dashed black line; red line indicates frequencies that significantly differ from baseline) for both target (A) and competitor (B) memories. **C)** The hippocampal region of interest used as the source for theta phase, shown in sagittal and (left) coronal (right) plane**. D)** Distribution of each individual's modulation index (MI) with the 95th percentile subtracted. The third retrieval repetition shows significant modulation exceeding the 95th percentile (red line). **E) and F)** Results of the phase modulation analysis linking decoding fidelity to the phase of the hippocampal 3Hz oscillation. Fidelity values for target (E) and competitor memories (F) were significantly modulated by the hippocampal theta rhythm at 3Hz. The histogram shows the chance distribution of the MI, together with the 95th percentile in grey vertical bars and empirical value in blue for target memories and orange for competitor memories.

Having established that the fidelity values of the classifiers contained a significant rhythmic component at 3Hz, we next tested more specifically if the classifier-based indices of memory reinstatement were modulated by the phase of the hippocampal 3Hz theta rhythm. A linear constrained minimum variance (LCMV) beamforming approach (Gross et al., 2007) was used to reconstruct the activity of virtual hippocampal channels in source space. Again, the analysis procedure closely followed our previous study (Kerrén et al., 2018; see Methods for details). The modulation index (MI; Tort et al., 2010) measures the extent to which the classifier fidelity values are modulated by the hippocampal theta phase. The MI was calculated separately for each retrieval repetition, and we contrasted phase modulation of the fidelity values from the real-label classifier to the 95th percentile of an empirical chance distribution derived from surrogate-label classifiers. For both target and competitor memories, we found that fidelity values were only significantly modulated ($p < .05$) by the hippocampal 3Hz phase in the third and final retrieval repetition (Figure 3D-F). Note that this observation partly deviates from our preregistered hypotheses, where we expected phase modulation throughout the recall phase, accompanied by an increase in phase separation of targets and competitors across the three repetitions (see next paragraph).

Target and competing reactivations peak at distinct theta phases

The key hypothesis was that over time, the reactivated representations of target and competitor memories would become optimally separated along the phase of the theta oscillation. This hypothesis follows from the oscillating interference resolution model motivating our work (Norman et al., 2006), where competitors initially have a strong tendency to co-activate during the high inhibition phase of the theta cycle, but such co-activation is punished such that during later cycles, irrelevant competitors should only resurface during lower-inhibition portions of the theta cycle. Phase separation according to goal-relevance is also a central feature of other

prominent theta phase coding models (see: Hasselmo et al., 2002; Lisman & Jensen, 2013). Note that the original model (Norman et el., 2006) predicts a gradual increase of target-competitor separation with each oscillatory cycle of excitation-inhibition, whereas we hypothesized in our preregistration that this effect would be measurable when comparing early with late recall repetitions. As also noted in our preregistration, there is an influential theta phase model (Hasselmo et al., 2002) suggesting that encoding and retrieval operations in the hippocampal circuit are prioritised at opposite phases of the theta rhythm (for evidence in humans, see Kerrén et al., 2018). We therefore hypothesised that the target-competitor phase segregation would occur within the retrieval portion (e.g., half) of the theta cycle, rather than spread out across the entire cycle.

The phase binning method used to calculate the modulation index (see above) also allowed us to determine the theta phase bin at which target and competitor reinstatement was maximal in a given trial and condition. Importantly, the absolute phase of reinstatement in terms of its angle is likely to vary across participants (e.g., due to individual differences in anatomy). We therefore contrasted the phase of maximal target and competitor reinstatement within each individual participant, computing their phase distance as an index of phase separation. This phase distance is expected to be consistent across participants irrespective of the absolute angle of target and competitor reactivation, and can thus be subjected to group-level statistics. We used a Rayleigh test for non-uniformity (i.e., clustering) to test how coherent the phase separation angle was across participants; and a circular v-test to establish if the mean separation angle significantly deviated from zero. Note that without significant clustering, the mean angle is meaningless in such a phase analysis. It is possible to obtain significant clustering but no difference from zero, which would indicate that consistently across participants, targets and competitors reactivate at the same theta phase. Figure 4A shows the results of this phase

separation analysis. We found a consistent phase difference between target and competitor reactivation exclusively in the last repetition (Rayleigh test for non-uniformity, $z(20) = 5.03$, $p = .005$), clustered around a mean angular difference of 34 degrees. Furthermore, a test against zero phase shift confirmed that this target-competitor distance was significantly different from zero difference ($z(20) = 8.48$, $p = .004$). This was not the case for all repetitions averaged, $z(20) = .84$, $p = .44$, nor for the first, $z(20) = 2.19$, $p = .11$, or second, $z(20) = 1.76$, $p = .17$) repetition separately (Figure 4A). Furthermore, there was no significant difference in target-competitor phase separation between the first and third repetition when comparing the difference in mean angle ($z(20) = 1.09$, $p = .29$). Note, however, that there was no significant clustering around a stable mean angle in the first repetition (see above), and this statistical comparison is therefore inconclusive. Finding the separation between target and competitor memories only in the last repetition, rather than throughout the entire recall phase, deviates from our preregistered hypothesis, although paralleling the phase modulation results reported above. Overall, the phase distance results thus partly confirm our hypotheses, indicating that competing memories become increasingly separated along the hippocampal theta rhythm over time, with significant phase separation after several recall cycles.

The next set of analyses was not pre-registered, and tested for a relationship between target-competitor phase distance and behaviour. The sample was split according to each individual's intrusion score in the third repetition, dividing participants into a low intrusion and a high intrusion group (Figure 4B). We found that the high intrusion group had a mean separation angle of 7 degrees that was not significantly different from zero (Rayleigh test for non-uniformity, $z(9) = 1.77$, $p = .17$). The low intrusion group, by contrast, showed a mean phase separation of 57 degrees that significantly clustered around this angle (Rayleigh test for non-uniformity, $z(9) = 3.74$, $p = .02$), with a trend but no significant deviation from zero phase shift

in this sub-group of participants (Rayleigh test for non-uniformity with a specified mean angle of 0, $z(9) = 3.32$, $p = .069$). A statistical comparison of the phase separation angle in the high and low intrusion groups (Wilcoxon signed-rank test in non-circular space, see Methods) did not indicate a difference between the mean of the high and low intrusion groups' separation angle ($Z = 43$, $p = .85$). Note that each sub-sample in this comparison only contains 10 participants, and the statistical comparison is thus likely underpowered. In sum, this analysis suggests that only low intrusion participants exhibit significant phase separation by the end of the task, which can be taken as an indication that more neural differentiation of overlapping memories along the theta cycle relates to more successful behavioural differentiation.

