Study-Phase Reinstatement: Encoding Spontaneous Thoughts as Memories

David J. Halpern[1], Bradley C. Lega[2], Robert E. Gross[3], Chengyuan Wu[4], Michael R. Sperling[5], Joshua P. Aronson[6], Barbara C. Jobst[7], Michael J. Kahana[1*]

[1*]Dept. of Psychology, Univ. of Pennsylvania, 425 S. Univ. Ave., Philadelphia, 19104, PA, USA.
[2]Dept. of Neurosurgery, Univ. of Texas Southwestern, Dallas, TX, USA.
[3]Dept. of Neurosurgery, Emory School of Medicine, Atlanta, GA, USA.
[4]Dept. of Neurosurgery, Thomas Jefferson Univ., Philadelphia, PA, USA.
[5]Dept. of Neurology, Thomas Jefferson Univ., Philadelphia, PA, USA.
[6]Dept. of Neurosurgery, Beth Israel Deaconess Med. Ctr., Boston, MA, USA.
[7]Dept. of Neurology, Dartmouth-Hitchcock Med. Ctr., Lebanon, NH, USA.

*Corresponding author(s). E-mail(s): kahana@psych.upenn.edu;

Abstract

Can the brain improve the retrievability of an experience after it has occurred? Systems consolidation theory proposes that item-specific cortical reactivation during extended post-encoding rest periods facilitates the formation of stable memory representations, a prediction supported by neural evidence in humans [1, 2]and animals [3, 4]. Such reactivation may also occur on short time scales as spontaneous thoughts come to mind during encoding, offering a potential account of classic list memory phenomena [5–9] but lacking in support from neural data. Leveraging the high-temporal specificity of intracranial electroencephalography (iEEG), we investigate spontaneous reactivation of previously experienced items during brief sub-second intervals between individual encoding events. Across two large-scale free recall experiments, we show that reactivation, as measured by spectral iEEG similarity, during these periods predicts subsequent recall. In a third experiment, we show that the same methodology can identify post-encoding reactivation that correlates with subsequent memory, consistent with previous results. Thus, spontaneous study-phase reinstatement reliably predicts memory

behavior, linking psychological accounts to neural mechanisms and providing the first such evidence for rapid consolidation processes during encoding.

Keywords: Memory, Reinstatement, intracranial EEG, Psychology, Neuroscience

1 Introduction

Recent evidence from functional magnetic resonance imaging (fMRI) and intracranial electroencephalography (iEEG) suggests that spontaneous neural reactivation of specific previously encoded content during sleep or awake rest predicts subsequent item recognition [10], cued recall [1, 2] and reconstruction [11] of that content [1] . Theoretically, neuroscientific two-stage models of memory [14, 15] propose that reactivation during post-encoding rest periods improves memory via a consolidation mechanism that turns fragile hippocampal memories into stable cortical representations. Consolidation can be a spontaneous process, happening even or especially without cues from the external environment. Psychological theories of memory encoding [e.g. 5–7, 16] and data on overt rehearsal [17–27] suggest that reactivation during or between encoding other experiences also determines subsequent recall probability and organization.

If the psychological construct of rehearsal, covert repetition of previously encoded items, is a type of consolidation, this implies that consolidation-like processes may actually be happening during a broader range of time periods and shorter time scales than previously studied. Indeed, the possibility of more frequent or opportunistic consolidation periods has been recognized by some theorists [28, 29] but the consensus in the field appears to be that sleep or rest remain special times for consolidation [14, 15, 30, 31] and thus the existing neural evidence for reactivation involves these time periods. Recent work suggests that the brain can transition rapidly between "online" externally-driven states and "offline" consolidation-promoting states on the time-scale of seconds [32, 33] but no study to date has linked this to content reinstatement. It, therefore, remains unknown whether such rapid, spontaneous item-specific neural reinstatement during a study phase predicts subsequent recall much as reinstatement during post-encoding periods does. Furthermore, it is unclear whether neural measurements of reinstatement play the same role in memory that the psychological construct of rehearsal does. In addition to the overall amount of rehearsal correlating with recall probability, work on overt rehearsal demonstrates that rehearsal predicts output order (more rehearsal predicts earlier recall [17]) and organization (items rehearsed together are also recalled together [26]). The present work leverages the high temporal precision of intracranial EEG and three large-scale free recall experiments to address these gaps in our knowledge.

Some neuroscientific studies have investigated reactivation correlates of instructed rehearsal. Fellner and colleagues [34], using a directed forgetting paradigm, showed that activity after the forgetting cue showed greater reinstatement on to-be-remembered trials that were later remembered relative to to-be-forgotten trials.

[1] In addition, Tambini and colleagues demonstrated that reinstatement of overall task-related activity predicted subsequent associative recognition [12, 13]

The authors therefore interpreted this reinstatement as corresponding to deliberate rehearsal. In addition, Bird et al. [35] instructed participants to rehearse previously seen videos when cued with the video title while in an fMRI scanner. They showed that the degree of reinstatement during these rehearsal periods correlated with subsequent memory for video details. However, neither of these directed rehearsal experiments investigated reinstatement signatures of *spontaneous* rehearsal.

Additionally, a number of recent studies have shown that neural pattern similarity of two repeated item presentations predicts subsequent memory for those items [36–39] and have argued that this is an index of covert study-phase retrieval of previous presentations. The spacing effect [40–42], the finding that repetition is more effective for memory when repetitions are spaced out, has been a key source of behavioral evidence for study-phase retrieval [6, 8]. Feng and colleagues [43] show that the degree of reinstatement depends on whether the repetitions are massed or spaced. However, reinstatement in these studies resulted from the presentation of a repeated item rather than occurring spontaneously.

In what may be the most direct evidence for spontaneous study-phase reinstatement, Wu and colleagues [44, 45] investigated how reactivation during a delay period immediately following the presentation of a sequence of images affected subsequent memory for details of the sequence. In one experiment with semantically coherent sequences, they found that the degree of reinstatement of images in the sequence differed between sequences where subjects recalled an above-average or below-average number of details in a subsequent cued recall test. In contrast, there was no similar effect when sequences were not coherent, leading Wu et al. to conclude that people reinstate sequences during a study phase to extract higher-level semantic information. Unlike the current study, Wu and colleagues did not investigate whether reinstatement of specific images led to improved memory for that specific image. Thus, the reinstatement they found could represent a reinstatement of the entire list context leaving open the question of whether item-specific study-phase reinstatement predicts future memory.

In the following work, we ask whether spontaneous reinstatement during a study phase can impact future memory behavior, as suggested by psychological explanations of memory behavior. These theories suggest that, in addition to externally-driven perceptual experiences being encoded in memory, internally-driven thoughts, reminders, retrieved memories and imagined scenarios can themselves become encoded in memory, greatly expanding the role of neural reactivation beyond that examined in previous studies. We use two two large, independent data sets with intracranial EEG to relate the degree of spontaneous item-specific reinstatement while encoding other experiences (study-phase reinstatement) to subsequent recall probability and organization for that item. In a separate third dataset using the same methodological approach, we ask whether we observe similar reinstatement during an unfilled delay interval between the initial encoding period and later recall, as in previous studies of reactivation. Overall, these analyses aim to answer the question of whether rapid, spontaneous study-phase reinstatement performs a similar role in memory behavior as past studies of post-encoding reinstatement.

2 Results

All three data sets we analyzed involved intracranial EEG recorded during variations on free recall tasks (see Figures 1 and 4). Subjects studied twelve words and, following a brief delay, attempted to recall as many as they could remember. In Experiments 1 (N = 217) and 2 (N = 260), subjects performed a math distractor task during the delay interval; in Experiment 3 (N = 49), the interval remained unfilled. In Experiment 3, some words repeated two or three times in the same list, resulting in 27 word presentations in total during the encoding interval. In experiment one, words on the list were drawn from three semantic categories while in the other two experiments, words were drawn at random from a wordpool. In all other respects, the three experiments did not differ, providing the opportunity to replicate our key results several times. Neurosurgical patients performed all three experiments while intracranial EEG was being recorded. These neural recordings provide the basis for identifying latent spontaneous reinstatement

We used a spectral representational similarity analysis (RSA) [46–50] approach to studying reinstatement. This method compares a representation of the neural signals at two time points and allows us to asses the degree of similarity between the neural representations. We first decomposed the neural signals from each bipolar pair of electrodes into power in eight log-spaced frequency bands. We then compute the z-transformed [51] Pearson correlation between the signals at the two time points across all electrodes and frequencies. This correlation measure is our dependent metric of reinstatement in the following analyses.

