Neural dynamics of retrieval suppression in abolishing item-specific cortical pattern of unwanted emotional memories

Xuanyi Lin[1], Danni Chen[1], Ziqing Yao[1],

Michael C. Anderson[2], Xiaoqing Hu[1]

1, Department of Psychology, The State Key Laboratory of Brain and Cognitive Sciences University of Hong Kong

2, MRC Cognition & Brain Sciences Unit, Behavioural and Clinical Neuroscience Institute, University of Cambridge

Correspondence

Michael C Anderson
MRC Cognition & Brain Sciences Unit, University of Cambridge,
Email: michael.anderson@mrc-cbu.cam.ac.uk

Xiaoqing Hu
Department of Psychology, The University of Hong Kong,
Email: xiaoqinghu@hku.hk

Summary

When reminded of an unpleasant experience, people often try to exclude the unwanted memory from awareness in an effort to forget it, a process known as retrieval suppression. Yet, how fast can individual memories be targeted and controlled, and the neural dynamics in modulating cortical traces of individual memories, remain elusive. Here, using multivariate decoding analyses on time-domain and time-frequency-domain EEGs, we found that retrieval suppression of aversive memories was distinct from retrieval and passive viewing, when given a reminder. Specifically, early elevation of mid-frontal theta power during the first 500 ms distinguished retrieval suppression from passive viewing, suggesting that suppression recruited early active control processes. On an item-level, we could discern activities relating to individual memories during active retrieval - initially, based on perceptual responses to reminders (0-500 ms) and later, via the reinstatement and maintenance of the target aversive scenes (500-3000 ms). Critically, suppressing retrieval significantly weakened (during 420-600 ms) and eventually abolished these item-specific cortical patterns till cue disappeared (1200-3000 ms), suggesting the successful exclusion of the unwelcome memory from awareness. Suppression of item-specific cortical patterns bore behavioral consequences in predicting subsequent episodic forgetting. These findings provide unique insight into the neural dynamics underlying the control of unwelcome memories: upon perceiving an unwelcome reminder, people rapidly deploy inhibitory control to truncate retrieval within 500 ms, which likely terminate the reminder-to-memory conversion at around 500 ms that would ordinarily arise through hippocampal pattern completion. We concluded that both rapid and sustained control are critical in abolishing cortical patterns of individual memories, limiting unwelcome awareness, and precipitating later forgetting.

Introduction

Following a painful or traumatic event, the past may come back to mind uninvitedly. Oftentimes, even seemingly innocuous objects can remind us of the trauma, triggering intrusive images, fear and avoidance behaviors. When this happens, people tend to recruit top-down control processes to terminate unwelcome retrieval, a process known as retrieval suppression. Ideally, control processes need to be fast and to target individual memories well before they fully unfold in our mnemonic awareness. However, the precise timing and neural dynamics of retrieval suppression in weakening individual memories remains elusive. Specifically, how fast can people stop retrieving a specific memory, and how timely retrieval suppression contributes to successful forgetting?

Employing techniques bearing millisecond temporal resolution including M/EEG, intracranial EEG and single-unit recording in humans, research on cued memory recall suggests a staged cued recall process: initially, a memory reminder undergoes perceptual analysis within 500 ms. After the perceptual information reaches the hippocampus, pattern completion processes occur at around 500 ms, driving cortical reinstatements of the target memory during the 500-1500 ms time window [1]. In particular, successful recall was associated with enhanced encoding-retrieval neural pattern similarities in the hippocampus during 1000-1500 ms [2]. Moreover, neural firing in the hippocampus preceded and predicted spikes in the adjacent entorhinal cortex (EC) between 500-1500 ms, during which memory traces can be identified [3]. These results provide strong evidence suggesting that hippocampus pattern completion and subsequent cortical reinstatement occur during the 500-1500 ms time window, supporting successful episodic recall.

While these findings delineate how a simple reminder gives rise to vivid remembering, there are scenarios when retrieval is unwelcome and needs to be stopped. Neuroimaging evidence suggests that when seeing cues of unwanted memories, the prefrontal cortex exerts top-down inhibitory control over the hippocampus and the medial temporal lobes to down-regulate unwanted retrieval [4, 5]. Furthermore, retrieval suppression weakens neural activities in the neocortex that are implicated in reinstating original memories [5-8]. However, it remains elusive exactly *when* top-down control processes weaken individual memories. To achieve timely control of unwanted memories, we hypothesize that inhibitory control needs to be engaged during or before the time window when cue-to-memory conversion occurs, i.e., within 500 ms after the cues. This early inhibitory control process should disrupt pattern completion to prevent the full-blown recollection experience, via weakening and eventually abolishing item-specific cortical patterns during the memory reinstatement time window (500-1500 ms) and till the memory cue disappears.

Despite EEG's unparalleled temporal resolution, its relatively poor spatial resolution posits particular challenges in isolating item-specific memory representations. To tackle this challenge, we applied a relatively new multivariate pattern analysis method to scalp EEGs [9] when participants voluntarily retrieve or suppress the retrieval of unwanted memories (aversive scenes) in an emotional think/no-think paradigm [5, 10, 11]. Using data from all EEG sensors at once, we first applied multivariate EEG analysis to distinguish between different conditions, to identify suppression-related activity and its time course. In particular, we not only compared voluntary retrieval (Think) versus retrieval suppression (No-Think), but also compared think/no-think manipulations with a perceptual baseline condition, in which no retrieval was involved. Pairwise condition-level decoding could unravel neural dynamics associated with retrieval and retrieval suppression, relative to the no-retrieval

baseline. Driven by our research question, we are particularly interested in the role of early frontal theta within the first 500 ms, given frontal theta power increase has been related to top-down inhibitory control processes [12-14].

To examine time-dependent evolution of item-specific neural representations, we next employed item-level multivariate pattern analyses within each condition. We then compared item-level decoding in the think condition with that from the perceptual baseline, no retrieval condition, to establish cue-to-memory processing during voluntary retrieval. Afterwards, we delineated how item-specific cortical patterns may evolve during retrieval suppression, particularly focusing on the 0-500 ms and the 500-1500 ms windows. We focused on theta activity during the early time window, given its roles in sensory intake and feedforward information flow originating from the sensory cortex [15, 16]. For memory reinstatement, we examined alpha activity given it has been implied in working memory maintenance and reinstatement [17, 18]. Comparing time-dependent evolution of item-specific cortical patterns between retrieval and retrieval suppression conditions, we could establish the timeline of inhibitory control in truncating individual memory traces.

To anticipate, we found that for successful forgetting, retrieval suppression enhanced early control and attenuated item-specific cortical patterns within the first 500 ms, probably disrupting the perception-to-memory conversion processes. Retrieval suppression then weakened and abolished item-specific cortical patterns during the 500-1500 ms memory reinstatement window in a sustained manner. In contrast, less successful forgetting was associated with insufficient mobilization of early control, and relapse of the unwanted memory during retrieval suppression.

Results

Suppressing Retrieval Induces Forgetting of Emotional Memories

Following the emotional Think/No-Think (TNT) task, participants completed a cued recall test during which they verbally described the aversive scene that they thought was linked to that cue. We coded and scored verbal descriptions of negative emotional scenes on *Identification, Gist and Detail*. Each of the three memory scores was submitted to a one-way repeated-measure (Think, No-Think and Baseline) analysis of variance (ANOVA). Results showed a significant condition effect on *Identification* $F(1.87,72.93) = 7.35$, $p = .002$, $\eta_p^2 = .159$; *Detail* ($F(1.93,75.2) = 13.79$, $p < .001$, $\eta_p^2 = .261$ and *Gist* ($F(1.92,74.95) = 6.22$, $p = .004$, $\eta_p^2 = .138$). Planned contrasts showed that, confirming our hypotheses, participants showed significant below-baseline, suppression-induced forgetting on *Identification*, $t(39) = -2.07$, $p = .045$, $dz = 0.33$, and *Details*, $t(39) = -2.16$, $p = .037$, $dz = 0.34$; whereas the forgetting effect on *Gist* was not significant $t(39) = -1.58$, $p = .123$, $dz = 0.25$, see Figure 1B).