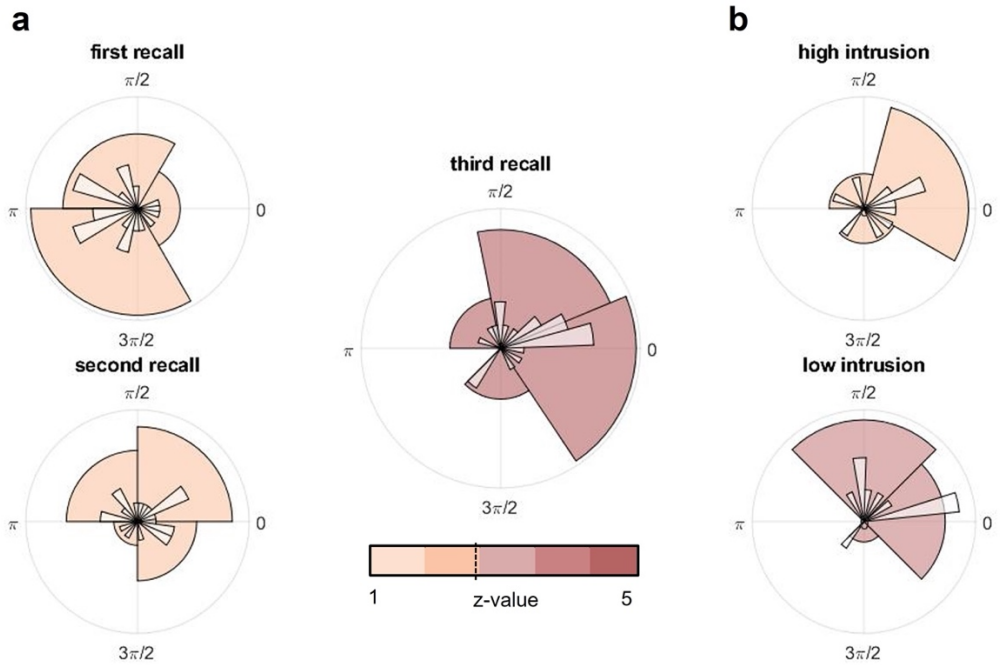


***Figure 4*. Target-competitor phase difference results. A)** Phase difference of maximal target minus competitor reactivation for each participant, separately for each recall repetition. At repetition 3 (middle circular plot) there was a phase shift of on average 34° that was statistically coherent across subjects and significantly different from zero, indicative of robust target-competitor phase segregation. **B)** When splitting the phase difference in the third repetition by behavioural intrusion score, the high intrusion group had a mean phase separation angle close to zero (top), while the low intrusion group had a mean separation angle of 57 degrees (bottom). Z-values (colour code) indicate strength of Rayleigh test for non-uniformity, with z-values/colours exceeding the dashed line on the colour bar indicating significant clustering of the phase difference angle.

Discussion

To remember a specific experience, the respective memory often needs to be selected against overlapping, competing memories. The processes by which the brain achieves such prioritisation are still poorly understood. We here tested a number of predictions derived from the oscillating interference resolution model (Norman et al., 2006). In line with its predictions, the present results demonstrate that the neural signatures of two competing memories become increasingly phase separated over time along the hippocampal theta rhythm. Furthermore, larger phase differences were associated with fewer intrusion errors from the interfering memories. These findings support the existence of a phase-coding mechanism along the cycle of a slow oscillation that adaptively separates competing mnemonic representations, minimising their temporal overlap (Lisman & Idiart, 1995; Norman et al., 2006).

Based on the oscillating interference resolution model (Norman et al., 2006), we hypothesised that target memories would be strengthened and become decodable at an earlier (higher inhibition) phase, while competing associates would be weakened, with time windows of their reactivation gradually shifting to a later (low inhibition) phase. The phase separation between the two competing memories should thus increase the more often the competing memories are being reactivated, and we expected to find an observable difference in phase separation after a number of repeated recalls. To further test whether a large phase segregation is beneficial for memory performance we related it to the number of behavioural intrusion errors. Supporting our hypotheses, in the third and final recall repetition target and competitor memories showed a significant phase-shift with an average distance of approximately 35 degrees (Figure 4A). Larger phase separation angles in individual participants were related to lower levels of behavioural interference. Participants with low levels of memory intrusions showed a significant phase difference of 57 degrees, while high-intrusion participants showed a non-

significant 7-degree phase shift on average (Figure 4B). Also in the last repetition, a significant phase-modulation was found for both target and competitor memories such that the fidelity of their neural reactivation was rhythmically modulated by the dominant 3Hz rhythm (Figure 3E-F). Though deviating from our preregistered hypotheses in some respects as discussed below, these results largely confirm our predictions derived from the oscillating interference resolution model (Norman et al., 2006), supporting a central role for slow oscillations in orchestrating memory recall by temporally segregating potentially interfering memories.

Based on the same model (Norman et al., 2006), we also predicted that across repeated retrieval attempts, target memories would be strengthened and competitors become less intrusive. Though not the focus of the present investigation, repeated recall is known to induce enhancement of the retrieved memories (Karpicke & Roediger, 2008; Rowland, 2014) and forgetting of non-retrieved, competing memories (Anderson et al., 1994) behaviourally. Evidence for the up- and down-regulation of neural target and competitor patterns, respectively, has previously been shown in fMRI studies (e.g., Kuhl et al., 2007; Wimber et al., 2015). In the present work, we found neural evidence for both target strengthening and competitor weakening within the time window of maximum memory reactivation (Figure 2F). The increase in target evidence was more robust than the decrease in competitor evidence, with the latter not surviving multiple comparisons correction across the entire time course of a recall trial. It is possible that competitor decoding was too noisy in the first place, and the lack of feedback in our task may have led to little decrease in the level of interference over time, as also reflected in the flat behavioural intrusion index (Figure 2B). Overall, however, our results suggest that dynamic changes in overlapping memories can be tracked using MEG.

The functional relevance of slow oscillations in regulating the excitation-inhibition balance of specific neural assemblies has featured strongly in other computational models too, and these models are supported by our findings at various levels. Most relevant in the present context is the theta-gamma model (Lisman & Idiart, 1995; Lisman & Jensen, 2013), which assumes that the coupling of gamma bursts to specific theta phases provides a timing mechanism for ordering sequences of items in working memory. Here, a gamma burst reflects the firing of a local cell assembly representing a distinct unit of information, while the phase of the slower theta oscillation rhythmically modulates the excitability of these local neural assemblies, and feedback inhibition allows for an organised, non-overlapping firing sequence across the theta cycle. In humans, studies using episodic and working memory tasks have provided evidence for such a theta-gamma code (Bahramisharif et al., 2018; Fuentemilla et al., 2010; Griffiths et al., 2021; Heusser et al., 2016; Reddy et al., 2021; ter Wal et al., 2021). The oscillating interference resolution model (Norman et al., 2006) and theta-gamma phase-coding models share the fundamental assumption that slow oscillations regulate the excitation-inhibition balance of local neural assemblies (Buzsáki, 2006). However, they do differ in their theoretical scopes. Theta-gamma code models were developed to explain how the brain can handle ongoing states of high attentional or working memory load, where multiple distinct items need to be kept separated and organised. The oscillating interference resolution model (Norman et al., 2006) is a learning model and thus explicitly oriented towards optimising future network states. It employs fluctuating inhibition to actively shape memory representations, with the goal to minimize future overlap and interference.

To our knowledge, no study in humans has explicitly tested for phase coding when several overlapping memories compete for retrieval. However, one intracranial study using a virtual navigation paradigm showed that the neural representations of potentially interfering spatial

goal locations are coupled to distinct phases of the hippocampal theta rhythm (Kunz et al., 2019). Phase precession and theta-coupling of neuronal firing during goal-directed navigation have also been shown in human hippocampal single unit recordings (Qasim et al., 2021; Watrous et al., 2018). The present results directly support the idea that phase coding can temporally separate co-active, competing memory representations in the human brain. They additionally suggest that such a code adaptively evolves across repeated target retrievals, in line with the idea that oscillating excitation-inhibition can optimise learning with respect to future access to the relevant target memory.

A second prominent computational model focuses on different functional operations within the hippocampal circuit at opposing phases of the theta rhythm, with one half of the hippocampal theta cycle devoted to the encoding of new incoming information, and the other half to the recovery of internally stored information (Hasselmo et al., 2002; Kunec et al., 2005). In humans, previous EEG work (Kerrén et al., 2018), and a recent study investigating rhythmic modulations of reaction times (ter Wal et al., 2021) have provided evidence for such process separation. In the present study, target and competitor reactivations peaked at distinct theta phases, with an average separation angle of 35 degrees, and 57 degrees in the low intrusion group that presumably reached optimal interference resolution. This observation is consistent with a framework where both targets and competitors are reactivated during the retrieval portion of a theta cycle (Hasselmo et al., 2002; Kunec et al., 2005), but with sufficient phase separation to allow for differentiation in time. Note, however, that the oscillating interference resolution model (Norman et al., 2006) dedicates the entire theta cycle to retrieval operations and would thus predict temporal distances of up to 180 degrees between targets and competitors. The two models may not necessarily be incompatible, especially when considering that encoding and retrieval computations take place in different subcircuits of the

hippocampus (Hasselmo et al., 2002). Future modelling work is needed, however, to better understand the learning dynamics across the full theta cycle, and to reconcile the apparent theoretical conflict between the two models.