2.1 Semantic pattern similarity

We first validate our RSA-based [46–48] approach by examining the degree to which time and semantic content determined spectral power similarity during the encoding periods in Experiment 1. Our future analyses, aimed at detecting item reactivation during inter-stimulus intervals, rely on the neural patterns representing the semantic content of the items. This Experiment provides an ideal setting to investigate this because lists comprised items from three different categories. By comparing the similarity of items from the same category to items from different categories, we can determine whether semantic content drives spectral pattern similarity. Figure 2 shows the correlation between the spectral power representations at two encoding time points on the same list as a function of their absolute serial position difference and whether or not the words come from the same semantic category. The dominant effect on pattern similarity is distance in time (as represented by an item's serial position). However, semantics additionally determines a component of similarity as items from the same category are more similar than items from different categories across all serial position distances. These analyses conceptually replicate effects demonstrated in [49] and [50], using a different data set and method. In addition, [52] demonstrated an ability to predict serial position and semantic category using the same spectral power components using a subset of 69 subjects from the current dataset. However, neither of those papers demonstrates these two effects simultaneously in the same neural signals, as shown here. Finally, a key difference between our analyses and past work

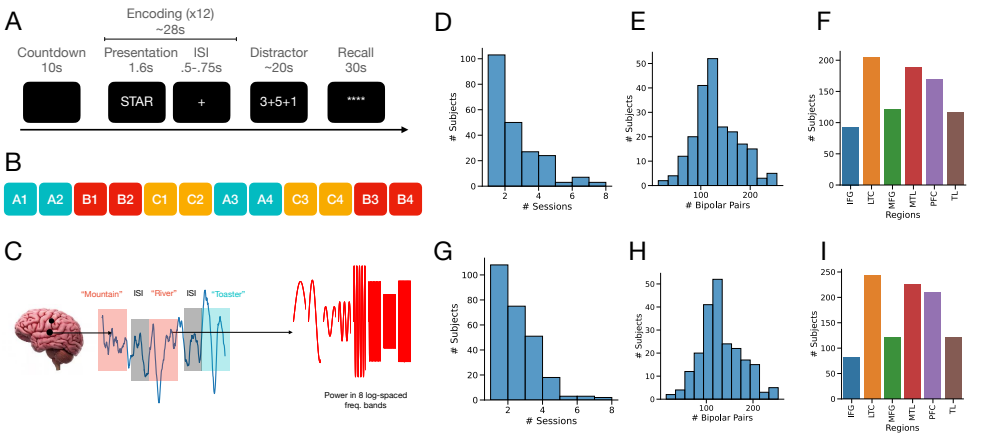

Fig. 1 Experimental tasks and methods for Experiments 1 and 2. A. Timeline for a single trial of free recall. Following a ten second countdown, twelve words appear sequentially on screen for 1.6 seconds each with .5 to .75 second inter-stimulus intervals. Following this encoding period, subjects perform twenty second math distractor task in order to prevent rehearsal. Finally, during the recall period, subjects have 30 seconds to recall as many words as possible. B. Category structure of Experiment 1 on a given twelve word list. Four words are drawn from three categories and presented in a block structure as shown. C. Methods used for constructing spectral features. EEG signals are taken each electrode from specific time period, e.g., while a word is on the screen or during an inter-stimulus interval, and then decomposed into power in 8 log-spaced frequency bands. They are then compared to this spectral representation at other time points, e.g., during the inter-stimulus interval. D. Number of sessions collected per subject in Experiment 1. E. Number of bipolar pairs per subject in Experiment 1. F. Number of subjects with bipolar pairs in each region for Experiment 1. G. Number of sessions collected per subject in Experiment 2. H. Number of bipolar pairs per subject in Experiment 2. I. Number of subjects with bipolar pairs in each region for Experiment 2.

using intracranial EEG (including the papers previously mentioned as well as [10]) is that we did not perform any selection of electrodes or time points prior to analyzing pattern similarity.

To test these differences statistically, we fit a mixed effect model predicting the pattern similarity between signals associated with two items as a function of whether the items were from the same category and the temporal distance between the items in terms of serial position. In this model and all other models we fit, we also included varying intercepts and slopes as a function of subject, session, list and the serial positions of the two items as well as their interactions. Because the large number of random effects could lead to overfitting, we reduced the model iteratively to account for the structure present in the data using the procedure from [53]. The full model is described in detail in the *Methods* section. We found significant main effects of serial position distance ($F(1, 88.62) = 151.65, p \leq 0.001$) and the two items being from the same category (($F(1, 3.03) = 20.93, p = 0.02$). We did not find a significant interaction between the two ($F(1, 22.31) = 1.72, p = .2$). Examining the point estimates, we found that, as expected, similarity decreases as a function of temporal distance and increases if the items are from the same category.

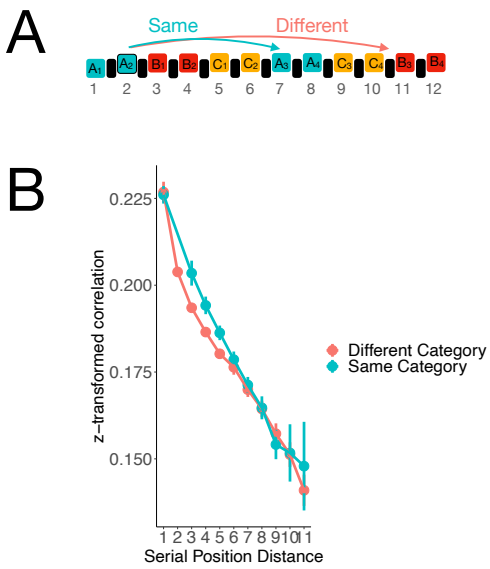

Fig. 2 Similarity of two time points during encoding word presentations on the same list as a function of the absolute serial position distance and whether the two words came from the same category or not. Because of the constraints on list order in the experimental design, it was impossible for words from the same category to appear at a distance of two serial positions away. A. Schematic of comparison, indicating which items are from the same category and which are not. B. Similarity computed from Experiment 1. Error bars reflect variation within list using an approach similar to [54]

2.2 Study-phase reinstatement

Having (1) validated the pattern similarity measure during encoding and (2) replicated prior context reinstatement results, we now use the same measure to examine study-phase reinstatement. For this analysis, we investigate whether neural patterns associated with a particular encoded item reactivate during the inter-stimulus intervals following future items on the list. We compared patterns of spectral power between item presentations (during the 750-1000ms ISI period, the *ISI*) with power while a word was presented on the screen (the *initial encoding presentation*). Figure 3 shows similarity as a function of whether the initially presented word was itself subsequently recalled (**Recalled**), whether the word prior to the ISI was recalled and whether the two words were from the same category. We include these variables in the analysis because we would expect that both of these variables would increase similarity between the initial encoding presentation and the ISI even without spontaneous reinstatement. Qualitatively, we find that similarity increases when the encoded item was subsequently recalled, regardless of the values of the other two variables. We additionally examined whether reinstatement differed as a function of recall organization in two ways, based on results from the overt rehearsal literature [17, 26]. First, we would expect a greater degree of reinstatement for the first recalled item than items recalled later. Second, we would expect that the degree of reinstatement during a particular ISI would predict the strength of associations between the item before the ISI and the

reinstated item and therefore the likelihood that items will later be recalled sequentially. To test these differences statistically in Experiment 1, we used a mixed effects model in `lme4` [55] with the following specification for the fixed effects:

$$\texttt{similarity} \sim (\texttt{item_recalled} + \texttt{item_first}) * \texttt{item_before_isi_recalled} * \texttt{same_category} \\ + \texttt{recall_neighbor} + \texttt{recall_neighbor} : \texttt{same_category}$$

`item_recalled` and `item_first` reflect whether the specific item was recalled and whether it was recalled first. `item_before_isi_recalled` indicates whether the item before the ISI was recalled and `same_category` indicates that the initial encoding item and the item before the ISI are drawn from the same category. Finally, `recall_neighbor` indicates whether the encoding item and the item before the ISI were recalled next to each other during the recall period, given that both items were recalled. The model has varying intercepts and slopes for each term as a function of nested grouping variables subject, session, list and the serial positions of both the encoded item and the item before the ISI. In addition, the model included random effects for across-subject effects of list ID, serial positions of the encoded item and the ISI and the interaction between the pair of serial positions. For Experiment 2, we used the same model but without the terms involving `same_category` because items in this Experiment were randomly sampled from a common wordpool.

Given the results from our first analysis, it is likely that the effects of recency and categorical similarity drive much of pattern similarity. To focus on spontaneous reinstatement, we therefore limit the analyses to pattern similarity comparisons involving initial encoding presentations from the first half of the list (the first six serial positions) and ISIs from the latter half (the five intervals between the last six serial positions). In the case of Experiment 1, we also ensured that there was at least one category pair between the initial encoded item and the comparison ISI. This ensures that any observed reactivation is unlikely to be due to recency and similarity alone.