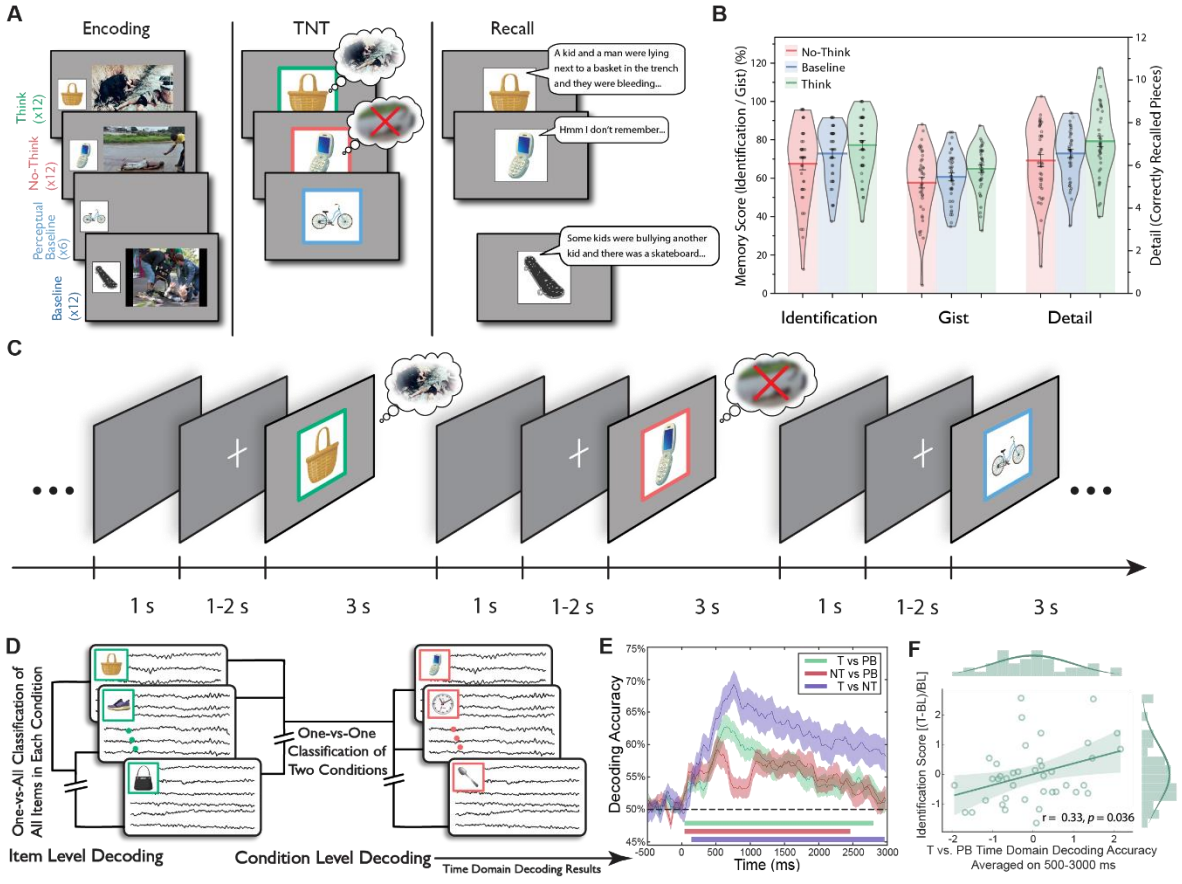

Figure 1. Experimental Procedure, Suppression-Induced Forgetting, Decoding Approaches and Condition-level Time-domain EEG Decoding Results.
(A) The emotional Think/No-Think task (eTNT) included three sessions. 1) Encoding: Participants first learnt object-aversive scene stimuli pairings; and they also viewed object without any scene pairings (i.e., Perceptual Baseline); 2) Think/No-Think (TNT): Participants either retrieve (Think) or suppress the retrieval (No-Think) of negative scene memories. Participants were also presented with Perceptual Baseline trials without any memory retrieval; 3) Cued Recall: Participants viewed object cues and verbally described their associated scenes.
(B) Suppression-Induced Forgetting on *Identification*, *Gist* and *Detail* from the Cued Recall.
(C-D) An illustration of trial flow in the EEG-based eTNT, and the logic of decoding analyses.
(E) Condition-level decoding based on time domain EEG revealed significant differences in all three pairwise comparisons. Colored disks along x-axis indicate significant clusters (permutation cluster corrected): No-Think

vs Perceptual Baseline, 40-2460 ms, $p_{corrected} < .001$; Think vs Perceptual Baseline, 40-2800 ms, $p_{corrected} < .001$; Think vs No-Think, 140-2960 ms, $p_{corrected} < .001$. Shaded areas indicate standard errors of the mean (S.E.M). (F) Time domain Think vs. Perceptual Baseline decoding accuracies during 500-3000 ms was positively correlated with subsequent memory recall on Identification score (Think normalized by Baseline).

Stopping Retrieval is Distinct From Not-Retrieving

Retrieval suppression impaired unwanted memories, inducing forgetting of the aversive scenes and their episodic details. We next examined EEG activities that could distinguish between No-Think, Think, and Perceptual Baseline (i.e., no-retrieval) conditions. Based on time-domain EEGs, condition-level multivariate decoding analysis successfully distinguished retrieval from no-retrieval (T vs. PB, $p_{corrected} < .001$, Figure 1E, green). Confirming participants' engagement in retrieving scene memories during Think trials, the Think vs. Perceptual Baseline decoding accuracies during 500-3000 ms predicted retrieval-induced facilitation in *Identification*: r = 0.33, p =.036; and *Detail*: r = 0.33, p =.041 (Figure 1F, also see Figure S2A). Furthermore, multivariate decoding not only distinguished retrieval suppression from voluntary retrieval (NT vs. T, $p_{corrected} < .001$, Figure 1E, purple), but also from perceptual baseline in which no cue-elicited retrieval was involved (NT vs. PB, $p_{corrected} < .001$, Figure 1E, red). Strikingly, the Think vs. No-Think differences started as early as 140 ms and persisted during the entire epoch until ~3000 ms.

Similarly, when using time-frequency-domain EEGs, between-condition decoding revealed significant differences among all pairwise comparisons (Figure 2A-F). Seeking EEG evidence for an early, active control process, we found that within the first 500 ms, significant NT vs. PB decoding was driven by 4-8 Hz theta activities over the frontal and posterior brain regions (Figure 2E, 2H), which continued throughout the 3000 ms. Since 500 ms, 9-15 Hz alpha/beta activities drove significant condition-level decoding performances till 3000 ms in addition to theta (Figure 2D-F).