Other predictions of the oscillating interference resolution model (Norman et al., 2006) relate to the neural overlap in spatial patterns, and have been tested primarily in fMRI studies. One such prediction is the non-monotonic plasticity hypothesis (NMPH), derived from the same learning dynamics along the oscillatory cycle as described above. If a competitor is sufficiently strong to be co-activated in the high inhibition (target) phase, it will benefit from synaptic strengthening of common connections and thus be integrated with the target memory. Moderately co-active competitors, on the other hand, will be subject to synaptic depression and as a result become differentiated from the target memory (Ritvo et al., 2019). Evidence for non-monotonic plasticity comes from fMRI work that directly tracks changes in the similarity of target and competitor representations in terms of their spatial patterns (Hulbert & Norman, 2015; Wammes et al., 2022), and the representational changes dependent on the level of (co)activation of overlapping memories (Detre et al., 2013; Lewis-Peacock & Norman, 2014; Poppenk & Norman, 2014; Wang et al., 2019). Indirectly related, fMRI studies investigating representational change resulting from repeated learning or recall have also found evidence for differentiation or integration of overlapping memories in hippocampus, late sensory cortex and prefrontal cortex (Favila et al., 2016; Hulbert & Norman, 2015; Kim et al., 2017; Kuhl et al., 2010; Schlichting & Preston, 2015; Wimber et al., 2015), and sometimes observed both in the same study but in different regions (Molitor et al., 2021; Schlichting & Preston, 2015). The present study was not designed to track representational changes, for example in terms of comparing target and competitor representations before and after repeated recall. Arguably, however, there is a straight-forward agreement between spatial and temporal separation in this

network model. Strong competitors would tend to co-activate in the high inhibition phase and thus be co-strengthened and integrated via long-term potentiation (LTP; Bliss & Lømo, 1973; Hebb, 1949), whereas moderately co-active competitor memories would tend to co-activate only at lower levels of inhibition, which according to the model would lead to long-term depression. Our data support the idea that low temporal separation is related to integration, behaviourally. The high-intrusion group had a mean phase difference of 7 degrees between target and competitor memories, theoretically in line with a time window of strong co-firing hence synaptic strengthening. Integration might thus have led to higher amounts of intrusions (Brunec et al., 2020). In the low-intrusion group, more pronounced phase-separation between target and competitor memories (57 degrees on average) might have led to only moderate co-firing and hence promoted differentiation, resulting in lower levels of behavioural interference. Although speculative, temporal separation could be a prerequisite for spatial differentiation. Future studies, combining high temporal with high spatial resolution, and paradigms to track the representational distance of overlapping memories, are needed to fully understand these dynamics.

A priori we included the entire theta band in our pre-registration, as previous studies have indicated large variations in the peak theta frequency in humans (Goyal et al., 2020; Jacobs, 2014; Kerrén et al., 2018; B. Lega et al., 2016; ter Wal et al., 2021). For example, in our previous study we found that reactivations of target memories were most consistently coupled to a hippocampal 7-8 Hz oscillation, even though a slower 3-4 Hz peak was also present (Kerrén et al., 2018). In another study, ter Wal and colleagues (2021) found that slower 2-3 Hz oscillations dominated behaviour and hippocampal signals, specifically when responses depended on memory, consistent with other intracranial work reporting pronounced slow theta oscillations in the human hippocampus during memory tasks (Goyal et al., 2020; Lega et al.,

2012). In working memory, it has even been demonstrated that theta oscillations slow down with increasing memory load (Fell et al., 2011; Wolinski et al., 2018), in line with the theoretical idea that slower rhythms accommodate better phase separation (Lisman & Idiart, 1995; Lisman & Jensen, 2013). While offering a possible explanation why slower oscillations are often observed in tasks with high memory demands, like the present experiment, such posthoc interpretations will need to be corroborated by empirical evidence in future studies.

Phase coding offers an elegant theoretical solution to temporally segregate potentially interfering information processing. Much empirical evidence for phase sequencing along the hippocampal theta oscillation comes from spatial navigation work, both in rodents and humans (O'Keefe & Recce, 1993; Reddy et al., 2021). We here demonstrated that phase coding facilitates the separation of overlapping, associatively linked memories. Together with the computational model used to derive these predictions, these findings offer a possible mechanism utilized by the human brain to resolve competition between simultaneously active memories. More generally, they add to a growing literature showing that slow oscillations orchestrate the intricate timing of neural processing with direct, observable effects on behaviour.

Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K. and a starting grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714). We also thank Sebastian Michelmann for their helpful conceptual input during data analysis.

Competing interests

The authors declare that no competing interests exist.

Materials and Methods

Subjects

Twenty-six right-handed participants (18 female, 8 male) took part for financial or course credit compensation (mean = 24.1 years, SD = 5.73 years, range 18 to 33). They all had normal or corrected-to-normal vision and reported no history of neurological disorders. All experimental procedures in the present study were approved by and conducted in accordance with the University of Birmingham's STEM Research Ethics Committee. Written informed consent was obtained from participants before they took part in the experiment.

Stimuli and task

The material consisted of 72 images depicting animate and inanimate objects (equal number of mammals, birds, insects, and marine animals, electronic devices, clothes, fruits, and vegetables), and 72 images depicting indoor and outdoor scenes, taken from the BOSS database (Brodeur et al., 2010) and online royalty-free databases. Stimulus selection was motivated by previous success at distinguishing these categories using time-resolved multi-variate pattern analysis (Kerrén et al., 2018). Images from both object and scene classes were pseudo-randomly split for each participant into 6 sets, so that each set consisted of 20 objects and 20 scenes, 10 animate and 10 inanimate, 10 indoor and 10 outdoor scenes. Each set constituted one learning block. Also, a list of 144 action verbs, largely overlapping with those used in (Linde-Domingo et al., 2019), served as cue words in the cued recall task. These words were

randomly assigned to images and conditions for each participant for creating the relevant word-image associations. Three additional associations were used for demonstrative purposes.

Participants received task instructions and first performed one short practice block. All participants (see exclusion above) then performed 6 experimental blocks (40 encoding trials and 72 (24x3) retrieval trials per block), each consisting of an associative learning phase, a distractor task, and a retrieval test with 3 repetitions per target item (Figure 1). At encoding, participants were asked on each trial to encode a word together with an image associate. In the competitive condition (CC), a word was encoded together with two associates, separated by at least three intervening trials. The instruction was to always memorise the most recent associate that was presented together with a given word for the subsequent memory test. Therefore, the second associate in the CC always served as the target, with the previously learned first associate (i.e., competitor) assumed to elicit proactive interference. In the non-competitive single exposure condition (NC1), a word was encoded together with only one associate, and these associations were never repeated during encoding. This condition served as the behavioural baseline for measuring the effect of proactive interference on memory performance (i.e., having previously encoded a competing associate compared with only one associate). In the non-competitive double exposure condition (NC2), participants also encoded a word together with only one associate, but these associations were presented twice. This condition served as the neurophysiological baseline and was specifically designed to control for neural effects induced by the repetition of the word cue (including but not limited to repetition suppression; Epstein et al., 2008; Kristjánsson & Campana, 2010). The conditions were shown in pseudo-randomised order within each block, where 1/3 of the associations were in NC1, 1/3 in NC2, and 1/3 in CC [NC1 = 8 associations, NC2 = 8 associations, CC = 8 associations]. In NC1 and NC2, there were an equal number of objects and scenes (4 objects and 4 scenes per

block, but shown twice each for NC2 hence resulting in 8 object and 8 scene trials), whereas in CC there were 8 object and 8 scene trials, with one pair each linked to the same cue word. In total this summed up to 40 trials for one block of learning (NC1 = 8 trials, NC2 = 16 trials, CC = 16 trials). The order of the trials belonging to the 3 conditions was pseudo-randomized such that the average serial position of each condition within a block was equal. Images were pseudo-randomly assigned to these conditions for each participant, with the constraint that in the CC, the associates needed to be from different image categories (one object and one scene, split such that on half of the CC trials, the target was an object and the competitor a scene image, and vice versa for the remaining half).