In Experiment 1, we found a main effect of `item_recalled` such that items that were subsequently recalled were more similar to activity during subsequent ISIs during encoding than items that were not subsequently recalled (Fig. 3, $F(1, 15.5) = 13.66, p = .002$). Post-hoc comparisons revealed significant effects of recall status, regardless of whether the item before the ISI was from the same category or was itself recalled. However, we could not reliably identify whether the items recalled first had consistently different levels of reactivation than other recalled items ($F(1, 52.2) = 1.43, p = .24$). we also did not see consistent differences between items that were sequentially recalled and pairs of items that were both recalled but not sequentially ($F(1, 16.4) = 0.01, p = .94$). Additionally, we find that the recall status of the item before the ISI is a consistently positive predictor of similarity ($F(1, 11.2) = 5.43, p = .04$), indicating that some component of the similarity may be due to cognitive processes that occur when viewing items that subjects subsequently recall. However, two items being from the same category is not a significant predictor ($F(1, 25.5) = 1.12, p = .3$), although, as can be seen in Fig. 3, the estimates of

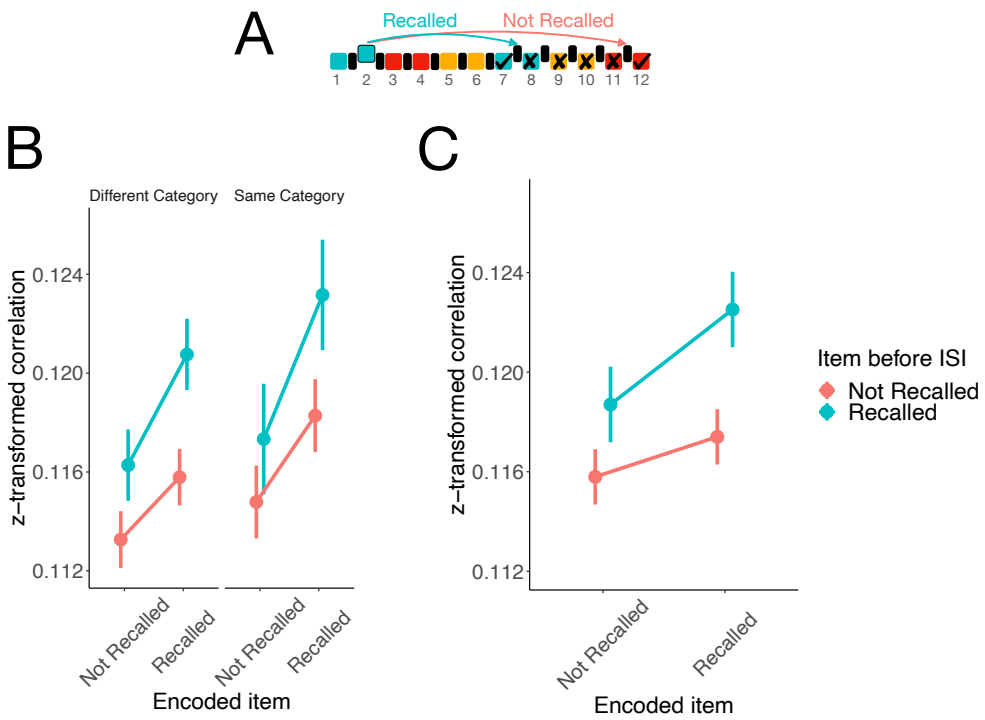

Fig. 3 Similarity of time points during encoding word presentations and time points during inter-item intervals as a function of whether the word was itself subsequently recalled (**Recalled**). Estimates are derived from the mixed effect model described in the main text. Error bars reflect 95% confidence intervals on the difference between remembered and forgotten, using an algorithm implemented by the R package `emmeans` and described in [56] that is a generalization of Loftus-Masson intervals [57, 58]. A. Schematic showing time points being compared when computing RSA, encoding time point from first half of list and inter-stimulus interval from second half. Checks and crosses indicate whether the item was subsequently recalled and colors indicate category, indicating the relevant control variables. B. results from Experiment 1. C. results from Experiment 2.

the difference between recalled and not recalled are positive, the direction we would predict.

We fit the same model to Experiment 2, dropping the terms involving `same_category`. We replicated the main effect of subsequent recall on study-phase spectral pattern similarity, i.e. items with more similar activity during subsequent ISIs during encoding than items that were not subsequently recalled (Fig. 3, $F(1, 12.74) = 6.39, p = .02$). In a post-hoc comparison, we could not establish a reliable difference when the item before the ISI was itself recalled ($\beta = .01, t(10) = 1.6, p = .14$). However, the estimates were positive qualitatively, replicating the results from Experiment 1. In addition, the interaction between recall status of the encoding item and the item before the ISI was only marginally significant, ($F(1, 11.43) = 3.29, p = .1$), suggesting that we cannot reliably determine whether the specific pairwise comparison is different from the overall main effect. Again, we could not establish the

direction of effects of first recall ($F(1, 52.2) = 1.43, p = .24$) or being a recall neighbor ($F(1, 13.36) = .67, p = .43$) on pattern similarity. Replicating results from Experiment 1, we find a main effect of the recall status of the item before the ISI ($F(1, 45.4) = 8.0, p = .007$).

To gain a more mechanistic understanding of these results, we asked whether our results were driven by activity in specific frequency bands. In the results just described, we computed correlations across eight frequency bands. We therefore repeated the statistical analysis while dropping a single frequency band from the similarity computation. This approach preserves the possibility that correlations were driven by across-frequency relationships between spectral signals. We found that the results were not particularly sensitive to any one frequency band being removed (Fig. A1) suggesting that the measured reinstatement is occurring generally at all spectral frequencies. We might be concerned that the analysis approach is more sensitive to variation in lower frequency bands because they do not have the opportunity to vary as much over the encoding interval. However, we don't see any consistent trends across the frequency spectrum in the magnitude of results, suggesting that our results were not driven by any particular range of frequencies.

Perception of items often doesn't require the full 1600 ms stimulus presentation interval. If subjects rapidly switch between online and offline states, it is possible that subjects can switch into an offline state even while other items are on the screen. We therefore wanted to check if our findings were specific to inter-stimulus intervals or whether reinstatement may be happening during the presentation of other items on the list. We replicated our analyses above using the last 500 ms of the 1600 ms stimulus presentation interval as the reinstatement period. We find overall replicate the main effect of recall status across both experiments and all of our control conditions (Fig. B2). This suggests that switches between encoding and retrieval states may be happening continuously throughtout the task phases.

2.3 Post-encoding reinstatement

Having identified study-phase reinstatement in two experiments, we sought to confirm that similar methods could identify the previously found post-encoding reinstatement. Therefore, in a third experiment, we investigated pattern similarity between initial encoding presentations and a subsequent unfilled 7s delay interval prior to recall. To test differences statistically, we fit the following statistical model:

$$\texttt{similarity} \sim (\texttt{item_recalled} + \texttt{item_first}) * \texttt{serial_position}$$

In addition to predicting reinstatement as a function of recall status, this model allowed for interactions with serial position in order to test the possibility that earlier serial positions showed greater effects of reinstatement (Figure 4)E, as suggested by the literature on overt rehearsal [27]. This analysis conceptually replicates past work on reinstatement (such as [2] and [44]) but applied to individual lists in a free recall task. As in the analyses of the inter-stimulus intervals, we find a main effect of recall status such that remembered items are more similar to the delay interval than forgotten items

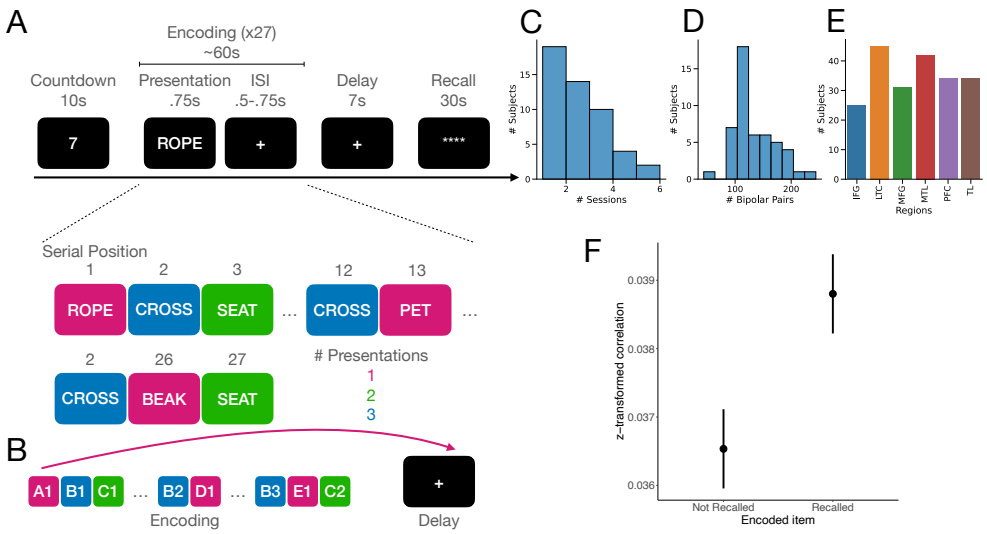

Fig. 4 A. Task design for Experiment 3. Task is similar to Experiments 1 and 2 except that lists include repeated items (some of which are presented twice and some presented three times) and the distractor task is replaced with an unfilled seven second delay interval. B. Schematic of comparisons used in analysis in panel E. C. Distribution of sessions across subjects. D. Number of bipolar pairs per subject E. Number of subjects with bipolar pairs in each memory-related region. F. Similarity of time points during encoding word presentations and time points during a delay period as a function of whether the word was itself subsequently recalled (**Recalled**). Estimates are derived from the mixed effect model described in the main text. Error bars reflect error bars of the difference between remembered and forgotten, using an algorithm implemented by the R package `emmeans` and described in [56] that is a generalization of Loftus-Masson intervals [57, 58].