More specifically, retrieval suppression (vs. retrieval or no-retrieval) enhanced midline and right prefrontal theta activity between 200-400 ms after the onset of the cue (NT > T, $p_{corrected} = .007$, Figure 2I; NT > PB, $p_{corrected} = .002$, Figure S1G). After this early theta enhancement, retrieval suppression reduced theta and alpha/beta power from 500 to 3000 ms (NT < T, theta: $p_{corrected} = .004$; alpha/beta: $p_{corrected} < .001$; NT < PB, theta: $p_{corrected} < .001$; alpha/beta: $p_{corrected} = .002$, Figure S1A-F). Specifically, during the 1000-2000 ms time window that was selected based on a recent study [17], we found that alpha-based decoding accuracies (No-Think vs. Perceptual Baseline) significantly predicted suppression-induced forgetting on *Identification* (r = -0.34, p =.034, Figure 2G, also see Figure S2B). This negative correlation, together with significant NT vs. PB decoding, suggest that reduced alpha power during 1000-2000 ms contributed to forgetting. Intriguingly, while alpha-based NT vs. PB decoding accuracies predicted suppression-induced forgetting, alpha-based T vs. PB decoding accuracies predicted retrieval-induced facilitation, with the difference being significant (*Detail*: z = 2.06, p =.039; Figure S2C). Together, these evidences suggested that early theta power elevation and subsequent theta/alpha power reduction supported active suppression that is distinct from not-retrieving.

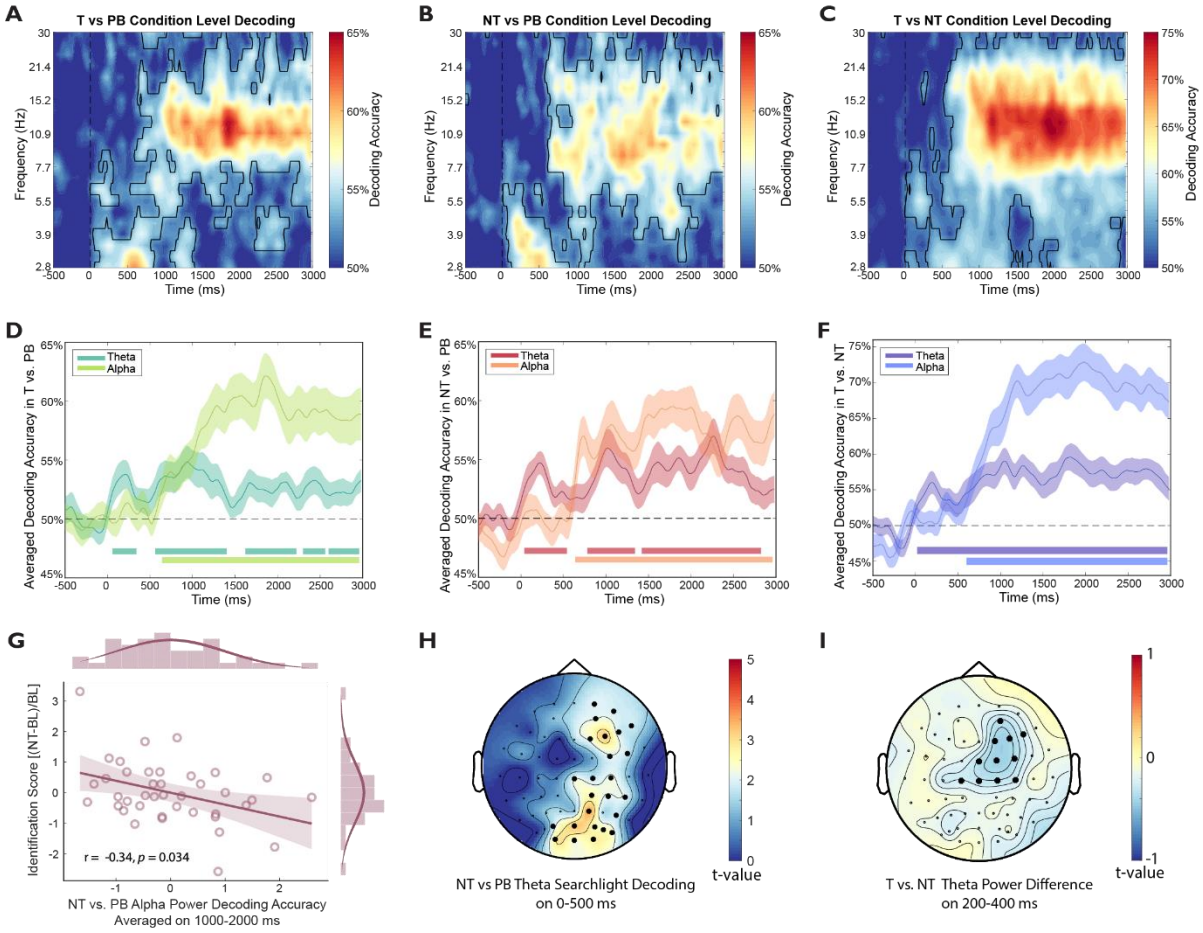

Figure 2. The Condition-Level Time-Frequency Domain Decoding

(A-C) Condition-level time-frequency decoding results. Frequency is log scaled with colorbar denoting decoding accuracy. Black outline highlights significant clusters against chance level (both cluster and permutation α are set at 0.05).

(D-F) Decoding accuracies in A-C are averaged on theta (4-8 Hz) and alpha (9-12 Hz) bands. Disks at the bottom denote significant clusters of averaged accuracy against chance level (50%) with permutation correction.

(G) The alpha-based No-Think vs. Perceptual Baseline decoding accuracies during 1,000-2,000 ms negatively predicted subsequent memory recall.

(H) Theta power on 0-500 ms distinguished NT vs. PB over frontal and posterior brain regions in channel searchlight decoding. Significant electrodes were cluster corrected and highlighted.

(I) Theta power averaged on 200-400 ms was higher in NT than T. The increased theta power showed a frontal-central distribution. Significant electrodes were cluster corrected and highlighted.

Spatial Patterns in EEG Discern Individual Memories During Retrieval

Whereas significant condition-level decoding indicates that retrieval, retrieval suppression and no-retrieval engaged distinct spatial-temporal EEG patterns, it remains unknown whether the scalp distribution of EEGs can discern individual memory traces of object-aversive scene pairings. We approached this question by conducting multivariate item-level decoding analyses in each condition, respectively.

We first sought to establish whether scalp-EEG patterns can distinguish among individual items during retrieval. Indeed, time-domain EEG significantly distinguished between individual memories across the entire 0-3000 ms window (Figure 3A, $p_{corrected} < .001$). In sharp contrast, for Perceptual Baseline trials, above-chance decoding arose only in the 0-500

ms (to be precise, 60-640 ms, $p_{corrected} < .001$), but not in the subsequent 500-3000 ms time window (Figure 3C). To directly compare item-level decoding between retrieval and no-retrieval, we repeated the analyses with 6 randomly sampled items from the Think condition, to match the item number in Perceptual Baseline (see Methods). We found that Think trials showed significantly higher item-level decoding accuracies than Perceptual Baseline trials during 360-1180 ms ($p_{corrected} < .001$) and 1220-1540 ms ($p_{corrected} = .022$, Figure 3K, purple disks).

Given that participants learnt about object-scene pairings during Think trials but not during Perceptual Baseline trials, the pre 360 ms significant item-level decoding across both Think and Perceptual Baseline trials may reflect visual-perceptual processing of object cues. In the later 360-1540 ms window, significantly higher decoding during Think trials than during Perceptual Baseline trials may reflect reinstatement and maintenance of distinct unpleasant scenes in mnemonic awareness. Another possibility, however, is that item-level decoding in the Think condition may simply reflect sustained attention to the unique object cues.