A learning trial consisted of a jittered fixation cross (between 500 and 1500ms), a unique action verb (1500ms), a fixation cross (1000ms), followed by a picture of an object or scene that was presented in the centre of the screen for 4 seconds. Participants were asked to come up with a vivid mental image that linked the image and the word presented in the current trial. As soon as they had a clear association in mind, they pressed the right-thumb key on the button box. Participants were aware of the later memory test.

A distractor task followed each learning phase. Here participants had to indicate if a given random number (between 1 and 99) presented on the screen was odd or even. They were instructed to accomplish as many trials as they could in 45 seconds and received feedback about their accuracy at the end of each distractor block.

After the distractor task, participants' memory for the 40 verb-object associations memorized in the immediately preceding learning phase was tested in pseudo-random order, with three repetitions per relevant association. The order was pseudo-randomized such that all

associations were first tested once in random order, before testing all associations again in a new random order, and again a third time. No significant difference in serial position resulted when comparing CC and NC2 items [t(23) = -1.69, p = .1], CC and NC1 items [t(23) = -1.18, p = .25], or NC1 and NC2 items [t(23) = -.43, p = .66]). One repetition consisted of 24 trials (8 for NC1, 8 for NC2 and 8 for CC), with the constraint that a given association could only be tested once per repetition. Each trial consisted of a jittered fixation cross (500-1500ms), followed by one of the words as a reminder for the association. Participants were asked to bring back to mind the most recent associate of this word as vividly as possible. The cue was presented on the screen for 500ms and thereafter a blank screen with a black empty frame was presented. To capture the particular moment when participants consciously recalled a specific association, they were asked to press the right-thumb key as soon as they had a vivid image of the associated memory in mind. If they did, the frame flashed once, and participants were presented for 4 seconds with a blank screen and asked to hold the image in mind. A question then appeared on the screen asking if the retrieved item was an object, a scene, or they were unable to remember. Across trials, the object and scene options randomly shifted between the left and right sides of the screen. If the participant did not remember the association, they were told to press the left-thumb button. If participants selected "object" or "scene", a follow-up question appeared (dependent on the response to the first question), asking if the retrieved associate was an inanimate or animate object, or whether it was an indoor or outdoor scene. The two follow-up questions were self-paced, and there was no feedback.

The experiment was set up via custom-written code in MATLAB 2016a (©The Mathworks, Munich, Germany) using functions from the Psychophysics Toolbox Version 3 (Brainard, 1997). The presentation was projected onto a screen located 1.5 meters away from the participant, using Windows 64 bit. Images were 2000x2000 pixels large.

Analysis

The general hypotheses and analysis steps were pre-registered and can be found on OSF (https://osf.io/evgaw/?view_only=1db46b44a71a47a3926c36bad017cb25).

After excluding participants based on the criteria stated in the pre-registration (more than 2 SD deviation from the mean accuracy for each condition separately), 24 participants (16 female, 8 male) remained, with an average age of 24.5 years (SD = 5.73). For the magnetoencephalography (MEG) analysis, a further 3 participants were excluded due to noisy data, which resulted in 21 participants included in the analyses. All participants performed all six blocks except for two, for whom time limit and button box errors occurred, resulting in only five blocks for these participants for MEG analysis.

Statistical analysis

In view of the pre-registration on OSF, all statistics were one-sided and conducted on a group level. Non-parametric cluster-based permutation tests were conducted to test classification performance against chance across multiple time points, and for time-frequency analyses comparing competitive and non-competitive conditions (see below). Only correct trials were used for all MEG analyses. Where a specific frequency band had to be preselected, the analyses were limited to 3 to 8 Hz. This, however, deviated from the frequency range of interest in the pre-registration, which was 4 to 8 Hz. Since the study was pre-registered, several studies have shown slower theta oscillation specifically related to human episodic memory (Goyal et al., 2020; ter Wal et al., 2021) and we therefore decided to include also 3Hz in the frequency band of interest.

MEG data analysis

The MEG was recorded at the Centre for Human Brain Health (CHBH), Birmingham, UK, using an Elekta Neuromag TRIUX system, with 306 channels (204 planar gradiometers and 102 magnetometers; only gradiometers are used for all analyses reported here), sampled at 1000Hz (Elekta, Stockholm, Sweden). EEG was recorded with a 64-channel electrode cap in the initial 10 participants as a sanity check, in order to verify that a strong mid-frontal theta signal can be observed in the CC > NC2 condition, and how its topography compares between MEG and EEG (Cavanagh & Frank, 2014). Since we found a comparably strong theta increase over EEG and MEG (gradiometer) sensors, the EEG data is not reported in any of the analyses presented here. The experiment was shown on a projector screen using a PROPixx projector (VPixx Technologies, Saint-Bruno, Canada) with a 1440Hz refresh rate, and participants' responses were collected using two button response boxes (fMRI Button Pad (2-Hand) System, NAtA Technologies, Coquitlam, Canada).

For source localisation, individual anatomical MRI scans (T1-weighted; 1x1x1 mm voxels; TR = 7.4 ms; TE = 3.5 ms; flip angle = 7°, field of view = 256x256x176 mm) were acquired at the CHBH (3T Achieva scanner; Philips, Eindhoven, the Netherlands).

Preprocessing

Preprocessing was done using the FieldTrip toolbox version 2019 and 2021 (Oostenveld et al., 2011) and custom-written MATLAB code. A Butterworth low-pass filter of 200Hz, and a band-stop filter (50Hz; 100Hz, and 150Hz), were applied to the data. To avoid temporal displacement of information, while still removing slow drifts, trial-masked robust detrending was applied to the retrieval data, in accordance with recent recommendations (van Driel, Olivers, & Fahrenfort, 2021). Briefly, the data for this correction step were divided into epochs spanning

15 seconds before cue onset and 15 seconds after cue onset. The experimentally relevant events started 1000ms before cue onset until 4000ms after cue onset. This time-window was masked out from each trial. To make sure all data would be included, the continuous data were symmetrically mirror-padded with 15 seconds prior to segmentation. To improve the fit of the higher order polynomial, a 1st order polynomial was used to detrend the entire epoch (in accordance with de Cheveigné and Arzounian (2018). Thereafter, a 30th order polynomial was fitted and removed from the data. Note that the events of interest were not a part of the fitting procedure, making sure the fit was not being influenced by cognitively relevant processing. The method was partly implemented using the Noise Tools toolbox (http://audition.ens.fr/adc/NoiseTools) together with custom-written MATLAB-code. After this step, the detrended data were cut into the experimentally relevant epochs (-1000 to 4000ms around cue onset at retrieval).

An automatized trial and component rejection was applied to the data. In a first step, to remove high-frequency bursts, data were high-pass filtered at 100Hz, and trials that exceeded 4 times the median absolute deviation of the amplitude distribution across trials were automatically removed. In a second step, independent component analysis (ICA) was used to detect artifacts to be removed in the data. To this end, MEG data were downsampled to 250Hz, and only the first 1.5 seconds after cue onset were used for ICA (to reduce computational load). The unmixing matrix was then applied to the entire epoch of non-downsampled data. To remove blink and cardiac artifacts, a template was created by running an ICA on the six first participants. The components for the two artifacts were then averaged independently and used as a template for all participants. For the blink template, the algorithm correlated the topography of every component with the blink template and then found the best match, and subsequently removed it. For the cardiac component, which is difficult to identify based on

topography, the algorithm filtered the component time-series to between 10 and 20 Hz and correlated the peak amplitude of each trial with the cardiac time-series template. Any component which correlated to a greater extent than 4 times the median absolute deviation of the correlation was classed as a cardiac component. All removed components were visually inspected for validation.

