Tracking selective rehearsal and active inhibition of memory traces in directed forgetting

Marie-Christin Fellner*, Gerd T. Waldhauser, & Nikolai Axmacher

*Corresponding author:
Dr. Marie-Christin Fellner
Department of Neuropsychology
Institute of Cognitive Neuroscience
Ruhr University Bochum
Universitätsstrasse 150
44801 Bochum
Germany

E-mail: marie-christin.fellner@rub.de
Phone: +49 234 32-28066
Fax : +49 234 32-14622

Abstract

Selectively remembering or forgetting newly encountered information is essential for goal-directed behavior. It is still an open question, however, whether intentional forgetting is an active process based on the inhibition of unwanted memory traces or whether it occurs passively through reduced recruitment of selective rehearsal [1,2]. Here we show that intentional control of memory encoding relies on both, enhanced active inhibition and decreased selective rehearsal, and that these two processes can be separated in time and space. We applied representational similarity analysis (RSA [3]) and time-frequency analysis to EEG data during an item-method directed forgetting experiment [4]. We identified neural signatures of both the intentional suppression and the voluntary upregulation of item-specific representations. Successful active forgetting was associated with a downregulation of item-specific representations in an early time window, 500ms after the instruction. This process was initiated by an increase in oscillatory alpha (8-13 Hz) power, a well-established signature of neural inhibition [5,6], in occipital brain areas. During a later time window, 1500ms after the cue, intentional forgetting was associated with reduced employment of active rehearsal processes, as reflected by an attenuated upregulation of item-specific representations as compared to intentionally encoded items. Our data show that active inhibition and selective rehearsal are two separate mechanisms whose consecutive employment allows for a voluntary control of memory formation.

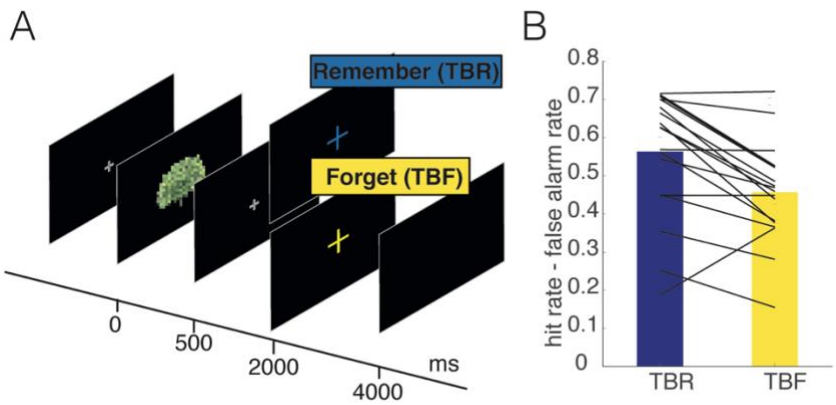


Figure 1: Paradigm and behavioral results. (A) During each encoding trial, a picture of an everyday object was shown that was followed by either a TBR (to be remembered) or a TBF (to be forgotten) cue. (B) TBF cues led to reduced recognition memory performance relative to TBR cues in almost all participants. Bar plot shows mean recognition memory performance, lines indicate performance of single subjects.

Results & Discussion

Forgetting instructions reduce subsequent memory performance

We employed the well-established item-method directed forgetting paradigm [1,2,4] in order to investigate the mechanisms underlying intentional forgetting (Figure 1A). Pictures of everyday objects were presented, followed by cues indicating whether the preceding item was to be forgotten (TBF) or to be remembered (TBR). Participants were instructed not to remember the TBF items in order to improve remembering the crucial TBR items. After a brief distractor phase, they performed an item recognition task which nevertheless contained all TBF and TBR items that had been presented during encoding, randomly intermixed with new pictures. Participants were now explicitly asked to remember both, TBF and TBR items. As expected, TBF items were remembered worse than TBR items (recognition performance measure Pr: hit rate - false alarm rate, $M_{TBF}=0.46$, $M_{TBR}=0.56$, $t(17)=4.19$, $p=0.0006$, Figure 1B), with all but three participants showing the expected directed forgetting effect. The reduced recognition memory performance suggests that the intention to forget or remember indeed leads to the down- or upregulation of individual memory traces, i.e. successful forgetting [1,7].

Item-cue similarity as a measure of active inhibition and active rehearsal

The processes underlying directed forgetting are heavily debated: It is unclear whether reduced memory for TBF items is due to their active inhibition – i.e., a specific downregulation of their neural representations (active inhibition account; Figure 2A, pink background) – or caused by reduced recruitment of rehearsal processes (selective rehearsal account; Figure 2A, green background). On a behavioral level, both accounts predict reduced memory for TBF items and are therefore indistinguishable. However, the two accounts make dissociable and testable predictions regarding the fate of item-specific representations after a TBF and a TBR cue (Figure 2A). According to the active inhibition account, forgetting after TBF cues results from successfully applied inhibitory control. The representations of forgotten TBF items (TBF-f) should thus be attenuated below the level which occurs due to passive forgetting, i.e., they should be reduced as compared to the representations of forgotten TBR items (TBR-f). This effect should be specific to the contrast of active versus passive forgotten items and not occur for later remembered TBF and TBR items (TBF-r and TBR-r, respectively).

By contrast, the selective rehearsal account proposes that reduced memory for TBF items results passively from reduced recruitment of rehearsal processes for these items. This account assumes that item representations of subsequently remembered TBR items (TBR-r) are actively upregulated as compared to unintentionally remembered TBF items (TBF-r). This rehearsal-related upregulation of item representations should drive actual remembering of the items and should thus not occur for later forgotten TBR and TBF items (TBR-f and TBF-f, respectively).

The two accounts thus make clearly distinguishable predictions for the neural representations of items in the different conditions: The active inhibition account predicts that representations of TBF-f items are suppressed below the level of TBR-f items (i.e., TBF-f < TBR-f), and that this suppression does not occur for TBF-r or TBR-r items. The reduced rehearsal account predicts decreased representations of TBF-r as compared to TBR-r items (i.e., TBF-r < TBR-r), an effect that again should be specific and not occur for later forgotten items. In order to isolate these two effects, we thus tested the interaction between the two contrasts [8,9]: We compared the difference in representational EEG similarity between representations of forgotten and remembered TBF items versus forgotten and remembered TBR items (TBF-f vs. TBR-f) - (TBF-r vs. TBR-r). In this interaction contrast, active inhibition should result in lower RSA values (because the first term of the interaction contrast is more negative), while active rehearsal results in higher RSA values (as the second term is more negative; Figure 2A&B).