To disambiguate these two possibilities, we examined brain regions giving rise to above-chance decoding in Think trials using searchlight decoding (see Methods). Results showed that during the early 0-500 ms time window, occipital EEGs primarily drove the significant decoding, suggesting that visual-perceptual processing of the cue was the basis for item distinction (Figure 3D). In contrast, during the subsequent 500-3000 ms, significant decoding involved the contributions of a distributed set of regions implicated in memory retrieval such as the right prefrontal and parietal-occipital cortex (Figure 3E). This finding suggests that decoding beyond the first 500 ms may be dominated not by object cue perception, but rather the reinstatement of the associated scene memories implicated by the involvement of frontal-parietal-occipital network. Consistent with this explanation, item-level decoding performance during the latter 500-3000 ms time window predicted *Detail* measure of scene memory ($r = 0.34$, $p = .034$, Figure 3J), whereas the early 0-500 ms time window did not ($r = 0.01$, $p = .946$).

In Perceptual Baseline trials, the same searchlight analysis showed that significant 0-500 ms decoding arose over a small cluster of occipital electrodes, which suggested that the classification relied on visual object processing (Figure 3H). In contrast, during the 500-3000 ms time window, no significant decoding was found at any electrode (Figure 3I, note that similar searchlight results were obtained when using 0-360 and 360-1540 ms time windows, see Figure S3A).

In sum, during retrieval, the spatial patterns of time domain EEG showed a staged cued-recall processing: during 0-500 ms, EEGs could discern perceived items over occipital region; during 500-3000 ms, EEGs could distinguish among retrieved items over fronto-parietal-occipital regions, with the item-level decoding accuracies predicting memories only in this later, 500-3000 ms time window.

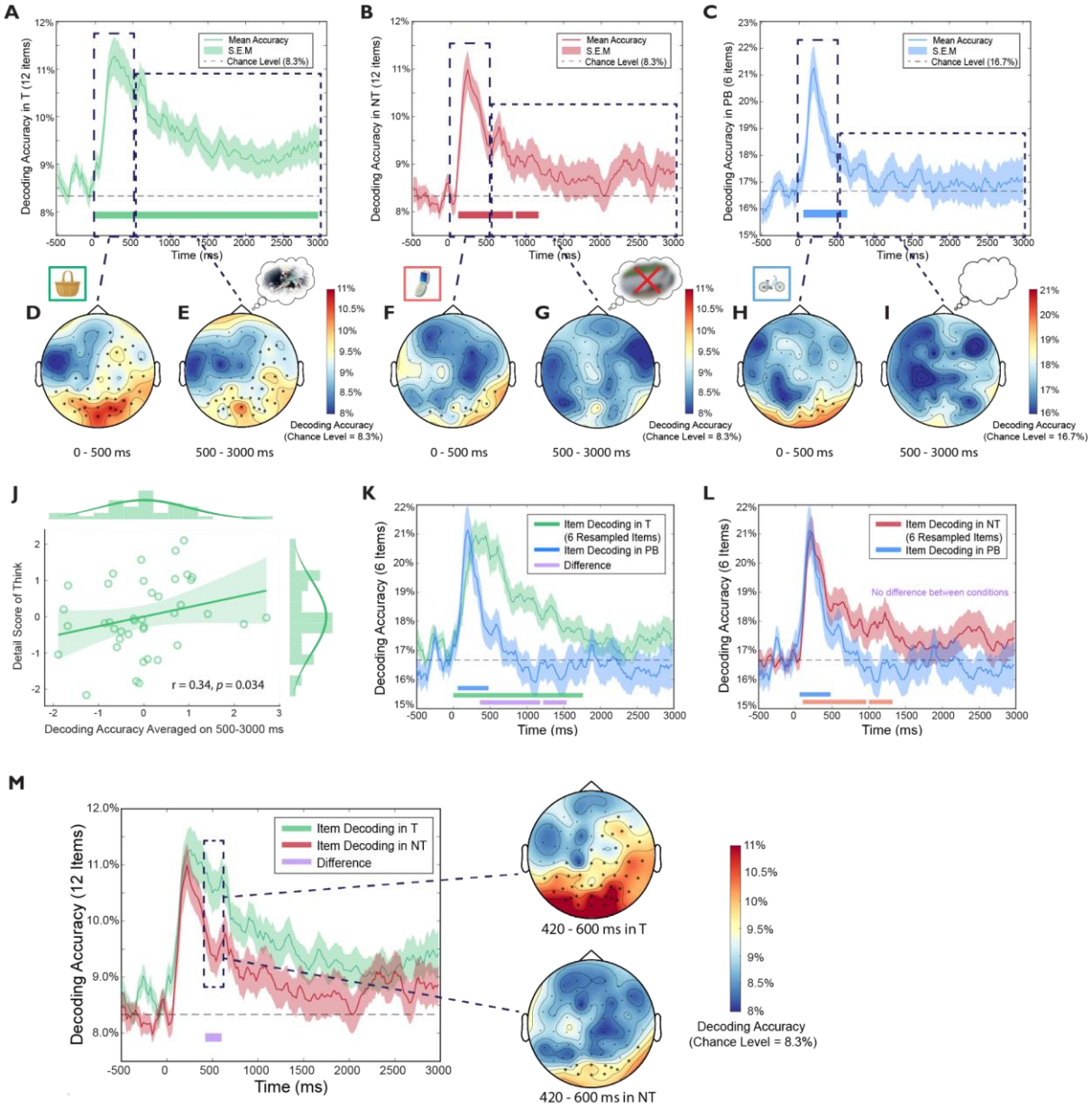

Figure 3. Item-level Time Domain Decoding

(A-C) The item-level decoding patterns (averaged across participants) in each retrieval condition. Disks at the bottom indicate significant time clusters against chance level, with permutation cluster correction (αs = 0.05).

(D-I) Channel searchlight analyses of time domain decoding during an early (0-500 ms) and a later time window (500-3000 ms). Colorbar indicates decoding accuracy. Electrodes with significant decoding accuracies are highlighted (permutation cluster corrected, αs = 0.05).

(J) During Think trials, decoding accuracies averaged on 500-3000 ms predicted details of recalled emotional scenes.

(K) Item-level decoding in Think (using 6 resampled items) is higher than Perceptual Baseline on 360-1180 ms, $p_{corrected} < .001$; 1220-1540 ms, $p_{corrected} = .022$. Disks at the bottom indicate cluster-corrected significant time clusters against chance level (green and blue for Think and Perceptual Baseline) or difference between the two conditions (purple).

(L) Item-level decoding in No-Think (using 6 resampled items) is not significantly different from Perceptual Baseline. Disks at the bottom indicate significant time clusters against chance level (red and blue for No-Think and Perceptual Baseline).

(M) Retrieval suppression significantly reduced item-level decoding accuracies on 420-600 ms compared to retrieval, with the right panel showing channel searchlight analyses on this time window.

Suppressing Retrieval Weakens and Abolishes Item-specific Cortical Patterns

Building on these results, we next examined our key question: when and how does retrieval suppression modulate item-specific cortical EEG pattern?

Examining time domain item-level decoding patterns in the No-Think trials revealed that decoding accuracy was significantly above-chance until 1160 ms ($ps_{corrected} < .028$). However, item-level decoding accuracies then became non-significant till 3000 ms when the cue disappeared. Consistent with Think and Perceptual Baseline analyses, we used *a priori* defined time window 0-500 vs. 500-3000 ms to examine EEGs scalp distributions that contributed to decoding. We found that during 0-500 ms, item-level decoding was driven by occipital region activities, which resembled scalp distributions of Perceptual Baseline EEGs during the same 0-500 ms window (Figure 3F, 3H). During the subsequent 500-3000 ms, no brain regions played a significant role in item-level decoding (Figure 3G).