After these components were removed, an additional automatized artifact rejection was conducted, similar to the initial step. The same procedure as in the first rejection round was followed but was done on channels instead of trials. Again, data were band-pass filtered between 1 and 100Hz, and the channels that exceeded 3 times the median absolute deviation of the channel distribution were rejected. We chose three times the median absolute deviation for the last two rounds of rejection to be more conservative. Bad channels were interpolated using the triangulation method (Oostenveld et al., 2011). After components and trials were removed, all trials and components were again visually inspected, and trials and components still containing artifacts were manually removed. Lastly, data were downsampled to 250Hz. Due to poor data quality in 3 participants, remaining data from 21 participants were used for MEG analyses. On average 353 out of 432 retrieval trials were kept (min = 290, max = 397, SD = 27) for analysis. In the competitive condition the number of trials on average was 118 (max = 140, min = 95, SD = 10.24). Per retrieval repetition there was an average of 40 (SD = 3.81), 39 (SD = 3.63) and 38 (SD = 4.31) trials for each respective repetition. For the NC1 and NC2 collapsed, the average number of trials per retrieval repetition were 79 (SD = 7.38), 77 (SD = 6.36) and 77 (SD = 7.29). At the end of preprocessing, a sanity ERF-check on occipital channels of the retrieval data was conducted and the average waveform can be found in Supplemental Figure 2A.

Time frequency decomposition

The spectral difference between the CC and neural baseline (NC2) during retrieval was calculated by convolving the combined activity of each planar channel (summing the magnitude of the planar gradient over both directions at each sensor) with a complex Morlet wavelet of a minimum of 5 cycles (increasing with frequency to cover approximately 500ms in length [for example, 10Hz: 10*.5 = 5 cycles]) from 1 to 20 Hz for each condition. First, paired-samples t-tests were computed between the two conditions to investigate the difference in spectral power. To account for multiple comparisons across time points (0ms to 1000ms post-cue), frequencies (3 to 8 Hz) and sensor (102 combined gradiometers), the t-statistics were subjected to nonparametric cluster-based permutation testing, as implemented in the FieldTrip software. The threshold for statistical testing was set to a cluster alpha level of 0.05. The minimum number of neighbouring channels that were considered a cluster was set to three. T-values above the threshold of 0.1 were then summed up in a cluster and compared against a distribution where condition labels were randomly assigned 1000 times with the Monte-Carlo method, following the default method implemented in FieldTrip.

Multivariate pattern analysis

Two separate Linear Discriminant Analysis (LDA) based classifiers were trained on discriminating sub-ordinate picture class, i.e., animate vs inanimate for objects, and indoor vs outdoor for scenes. This important feature of the design allowed us to have two classifiers that provide independent measures of target and competitor reactivation. The two classifiers were trained on the non-competitive ("pure") conditions (i.e., retrieval of objects and scenes in the NC1 and NC2 conditions), and were then tested on both the non-competitive and the competitive conditions. The time interval of interest for these multivariate analyses started 500ms before onset of the retrieval cue and lasted up to 3000ms post-cue. This time window

was selected because no memory reactivation is expected earlier than 500msec (Staresina & Wimber, 2019), and previous work using a similar (though non-competitive) cued recall paradigm suggests that participants need approximately 3 sec to mentally reinstate a image (Linde-Domingo et al., 2019; Kerrén et al., 2018). A sliding Gaussian window with a full-width at half maximum (FWHM) in the time-domain of 40ms was applied to the trial time series before classification analysis. Each retrieval trial was then baseline corrected by subtracting the mean amplitude in a time window from 400 to 50ms pre-cue, separately per sensor. Training and testing were conducted time point per time point, in steps of 8 msec, each time point centred at the centre of the Gaussian smoothing window. The MEG gradiometer sensor patterns (amplitudes on each of the 204 channels at a given time point) were used as classifier features, independently per participant.

Training/testing in the non-competitive condition was done as a sanity check for classifier performance in a memory retrieval situation with no interference. Since the same data was used for training and testing in this case, a 10-fold cross-validation was used and repeated 5 times. To test for reactivation of target and competitor memories in the competitive condition, trials were split into those where the target was an object and the competitor was a scene, and vice versa, and the corresponding object and scene classifiers (trained on non-competitive retrieval) were then used to separately indicate evidence for target and competitor reactivation on each single trial. Only correct trials were used in the competitive condition. As the training and testing data came from different trials, cross-validation in the competitive condition was not necessary. To avoid overfitting, the covariance matrix was regularized using shrinkage regularization, with the lambda set to automatic (Blankertz et al., 2011)

The LDA, as used here, reduces the data from 204 channels into a single decoding time course per trial, and we used these single-trial, time-resolved decision values (d-values or fidelity values) of the classifier as an index of memory reinstatement (Carlson et al., 2014). Classifier accuracy was derived by calculating the fraction correctly predicted labels, whereas chance was defined as 50% for a binomial classifier. More specifically, during training, the classifier found the decision boundary that could best separate the patterns of activity from the two classes (animate vs inanimate for objects, indoor vs outdoor for scenes) in a high-dimensional space. The classifier was then asked to estimate whether the unlabelled pattern of brain activity in any given retrieval trial and at each time point was more similar to one or the other class. This training-test procedure was repeated until every single retrieval trial had been classified. A larger smoothing kernel of 200ms FWHM in the time domain was applied to the decoding time courses purely for visualisation but not for statistical analysis purposes.

Determine peak frequency of fidelity values using IRASA

The Irregular-Resampling Auto-Spectral Analysis (IRASA) method has been shown to robustly find and separate the oscillatory from the fractal signal both in ECoG and MEG data, and was here used to quantify the oscillatory signal component of the fidelity values (Wen & Liu, 2016). More specifically, the brain produces task-related rhythmic (oscillatory) components, but also arrhythmic scale-free (fractal) components (Buzsáki & Draguhn, 2004). The rhythmic oscillatory components are regular across time, whereas the fractal components are irregular (Wen & Liu, 2016). In short, IRASA resamples a time-series signal and computes a geometric mean of every pair (oscillatory and fractal) of the resampled signal. The median of the geometric mean is then used to extract the fractal power spectrum. The difference between the original power spectrum and the fractal power spectrum is the estimate of the power spectrum of the oscillatory component of the signal (Wen & Liu, 2016). In the present study,

IRASA (as implemented in the FieldTrip toolbox) was applied to the fidelity values, in a time window from 500 to 3000ms post-cue, padding each trial length up to the next power of 2. Apart from the reasons mentioned above, this large time window also assures that low frequencies can be properly estimated.

To test for significance in the frequency-transformed decoding time courses, the LDA as described above was repeated 15 times per subject, now trained on random labels in each iteration. The fidelity values from each iteration were then subjected to the same IRASA-method as the non-shuffled data. An empirical null distribution was created on the second (i.e., participant) level as follows. In each of 1000 repetitions, one of the 15 surrogate-label classifier outputs (i.e., the oscillatory component output by IRASA) per participant was randomly selected, and an empirical group average was computed for this random selection. This procedure resulted in 1000 group averages of oscillatory strength estimates for each given frequency of interest, representing the chance distribution at this frequency. The IRASA output of from the real-label classifier could then be contrasted against this chance distribution (Cohen, 2014). Note that due to the algorithm for separating the fractal components from the oscillatory components, the output from IRASA yields a higher frequency resolution than 1 point per frequency. However, our a priori frequency range was set to 3 to 8 Hz and we therefore tested for significant oscillatory components in each 1Hz frequency bin between 3 and 8 Hz, with the estimated chance distribution subtracted from the real value and subsequently divided by the standard deviation of the estimated chance distribution (Figure 3A-B). This gave a z-value, which was compared to the critical threshold of $z = 2.32$ at $p = .01$, correcting for 5 multiple comparisons (3-4 Hz, 4-5 Hz, 5-6 Hz, 6-7 Hz, 7-8 Hz). Again, this procedure replicated the one used in Kerrén et al. (2018).