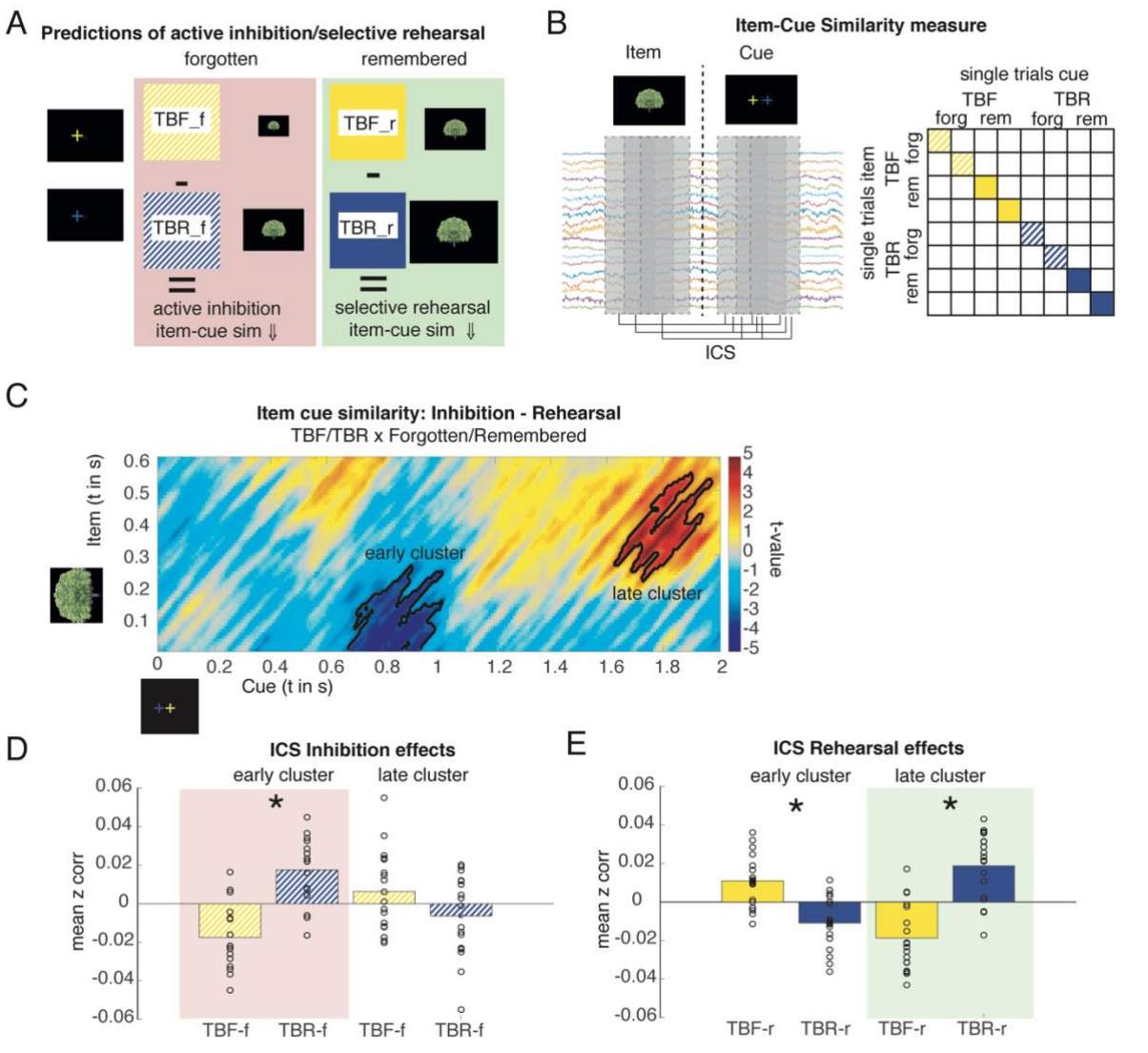


Figure 2: Transformation of item-specific representations during directed forgetting. (A) Predictions of active inhibition and selective rehearsal accounts on item-cue similarity (ICS) in the different experimental conditions. (B) Correlations between EEG activities during the item and the cue interval in every trial were calculated as single-trial ICS measures. These trial-based ICS values were then contrasted depending on the cue (TBF or TBR) and depending on subsequent memory (remembered or forgotten). The matrix shown on the right depicts the correlations that were included in the analysis. (C) ICS differences for all combinations of item-time x cue-time intervals. We tested for the differential employment of active inhibition (i.e., TBF-f vs. TBR-f) or reduced rehearsal (i.e., TBF-r vs. TBR-r), corresponding to a TBF/TBR cue x memory interaction. Colors depict t-values of ICS differences; significant clusters are highlighted by black contours (permutation corrected for multiple comparisons). The early negative cluster indicates active inhibition, while the late positive cluster reflects reduced rehearsal. (D,E) Mean ICS values of forgotten (D) and remembered (E) items in the two interaction clusters shown in (C). ICS values were mean centered to subject-specific mean ICS values for forgotten or remembered items to depict within-subject differences. Stars denote significant differences ($p<.05$). Predictions for active inhibition and selective rehearsal effects are highlighted in pink and green, respectively. Please note that the selective rehearsal contrast is always plotted as reduced rehearsal (i.e., TBF-r < TBR-r), i.e. more pronounced representations of voluntarily remembered items in contrast to incidentally remembered items is indicated by negative values. Results show that the interaction contrast in the early cluster relies on active inhibition of TBF-f items and even slightly increased rehearsal of TBF-r items, while the contrast in the later cluster is specifically due to reduced rehearsal of TBF-r items.

Item-cue similarities provide evidence for early active forgetting and later reduced rehearsal

To estimate item representations after the cue, item-cue similarities (ICS values) were estimated in each trial by calculating Spearman's rank correlations between spatiotemporal EEG activity patterns during item presentation and those following the TBF/TBR cue presentation. Spatiotemporal patterns were defined by cutting the preprocessed time-domain EEG data in 200ms windows across all electrodes (Figure 2B) and concatenating these two-dimensional electrode x time patterns into one-dimensional representations of spatiotemporal features [10–12]. ICS values were calculated for all combinations of item windows and cue windows in steps of 10ms, resulting in a time-resolved matrix of item representations during the cue interval. This analysis allows for a time-resolved assessment [13] of the specific predictions of the active inhibition and the reduced rehearsal accounts, a clear advantage over comparable analyses with functional magnetic resonance imaging [12].

Cluster-based permutation statistics applied on the ICS matrix for all time bins revealed an early negative (p_{corr}=0.004) and a late positive cluster (p_{corr}=0.009, see Figure 2C). The early negative cluster corresponds to EEG activity around 600-1200ms after the onset of TBF/TBR cues and 0-300ms after item onset. Follow-up analyses of simple contrasts in this early cluster revealed a significant decrease of ICS values for forgotten TBF vs. forgotten TBR items (t(17)=-4.52, p=0.0003, Figure 2D), i.e., item-related patterns were reduced during intentional as compared to unintentional forgetting. This effect was not observed when comparing later remembered TBF and TBR items (see below), in line with the predictions of the active inhibition account.

The second significant cluster occurred later, in a time window corresponding to 1500-2000ms after cue onset and 250-700ms after item onset. In this late time window cluster, ICS values were significantly higher for remembered TBR as compared to remembered TBF items (t(17)=-4.5941, p=0.0002, Figure 2E). Follow-up analyses did not reveal active inhibition in this time period (TBF-f vs. TBR-f items; t(17)=1.287, p=0.22, Figure 2D). This result suggests that remembering of TBR items is driven by an upregulation of the corresponding memory traces, i.e. active rehearsal, and that this process is less recruited for incidentally remembered TBF items.