In addition to scalp EEG distributions revealed by channel searchlight, confusion matrices of item-level decoding provided consistent evidence supporting the hypothesized staged retrieval suppression: while we found significant above-chance classifications among items in all three conditions during the 0-500 ms time window, distinctive classification patterns only remained in the Think condition in later time windows (Figure S3C-E).

To gain a more precise understanding of the neural dynamics in suppressing individual memories, it is crucial to compare time-dependent evolution of item-specific cortical patterns between retrieval suppression and retrieval/no-retrieval conditions. A direct Think vs. No-Think comparison of item-level decoding revealed that the retrieval suppression significantly reduced decoding accuracies on 420 to 600 ms ($p_{corrected} < .05$, Figure 3M left panel). Searchlight channel analyses during 420-600 ms revealed that, while voluntary retrieval engaged brain activities over frontal-parietal-occipital regions, retrieval suppression was only associated with occipital activity (Figure 3M right panel). When No-Think was directly compared to Perceptual Baseline (using 6 randomly sampled items from the No-Think condition), there were no significant differences in terms of item-level decoding during the entire 0-3000 epoch (none of the differences survived permutation correction, see Figure 3L).

Linking weakened item-level decoding with the early active control processes, we found that in the No-Think (vs. Think) trials, reduction of item-level decoding during 420-600 ms was preceded by enhanced 200-400 ms theta power over midline and right prefrontal cortex (Figure 2I). Critically, theta power elevation across this region were positively correlated with the 420-600 ms decoding accuracy reduction ($r = 0.30$, $p = .064$, Figure S3F), suggesting that higher theta power (No-Think > Think) was associated with lower item-specific decoding accuracies (No-Think < Think).

Together, beyond the active suppression evidence found on condition level, these item-level decoding results revealed a precise timeline on how retrieval suppression unfolded: inhibitory control was engaged within the first 500 ms upon encountering a cue object, presumably before the cue-to-memory conversion process to obstruct retrieval, resulting in a weakened and eventually abolished memory-specific cortical pattern during 500-3000 ms.

Rapid and Sustained Suppression Led to Successful Epidosic Forgetting

To understand how timing of suppression contributed to subsequent forgetting, we divided our participants into *High*- vs. *Low-Suppression Groups* based on the median of NT-minus-BL *Detail* scores (we used *Detail* given that it is a continuous measure, see Methods).

We then compared the item-level decoding accuracies between Think and No-Think in *High-Suppression Group* (Figure 4A). This comparison revealed that significant reductions of decoding during No-Think (vs. Think) trials emerged on two time windows: during 300-680 ms ($p_{corrected}$ = .006) and 1140-1400 ms ($p_{corrected}$ = .031). These differences may reflect the early top-down disruption of cue-to-memory conversion process around 500 ms, and the later weakening of item-specific cortical reinstatement patterns between 1000-1500 ms. In contrast, the same comparison in the *Low-Suppression Group* revealed no significant NT vs. T difference (Figure 4B), suggesting comparable item-level decoding efficiencies between retrieval and retrieval suppression in this group. Corroborating the putative role of early and timely suppression in forgetting, we found that item-level decoding accuracy during 300-680 ms was correlated with subsequent suppression-induced forgetting across all participants (r = 0.35, p = .027, Figure 4C), suggesting that the more effectively the participants suppressed unwanted memories during 300-680 ms, the more likely suppression would cause later forgetting.

We next compared item-level decoding between No-Think (using 6 randomly sampled items) and Perceptual Baseline, in the *High*- and *Low-Suppression Groups* respectively. While no differences emerged in the *High-Suppression Group* (Figure 4D), we found that low-suppression participants showed significantly higher item-level decoding accuracies in No-Think trials than Perceptual Baseline trials during 2300-2560 ms ($p_{corrected}$ = .029, Figure 4E, purple dashed outline). Thus, less successful forgetting was associated with relapses during sustained control of unwanted memories. Together, these results provided intriguing evidence that both early rapid, and later sustained control may be necessary in successful forgetting.

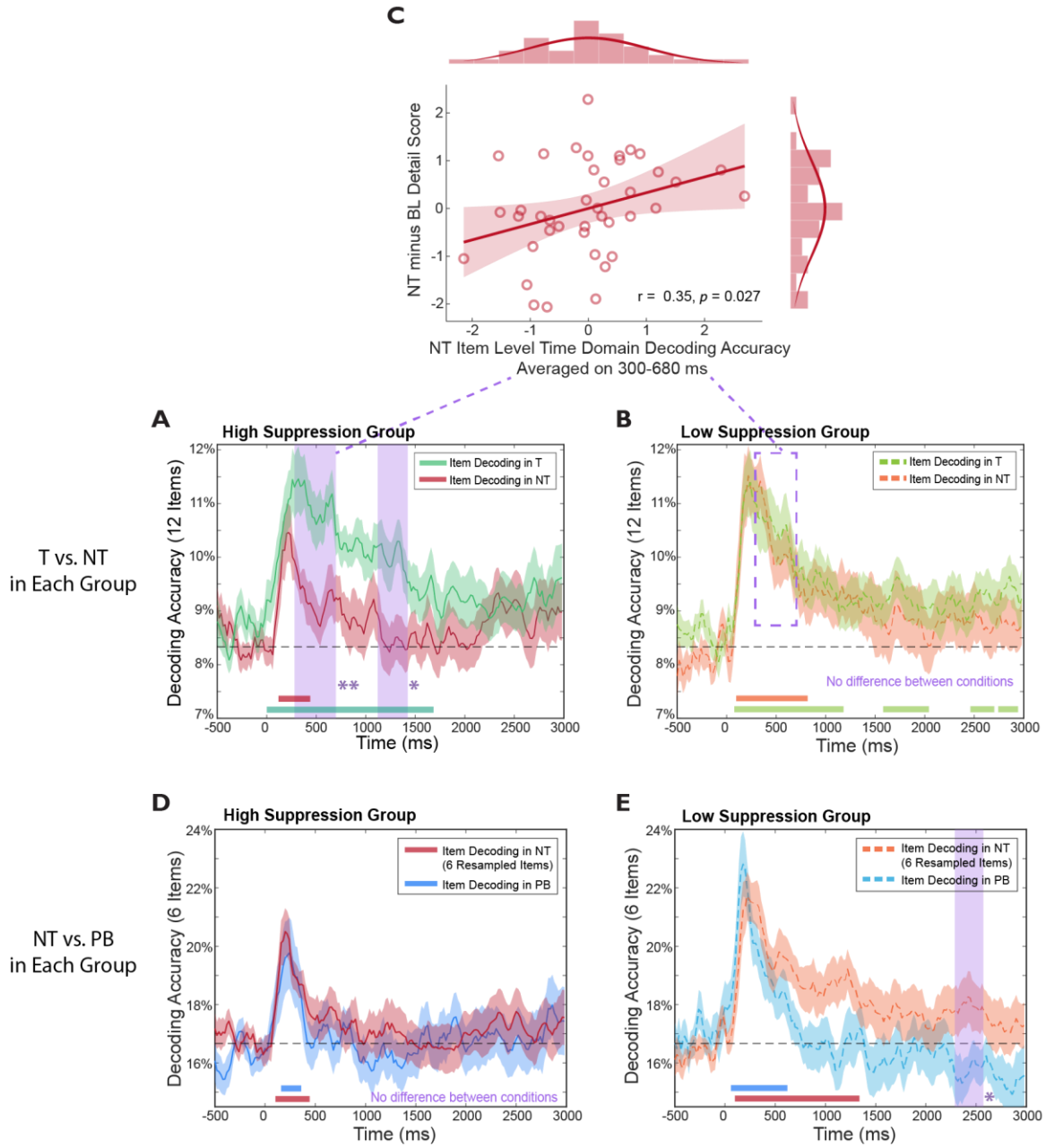

Figure 4. Item-level Decoding Results in *High-* and *Low-Suppression Group*

(A, B) Comparisons between Think and No-Think item-level decoding in High-/Low-Suppression Group, respectively. In the *High-Suppression Group*, Think vs. No-Think difference was significant on 300-680 ms and 1140-1400 ms, while no difference was found in the *Low-Suppression Group*.