Source Analysis

The aim was to extract activity from hippocampal virtual channels; hence the magnetometers were used and their data projected into source space using a linear constrained minimum variance (LCMV) beamforming approach (Gross et al., 2001). Individual MRI scans (anatomical data) were available for 18 participants, and for the remaining 3 participants the standard head model as implemented in FieldTrip was used. Anatomical scans were aligned based on the sensor position obtained from a Polhemus system (Colchester, Vermont, USA). This was done in three steps: the first step was done manually by adjusting the alignment between the anatomical data and the head position; the second was done by the Iterative Closest Point algorithm implemented in the Fieldtrip toolbox (Oostenveld et al., 2011); the third step was again done manually to check that the alignment worked, and adjusted if it did not. The realigned model was used to reslice and segment the brain to make the axes of the voxels consistent with the head position and subsequently to extract the brain compartments. The segmented brain was then used to create a forward model (head model). In our case, we used a semi-realistic forward model (Nolte, 2003). The forward model was used to create the source model (lead field), where for each grid point the source model matrix was calculated with a 1 cm resolution, and the virtual sensors were placed 10mm below the cortical surface, and subsequently warped into each brain. In total we modelled 3294 virtual sensors for each participant with whole-brain coverage. LCMV beamforming was used to reconstruct the activity of all virtual channels in source space (see below for selecting hippocampal channels) with twenty percent regularisation. To confirm that the source localization provided reliable results, we checked sources for the early visually elicited response to the cue word in an early time window, by calculating the pre/post variance around cue onset (-200ms to 0 and 0 to 200ms). Because of arbitrary orientation of the dipole in source space, variance is an unbiased

measure to investigate where in space and when in time activity fluctuates (see Supplemental Figure 2B).

For the phase-amplitude coupling between source-localized hippocampal 3Hz phase and fidelity values, the filters were computed on 4Hz lowpass-filtered data, from -1000ms to 3000ms after cue onset (baseline-corrected [-400 to -50]). The filters were obtained from correct trials in the competitive condition. The left and right hippocampus were chosen as a Region of Interest (ROI), derived from the AAL atlas (Oostenveld et al., 2011). The individual MRI and the atlas were interpolated using the interpolation method *nearest*, finding the nearest region in the AAL atlas based on the Euclidean distance. Single-trial time courses were extracted from the corresponding virtual sensors, and activity averaged across those sensors in the left and right hippocampus, resulting in one activity time course per trial for left and right hippocampus, collapsed.

Phase-amplitude coupling between MEG data and fidelity values

The modulation index (MI; Tort et al., 2010) was calculated to test for a relationship between the fidelity values and a hippocampal theta oscillation, with 3Hz selected as the frequency of interest based on the results from previous steps (IRASA, see above). The signal from virtual channels was transformed to an analytic signal by convolving the raw data with a complex Morlet wavelet with cycles increasing with frequency to be approximately 500ms in length for each frequency as implemented in Fieldtrip (Oostenveld et al., 2011). Each complex value data point was then point-wise divided by its magnitude, resulting in a 4D-matrix of phase values, containing trials*channels*frequencies*time. Phase values at 3Hz were divided into 10 adjacent bins, ranging from – pi to pi, following the pre-registered procedure. To link classifier-based memory reinstatement indices to the hippocampal phase, the amplitudes of the single

trial fidelity values from corresponding time points were sorted into their respective phase bins, and the average classifier amplitude of each phase bin was calculated.

Following this binning procedure, two metrics were of relevance: (i) the MI showing the extent to which the fidelity values are modulated by the hippocampal phase; and (ii) the maximum bin for target and competitor reinstatement. The MI was obtained by creating a uniform distribution and calculating the Kullback-Leibler distance and subsequently the MI, in accordance with the approach originally proposed by Tort et al., (2010). To calculate the pairwise phase difference between the target and competitor reinstatement, the absolute value of the complex number (the real and the imaginary part) representing the average maximum reinstatement peak for each participant was obtained by taking the square root of the sum of the squares of the parts (using the Pythagorean Theorem). The amplitude was arbitrarily set to 1 because we were only interested in phase angle. The phase vectors were then point-wise divided with the complex modulus. And finally, to subtract the angles from each other, one vector was rotated by multiplying it with the other's complex conjugate. To obtain the angle, the inverse tangent of the ratio was taken of the product between the vectors. In MATLAB the following code was used:

$$\theta = angle(\exp(1i \times \boldsymbol{X}) .\div \exp(1i \times \boldsymbol{Y}))$$

where θ is the vector of angle difference.

We used the CircStat toolbox (Berens, 2009), as implemented in MATLAB, to statistically test the phase difference between target and competitor memories. More specifically, to test the difference in each repetition, as well as the average distance collapsed across all repetitions, we used a Rayleigh test for non-uniformity (*circ_rtest* function in the toolbox). This test will show significant clustering if the phase differences between target and competitor memories

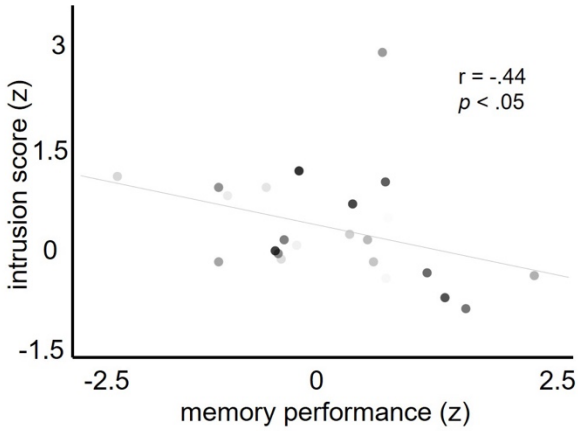
are non-uniform, independent of their absolute mean angle. To test whether the mean phase difference we obtained was significantly different from zero phase shift, we used the *circ_vtest* function with a specified mean angle of 0. This test indicates whether the mean angle of the phase distance is significantly different from zero. Lastly, we wanted to test the difference between the high and low intrusion groups' mean angle. To do so, we calculated the mean phase angle of target and competitor reactivation for each participant. We then calculated the circular mean of the two vectors within each participant, where the length of this vector defines the similarity between target and competitor phase angles, independent of absolute angle. This vector is non-circular and can be subjected to a Wilcoxon sign-rank test. We choose a Wilcoxon sign-rank test as the sample size was low in each sub-group (n = 10 in each group, removing one participant from the larger group randomly to equate group size).

Additional information

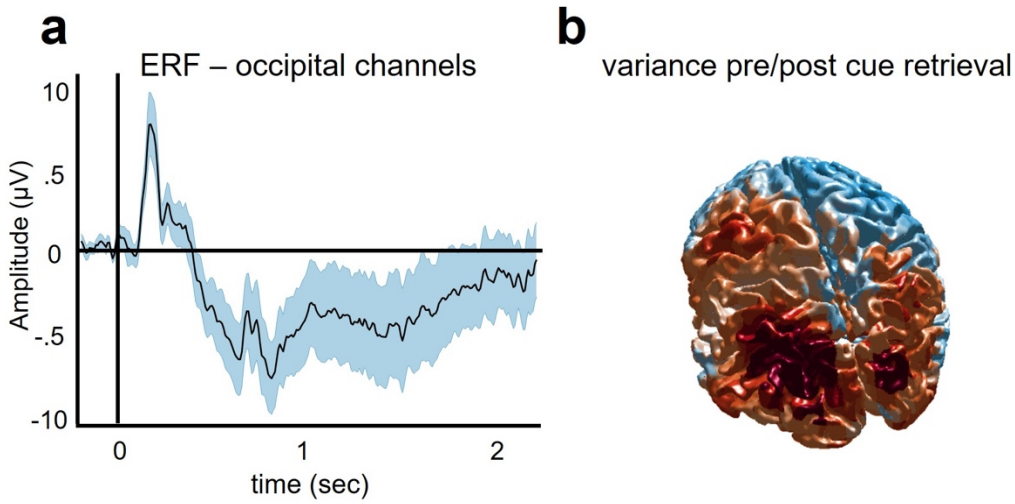
Author contributions

Conceptualization, C.K. and M.W.; Methodology, C.K., B.G. and M.W.; Investigation, C.K.; Formal Analysis, C.K., S.B and B.G.; Writing Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K., B.G., S.v.B and M.W.; Visualization, C.K.; Funding Acquisition, C.K. and M.W.