($F(1, 131.55) = 14.98, p < .001$. We additionally find a main effect of serial position ($F(26, 302.88) = 24.0, p < .001$) but do not find any interactions between the recall effect and serial position ($F(26, 475.83) = 0.72, p = .84$). As before, we do not find any reliable differences between first recalls and other recalls ($F(1, 1553.43) = 2.88, p = .09$) or any interaction with serial position ($F(26, 257.88) = .8, p = .75$). Examination of the point estimates suggests that similarity increases with later serial positions, likely due to auto-correlated noise in the signal.

To relate our findings to those of **(author?)** [2], who found that reinstatement during a post-encoding math task correlates with subsequent cued recall, we asked whether we could identify memory-related reinstatement during the math distractor tasks in Experiments 1 and 2. We replicate the main effect of recall status on reinstatement during the post-encoding interval in Experiment 3 (Fig. C3). Studies on the spacing of overt rehearsal suggest that distributed rehearsal should increase the effect on retrieval [21]. Interestingly, we find an interaction with serial position but in the opposite direction such that the difference between remembered and forgotten items is greater for items towards the end of the list. Thus rehearsal of items closer in time to their initial presentation showed greater impacts on subsequent recall.

3 Discussion

Three separate experiments involving intracranial EEG recordings from large samples reveal that subsequently recalled items reinstate during unfilled intervals to a greater degree than forgotten items. In the first two experiments, these intervals occurred between study opportunities, a time period not investigated in past work on reinstatement and yet one that figures prominently in psychological theorizing and data on rehearsal processes. In addition, reinstatement effects were item-specific and relevant for recall of a specific list, even in an experimental context where there were several lists.

Overt rehearsal studies show that when subjects rehearse items together they also recall them together [26]. Similarly, subjects tend to recall items they rehearsed more earlier in their output [17]. Here, we consider the possibility that study-phase reinstatement and rehearsal measure similar processes. The above results make two key predictions. First, finding greater study-phase reinstatement of an item from an earlier list position in the inter-stimulus interval immediately following a later item should increase the likelihood of those two items being successively recalled (as compared with being recalled in non-sequential output positions). Second, items with greater reinstatement should be more likely to be recalled first. However, we failed to find reliable evidence for either of these effect. One possibility is that this indicates a difference between overt rehearsal and the covert rehearsal processes signaled by our measures of neural reinstatement. However, another possibility is that detecting such effects requires a high degree of precision relative to reinstatement effects on overall recall probability. This is due to the organization effects making predictions about specific inter-stimulus intervals rather than integrated over all ISIs throughout the learning period.

In seeking theoretical models of behavioral memory data, psychologists have frequently turned to study-phase retrieval as an explanatory construct [5–7, 9, 16, 59]. This has been used to explain why reliable false memory effects exist [60] as well as why rehearsal, repetitions, repeated testing or even presentations of similar items can improve memory for a past experience. While theoretically compelling, it remains challenging to obtain direct evidence for study-phase retrieval because of the lack of corresponding overt behavioral correlates. Here, we provide evidence of a potential neural foundation for such retrieval processes, both bolstering such claims and also allowing for future work to investigate specific predictions in more detail. This work also broadens the scope of previous work on neural reactivation and its effects on subsequent memory to include time periods during encoding events, rather than only potential consolidation periods.

Overall, we believe this work contributes to the greater cognitive neuroscience literature on memory consolidation in three ways. Firstly, we show evidence of rapid, sub-second, switches between online encoding and offline consolidation states. While switches on this timescale had been theorized in the literature, we believe this is the first empirical evidence. This provides constraints on biological theories of memory consolidation, some of which postulated special roles for longer offline intervals such as sleep. Secondly, as mentioned above, we provide direct neuroscientific evidence of spontaneous study-phase retrieval, a mechanism that had been theorized based on

behavioral data but with little data on biological substrates. Finally, we provide the first comparison of stylized facts about overt rehearsal, a proposed psychological mechanism for rapid consolidation, to neural reinstatement. We fail to find evidence that neural reinstatement is related to memory organization in the same way as rehearsal, even with multiple highly-powered datasets, by the standards of the extant neuroimaging literature. We believe these three contributions provide novel constraints on theories of memory consolidation, a fundamental area of inquiry in the cognitive neurosciences.

How can we be confident that the reinstatement effects we uncovered reflect true spontaneous reactivation? In the analyses of study-phase reinstatement, one potential issue is that activity driven by the item prior to the inter-stimulus interval could carry over into the delay period. This activity could be similar to prior items for two reasons unrelated to spontaneous reactivation. One reason is that semantically similar items will have more similar neural representations than different ones [61]. Semantic similarity effects [62] imply that the probability of recalling an item increases when semantically similar items are also recalled. Thus, strong semantic representations of the item prior to the ISI could predict greater recall probabilities. A second reason is that similar cognitive processes are involved in processing the item. Subsequent memory effects [63–66] show that neural activity when viewing stimuli that will be recalled consistently differ from neural activity during forgotten stimuli. Thus, the similarity of neural responses to an item and other recalled items on a list should predict the item's probability of recall. To handle both of these potential confounds, we include terms in the model for whether the item before the ISI was itself recalled and, in the case of Experiment 1, whether it is from the same category as the earlier item. We show that the difference in reactivation exists even in the case where the item before the ISI was not recalled and from a different category in the case of Experiment 1. In Experiment 2, the pairwise comparison when the item before the ISI is not recalled is marginally significant and in the same direction as all of the other pairwise comparisons. In addition, the interaction between the recall status of the encoded item and the item before the ISI is not significant in an analysis of variance, suggesting that there is not reliable evidence for a difference in effects depending on whether the item before the ISI is recalled. Thus, we conclude that our findings likely reflect spontaneous reactivation.

Beyond identifying rapid spontaneous reinstatement, we might wonder whether it reflects conscious, controlled rehearsal or a more unconscious and automatic process [67, 68]. We conducted an analysis of reinstatement during the math distractor intervals in Experiments 1 and 2 (Fig. C3) and found evidence of reinstatement, replicating findings from **(author?)** [2]. This suggests that the measured reinstatement is likely, at least in part, automatic. We additionally found that the difference in reinstatement for remembered items compared to forgotten items was greater for items towards the end of the list, which goes against findings from overt rehearsal studies suggesting that conscious controlled rehearsal primarily supports early items [18, 27].

Our results provide consistent evidence across multiple large studies for a relation between spontaneous, sub-second, study-phase reinstatement and subsequent recall. Whereas intracranial EEG recordings provide high spatial and temporal precision,

there is frequently little overlap in electrode locations across patients. To maximize the inclusion of electrodes and measure the similarity of neural activity across as wide of a span of cortex as possible, we include all available electrodes in our representational similarity analyses. However, this makes it difficult to investigate in great detail the anatomical substrates of such reactivation and their consistency across people. We therefore leave this investigation to future work, either with more regionally specific data sets or using recording methodologies that allow for broad coverage and alignment across subjects (e.g. fMRI). As shown in Figures 1E and G and 4D, many subjects have at least one electrode in regions thought to be relevant for memory. In this work, we focused on a single memory task with relatively short study-test delays. While this allowed us to easily test generalization across variants of this task (with surprisingly close quantitative estimates in figures 3B and C), one may ask whether the effects here generalize to the longer time scales studied in studies involving sleep [1, 11]. Although we demonstrate covert reactivation within the short interstimulus intervals and we measure recall following a 30 second filled distractor period, the effects of study-phase reactivation may be different after hours or days. Future research examining the relation between sub-second covert reactivation and longer term retrieval will be required a more complete understanding of the relation between rapid consolidation and consolidation over longer periods.

One might also ask whether the identified reinstatement is reinstatement of specific content or whether it reflects reinstatement of the context associated with a retrieved episode. Studies of reinstatement during delay periods [1, 2, 10], including the present one, have focused on between neural similarity between a specific episode and the delay. Recent work [49, 69–71] has identified a neural contiguity effect during retrieval such that neural activity prior to retrieval is similar to items nearby the to-be-retrieved item. Inspired by theoretical models of memory [72–74], these papers have claimed that this is evidence for reinstatement of the surrounding context. This literature has largely focused on retrieval and we do not know of any papers investigating the relationship between reinstatement of context during a delay interval and subsequent memory (although perhaps the work by Wu and colleagues [44, 45] could be interpreted in this light). Without analyzing neural similarity to episodes near a target episode, it is impossible to differentiate these two accounts of neural reinstatement so we leave this important question to future work.

The present findings show that covertly and spontaneously reactivated neural traces while encoding other experiences predict future memory. This provides direct electrophysiological evidence for cognitive theories which posit such latent retrieval as a fundamental mechanism of human (and animal) learning. That we can observe the memorial consequences of study-phase reactivation suggests that memory consolidation, hypothesized to underlie the formation of long-term memory representations, is a much more general phenomenon than many previously thought. Indeed, it likely occurs throughout awake experience with rapid switches between externally- and internally-focused cognitive states. While several theorists have previously suggested this opportunistic view on consolidation and reinstatement [28, 29, 33], this work provides the first direct neural evidence. Many open questions remain about the psychological and neural mechanisms. If the measured reinstatement corresponds to

rehearsal processes, to what extent is it under cognitive control? Can rehearsal and reinstatement be targeted in a way to optimize for goal-directed behavior [11, 75]? And if this is the same process as consolidation during longer rest and sleep periods, are the physiological mechanisms are the same? Although there are several possibilities that may operate on different time scales [28], future work will have to determine whether the behavioral consequences of consolidation depend on the neural process or whether they are functionally equivalent. With this ability to directly measure these processes with neural data, we can now begin to answer these questions and many others. Overall, identifying the connection between neural consolidation processes and psychological constructs developed to explain behavioral data will allow for a tighter link in theory development across cognitive and neural levels of analysis, a central goal of the cognitive neuroscience of memory.