(C) Across both groups, the averaged decoding accuracy on 300-680 ms positively correlated with participant's suppression-induced forgetting, i.e. No-Think minus Baseline Detail score.

(D, E) Resampled item-level decoding comparisons between No-Think and Perceptual Baseline in High- and Low-Suppression Group, respectively. In the High-Suppression Group, No-Think did not differ from Perceptual Baseline in item-level decoding accuracy, despite both showing above chance decoding within 0-500 ms. In the Low-Suppression Group, a significant difference between No-Think and Perceptual Baseline was observed on 2300-2560 ms.

Color disks at the bottom of each figure denote time clusters significantly above chance (permutation corrected, one-sided αs = 0.05). Purple dashed outlines denote significant time clusters between conditions/groups (permutation corrected, two-sided αs = 0.05).

Theta and Alpha Oscillations Track Item-Level Perception and Reinstatement Processes, Respectively

While theta and alpha/beta activities were associated with top-down retrieval suppression vs. retrieval and no-retrieval, it remains unclear how retrieval suppression modulates item-specific EEG activity. We found that within all three conditions, theta activity during 0-500 ms significantly distinguished among individual items ($ps_{corrected} < .001$, Figure 5 A-C, also see Figure 5 D-F). Channel searchlight analyses during 0-500 ms revealed that significant decoding was driven by theta activity over the occipital cortex, suggesting theta's role in visual processing of individual items (Figure 5G). During 500-3000 ms, we found that both theta and alpha power drove significant above-chance decoding during voluntary retrieval (theta: $ps_{corrected} < .027$; alpha: $ps_{corrected} < .039$, Figure 5D), but not during retrieval suppression (Figure 5E). There was short-lived late theta-driven decoding in Perceptual Baseline trials, which may reflect occasional perceptual processing of object cues (theta: $ps_{corrected} < .011$, Figure 5F). Channel searchlight analyses during 500-3000 ms revealed that *alpha* activity over the posterior regions contributed to decoding performance only in Think, but not in the other conditions (see Figure 5H), further suggesting that alpha activity is linked with item-specific memory reinstatement processes. Hence, the lack of significant alpha-based decoding in No-Think might reflect a suppression-induced abolition of reinstatement. Together, on an item level, occipital theta and posterior alpha activities may support visual sensory intake and memory reinstatement, respectively.

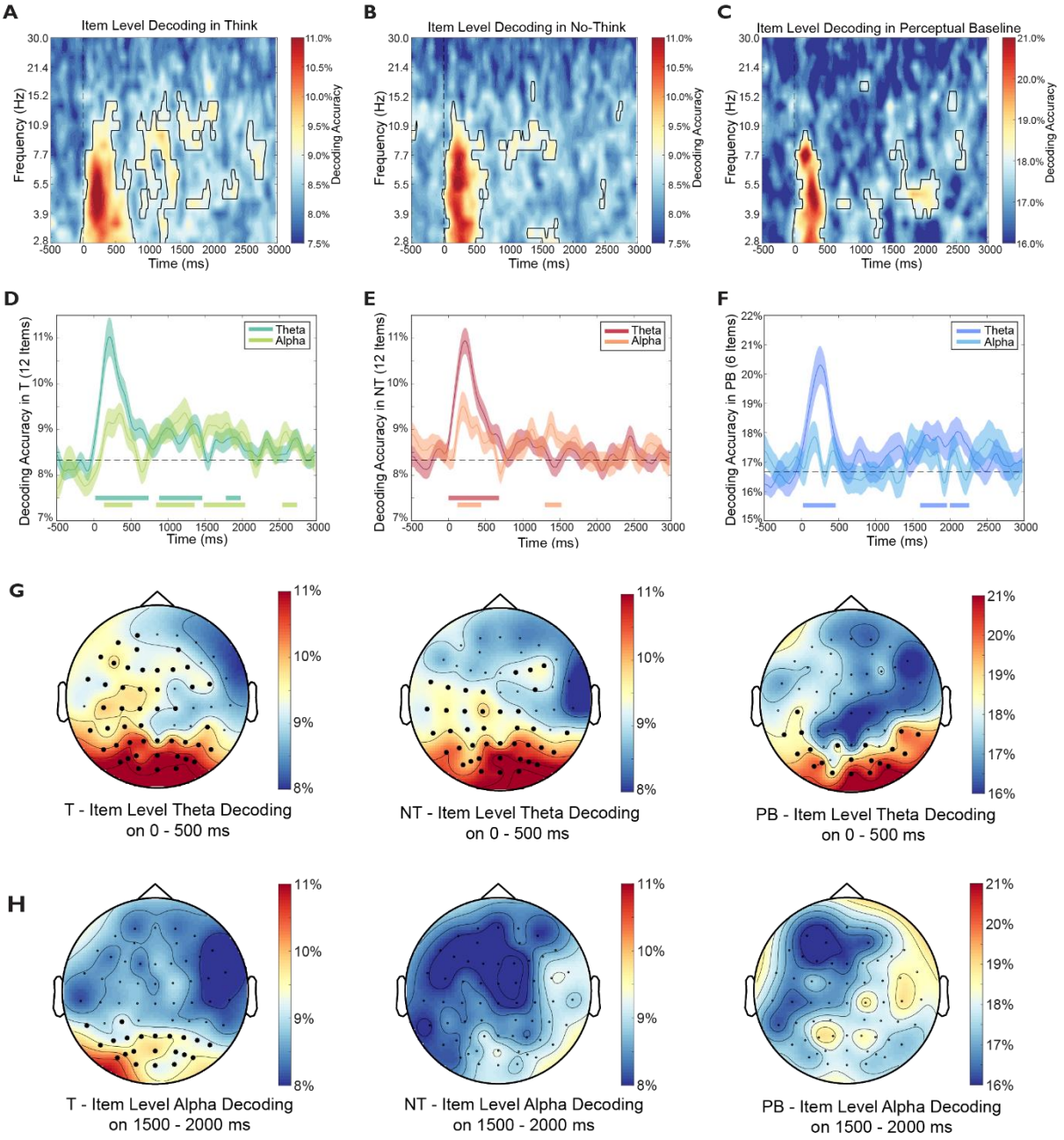

Figure 5. The Item-level Time-Frequency Domain Decoding

(A-C) Item-level time-frequency decoding results. Frequency is log scaled and colorbar denotes decoding accuracy. Black outline highlights significant clusters against chance levels (both cluster alpha and permutation α are 0.05, one-sided).

(D-F) Decoding accuracies in A-C are averaged on theta and alpha bands. Disks denote significant clusters of the band-averaged accuracies against chance level (cluster corrected, one-sided αs = 0.05).

(G) Item-level theta searchlight on 0-500 ms showed an occipital distribution in all three conditions. Significant channels are highlighted (permutation cluster corrected with one-sided αs = 0.05).

(H) Item-level alpha searchlight on 1500-2000 ms showed that only in Think was alpha power able to distinguish among items. The alpha searchlight decoding in Think originated from the posterior region. Significant channels are highlighted (permutation cluster corrected with one-sided αs = 0.05).