Interestingly, in the early cluster, also an "inverted" active rehearsal effect was present: we observed higher item-cue similarity for incidentally remembered TBF-r items as compared to intentionally remembered TBR-r items (t(17)=3.4308, p=0.0032, Figure 2E). Possibly, this effect indicates that voluntary inhibition attempts are not successful for items with particularly strong memory representations [14]. Alternatively, inhibition attempts may induce an early and seemingly paradoxical temporary enhancement of memory before its later suppression [12,15,16]. Since this effect occurred early after the cue, it may reflect an automatic reactivation of a strong item representation acting beyond inhibitory reach [17]. Importantly, the post-hoc tests show that both the active inhibition effect and the

paradoxical rehearsal effect reach significance, i.e. the interaction contrast (Figure 2C) is indeed driven by both effects.

We conducted several control analyses in order to exclude possible confounds in our assessment of ICS. These control analyses showed that condition-specific differences in ICS were indeed driven by item-specific correlations between item and cue representations and do not reflect unspecific condition effects on EEG activity (see Supplementary Material and Supplementary Figure 1 & 2).

Taken together, assessing ICS revealed evidence for both accounts of directed forgetting, and showed that active inhibition and selective rehearsal both contribute to the difference in memory performance between TBF and TBR cued items.

Increased posterior alpha oscillations precede active inhibition of item-specific representations

Numerous theoretical frameworks and empirical results have established the role of low-frequency brain oscillations in routing cortical information flow within and between brain regions. Specifically, alpha oscillations (~8-13Hz) have been described as a signature of active inhibition of information processing within a brain area [6,18–20]. In order to reveal the mechanisms underlying the up- and downregulation of memory traces, we thus investigated brain oscillatory dynamics related to active inhibition and selective rehearsal. Again, we contrasted activity during forgotten TBF vs. forgotten TBR items as an index of active inhibition, and during remembered TBF vs. remembered TBR items reflecting selective rehearsal.

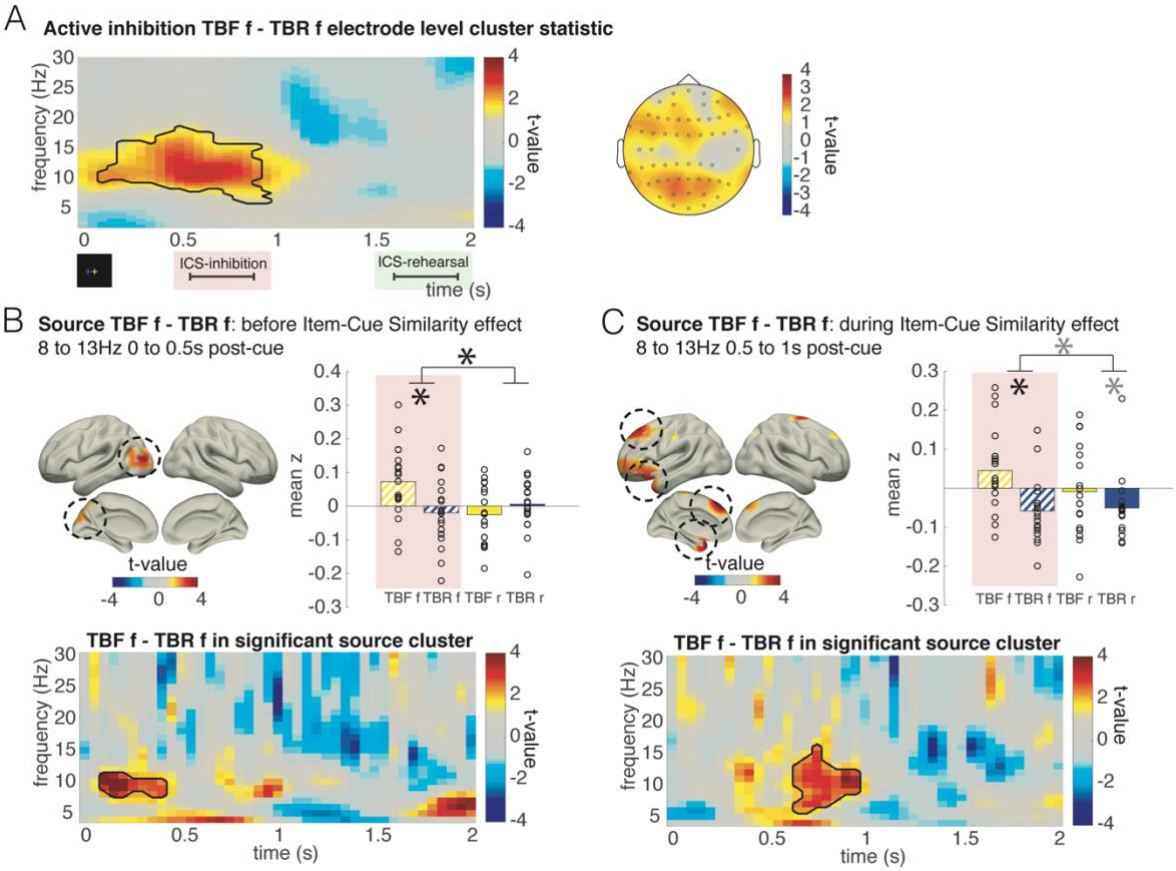


Figure 3: Brain oscillatory correlates of active inhibition. (A) Contrast of intentionally forgotten trials (TBF-f) vs. incidentally forgotten trials (TBR-f). Results are corrected for multiple comparisons using a cluster-based permutation test across all electrodes, frequencies (2-30Hz), and time points (0-2s post cue). Time-frequency plots show average t-values averaged across all significant electrodes. Right side: topographical plot highlighting the respective electrodes. (B) Oscillatory active inhibition effects preceding active inhibition-related ICS changes (time window: 0-0.5s post-cue). (C) Oscillatory active inhibition effects during the inhibition-related ICS changes (time window: 0.5-1s post-cue). Source plots show differences between TBF-f and TBR-f trials, thresholded at $p<0.01$, circles highlight significant clusters ($p<0.05$). Bar plots to the right depict average power in the significant source clusters, black stars indicate $p<0.05$, grey stars $p<0.1$, black circles denote single subject trial averages. Time-frequency plots below show average power changes in the respective significant source clusters, black contours highlight significant clusters ($p<0.05$).