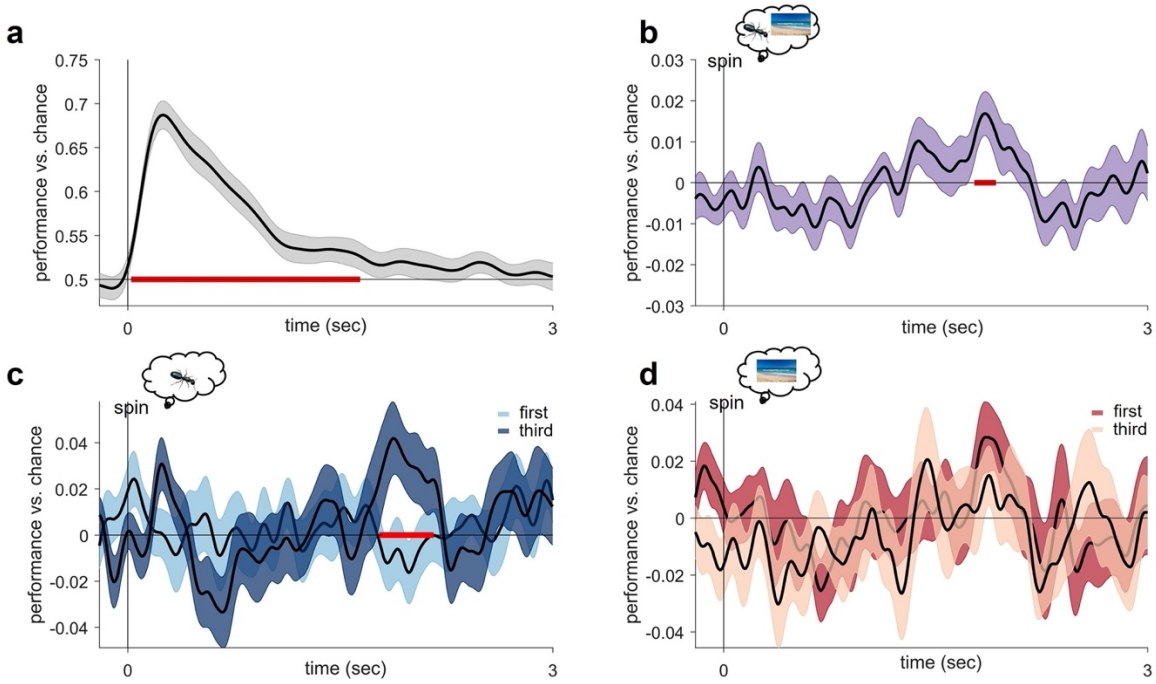
Supplemental Material



Supplemental Figure 1. Correlation between memory performance in the non-competitive condition (NC1 and NC2) and the number of intrusion errors in the competitive condition (CC), with each dot representing one participant's z-scored performance. A correlation of r = -.44 ($p < .05$) was found such that more intrusions in the competitive condition were related to lower memory performance in the baseline conditions.



Supplemental Figure 2. **A)** An ERF analysis was conducted after preprocessing to ensure that a clean visually-evoked potential was obtained from posterior sensors. The data were low-pass filtered at 30Hz, and the vertical and horizontal gradients were summed for occipital channels. Before averaging the trials, the data were baseline corrected based on the activity at each channel from -200 to -50ms before cue onset at retrieval. Result show a typical ERF with a maximum around 200ms after cue onset. **B)** A variance index was calculated to check the source distribution in response to a visual stimulus. The data were source localised using a LCMV beamforming approach (Gross et al., 2001). We then subtracted the temporal variance from -200 to 0 ms pre-cue from the – to 200 ms post-cue variance, and subsequently divided it with the pre-cue variance. We found the expected source distribution in this early time window with an occipital maximum.



Supplemental Figure 3. **A)** LDA-based decoding of image class during visual exposure, conducted as a sanity check. MEG trial timecourses were smoothed with a 40ms gaussian window and thereafter baseline corrected from -200 to -50ms before cue onset. The spatial pattern (i.e., sensor amplitudes) of all gradiometers at a given time point was used as feature vector to train the classifier. A k-fold-10 cross-validation repeated 5 times was used to minimize dependencies between training and testing data. This analysis revealed a cluster of significantly above-chance decoding performance peaking around 150ms after image onset. **B)** Decoding collapsed over target and competitor memories averaged across all retrieval attempts, in order to identify time window(s) of maximum decodability common to targets and competitors, to then test for changes over repetitions in this select time window. We found significant decoding (uncorrected $p < .05$, with cluster not surviving stringent cluster-correction, $p = .15$) between 1.77 and 1.93 seconds after cue onset. **C)** Contrasting target decoding for first and third repetition revealed significantly higher decoding (at $p_{cluster} < .05$) in a similar time window as in B), in line with the pre-registered hypothesis. **D)** When instead testing for a down-regulation of competitor memories across retrieval attempts, despite a qualitative difference in the expected direction around 2 sec, the contrast between first and third repetition did not reveal any statistically significant clusters (at $p_{cluster} < .05$), contrary to our preregistered hypotheses.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In *Memory* (pp. 237–313). Academic Press. https://doi.org/10.1016/B978-012102570-0/50010-0

Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2018). Serial representation of items during working memory maintenance at letter-selective cortical sites. *PLOS Biology*, *16*(8), e2003805. https://doi.org/10.1371/journal.pbio.2003805

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*, 1–21. https://doi.org/10.18637/jss.v031.i10

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—A tutorial. *NeuroImage*, *56*(2), 814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048

Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, *232*(2), 331–356.

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., & Buzsáki, G. (1995). Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *15*(1 Pt 1), 47–60.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS One*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brunec, I. K., Robin, J., Olsen, R. K., Moscovitch, M., & Barense, M. D. (2020). Integration and differentiation of hippocampal memory traces. *Neuroscience and Biobehavioral Reviews*, *118*, 196–208. https://doi.org/10.1016/j.neubiorev.2020.07.024

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301069.001.0001

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science (New York, N.Y.)*, *304*(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science (New York, N.Y.)*, *313*(5793), 1626–1628. https://doi.org/10.1126/science.1128115

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience*, *26*(1), 132–142. https://doi.org/10.1162/jocn_a_00476

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M.-B., & Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, *462*(7271), 353–357. https://doi.org/10.1038/nature08573

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, *51*(12), 2371–2388. https://doi.org/10.1016/j.neuropsychologia.2013.02.017

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two Kinds of fMRI Repetition Suppression? Evidence for Dissociable Neural Mechanisms. *Journal of Neurophysiology*, *99*(6), 2877–2886. https://doi.org/10.1152/jn.90376.2008

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1), 11066. https://doi.org/10.1038/ncomms11066

Fell, J., Ludowig, E., Staresina, B. P., Wagner, T., Kranz, T., Elger, C. E., & Axmacher, N. (2011). Medial Temporal Theta/Alpha Power Enhancement Precedes Successful Memory Encoding: Evidence Based on Intracranial EEG. *Journal of Neuroscience*, *31*(14), 5392–5397. https://doi.org/10.1523/JNEUROSCI.3668-10.2011

Ferreira, C. S., Marful, A., Staudigl, T., Bajo, T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *Journal of Cognitive Neuroscience*, *26*(4), 777–791. https://doi.org/10.1162/jocn_a_00523

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-Coupled Periodic Replay in Working Memory. *Current Biology*, *20*(7), 606–612. https://doi.org/10.1016/j.cub.2010.01.057

Goyal, A., Miller, J., Qasim, S. E., Watrous, A. J., Zhang, H., Stein, J. M., Inman, C. S., Gross, R. E., Willie, J. T., Lega, B., Lin, J.-J., Sharan, A., Wu, C., Sperling, M. R., Sheth, S. A., McKhann, G. M., Smith, E. H., Schevon, C., & Jacobs, J. (2020). Functionally distinct high and low theta oscillations in the human hippocampus. *Nature Communications*, *11*(1), 2469. https://doi.org/10.1038/s41467-020-15670-6

Griffiths, B. J., Martín-Buro, M. C., Staresina, B. P., & Hanslmayr, S. (2021). Disentangling neocortical alpha/beta and hippocampal theta/gamma oscillations in human episodic

memory formation. *NeuroImage*, *242*, 118454. https://doi.org/10.1016/j.neuroimage.2021.118454

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2007, November). *Dynamic imaging of coherent sources: Studying neural interactions in the human brain*. https://doi.org/10.1073/pnas.98.2.694

Hanslmayr, S., Staudigl, T., Aslan, A., & Bäuml, K.-H. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive, Affective & Behavioral Neuroscience*, *10*(3), 329–338. https://doi.org/10.3758/CABN.10.3.329

Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, *14*(4), 793–817. https://doi.org/10.1162/089976602317318965

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory* (pp. xix, 335). Wiley.

Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nature Neuroscience*, *19*(10), 1374–1380. https://doi.org/10.1038/nn.4374

Hulbert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents: Implications for models of spatial navigation and memory. *Philosophical Transactions*

of the Royal Society B: Biological Sciences, *369*(1635), 20130304. https://doi.org/10.1098/rstb.2013.0304

Johansson, M., Aslan, A., Bäuml, K.-H., Gäbel, A., & Mecklinger, A. (2007). When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(6), 1335–1341. https://doi.org/10.1093/cercor/bhl044

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science (New York, N.Y.)*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kerrén, C., Linde-Domingo, J., Hanslmayr, S., & Wimber, M. (2018). An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Current Biology: CB*, *28*(21), 3383-3392.e6. https://doi.org/10.1016/j.cub.2018.08.065

Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural Differentiation of Incorrectly Predicted Memories. *The Journal of Neuroscience*, *37*(8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017

Kliegl, O., & Bäuml, K.-H. T. (2021). Buildup and release from proactive interference—Cognitive and neural mechanisms. *Neuroscience and Biobehavioral Reviews*, *120*, 264–278. https://doi.org/10.1016/j.neubiorev.2020.10.028

Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, *72*(1), 5–18. https://doi.org/10.3758/APP.72.1.5

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, *10*(7), 908–914. https://doi.org/10.1038/nn1918

Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature Neuroscience*, *13*(4), 501–506. https://doi.org/10.1038/nn.2498

Kunec, S., Hasselmo, M. E., & Kopell, N. (2005). Encoding and Retrieval in the CA3 Region of the Hippocampus: A Model of Theta-Phase Separation. *Journal of Neurophysiology*, *94*(1), 70–82. https://doi.org/10.1152/jn.00731.2004

Kunz, L., Wang, L., Lachner-Piza, D., Zhang, H., Brandt, A., Dümpelmann, M., Reinacher, P. C., Coenen, V. A., Chen, D., Wang, W.-X., Zhou, W., Liang, S., Grewe, P., Bien, C. G., Bierbrauer, A., Navarro Schröder, T., Schulze-Bonhage, A., & Axmacher, N. (2019). Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation. *Science Advances*, *5*(7), eaav8192. https://doi.org/10.1126/sciadv.aav8192

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, *94*(3), 1904–1911. https://doi.org/10.1152/jn.00263.2005

Lega, B., Burke, J., Jacobs, J., & Kahana, M. J. (2016). Slow-Theta-to-Gamma Phase–Amplitude Coupling in Human Hippocampus Supports the Formation of New Episodic Memories. *Cerebral Cortex*, *26*(1), 268–278. https://doi.org/10.1093/cercor/bhu232

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, *22*(4), 748–761. https://doi.org/10.1002/hipo.20937

Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. *Nature Communications*, *5*(1), 5768. https://doi.org/10.1038/ncomms6768

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179. https://doi.org/10.1038/s41467-018-08080-2

Lisman, J. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, *15*(7), 913–922. https://doi.org/10.1002/hipo.20121

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, *267*(5203), 1512–1515. https://doi.org/10.1126/science.7878473

Lisman, J. E., & Jensen, O. (2013). The θ-γ neural code. *Neuron*, *77*(6), 1002–1016. https://doi.org/10.1016/j.neuron.2013.03.007

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., & Preston, A. R. (2021). Memory Reactivation during Learning Simultaneously Promotes Dentate Gyrus/CA2,3 Pattern Differentiation and CA1 Memory Integration. *Journal of Neuroscience*, *41*(4), 726–738. https://doi.org/10.1523/JNEUROSCI.0394-20.2020

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors.

Physics in Medicine and Biology, *48*(22), 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002

Norman, K. A., Newman, E., Detre, G., & Polyn, S. (2006). How Inhibitory Oscillations Can Train Neural Networks and Punish Competitors. *Neural Computation*, *18*(7), 1577–1610. https://doi.org/10.1162/neco.2006.18.7.1577

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1023–1035. https://doi.org/10.1016/j.neubiorev.2009.12.014

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317–330. https://doi.org/10.1002/hipo.450030307

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Poppenk, J., & Norman, K. A. (2014). Briefly cuing memories leads to suppression of their neural representations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(23), 8010–8020. https://doi.org/10.1523/JNEUROSCI.4584-13.2014

Qasim, S. E., Fried, I., & Jacobs, J. (2021). Phase precession in the human hippocampus and entorhinal cortex. *Cell*, *184*(12), 3242-3255.e10. https://doi.org/10.1016/j.cell.2021.04.017

Reddy, L., Self, M. W., Zoefel, B., Poncet, M., Possel, J. K., Peters, J. C., Baayen, J. C., Idema, S., VanRullen, R., & Roelfsema, P. R. (2021). Theta-phase dependent neuronal coding during sequence learning in human single neurons. *Nature Communications*, *12*(1), 4839. https://doi.org/10.1038/s41467-021-25150-0

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences*, *23*(9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005

Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. *Trends in Cognitive Sciences*, *23*(12), 1071–1085. https://doi.org/10.1016/j.tics.2019.09.011

Staudigl, T., Hanslmayr, S., & Bäuml, K.-H. T. (2010). Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(34), 11356–11362. https://doi.org/10.1523/JNEUROSCI.0637-10.2010

ten Oever, S., De Weerd, P., & Sack, A. T. (2020). Phase-dependent amplification of working memory content and performance. *Nature Communications*, *11*(1), 1832. https://doi.org/10.1038/s41467-020-15629-7

ter Wal, M., Linde-Domingo, J., Lifanov, J., Roux, F., Kolibius, L. D., Gollwitzer, S., Lang, J., Hamer, H., Rollings, D., Sawlani, V., Chelvarajah, R., Staresina, B., Hanslmayr, S., & Wimber, M. (2021). Theta rhythmicity governs human behavior and hippocampal signals during memory-dependent tasks. *Nature Communications*, *12*(1), 7048. https://doi.org/10.1038/s41467-021-27323-3

Tort, A. B. L., Komorowski, R., Eichenbaum, H., & Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of Neurophysiology*, *104*(2), 1195–1210. https://doi.org/10.1152/jn.00106.2010

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning & Verbal Behavior*, *13*(2), 181–193. https://doi.org/10.1016/S0022-5371(74)80043-5

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60. https://doi.org/10.1037/h0044616

Wammes, J., Norman, K. A., & Turk-Browne, N. (2022). Increasing stimulus similarity drives nonmonotonic representational change in hippocampus. *ELife*, *11*, e68344. https://doi.org/10.7554/eLife.68344

Wang, T. H., Placek, K., & Lewis-Peacock, J. A. (2019). More Is Less: Increased Processing of Unwanted Memories Facilitates Forgetting. *Journal of Neuroscience*, *39*(18), 3551–3560. https://doi.org/10.1523/JNEUROSCI.2033-18.2019

Watrous, A. J., Miller, J., Qasim, S. E., Fried, I., & Jacobs, J. (2018). Phase-tuned neuronal firing encodes human contextual representations for navigational goals. *ELife*, *7*, e32554. https://doi.org/10.7554/eLife.32554

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

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589. https://doi.org/10.1038/nn.3973

Wolinski, N., Cooper, N. R., Sauseng, P., & Romei, V. (2018). The speed of parietal theta frequency drives visuospatial working memory capacity. *PLOS Biology*, *16*(3), e2005348. https://doi.org/10.1371/journal.pbio.2005348