Discussion

Oblivion can be a blessing: suppressing unwanted memories frees our minds from troubling past, facilitates subsequent learning and planning, and promotes resilience following trauma exposure [6-8, 19]. Particularly, avoiding retrieval of an unwelcome memory requires effort; it is not simply neglecting to engage an optional retrieval process when an unwelcome reminder appears, but rather requires an active inhibition mechanism that countermands automatic retrieval [14]. Ideally, the act of inhibition should happen rapidly, especially when automatic retrieval of the unwelcome memory is to be prevented. Our results support this idea: active forgetting requires 1) rapid deployment of inhibitory control and suppression of individual memory traces within the first 500 ms and 2) sustained control in weakening and abolishing item-specific cortical EEG patterns during 500-3000 ms.

Integrating unparalleled temporal resolution afforded by EEG, and enhanced spatial resolution offered by multivariate analyses, we provide three pieces of evidence suggesting an early, active control process was critical in truncating retrieval of highly specific, individual memories. First, on a condition level, spatial EEGs distinguish retrieval suppression from no-retrieval within the first 500 ms, with significant decoding performance contributed by enhanced midfrontal and right prefrontal theta activity during retrieval suppression. Given the well-established evidence linking frontal theta and inhibitory control processes, [12] [13], this result provides convergent evidence that retrieval suppression engaged early inhibitory control processes in the first 500 ms upon seeing an unwelcome memory cue. Substantiating theta's putative role in the early top-down inhibitory control, we found that this 200-400 ms frontal theta power elevation predicted subsequent reduction of item-level decoding accuracies during 420-600 ms, which we discuss below.

Second, retrieval suppression (vs. retrieval) significantly weakened item-level decoding during 420-600 ms. This result supports our hypothesis that retrieval suppression would disrupt the perception-to-memory conversion process at around 500 ms, when hippocampus-dependent pattern completion would otherwise occur. Indeed, given that hippocampus-dependent pattern completion would trigger reinstatement of target memories in the neocortex and give rise to vivid recollection [1, 20, 21], effective retrieval suppression should precisely target this process to truncate retrieval and limit mnemonic awareness from occurring.

Third, an early attenuation of item-specific cortical pattern was associated with later forgetting. Specifically, among High-Suppression participants, they showed significantly reduced item-level decoding accuracies (No-Think < Think) on 300-680 ms. Such reduction was not observed among low-suppression participants. Across all participants, reduction of No-Think item decoding accuracies within 300-680 ms time window were correlated with suppression-induced forgetting. This result provides direct evidence supporting the critical role of an early suppression effect for participants to forget unwanted memories. Given that the hippocampus-dependent pattern completion processes occur at around 500 ms [1], this finding also suggests that for successful forgetting, top-down inhibitory control shall be engaged well before episodic memories reinstatement during 500-1500 ms.

Examining time-dependent evolution of item-specific cortical patterns suggests that not only early, rapid control is important, but also sustained control is necessary for successful forgetting. While retrieval suppression significantly weakened item-specific cortical patterns starting from 400 ms, individual memories could still be identified till 1200 ms. Reduced yet still above-chance item-specific cortical patterns during 420-1200 ms may call for sustained

control processes to implement goal-directed suppression, supported by reduced condition-level alpha/beta power during later stages of retrieval suppression. Item-specific cortical patterns were eventually weakened to be indiscernible from 1200 ms, till the cue disappeared at 3000 ms. Together, these temporal characteristics revealed a fine-grained timeline in retrieval suppression of aversive scenes: early control processes truncated retrieval during the perception-to-memory conversion time window (e.g., ~420-600 ms), with sustained control processes down-regulating unwanted memories (e.g., ~1200 ms) and eventually abolishing item-specific cortical patterns (1200-3000 ms).

Intriguingly, during the time window of 500-3000 ms when cortical reinstatement would ordinarily occur and give rise to full-blown memories, two specific time windows bore relevance with active forgetting. First, during 1140-1400 ms, among high-suppression participants, retrieval suppression (vs. retrieval) significantly reduced item-level decoding accuracies. Second, during 2600-2800 ms, low-suppression participants showed an ironic rebound effect: retrieval suppression was associated with significantly higher decoding accuracies than no-retrieval perceptual baseline trials. This rebound effect suggests that participants who later showed less successful forgetting had relapses in controlling unwanted memories, particularly towards the end of retrieval suppression [22]. These results suggest that successful forgetting roots in sustained suppression of individual memories during the cortical reinstatement time window.

Our item-level decoding results of voluntary retrieval (i.e., during Think trials) provides further support to the staged model of cued memory recall. To rule out the possibility that sustained item-level decoding during retrieval may simply reflect sustained attention devoted to each individual object cue, we showed that 1) the early (0-500 ms) vs. late (500-3000 ms) decoding patterns are characterized by distinct spatial distributions of EEGs, and 2) only the 500-3000 ms decoding accuracy predicts retrieval-induced memory facilitation. These results suggest that the early vs. late decoding patterns reflect perceptual vs. retrieval processes, respectively. Consistent with these results, both theta and alpha/beta power contributed to item-level decoding throughout the entire epoch during voluntary retrieval, with an early onset of occipital theta activity followed by parietal-occipital alpha/beta activity. Theta and alpha/beta activity may reflect sensory intake [15, 16], hippocampo-cortical communication loops that support perception-to-memory conversion and neocortex-dependent memory reinstatement processes [1, 3]. Decoding patterns of Perceptual Baseline trials provided clear, additional support for this account: when participants viewed object cues that lacked any associated memory, decoding was significant only in the early 0-500 ms time window and was driven by occipital theta activity, ruling out any contribution of scene retrieval.

In addition to theta, we found that different retrieval conditions significantly modulated alpha power during the extended 500-3000 ms time window. Previous research showed that memory reinstatements are tightly associated with alpha oscillations. For example, Fellner, et al. [17] reported that alpha power increased during 1000-2000 ms following to-be-remember cues, which were associated with selective rehearsal [see also 23, 24, 25]. Consistent with these studies, we found that voluntary retrieval enhanced alpha power during the same 1000-2000 ms when memory reinstatement would be expected (Figure S1H-M). In contrast, retrieval suppression reduced alpha power and abolished alpha-based item-level decoding performance, presumably reflecting abolished memory reinstatement [23, 26]. Critically, between-condition (retrieval suppression vs. no-retrieval) alpha-based decoding accuracies predicted subsequent forgetting, highlighting the critical role of alpha power reduction in suppression-induced forgetting. Together, reduced theta/alpha power and abolished item-

level decoding during 1000-2000 ms suggested that retrieval suppression engaged active control processes to disrupt the feedforward/feedback cross-region information flow that would otherwise support cue-to-memory conversion and cortical reinstatement as in voluntary retrieval.

Collectively, we showed that for successful forgetting, top-down retrieval suppression needs to be fast and sustained: early frontal theta disrupted cue-to-memory conversion and truncated individual memory traces within the first 500 ms upon seeing the cues, preventing the aversive scenes from being fully reinstated in mnemonic awareness. Sustained control would then weaken and eventually abolish item-specific cortical EEG patterns during 500-3000 ms, supported by reduced alpha activity. In contrast, poor deployment of early control and relapses during sustained control resulted in less successful voluntary forgetting. By offering new insights into the precise timing and neural dynamics of retrieval suppression in modulating individual memories, our results may inform future research on when and how to intervene along the information processing stream to help people forget unwanted memories and have a spotless mind.