We conducted a three-dimensional cluster permutation statistic across frequencies, time points, and electrodes, contrasting TBF-f with TBR-f trials. This analysis revealed a significant increase of alpha power early after the TBF cue (~8-13Hz, 100-1000ms, $p_{corr}=0.001$, Figure 3A). This effect included a similar time window as the active inhibition effect on ICS (600-1200ms; see above). Importantly, however, it showed a markedly earlier onset (starting at 100ms rather than 600ms post cue onset). In a second step, we therefore investigated alpha power changes separately for the time windows preceding and during the ICS effect, by dividing the power effect time window into equal halves (0-500ms and 500-1000ms, respectively). We applied beamforming-based source estimation both to the early alpha power effect (preceding the active downregulation of item-specific memory traces) and to the late alpha power effect (during this downregulation). This analysis revealed a striking pattern: The early alpha power increase was localized to areas of the medial and lateral occipital cortex (Figure 3B, MNI peak -

37, -89, 12, left middle occipital gyrus; t(17)=4.08, p_{corr}=0.045). This early occipital alpha power increase preceded the active downregulation of memory traces. It was specific to TBF trials that were actually forgotten, and did not occur for later remembered trials (Figure 3B; interaction analysis, i.e. contrast of active inhibition vs rehearsal (see Methods): t(17)=4.30, p=0.0005; follow-up pairwise tests: TBF-f vs TBR-f: t(17)=3.52, p=0.0026; TBF-r vs TBR-f: t(17)=-1.42; p=0.17). To furthermore characterize the temporal dynamics of the occipital alpha increase, oscillatory power changes in this cluster are depicted (Figure 3b). A cluster-based permutation statistic showed the most pronounced alpha power increase in the early time window from 0-500ms post cue (Figure 3B, p_{corr}=0.045).

In contrast, the later alpha power effect that occurred in the same time period as the item-cue similarity effect (500 to 1000ms post-cue) was localized to two source clusters in temporo-frontal regions (Figure 3C, cluster 1: MNI peak -17, 1, -36, left parahippocampal gyrus/temporal pole; t(17)=3.59, p_{corr}=0.026; cluster 2: MNI peak -5, 43, -52, left superior medial gyrus; t(17)=4.65, p_{corr}=0.047). This later alpha power increase was not specifically indexing successful active forgetting, as indicated by a weaker interaction effect that did not reach significance (interaction analysis, i.e. contrast of inhibition vs rehearsal: p=0.099, t(17)=1.74). Analysis of this later time window showed a more unspecific main effect of increased alpha power for TBF vs. TBR items independent of later memory (Figure 3B, main effect of TBF vs TBR: t(17)=2.59, p=0.019; pairwise tests: TBF-f vs TBR-f: t(17)=4.024, p=0.0008; TBF-r vs TBR-f: t(17)=1.74, p=0.099). A follow-up analysis on the temporal evolution of active inhibition-related power changes in these frontal clusters revealed that the alpha power increase was limited to the later time window (500-1000ms post-cue, Figure 3E; p_{corr}=0.0048).

Taken together, these results show that early increases in occipital alpha crucially predict successful forgetting of a memory trace. This alpha power increase preceded the reduction of item representations that we found in the ICS analysis. Thus, early alpha power increases in occipital areas seem to be critically involved in the successful suppression of memory traces. This finding is in line with the role of occipital alpha power increases in anticipatory memory inhibition [21] and active inhibition of unattended distracting information in general [22]. More specifically, the result suggests that a reduction of memory representations in ventral visual stream leads to subsequent forgetting [23]. By contrast, the additional alpha power increase in frontal regions that we observed concurrent with the ICS effect was not specific to active inhibition, but occurred generally during TBF as compared to TBR trials – possibly indicating the employment of memory control processes irrespective of their success.

Late alpha/beta power increases index selective rehearsal

Interestingly, alpha power increases did not only precede and accompany active inhibition, but also selective rehearsal. We observed an increased alpha power for actively rehearsed and remembered items (TBR-r) as compared to incidentally remembered items (TBF-r), i.e. reduced rehearsal was associated

with a decrease in alpha power. This effect occurred in a cluster that again both, preceded and concurred with the time period of the ICS selective rehearsal effect (3-dimensional cluster statistics; 8-13Hz, 1300-1700ms post cue, p_{corr}=0.03; Figure 4A). Again, we separately analyzed effects in the two time windows (1000-1500ms and 1500-2000ms post cue) and observed a similar pattern of results as for the active inhibition alpha power changes: The source of the rehearsal-related alpha power effect again switched from posterior sources during the time window preceding the ICS effect (Figure 4B, peak MNI 57,-60, -19, right inferior temporal gyrus; t(17)=-4.97, p_{corr}=0.006) to frontal sources during the ICS effect (Figure 4C, two significant clusters: cluster 1: peak MNI -26, 32, 41, right middle frontal gyrus; t(17)= -3.37, p_{corr}=0.05; cluster 2: peak MNI -47, -17, 62, left precentral gyrus; t(17)=-3.41, p_{corr}=0.05).

In both time windows, alpha power changes were not specific to selective rehearsal: Alpha power was not only increased for remembered TBR vs. remembered TBF items, as expected (Figure 4B: 1000-1500ms, TBF-r vs TBR-r, t(17)=-3.85, p=0.0013; Figure 4C: 1500-2000ms, TBF-r vs TBR-r, t(17)=-3.59, p=0.0023), no significant difference between rehearsal and inhibition effects was found (Figure 4B: 1000-1500ms, t(17)=0.66, p=0.51; Figure 4C: 1500-2000ms, t(17)=1.42, p=0.17). Instead, there was a significant main effect of TBF vs TBR items in both time windows (Figure 4B: 1000-1500s, t(17)=-4.16, p=0.00065; Figure 4C: 1500-2000ms, t(17)=-3.22, p=0.0050). Time-frequency analyses in the two significant source clusters revealed that alpha power reductions for TBF-r as compared to TBR-r items appeared in similar time intervals in posterior and frontal regions (Figure 4B&C, early effect: 2 significant clusters: p_{corr}=0.001, p_{corr}=0.018; late effect: p_{corr}=0.0003).

The relative decreases in alpha power for reduced rehearsal (relative increases in alpha power for successfully rehearsed TBR-r items) are well in line with previously described increases in alpha power during working memory maintenance [24]. Interestingly, while the oscillatory power changes related to active inhibition in occipital areas were restricted to the alpha frequency range, the selective rehearsal effect extended to higher frequencies in the beta range when focusing on the significant source clusters (Figure 3B&C).