4 Methods

4.1 Task and participants

In Experiment 1, we recorded intracranial EEG from 217 neurosurgical patients while they performed a categorized free-recall task [76]. Lists consisted of 12 items, presented in same-category pairs, from three distinct categories with categories drawn from total set of 25. Each item was on the screen for 1600 ms followed by an inter-stimulus interval with the length randomly jittered across trials. The interval on each trial was drawn uniformly on an range with a minimum of 750ms and a maximum of between 1000 and 1180 ms depending on which version of the experiment was used. Presentation of words during encoding was followed by a 20s post-encoding delay, during which subjects performed an arithmetic task in order to disrupt memory for end-of-list items (Fig. 2). Finally, subjects attempted to freely recall as many words as possible during a 30 second interval.

In Experiment 2, we collected EEG from 260 neurosurgical patients while they performed a standard free-recall task. Lists consisted of 12 items selected so that the average pairwise semantic similarity based on latent semantic analysis [77, 78] was .2. Timing of the item presentations, inter-stimulus intervals, distractor task and recall period were the same as in Experiment 1.

In Experiment 3, we collected EEG from 49 neurosurgical patients while they performed a free-recall task with repeated items. Lists consisted of 12 unique items and used the same wordpool as Experiment 2. However, 3 items were repeated twice and six items were repeated three times resulting in twenty-seven total encoding presentations per list. For the first 16 subjects items were on the screen for 1600 ms followed by an inter-stimulus interval of 750-1000ms, with the length of the interval randomly jittered across trials. Due to a coding error, the final 33 subjects only saw items on the screen for 750ms. Presentation of words during encoding was followed by a 7s unfilled post-encoding delay. As the previous experiments, subjects then attempted to freely recall as many words as possible during a 30 second interval.

4.2 Intracranial EEG

As described in [76], we recorded from subdural grids and strips (space between adjacent contacts: 10 mm) and from depth electrodes (space between adjacent contacts: 5–10 mm) on a variety of recording systems across clinical sites with sampling rates varying between 500 and 2000 Hz to accommodate the local recording environment. We rereferenced all recordings using a bipolar referencing scheme [79] and applied a 4th order Butterworth filter with a 58-62 Hz stop-band to remove line noise. We then Fourier resampled the data to 1000 Hz and convolved the resulting signals with Morlet wavelets (wave number 5; 8 center frequencies, log-spaced between 3 and 180 Hz) to obtain a representation of spectral power in each bipolar pair. When computing power, we always include a buffer of 1 second on either side of the period of interest. This ensures that there are sufficient samples for estimating power in all frequency bands of interest. For the analysis of encoding data, we averaged power over the entire 1600 ms stimulus presentation interval (we use 15-1575ms following stimulus onset to avoid effects of luminance changes). We computed this average prior to running the similarity analyses. In Experiment 3, due to the coding error mentioned in the task description, we only used the first 750ms of the item presentation interval. For analysis of the inter-stimulus intervals, we use the same approach but for the 700ms interval immediately preceding the presentation of the subsequent stimulus. Similar to the approach for encoding periods, we compute this average prior to running the similarity analyses. For the analysis of the delay interval data in Experiment 3, we first preprocess all 7 seconds using the same approach as above. We then divide it into 7 one second bins and average over the 1 second, using a resampling technique (implemented in scipy [80]). For analyses presented here, we then average over the 1 second bins. We compute the similarity analyses within each 1 second bin and then average the 7 bins to obtain a single similarity value for each encoded item. For analysis of the end of the encoding interval in Appendix B, we followed the same approach as for the inter-stimulus interval but using the last 500ms of the stimulus presentation. For analysis of the math distractor task data from Experiments 1 and 2 in Appendix C, we followed the same approach by dividing up the interval in 20 one second bins and then averaging.

In order to compare activity across periods and subjects, we normalize all data relative to spectral signals collected during the 10 second countdown period prior to each list. We first preprocess the 10 seconds using the same approach as above. We then divide it into 10 one second bins and average over the 1 second, using the resampling technique mentioned above. We then compute the mean and standard deviation across lists for each session. All spectral power estimates in subsequent periods are then normalized by subtracting the countdown mean and dividing by the countdown standard deviation.

4.3 Analysis

Our main analysis approach relies on analyzing the similarity of spectral power patterns [46–48] in the eight frequency bands across all bipolar pairs, regardless of location in the brain. As is common in the literature, similarity was determined as the Fisher's

z-transformed Pearson correlation [51] between power at two time points. These analyses relied on `xarray` [81], `pandas` [82], `scipy` [80] and `numpy` [83] in `python 3.11`.

4.4 Modeling

For all models, we use mixed effect models to account for systematic variation unrelated to our effects of interest, as well as variability in these effects across the population and task structure. We fit the models using `lme4` [55] and `lmerTest` [84], with p values for all inferential t and F tests determined using the Satterthwaite approximation [85–87].

In our first model, we test whether pattern similarity between two word presentations at encoding is driven by both semantic and temporal distance. We allow these effects to vary by subject, session, list and serial position. The full model that we fit, as specified in `lme4`, is:

$$\begin{aligned}\texttt{similarity} \sim & \texttt{same_category} * \texttt{absolute_serial_position_distance} + \\ & (\cdot|\texttt{subject}) + \\ & (\cdot|\texttt{subject} : \texttt{session}) + \\ & (\cdot|\texttt{subject} : \texttt{session} : \texttt{list}) + \\ & (\cdot|\texttt{subject} : \texttt{session} : \texttt{list} : \texttt{serial_position_1}) + \\ & (\cdot|\texttt{subject} : \texttt{session} : \texttt{list} : \texttt{serial_position_2}) + \\ & (\cdot|\texttt{list}) + \\ & (\texttt{same_category}|\texttt{serial_position_1}) + \\ & (\texttt{same_category}|\texttt{serial_position_2}) + \\ & (\texttt{same_category}|\texttt{serial_position_1} : \texttt{serial_position_2})\end{aligned} \tag{1}$$

where `same_category` indicates that the two items are from the same category and `absolute_serial_position_distance` indicates the distance in terms of serial positions between the two items as a linear effect. The dot $(\cdot)$ indicates that all terms vary as a function of that random effect grouping variable. `list` indicates the list position within the session (from 1 to 25) and `serial_position_1` indicates the earlier serial position in the pair while `serial_position_2` indicates the later one. For this analysis we used all possible combinations of earlier and later serial positions.

In the second set of models, we test whether pattern similarity between a word presentation and a later inter-stimulus interval at encoding is driven by whether the initial word was recalled, whether it was recalled first and whether it was recalled next to the item before the ISI. For this analysis, we only used word presentations in the first half of the list and inter-stimulus intervals in the latter half. This ensured that each encoded item tested could be reinstated during each inter-stimulus interval. We also adjust for potential effects and interactions with semantic similarity and whether the item before the ISI was itself recalled. We again allow these effects to vary by subject, session, list and serial position. The full model that we fit, as specified in `lme4`, is:

$$\begin{aligned}\texttt{similarity} \sim &(\texttt{item_recalled} + \texttt{item_first}) * \texttt{item_before_isi_recalled} * \texttt{same_category} \\ &+ \texttt{recall_neighbor} + \texttt{recall_neighbor} : \texttt{same_category} + \\ &(\cdot|\texttt{subject}) + \\ &(\cdot|\texttt{subject} : \texttt{session}) + \\ &(\cdot|\texttt{subject} : \texttt{session} : \texttt{list}) + \\ &(\cdot|\texttt{subject} : \texttt{session} : \texttt{list} : \texttt{serial_position}) + \\ &(\cdot|\texttt{subject} : \texttt{session} : \texttt{list} : \texttt{serial_position_isi}) + \\ &(\cdot|\texttt{list}) + \\ &(\cdot|\texttt{serial_position}) + \\ &(\cdot|\texttt{serial_position_isi}) + \\ &(\cdot|\texttt{serial_position} : \texttt{serial_position_isi})\end{aligned} \tag{2}$$

`item_recalled` and `item_first` reflect whether the specific item was recalled and whether it was recalled first. `item_before_isi_recalled` indicates whether the item before the isi was recalled and `same_category` indicates that the initial encoding item and the item before the ISI are from the same category. Finally, `recall_neighbor` indicates that, if both items were recalled, whether they were recalled next to each other during the recall period. The dot $(\cdot)$ indicates that all terms vary as a function of that random effect grouping variable. `serial_position` indicates the serial position of the initial presentation while `serial_position_isi` indicates serial position preceding the ISI.