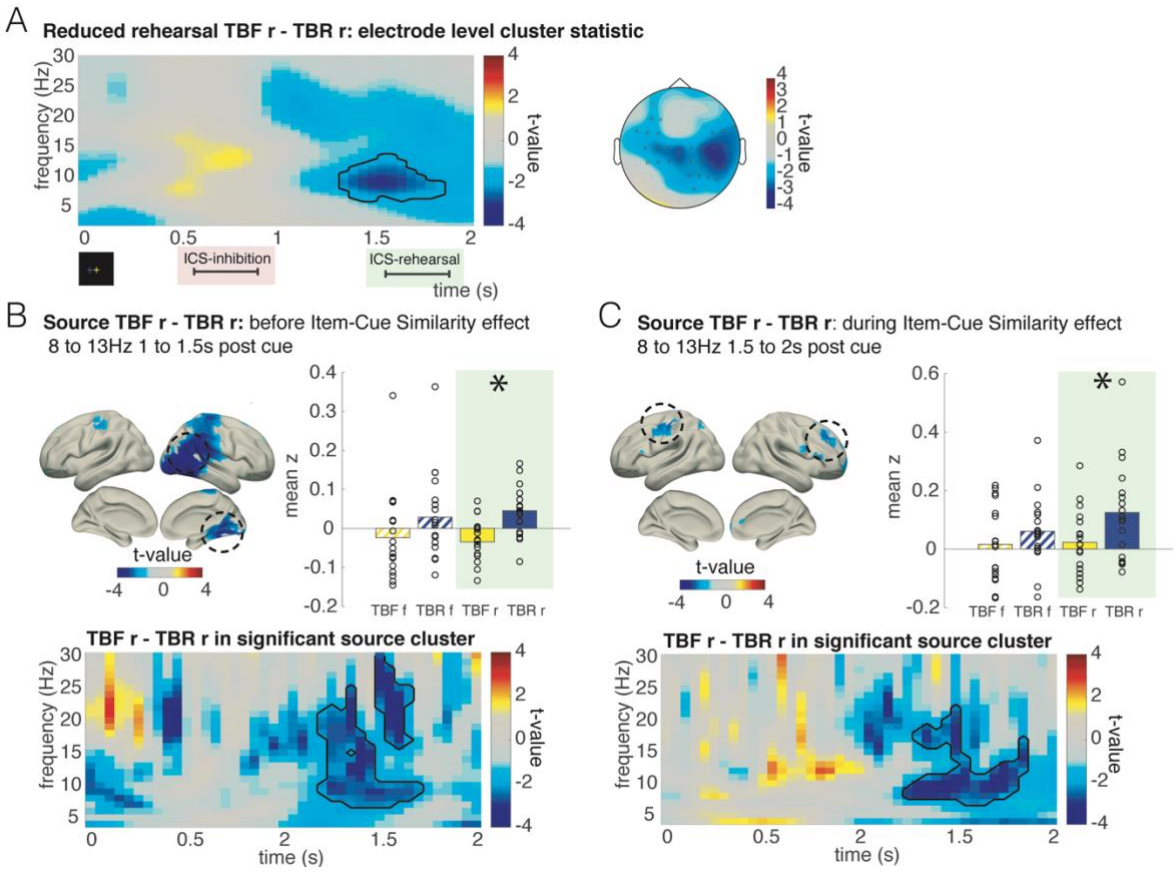


Figure 4: Brain oscillatory correlates of selective rehearsal. (A) Significant differences of oscillatory power between incidentally remembered (TBF-r) and actively rehearsed trials (TBR-r), as revealed by a cluster-based permutation test across electrodes, frequencies (2-30Hz), and time points (0-2s post cue). Left: Time-frequency plot showing average t-values across significant electrodes. Right: Topographical plot highlighting the significant electrodes. (B) Reduced oscillatory rehearsal effect preceding rehearsal-related ICS changes (time window: 1-1.5s post-cue). (C) Reduced oscillatory rehearsal effects concurrent with the rehearsal-related ICS changes (time window: 1.5-2s post-cue). Source plots show differences between TBF-r and TBR-r, thresholded at $p<0.01$, circles highlight significant clusters ($p<0.05$). Bar plots to the right depict average power in the significant source clusters, black stars indicate $p<0.05$, black circles denote single subject trial averages. Time-frequency plots below show average power changes in the respective significant source clusters, black contours highlight significant clusters ($p<0.05$).

Taken together, we found robust increases in alpha power related to both active inhibition and selective rehearsal that clearly precede the onset of item-cue similarity effects. Interestingly, alpha power increases were found during both, active inhibition and rehearsal. This finding is well in line with the proposed inhibitory function of the alpha band. We propose that alpha power increases during active inhibition reflect an active downregulation of unwanted memory traces, whereas active maintenance during selective rehearsal indicates a focus on internal representations which also requires a suppression of bottom-up information flow of other, possibly distracting information [6,25,26]. The topographical patterns showing different sources of the two early alpha effects fits in this explanation: While the early alpha inhibition effect was localized to occipital regions and thus may indicate a downregulation of visual representations, the early alpha power increases during selective rehearsal were found in right-

lateralized temporal regions. Since the alpha increase during selective rehearsal did not depend on the later memory success of TBR items, it may generally reflect an attempt to inhibit distracting inputs during active maintenance.

Active inhibition and selective rehearsal are dissociable processes contributing to directed forgetting of unwanted memory traces

Our data described so far provide evidence that directed forgetting relies on two temporally separable and consecutively employed processes, active inhibition and selective rehearsal. We next investigated whether recruitment of these two processes is inter-individually correlated, i.e. whether participants showing stronger active inhibition effects also exhibit more pronounced rehearsal effects. Alternatively, one may expect a negative relationship, i.e. a reliance on either but not both of these processes, or their independent recruitment. We correlated the respective item-cue similarity and alpha power effects. For none of these variables of interest did we observe significant correlations (item-cue similarity effects: $r_{spearman}$=-0.23, p=0.34; early alpha power inhibition and rehearsal effects: $r_{spearman}$=-0.30, p=0.22; late alpha power inhibition and rehearsal effects: $r_{spearman}$=0.05, p=0.83). This result suggests that active inhibition and selective rehearsal are indeed independently employed, as described before for the Think/No-Think paradigm, a different method to study voluntary memory suppression (Benoit and Anderson, 2012).

Intentional forgetting relies on flexible modulation of memory representations

The present results show that intentional forgetting of memories relies on two different processes, an active downregulation of unwanted memory traces followed by a reduced rehearsal of these traces. Alterations of alpha power accompanied both of these changes in memory representations, in line with the established roles of alpha oscillations for distractor inhibition and working memory maintenance [6,27]. Interestingly, a prior intracranial EEG study investigating directed forgetting found similar increases in alpha power in the hippocampus around the time window of the ICS inhibition effect described here [8]. Our results demonstrate that item representations are indeed actively inhibited at the time of these hippocampal effects. Together, these findings show that motivated forgetting attenuates memory traces, leading to their later forgetting.

We found evidence that active inhibition and selective rehearsal processes both contribute to directed forgetting, and that the processes are temporally dissociable and uncorrelated. The human brain can utilize different mechanisms to achieve the important task of motivated forgetting, an active attenuation processes as well as an active amplification mechanism.

Materials and Methods

Participants

Twenty-three healthy volunteers with normal or corrected-to-normal vision took part in the experiment. One dataset was excluded because of the age of the participant (48 years), another 4 datasets were excluded after artifact correction (less than 10 trials in one of the conditions), resulting in a final sample of 18 datasets (mean age: 23.4, age range: 18-32 years, 8 male). All subjects were right handed and reported no history of a neurologic or psychiatric disease. All participants gave their written informed consent, and the experimental protocol was approved by the ethical review board of the Faculty of Psychology at Ruhr University Bochum.

Experimental design

The experiment consisted of three parts: an encoding phase, a short distractor task (3min of backwards counting starting from random 3-digit numbers), and a recognition phase. During the paradigm, participants were seated in front of a computer screen. The experiment started with the instruction on screen and 4 practice trials. Participants were explicitly instructed that the experiment served to study the effect of forgetting on memory. In each trial a picture was presented that was followed by a colored fixation cross. Depending on the color of the fixation cross, the participants had to either rehearse and try to encode the previously presented item, or to voluntarily forget the item as these items would not be tested later. Prior to the recognition phase participants were told that they now should recognize the previously presented item irrespective of whether the item had been followed by a “forget” or “remember” instruction.

We used 288 picture stimuli depicting nameable objects from an existing database [28]. During encoding, 96 pictures were presented and followed by a TBF or TBR cue, respectively. The sequence of TBF and TBR trials was randomized. Each trial started with a fixation cross (500-1000ms jittered), then the picture was presented for 500ms, followed by another fixation cross (1500ms) and the TBF/TBR cue that remained on screen for 2000ms. During recognition, all old items (TBF&TBR) were shown mixed with 96 new items. Participants were instructed to respond to each item on a 1-6 confidence scale (i.e. 1= very sure old, 6 = very sure new). Participants used index, middle and ring fingers on both hands to respond. Each recognition trial started with a fixation cross (500-1000ms jittered), then the item was presented for 500ms, followed by another fixation cross (1500ms) and a response screen for 2000ms. During all phases, a blank screen was presented during the inter-trial interval (1250ms). Picture material, response hand and TBR/TBF cue colors were counterbalanced across the participant sample.

Behavioral Analysis

To analyze memory performance, we used a signal detection approach to obtain bias free measures of memory strength and to classify hits and misses relative to an individually defined neutral response criterion (for a similar procedure see [13,29]). As demonstrated previously, individually defining the confidence rating boundary between hits and misses enhances signal to noise ratio by taking into account individual differences in the use of confidence ratings [13].

First, the 1-6 confidence recognition responses of the participants were mapped to binary old/new judgements taking into account participants' individual response biases. Confidence ratings were defined as "old" judgements up to the highest rating that was more often used for new items than old items during recognition. According to this definition, for two participants only "1" indicated an old response, for 4 participants "1-2", for 9 participants "1-3" and for 3 participants "1-4". Memory performance was then assessed by calculating Pr, i.e. the difference between hit rate (percentage of old items remembered as old) and false alarms (percentage of new items incorrectly judged as old).

EEG recording and preprocessing

The EEG was recorded from 64 electrodes in an extended 10/20 montage (Brainamp Standard, EasyCap). Recordings were referenced to Fz and later re-referenced offline to average reference. Impedances were kept at below 10kΩ. The signals were amplified between 0.1Hz and 250Hz and recorded with a sampling rate of 500Hz.

All EEG data analyses were carried out using fieldtrip (http://www.fieldtriptoolbox.org, [30]) and custom MATLAB scripts. Data were epoched in trials from 1 second before an item to 5 seconds after item onset during encoding. Data were visually inspected to exclude trials with idiosyncratic artifacts (channel jumps, muscle artifacts, noisy channels) from further analysis. Noisy channels were excluded (in four datasets, up to three electrodes were excluded). Infomax independent component (IC) analysis was applied to correct for residual artifacts (e.g., eye blinks, eye movements, or tonic muscle activity). On average 29.39 TBF-f (range: 11-52), 52.61 TBF-r (range: 34-66), 20.11 TBR-f (range: 10-52), and 63.11 TBR-r trials (range: 30-83) passed artifact corrections.

Representational similarity analysis of item-cue similarity

Item-cue similarities were calculated by correlating EEG activity during item presentation (0-500ms) with EEG activity post-cue presentation (2000-4000ms relative to item, 0-2000ms post-cue) within each trial. To this end, artifact-corrected raw data was downsampled to 100Hz. To remove condition-specific ERP related activity, all trials were normalized using the mean and standard deviation across trials for each data point and channel, separately for each condition (TBF-f, TBF-r, TBR-f, TBR-r). EEG data

was cut into windows of 200ms in overlapping time windows with 10ms increment. These 200ms bins were then concatenated for each electrode, resulting in vectors with combined spatial and temporal information (vector dimensions: 21 time points x 64 electrodes). Item time x channel vectors during item presentation were then correlated with item time x channel vectors during the cue period using Spearman correlations. This results in item-cue similarity matrices for all combinations of item time bins and cue time bins.

For statistical testing, a cluster-based permutation approach was used to accommodate for potential biases due to different trial numbers in the different conditions and to correct for multiple comparisons. To this end, the trial labels were shuffled in each subject for 1000 times. This random data was then used to construct null distributions of effects under the existing bias in trial number. In a second, group statistical step, clusters of temporally adjacent significant differences (threshold $p<0.01$) were identified and the sum of t-values in each cluster was calculated in the original data and in the data based on the random permutations. If no t-value reached significance in one of the permutations, a cluster value of 0 was assigned. Significance of clusters was assessed by calculating the rank of the cluster t-values in the distribution of random data. A cluster was interpreted as significant if an absolutely higher cluster t-value was found in less than 5% of the random permutations.

EEG oscillatory power analysis

Data were filtered using wavelets with a length of 5 cycles to obtain oscillatory power between 2 and 30Hz. Single trial power values were z-transformed using frequency- and channel-specific means and standard deviations across all trials.

Source analysis was performed using a linearly constrained minimal variance (LCMV) beamformer [31], calculating a spatial filter based on the whole length of all trials. For all subjects, we used a standard boundary element source model with a grid resolution of 10mm based on the Montreal Neurological Institute (MNI) brain and standard electrode positions realigned to the MNI MRI. The source time-course for each grid point was calculated, subjected to a wavelet analysis (same settings as for the raw data) and z-transformed as for the electrode-level data. Statistical analysis was restricted to grid voxels inside aal-defined brain regions. For region of interest analyses, data across all grid voxels covering the region of interest (significant source) were averaged. Grid voxel data were interpolated to a 2mm resolution single-subject MNI brain for plotting and to define the locations of clusters and peaks.

For statistical analysis of EEG data, power spectra for each subject were collapsed and averaged across all trials for each “cell” of the design matrix (TBF-f, TBF-r, TBR-f, TBR-r). All of the following EEG analyses were based on these first-level averages (4 cells, 18 subjects). This prior averaging of data within each cell for each subject controls for possible biases of trial numbers in the analysis of main effects (e.g., there were more TBF-f trials than TBR-f trials, and thus, if all TBR trials irrespective of

memory would be pooled, condition effects would be confounded by subsequent memory differences). This analysis is essentially equivalent to a 2x2 repeated measurements ANOVA.

For statistical analysis, we again used nonparametric cluster-based permutation tests as implemented in fieldtrip [32]. The cluster-based permutation test consisted of the following two steps: first, clusters of coherent t-values exceeding a certain threshold (here, $p<0.01$) along selected dimensions (time, frequency, electrodes/grid voxels) were detected in the data. Second, summed t-values of these clusters were compared to a null distribution of t-sums of random clusters obtained by permuting condition labels across subjects. This procedure effectively controls for type I errors due to multiple testing. The clusters of t-values subjected to permutation testing can be built across different dimensions: clustering can be performed on non-averaged data across all dimensions (electrode, frequency, time) or in a specific dimension when averaging over certain dimensions (i.e., averaging in the time-frequency window and then clustering across the electrode dimension). Clustering was employed along different dimensions depending on the data. In order to investigate power changes on the source level, power values of significant source clusters were extracted, averaged across significant voxels, and subjected to further statistical analysis. In order to correlate the active inhibition and selective rehearsal power effects across subjects, average power in the reported peak voxel of the effect was extracted.