In the final set of models, we examine the degree to which items that are subsequently recalled are reinstated during a delay interval between the encoding phase and the test phase. To test this, allowing for potential interactions with serial position, we use the following specification:

$$\begin{aligned}\texttt{similarity} \sim &(\texttt{item_recalled} + \texttt{item_first}) * \texttt{serial_position} + \\ &(\cdot|\texttt{subject}) + \\ &(\cdot|\texttt{subject} : \texttt{session}) + \\ &(\cdot|\texttt{subject} : \texttt{session} : \texttt{list}) + \\ &(\cdot|\texttt{list}) +\end{aligned} \tag{3}$$

This model captures serial position effects with a categorical fixed effect. For this analysis, we use all encoded items, because they all have the potential to be reinstated during the delay.

Post-hoc analysis of models relied on `tidyverse` [88] and `emmeans` [56] in `R 4.3.2` [89].

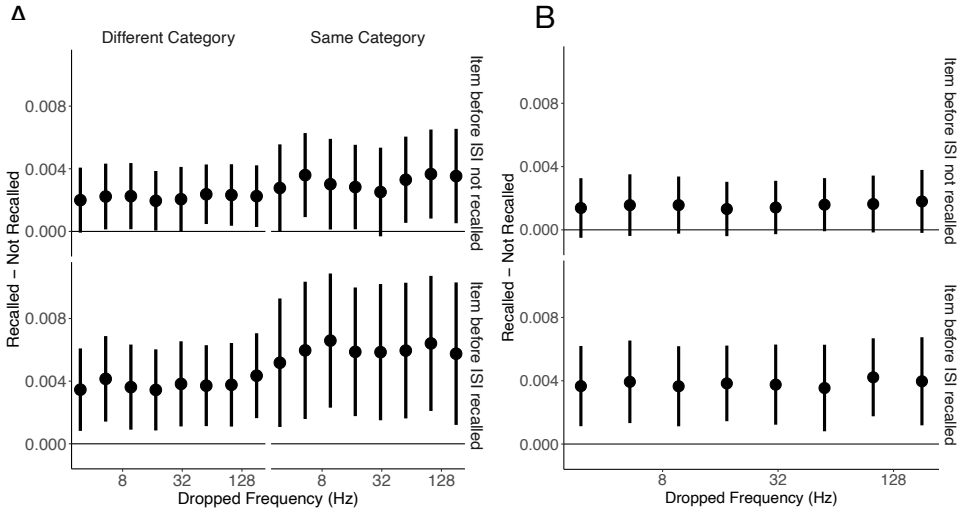

Fig. A1 Replication of Fig. 3 while dropping individual frequency bands from the similarity computation. X axis indicates which frequency is dropped and is shown on a log scale. A. results from Experiment 1. B. results from Experiment 2.

Appendix A Frequency Specificity

In order to obtain a more mechanistic understanding of the frequencies contributing to our observed rapid reinstatement results, we replicated our analysis dropping individual frequency bands when computing the correlation between timepoints. Fig. A1 replicates Fig. 3, showing the estimated difference in similarity to subsequent ISIs when the encoding trial is recalled vs. not recalled when each of the eight frequency bands are dropped in each subgroup of trials. The results indicate that the effect is largely unchanged when individual frequencies are dropped, suggesting that all frequencies contribute to the observed reinstatement.

Appendix B Generalization to end of encoding presentation

In Experiment 1, we replicate the results from our ISI analysis, finding a main effect of `item_recalled` such that items that were subsequently recalled were more similar to activity during the last 500ms of subsequent encoding items items that were not subsequently recalled (Fig. B2, $F(1, 8.6) = 16.09, p = .003$). Post-hoc comparisons revealed significant effects of recall status, regardless of whether the subsequent item was from the same category or was itself recalled. However, we could not reliably identify whether the items recalled first had consistently different levels of reactivation than other recalled items ($F(1, 7.6) = 1.02, p = .34$). we also did not see consistent differences between items that were sequentially recalled and pairs of items that were

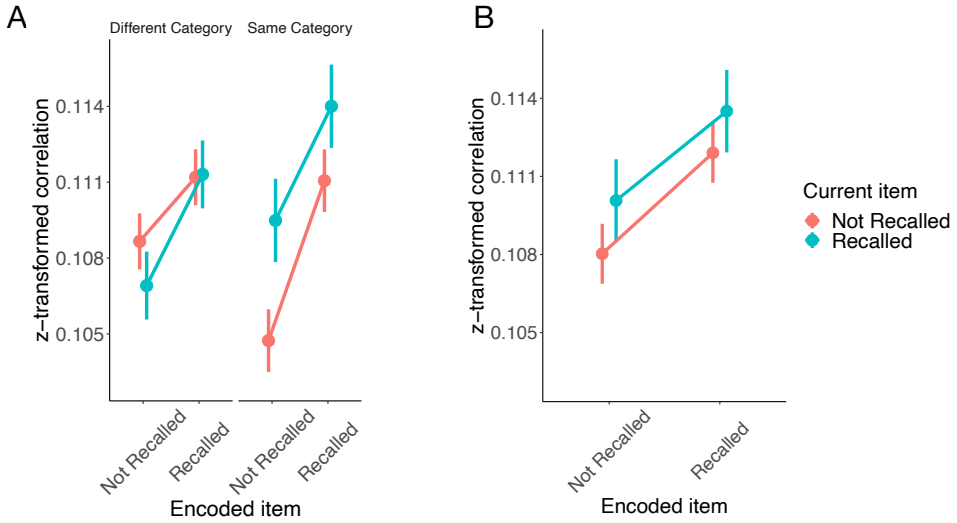

Fig. B2 Similarity of time points during encoding word presentations and time points during the last 500ms of subsequent encoding word presentations as a function of whether the word was itself subsequently recalled (**Recalled**). Estimates are derived from the mixed effect model described in the main text. Error bars reflect 95% confidence intervals on the difference between remembered and forgotten, using an algorithm implemented by the R package `emmeans` and described in [56] that is a generalization of Loftus-Masson intervals [57, 58]. A. results from Experiment 1. B. results from Experiment 2.

both recalled but not sequentially ($F(1, 6.2) = 0.34, p = .58$). Unlike in the ISI analysis, we find that the recall status of the currently encoded item does not in general predict similarity ($F(1, 14.5) = 1.22, p = .29$). As before, we find that the two items being from the same category is also not a significant predictor ($F(1, 24.6) = 0.03, p = .85$). Unlike above, we find significant interactions with the overall recalled effect. In particular, the recalled effect differed depending on whether the current item was from the same category or not ($F(1, 17013.9) = 4.98, p = .03$). Inspection of Fig. B2 suggests that this is due to the effect being larger when the two items are from the same category. There is an additional three-way interaction with the recall status of the current item, recall status of the earlier item and whether the two items were from the same category ($F(1, 17601.8) = 4.12, p = .04$). This is because when the items are from the same category, the current item being recalled makes the recalled effect smaller. We speculate that this may be due to limitations of the measurement in distinguishing between focus on the current item or reinstating a previous item when the two items are from the same category. Finally, we find a significant interaction between whether the two items are from the same category and the recall status of the later item ($F(1, 11.3) = 4.98, p = .03$). We speculate that this is due to the fact that when the later item is recalled, the semantic information is more strongly encoded, making its neural representation more similar to earlier same category items.

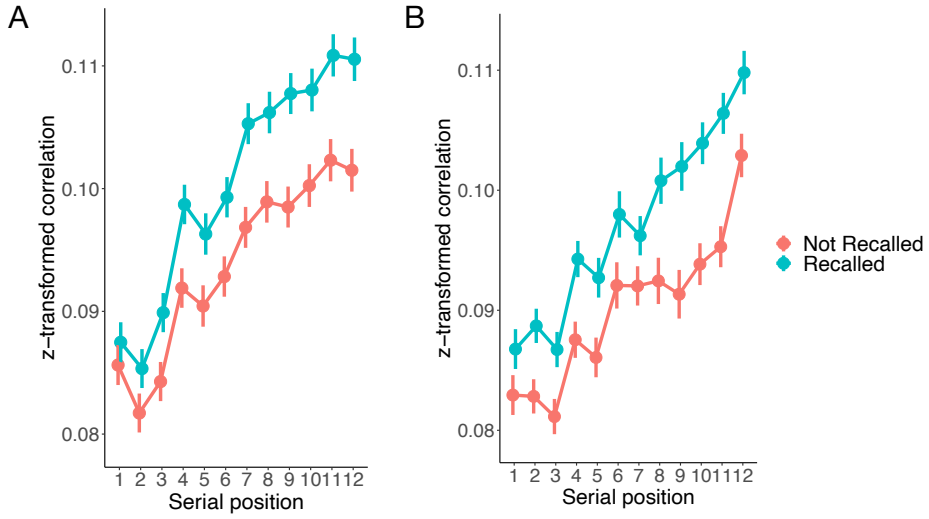

Fig. C3 Similarity of time points during encoding word presentations and timepoints during the math distractor task as a function of whether the word was itself subsequently recalled (**Recalled**). Estimates are derived from the mixed effect model described in the main text. Error bars reflect 95% confidence intervals on the difference between remembered and forgotten, using an algorithm implemented by the R package `emmeans` and described in [56] that is a generalization of Loftus-Masson intervals [57, 58]. A. results from Experiment 1. B. results from Experiment 2.