Supplemental Material

ICS control analysis

To ensure that the ICS effects indeed reflect reactivation of item-specific information, we conducted several control analyses. A possible confound of the reported ICS interaction effects reported in Figure 2C is that effects might not only reflect modulations of item-specific representations, but instead general condition-specific activity differences such as ERPs, power changes, or other unspecific differences.

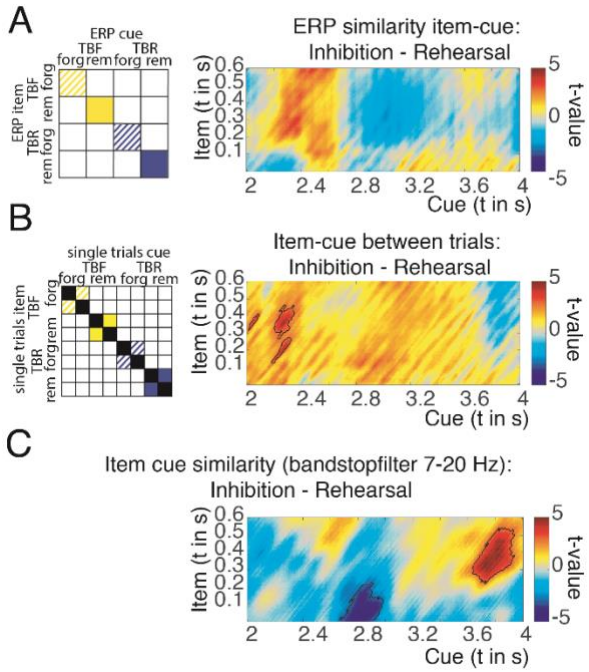
We first tested for a possible influence of ERPs on the reported ICS effects. ERPs were previously reported to show condition differences during directed forgetting [33]. These differences in average condition-specific waveforms reflecting attentional and mnemonic processes could potentially influence the ICS measure, which could substantially change the interpretation of our effects: ICS effects would then reflect condition-specific differences in neural activity and not up- or downregulations of item-specific memory traces. However, additional analyses demonstrated that ERP effects cannot explain the item-cue similarity pattern that we had observed: Conducting the same interaction analysis as reported in Figure 2C using trial-averaged ERPs rather than single-trial data did not yield any significant clusters. For this analysis single subject, condition specific ERPs were calculated and subjected to the same item-cue-correlation analysis (Supplementary Figure 1A). More importantly, the pattern of ERP similarity values showed no comparable increases and decreases as observed during the ICS interaction analysis. The pattern of (non-significant) effects is actually opposing the ICS effects: numerically positive differences were found early after the cue followed by numerically negative differences late post-cue (Supplementary Figure 1A). This control analysis shows that the ICS effects cannot be attributed to univariate condition differences.

A second, related confound is whether the reported effects – even if they rely on single trials rather than trial-averaged ERPs – are indeed attributable to item-specific representations or whether they reflect differences that are independent on item identity. If the reported ICS effects are indeed attributable to item-specific memory traces, these effects should be specific to correlation of activity during one item window and activity during the matching cue window (i.e., within-item correlation); they should not be present when correlating activity during one item window with activity during a non-matching cue window (i.e., between-item correlation; see model matrix in Supplementary Figure 1B). Item specificity is often ensured by calculating the difference of within-item correlations and between-item correlations [34,35]. As within-item correlations (i.e., matching item and cue) in our data are necessarily within-trial correlations, and between-item correlations are across trials, directly calculating this difference leads to biased results, as within trial correlations are contrasted with between trial correlations. However, we tested whether similar effects as observed within-items occur also between-items. We repeated the same analysis as reported in Figure 2C based on between-item correlations and tested for the reported interaction effects. Correlations of non-matching item-cue intervals were calculated and again the

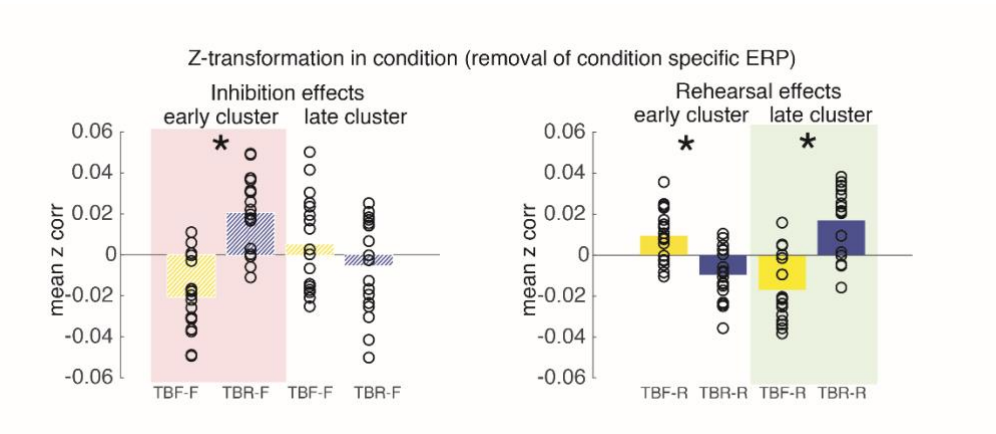
interaction between TBF/TBR x memory was assessed. This analysis yielded a very different pattern of results than the within-trial item-cue similarity analysis (Supplementary Figure 1B). Most importantly, the results are non-overlapping with the reported interaction effect (Figure 2C). Early after the cue, a significant positive effect was evident, an effect that is in stark contrast to the reported negative cluster in the within-item ICS analysis. This result provides further evidence that the reported effects are indeed attributable to an up- or downregulation of item-specific representations.

Third, we tested whether our ICS results were directly driven by the reported differences in alpha/beta power between conditions (see main Figure 2C). We thus repeated our main analysis of ICS after applying a bandstop-filter between 7-20Hz. This analysis yielded very similar results as our main analysis (Supplemental Figure 1C), with the exception of lacking aliasing effects that are caused by (usually prominent) alpha oscillations.

Fourth, we ensured that our normalization procedure did not induce any spurious effects. As described in the Experimental Procedures, we applied a z-transformation to the EEG activity at every time point across all trials of all conditions in order to remove ERP components prior to calculating correlations. As a control analysis, item-cue similarity was calculated using data that were not z-transformed across conditions but only within conditions (Supplementary Figure 2). This approach yielded the same pattern of results.



Supplementary Figure 1: Control analysis of item-cue similarity. To ensure that item-cue similarity differences presented in Figure 2C were not confounded by item-unspecific differences between conditions, several control analyses were carried out. (A) The interaction effect as in Figure 1C was calculated not based on single trials, but correlating ERPs during item presentation with ERPs during cue presentation in each condition. This analysis revealed a very different result than in the original analysis. (B) The same analysis was also conducted between item and cue periods of different trials (rather than within the same trials), again revealing a very different result. (C) The exact same interaction analysis as in Figure 1C was repeated based on data with an alpha 8-20Hz band stop filter, to ensure that alpha power changes were not driving the effects. We obtained similar results as in the original analysis. Contours highlight significant clusters ($p<0.05$).



Supplementary Figure 2: Control analysis for normalization procedure. To ensure that different normalization steps prior to correlations did not influence our results, the same analysis as presented in Figure 2 D&E was reproduced without employing condition-specific normalization prior to the item cue similarity analysis.

References

1. Basden BH, Basden DR, Gargano GJ. Directed forgetting in implicit and explicit memory tests: A comparison of methods. J Exp Psychology Learn Mem Cognition. 1993;19: 603. doi:10.1037/0278-7393.19.3.603

2. Anderson MC, Hanslmayr S. Neural mechanisms of motivated forgetting. Trends Cogn Sci. 2014;18: 279–292. doi:10.1016/j.tics.2014.03.002

3. Kriegeskorte N, Kievit RA. Representational geometry: integrating cognition, computation, and the brain. Trends in Cognitive Sciences. 2013;17: 401–12. doi:10.1016/j.tics.2013.06.007

4. Muther SW. Erasure or partitioning in short-term memory. Psychonomic Science. 1965;3: 429–430. doi:10.3758/bf03343215

5. Jensen O, Bonnefond M, Marshall TR, Tiesinga P. Oscillatory mechanisms of feedforward and feedback visual processing. Trends in Neurosciences. 2015;38: 192–194. doi:10.1016/j.tins.2015.02.006

6. Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences. 2012;16: 606–17. doi:10.1016/j.tics.2012.10.007

7. Tulving E, Pearlstone Z. Availability versus accessibility of information in memory for words. J Verb Learn Verb Be. 1966;5: 381–391. doi:10.1016/s0022-5371(66)80048-8

8. Oehrn CR, Fell J, Baumann C, Rosburg T, Ludowig E, Kessler H, et al. Direct Electrophysiological Evidence for Prefrontal Control of Hippocampal Processing during Voluntary Forgetting. Current Biology. 2018;28: 3016-3022.e4. doi:10.1016/j.cub.2018.07.042

9. Wylie GR, Foxe JJ, Taylor TL. Forgetting as an Active Process: An fMRI Investigation of Item-Method–Directed Forgetting. Cereb Cortex. 2008;18: 670–682. doi:10.1093/cercor/bhm101

10. Lu Y, Wang C, Chen C, Biology XG. Spatiotemporal neural pattern similarity supports episodic memory. 2015;

11. Sols I, DuBrow S, Davachi L, Fuentemilla L. Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. Current Biology. 2017;27: 3499-3504.e4. doi:10.1016/j.cub.2017.09.057

12. Wang TH, Placek K, Lewis-Peacock JA. More is less: increased processing of unwanted memories facilitates forgetting. The Journal of Neuroscience. 2019; 2033–18. doi:10.1523/jneurosci.2033-18.2019

13. Fellner M-C, Bäuml K-HT, Hanslmayr S. Brain oscillatory subsequent memory effects differ in power and long-range synchronization between semantic and survival processing. Neuroimage. 2013;79: 361–370. doi:10.1016/j.neuroimage.2013.04.121

14. Detre GJ, Natarajan A, Gershman SJ, Norman KA. Moderate levels of activation lead to forgetting in the think/no-think paradigm. Neuropsychologia. 2013;51: 2371–88. doi:10.1016/j.neuropsychologia.2013.02.017

15. Wenzlaff RM, Wegner DM. Thought Suppression. Annu Rev Psychol. 2000;51: 59–91. doi:10.1146/annurev.psych.51.1.59
16. Levy BJ, Anderson MC. Purging of Memories from Conscious Awareness Tracked in the Human Brain. The Journal of Neuroscience. 2012;32: 16785–94. doi:10.1523/jneurosci.2640-12.2012
17. Waldhauser GT, Braun V, Hanslmayr S. Episodic Memory Retrieval Functionally Relies on Very Rapid Reactivation of Sensory Information. The Journal of Neuroscience. 2016;36: 251–60. doi:10.1523/jneurosci.2101-15.2016
18. van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, et al. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proceedings of the National Academy of Sciences. 2014;111: 14332–14341. doi:10.1073/pnas.1402773111
19. Zumer JM, Scheeringa R, Schoffelen J-M, Norris DG, Jensen O. Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. PLoS biology. 2014;12: e1001965. doi:10.1371/journal.pbio.1001965
20. Bastos AM, Loonis R, Kornblith S, Lundqvist M, Miller EK. Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. Proceedings of the National Academy of Sciences of the United States of America. 2018;115: 1117–1122. doi:10.1073/pnas.1710323115
21. Park H, Lee DS, Kang E, Kang H, Hahm J, Kim JS, et al. Blocking of irrelevant memories by posterior alpha activity boosts memory encoding. Human brain mapping. 2014;35: 3972–87. doi:10.1002/hbm.22452
22. Bonnefond M, Jensen O. Alpha oscillations serve to protect working memory maintenance against anticipated distracters. Current biology : CB. 2012;22: 1969–74. doi:10.1016/j.cub.2012.08.029
23. Wimber M, Alink A, Charest I, Kriegeskorte N, Anderson MC. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. Nature neuroscience. 2015;18: 582–9. doi:10.1038/nn.3973
24. Jensen O, Gelfand J, Kounios J, Lisman JE. Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cerebral cortex. 2002;12: 877–882. doi:10.1093/cercor/12.8.877
25. Myers NE, Stokes MG, Nobre AC. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. Trends Cogn Sci. 2017;21: 449–461. doi:10.1016/j.tics.2017.03.010
26. Sadaghiani S, Kleinschmidt A. Brain Networks and α-Oscillations: Structural and Functional Foundations of Cognitive Control. Trends in cognitive sciences. 2016; doi:10.1016/j.tics.2016.09.004
27. Miller EK, Lundqvist M, Bastos AM. Working Memory 2.0. Neuron. 2018;100: 463–475. doi:10.1016/j.neuron.2018.09.023
28. Moreno-Martínez F, Montoro PR. An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. Plos One. 2012;7:

e37527. doi:10.1371/journal.pone.0037527
29. Hanslmayr S, Spitzer B, Bäuml K-H. Brain Oscillations Dissociate between Semantic and Nonsemantic Encoding of Episodic Memories. Cereb Cortex. 2009;19: 1631–1640. doi:10.1093/cercor/bhn197
30. Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci. 2011;2011: 156869. doi:10.1155/2011/156869
31. Veen BD, Drongelen VW, Yuchtman M, Suzuki A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Transactions on Biomedical Engineering. 1997;44: 867–880. doi:10.1109/10.623056
32. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Meth. 2007;164: 177–190. doi:10.1016/j.jneumeth.2007.03.024
33. Paz-Caballero DM, Menor J, Jiménez JM. Predictive validity of event-related potentials (ERPs) in relation to the directed forgetting effects. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology. 2004;115: 369–77.
34. Xiao X, Dong Q, Gao J, Men W, Poldrack RA, Xue G. Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. Journal of Neuroscience. 2017;37: 2986–2998. doi:10.1523/JNEUROSCI.2324-16.2017
35. Michelmann S, Bowman H, Hanslmayr S. The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. Plos Biol. 2016;14: e1002528. doi:10.1371/journal.pbio.1002528