We again fit the same model to Experiment 2, dropping the terms involving `same_category`. We replicated the main effect of subsequent recall on study-phase spectral pattern similarity, i.e. items with more similar activity during subsequent ISIs during encoding than items that were not subsequently recalled (Fig. 3, $F(1, 12.13) = 11.11, p = .006$). In a post-hoc comparison, unlike the results with ISIs, we find that this result is significant regardless of whether the subsequent item is itself recalled. Again, we could not establish the direction of effects of first recall ($F(1, 5.8) = .67, p = .44$) or being a recall neighbor ($F(1, 4.77) = .05, p = .83$) on pattern similarity.

Appendix C Generalization to distractor period

In order to more directly replicate [2] and investigate whether at least some of the measured reinstatement is likely to be unconscious, we analyzed data during the math distractor task in experiments 1 and 2 using the same model as used above in Experiment 3. We find a significant main effect of recall status on reinstatement (Experiment 1: $F(1, 204.81) = 66.56, p < .001$; Experiment 2: $F(1, 275.77) = 92.85, p < .001$). In both experiments, we also find a significant interaction with serial position (Experiment 1: $F(11, 326.01) = 2.11, p = .02$; Experiment 2: $F(11, 297.51) = 2.05, p = .02$). Visual inspection of Fig. C3 suggests that this is primarily due to the difference in reinstatement between remembered and forgotten items being larger in

later serial positions. Finally, we find a main effect of serial position (Experiment 1: $F(11, 212.09) = 25.93, p < .001$; Experiment 2: $F(11, 136.33) = 24.41, p < .001$).

References

[1] Deuker, L. *et al.* Memory Consolidation by Replay of Stimulus-Specific Neural Activity. *The Journal of Neuroscience* **33**, 19373–19383 (2013). URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0414-13.2013.

[2] Staresina, B. P., Alink, A., Kriegeskorte, N. & Henson, R. N. Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences* **110**, 21159–21164 (2013). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1311989110.

[3] Wilson, M. & McNaughton, B. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994). URL https://www.sciencemag.org/lookup/doi/10.1126/science.8036517.

[4] Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience* **14**, 147–153 (2011). URL http://www.nature.com/articles/nn.2732.

[5] Benjamin, A. S. & Tullis, J. What makes distributed practice effective? *Cognitive Psychology* **61**, 228–247 (2010). URL https://linkinghub.elsevier.com/retrieve/pii/S0010028510000332.

[6] Thios, S. J. & D'Agostino, P. R. Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior* **15**, 529–536 (1976). URL https://linkinghub.elsevier.com/retrieve/pii/0022537176900475.

[7] Siegel, L. L. & Kahana, M. J. A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **40**, 755–764 (2014). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0035585.

[8] Greene, R. L. Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 371–377 (1989). Place: US Publisher: American Psychological Association.

[9] Hintzman, D. L., Summers, J. J. & Block, R. A. Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory* **1**, 31–40 (1975). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.1.1.31.

[10] Zhang, H., Fell, J. & Axmacher, N. Electrophysiological mechanisms of human memory consolidation. *Nature Communications* **9**, 4103 (2018). URL http://www.nature.com/articles/s41467-018-06553-y.

[11] Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C. & Norman, K. A. Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications* **9**, 3920 (2018). URL http://www.nature.com/articles/s41467-018-06213-1.

[12] Tambini, A., Ketz, N. & Davachi, L. Enhanced Brain Correlations during Rest Are Related to Memory for Recent Experiences. *Neuron* **65**, 280–290 (2010). URL https://linkinghub.elsevier.com/retrieve/pii/S0896627310000061.

[13] Tambini, A. & Davachi, L. Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proceedings of the National Academy of Sciences* **110**, 19591–19596 (2013). URL https://pnas.org/doi/full/10.1073/pnas.1308499110.

[14] Buzsáki, G. Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience* **31**, 551–570 (1989). URL https://linkinghub.elsevier.com/retrieve/pii/0306452289904235.

[15] McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. 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**, 419–457 (1995). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.102.3.419.

[16] Ward, G. & Tan, L. The role of rehearsal and reminding in the recall of categorized word lists. *Cognitive Psychology* **143**, 101563 (2023). URL https://linkinghub.elsevier.com/retrieve/pii/S001002852300021X.

[17] Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology* **89**, 63–77 (1971). URL http://content.apa.org/journals/xge/89/1/63.

[18] Rundus, D. & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior* **9**, 99–105 (1970). URL https://linkinghub.elsevier.com/retrieve/pii/S0022537170800159.

[19] Brodie, D. A. & Prytulak, L. S. Free recall curves: Nothing but rehearsing some items more or recalling them sooner? *Journal of Verbal Learning and Verbal Behavior* **14**, 549–563 (1975). URL https://linkinghub.elsevier.com/retrieve/pii/S0022537175800326.

[20] Brodie, D. A. & Murdock, B. B. Effect of presentation time on nominal and functional serial-position curves of free recall. *Journal of Verbal Learning and Verbal Behavior* **16**, 185–200 (1977). URL https://linkinghub.elsevier.com/retrieve/pii/S0022537177800467.

[21] Modigliani, V. & Hedges, D. G. Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **13**, 426–436 (1987). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.13.3.426.

[22] McKinley, G. L. & Benjamin, A. S. The role of retrieval during study: Evidence of reminding from overt rehearsal. *Journal of Memory and Language* **114**, 104128 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S0749596X20300425.

[23] Laming, D. Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 1146–1163 (2006). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.32.5.1146.

[24] Ward, G. A recency-based account of the list length effect in free recall. *Memory & Cognition* **30**, 885–892 (2002). URL http://link.springer.com/10.3758/BF03195774.

[25] Ward, G. & Tan, L. The Effect of the Length of To-Be-Remembered Lists and Intervening Lists on Free Recall: A Reexamination Using Overt Rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 1196–1210 (2004). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.30.6.1196.

[26] Ward, G., Woodward, G., Stevens, A. & Stinson, C. Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 186–210 (2003). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.29.2.186.

[27] Tan, L. & Ward, G. A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1589–1625 (2000). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.26.6.1589.

[28] Honey, C. J., Newman, E. L. & Schapiro, A. C. Switching between internal and external modes: A multiscale learning principle. *Network Neuroscience* **1**, 339–356 (2017). URL https://direct.mit.edu/netn/article/1/4/339-356/5395.

[29] Mednick, S. C., Cai, D. J., Shuman, T., Anagnostaras, S. & Wixted, J. T. An opportunistic theory of cellular and systems consolidation. *Trends in Neurosciences* **34**, 504–514 (2011). URL https://linkinghub.elsevier.com/retrieve/pii/S0166223611000919.

[30] Norman, K. A., Newman, E. L. & Perotte, A. J. Methods for reducing interference in the Complementary Learning Systems model: Oscillating inhibition and autonomous memory rehearsal. *Neural Networks* **18**, 1212–1228 (2005). URL https://linkinghub.elsevier.com/retrieve/pii/S0893608005001930.

[31] Moscovitch, M. & Gilboa, A. Systems consolidation, transformation and reorganization: Multiple Trace Theory, Trace Transformation Theory and their Competitors. preprint, PsyArXiv (2021). URL https://osf.io/yxbrs.

[32] Wamsley, E. J., Arora, M., Gibson, H., Powell, P. & Collins, M. Memory Consolidation during Ultra-short Offline States. *Journal of Cognitive Neuroscience* **35**, 1617–1634 (2023). URL https://direct.mit.edu/jocn/article/35/10/1617/117068/Memory-Consolidation-during-Ultra-short-Offline.

[33] Wamsley, E. J. & Summer, T. Spontaneous Entry into an "Offline" State during Wakefulness: A Mechanism of Memory Consolidation? *Journal of Cognitive Neuroscience* **32**, 1714–1734 (2020). URL https://direct.mit.edu/jocn/article/32/9/1714/95497/Spontaneous-Entry-into-an-Offline-State-during.

[34] Fellner, M.-C., Waldhauser, G. T. & Axmacher, N. Tracking Selective Rehearsal and Active Inhibition of Memory Traces in Directed Forgetting. *Current Biology* **30**, 2638–2644.e4 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S0960982220306394.

[35] Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J. & Burgess, N. Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *The Journal of Neuroscience* **35**, 14426–14434 (2015). URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1774-15.2015.

[36] Xue, G. *et al.* Greater Neural Pattern Similarity Across Repetitions Is Associated with Better Memory. *Science* **330**, 97–101 (2010). URL https://www.sciencemag.org/lookup/doi/10.1126/science.1193125.

[37] Ward, E. J., Chun, M. M. & Kuhl, B. A. Repetition Suppression and Multi-Voxel Pattern Similarity Differentially Track Implicit and Explicit Visual Memory. *Journal of Neuroscience* **33**, 14749–14757 (2013). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4889-12.2013.

[38] Hasinski, A. E. & Sederberg, P. B. Trial-level information for individual faces in the fusiform face area depends on subsequent memory. *NeuroImage* **124**, 526–535 (2016). URL https://linkinghub.elsevier.com/retrieve/pii/S1053811915007867.

[39] Lu, Y., Wang, C., Chen, C. & Xue, G. Spatiotemporal Neural Pattern Similarity Supports Episodic Memory. *Current Biology* **25**, 780–785 (2015). URL https://linkinghub.elsevier.com/retrieve/pii/S0960982215000871.

[40] Ebbinghaus, H. *Memory: A contribution to experimental psychology.* (Teachers College Press, New York, 1913). URL http://content.apa.org/books/10011-000.

[41] Madigan, S. A. Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior* **8**, 828–835 (1969). URL https://linkinghub.elsevier.com/retrieve/pii/S0022537169800502.

[42] Melton, A. W. The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior* **9**, 596–606 (1970). URL https://linkinghub.elsevier.com/retrieve/pii/S0022537170801074.

[43] Feng, K. *et al.* Spaced Learning Enhances Episodic Memory by Increasing Neural Pattern Similarity Across Repetitions. *The Journal of Neuroscience* **39**, 5351–5360 (2019). URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2741-18.2019.

[44] Wu, X. & Fuentemilla, L. Distinct encoding and post-encoding representational formats contribute to episodic sequence memory formation. *Cerebral Cortex* **33**, 8534–8545 (2023). URL https://academic.oup.com/cercor/article/33/13/8534/7147876.

[45] Wu, X., Viñals, X., Ben-Yakov, A., Staresina, B. P. & Fuentemilla, L. Post-encoding Reactivation Is Related to Learning of Episodes in Humans. *Journal of Cognitive Neuroscience* **35**, 74–89 (2022). URL https://direct.mit.edu/jocn/article/35/1/74/113587/Post-encoding-Reactivation-Is-Related-to-Learning.

[46] Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* (2008). URL http://journal.frontiersin.org/article/10.3389/neuro.06.004.2008/abstract.

[47] Edelman, S., Grill-Spector, K., Kushnir, T. & Malach, R. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* **26**, 309–321 (1998). URL https://doi.org/10.3758/BF03330618.

[48] Sommer, V. R., Mount, L., Weigelt, S., Werkle-Bergner, M. & Sander, M. C. Spectral pattern similarity analysis: Tutorial and application in developmental cognitive neuroscience. *Developmental Cognitive Neuroscience* **54**, 101071 (2022). URL https://linkinghub.elsevier.com/retrieve/pii/S1878929322000159.

[49] Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B. & Kahana, M. J. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences* **108**, 12893–12897 (2011). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1015174108.

[50] Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A. & Kahana, M. J. Spontaneously Reactivated Patterns in Frontal and Temporal Lobe Predict Semantic Clustering during Memory Search. *Journal of Neuroscience* **32**, 8871–8878 (2012). URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5321-11.2012.

[51] Fisher, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* **10**, 507 (1915). URL https://www.jstor.org/stable/2331838?origin=crossref.

[52] Kragel, J. E. *et al.* Distinct cortical systems reinstate the content and context of episodic memories. *Nature Communications* **12**, 4444 (2021). URL http://www.nature.com/articles/s41467-021-24393-1.

[53] Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. Parsimonious Mixed Models. *arXiv:1506.04967 [stat]* (2018). URL http://arxiv.org/abs/1506.04967. ArXiv: 1506.04967.

[54] Cousineau, D. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology* **1**, 42–45 (2005). URL http://www.tqmp.org/RegularArticles/vol01-1/p042.

[55] Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* **67** (2015). URL http://www.jstatsoft.org/v67/i01/.

[56] Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means* (2019). URL https://CRAN.R-project.org/package=emmeans.

[57] Loftus, G. R. & Masson, M. E. J. Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review* **1**, 476–490 (1994). URL http://link.springer.com/10.3758/BF03210951.

[58] Franz, V. H. & Loftus, G. R. Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review* **19**, 395–404 (2012). URL http://link.springer.com/10.3758/s13423-012-0230-1.

[59] Braun, K. & Rubin, D. C. The Spacing Effect Depends on an Encoding Deficit, Retrieval, and Time in Working Memory: Evidence. *Memory* **6**, 37–66 (1998). URL http://www.tandfonline.com/doi/abs/10.1080/741941599.

[60] Kimball, D. R., Smith, T. A. & Kahana, M. J. The fSAM model of false recall. *Psychological Review* **114**, 954–993 (2007). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.114.4.954.

[61] Haxby, J. V. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science* **293**, 2425–2430 (2001). URL https://www.sciencemag.org/lookup/doi/10.1126/science.1063736.

[62] Howard, M. W. & Kahana, M. J. When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language* **46**, 85–98 (2002). URL https://linkinghub.elsevier.com/retrieve/pii/S0749596X01927986.

[63] Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K. & Lindsley, D. B. Electrocortical Signs of Levels of Processing: Perceptual Analysis and Recognition Memory. *Psychophysiology* **17**, 568–576 (1980). URL http://doi.wiley.com/10.

1111/j.1469-8986.1980.tb02299.x.

[64] Fernández, G. *et al.* Real-Time Tracking of Memory Formation in the Human Rhinal Cortex and Hippocampus. *Science* **285**, 1582–1585 (1999). URL https://www.science.org/doi/10.1126/science.285.5433.1582.

[65] Fell, J. *et al.* Human memory formation is accompanied by rhinal–hippocampal coupling and decoupling. *Nature Neuroscience* **4**, 1259–1264 (2001). URL https://www.nature.com/articles/nn759.

[66] Paller, K. A. & Wagner, A. D. Observing the transformation of experience into memory. *Trends in Cognitive Sciences* **6**, 93–102 (2002). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661300018453.

[67] Dewar, M., Alber, J., Cowan, N. & Della Sala, S. Boosting Long-Term Memory via Wakeful Rest: Intentional Rehearsal Is Not Necessary, Consolidation Is Sufficient. *PLoS ONE* **9**, e109542 (2014). URL https://dx.plos.org/10.1371/journal.pone.0109542.

[68] Tambini, A. & Davachi, L. Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. *Trends in Cognitive Sciences* **23**, 876–890 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661319301834.

[69] Howard, M. W., Viskontas, I. V., Shankar, K. H. & Fried, I. Ensembles of human MTL neurons "jump back in time" in response to a repeated stimulus. *Hippocampus* **22**, 1833–1847 (2012). URL https://onlinelibrary.wiley.com/doi/10.1002/hipo.22018.

[70] Folkerts, S., Rutishauser, U. & Howard, M. W. Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. *The Journal of Neuroscience* **38**, 4200–4211 (2018). URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2312-17.2018.

[71] Lohnas, L. J., Healey, M. K. & Davachi, L. Neural temporal context reinstatement of event structure during memory recall. *Journal of Experimental Psychology: General* **152**, 1840–1872 (2023). URL https://doi.apa.org/doi/10.1037/xge0001354.

[72] Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychological Review* **62**, 145–154 (1955). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/h0048509.

[73] Howard, M. W. & Kahana, M. J. A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology* **46**, 269–299 (2002). URL https://linkinghub.elsevier.com/retrieve/pii/S0022249601913884.

[74] Polyn, S. M., Norman, K. A. & Kahana, M. J. A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review* **116**, 129–156 (2009). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0014420.

[75] Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* **21**, 1609–1617 (2018). URL http://www.nature.com/articles/s41593-018-0232-z.

[76] Weidemann, C. T. *et al.* Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *Journal of Experimental Psychology: General* **148**, 1–12 (2019). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000480.

[77] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* **41**, 391–407 (1990). URL https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199009)41:6⟨391::AID-ASI1⟩3.0.CO;2-9.

[78] Landauer, T. K. & Dumais, S. T. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104**, 211–240 (1997). Place: US Publisher: American Psychological Association.

[79] Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J. & Ding, M. Theta Oscillations Mediate Interaction between Prefrontal Cortex and Medial Temporal Lobe in Human Memory. *Cerebral Cortex* **20**, 1604–1612 (2010). URL https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp223.

[80] SciPy 1.0 Contributors *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272 (2020). URL http://www.nature.com/articles/s41592-019-0686-2.

[81] Hoyer, S. & Hamman, J. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software* **5**, 10 (2017). URL https://openresearchsoftware.metajnl.com/article/10.5334/jors.148/.

[82] McKinney, W. *Data Structures for Statistical Computing in Python*, 56–61 (Austin, Texas, 2010). URL https://conference.scipy.org/proceedings/scipy2010/mckinney.html.

[83] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020). URL https://www.nature.com/articles/s41586-020-2649-2.

[84] Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82** (2017). URL http://www.jstatsoft.org/v82/i13/.

[85] Satterthwaite, F. E. An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin* **2**, 110 (1946). URL http://dx.doi.org/10.2307/3002019. Publisher: JSTOR.

[86] Giesbrecht, F. G. & Burns, J. C. Two-Stage Analysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics* **41**, 477 (1985). URL http://dx.doi.org/10.2307/2530872. Publisher: JSTOR.

[87] Hrong-Tai Fai, A. & Cornelius, P. L. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation* **54**, 363–378 (1996). URL http://dx.doi.org/10.1080/00949659608811740. Publisher: Informa UK Limited.

[88] Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019). URL https://joss.theoj.org/papers/10.21105/joss.01686.

[89] R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020). URL https://www.R-project.org.