Methods

Experimental Subject Details

Participants

41 participants (mean age = 19.57, age range: 18-23 years, 26 females) were recruited from The University of Hong Kong. One participant was excluded due to non-compliance of task instructions (details see ***Materials and Procedure***). Ethical approval was obtained from the Human Research Ethics Committee of The University of Hong Kong.

Method Details

Materials and Procedure

We used 42 object-scene picture pairs from Küpper, et al. [11]. Scenes depict aversive contents such as natural disasters, assault, injury, etc. Each object resembled an item from its paired negative scene, thus establishing naturalistic and strong associations. Six pairs were used for instruction and practice purposes. The remaining 36 pairs were equally divided into 3 sets, with 12 pairs in each of three following conditions: Think, No-Think, and Baseline. Picture pairs used in the three conditions were matched on valence and arousal, and were counterbalanced across participants. Another 6 objects without any paired scenes were used as Perceptual Baseline trials, which did not involve any memory retrieval. Participants completed the following sessions in order: Encoding, Think/No-think (TNT) and Cued Recall. Participants also completed a 3-item, instruction compliance questionnaire at the end of the TNT session (see the OSF for the questionnaire).

Encoding: Participants were presented with 42 object-scene pairings, plus 6 objects from Perceptual Baseline. Each object-scene pair was presented on an LCD monitor for 6 s with an inter-trial-interval (ITI) of 1 s. Participants were instructed to pay attention to all the details of each scene, and to associate the left-sided object and the right-sided scene. They were then given a test-feedback session, in which each object was presented up to 4 s until participants pressed a button indicating whether they could recall the associated scene or not. If participants gave a ‘yes’ response, they were presented with three scenes from the learning phase and needed to identify the correct one within another 4 s. Regardless of accuracy, the correct pairing would be presented again for 2.5 s. This test-feedback cycle repeated until participants reached 60% accuracy. Twenty-six participants reached this criterion in the first

cycle, 13 participants in two, and 1 in three. Following the test-feedback cycles, participants were given a recognition-without-feedback test, to assess their memory before the TNT session. Items from different conditions were encoded at comparable levels (*p*s > .104).

TNT: Participants were presented with 24 objects from the 36 object-scene pairings, with 12 objects in each of the Think or No-think conditions, respectively. The remaining 12 objects were not shown in the TNT and would be in the Baseline condition. These 24 objects were presented in either yellow- or blue-colored frames indicating think and no-think conditions, with colors counterbalanced across participants. The 6 objects (without any pairing scenes) were presented in white-colored frames and served as Perceptual Baseline trials. Thus, 30 unique objects were shown in the TNT session. Each object was presented for 10 times, resulting in a total of 300 trials. Each trial began with a fixation cross (2-3s), followed by the object in a colored frame for 3s. The ITI was 1 s.

For Think trials, participants were instructed to try their best to think about the objects' associated scenes in detail, and to keep the scenes in mind while the objects remained on the monitor. For No-Think trials, participants were given direct-suppression instructions: they were told to pay full attention to the objects while refraining from thinking about anything. If any thoughts or memories other than the objects come to mind, they need to try their best to push the intruding thoughts/memories out of their mind and re-focus on the objects. Participants were also prohibited from using any thought substitution strategies (i.e., thinking about a different scene). For Perceptual Baseline trials, participants were simply instructed to focus on the object.

Cued Recall: Following the TNT session, participants were presented with each of the 36 objects from *Think*, *No-Think* and *Baseline* conditions. Each object was presented at the center of the monitor, alongside a beep sound prompting participants to verbally describe the associated scenes within 15 s. The ITI was 3 s. Participants' verbal descriptions were recorded for later scoring. Perceptual Baseline objects were not shown in this recall test because they were not paired up with any scenes.

Cued Recall Analyses: Two trained raters who were blind to experimental conditions coded each of the verbal descriptions along three dimensions following the criteria used in Küpper et al., 2014, namely *Identification*, *Gist* and *Detail*. Each measure focused on different aspects of memories: Identification referred to whether the verbal description was clear enough to correctly identify the unique scene, and was scored as 1 or 0. Inconsistent ratings were resolved by averaging 0 and 1, resulting in a score of 0.5. Gist measured whether participants could correctly describe the scene's main themes, and was scored on how many correct gists were given. Detail measured how many correct meaningful segments were provided during the verbal description, and was scored on the number of details. Interrater agreement for the scoring of all three measures was high: Identification r = 0.71, Gist r = 0.90, Detail r = 0.86.

EEG Recording and Preprocessing: Continuous EEGs were recorded during the TNT session using ANT Neuro eego with a 500 Hz sampling rate (ANT B.V., Enschede, The Netherlands), from 64-channel ANT Neuro Waveguard caps with electrodes positioned according to the 10-5 system. The AFz served as the ground and CPz was used as the online reference. Electrode impedances were kept below 20 kilo-ohms before recording. Eye movements were monitored through EOG channels.

Raw EEG data were preprocessed in MATLAB using EEGlab Toolbox [27] and ERPlab Toolbox [28]: data were first downsampled to 250 Hz, and were band-passed from 0.1 to 60 Hz, followed by a notch filter of 50Hz to remove line noise. Bad channels were identified via visual inspection, and were removed and interpolated before re-referencing to common averages. Continuous EEG data were segmented into -1000 to 3500 ms epochs relative to the cue onset, and baseline corrected using -500 to 0 ms as baseline period. Next, independent component analyses (ICAs) were implemented to remove eye blinks and muscle artifacts. Epochs with remaining artifacts (exceeding ± 100 μV) were rejected. The numbers of accepted epochs used in all following analyses were comparable across *Think* (Mean ± SD, 100.33 ± 11.57) and *No-think* (103.18 ± 10.61) conditions. Valid trials number in *Perceptual Baseline* is 56.58 ± 3.23. All EEG analyses were based on 61 electrodes, excluding EOG, M1, M2, AFz (ground) and CPz (online reference).

Condition-/Item-level Decoding with Time Domain EEG: Decoding analyses were conducted in MATLAB using scripts adapted from [9], which used a support vector machine (SVM) and error-correcting output codes (ECOC). The ECOC model combined results from several binary classifiers for prediction output in multiclass classification.

In condition-level decoding, we used one-vs-one SVMs to perform pairwise decoding among the three conditions (Think vs. Perceptual Baseline, No-Think vs. Perceptual Baseline, and Think vs. No-Think). For Think vs. Perceptual Baseline and No-Think vs. Perceptual Baseline condition-level decoding, we first subsampled trials in T/NT to be comparable with Perceptual Baseline so that each condition had about 56 trials. We next divided EEG trials from each condition into 3 equal sets and averaged EEG epochs within each set into sub-ERPs to improve signal-to-noise ratio. The decoding was achieved within each participant from -500 to 3000 ms using these sub-ERPs in a 3-fold cross validation: each time 2 of the 3 sub-ERPs are used as training dataset with the condition labels, and the remaining one was used as testing dataset. After splitting training and testing datasets, sub-ERPs were both normalized using the mean and standard deviation of training dataset to remove ERP-related activity. This process was conducted on every 20 ms time point (subsampled to 50 Hz), and repeated for 10 iterations. We were comparing condition-level decoding accuracy against its chance level, 50%, given two conditions were involved in each pairwise decoding.

For item-level decoding, we used one-vs-all SVMs to decode each individual stimulus within each condition, separately. Decoding procedures were the same as condition-level decoding. Thus, the trial numbers of each stimulus are first matched to the least one within each participant (at most 10 trials, if no trial was rejected). Then, all trials of each stimulus were divided into 3 sets before averaging and the 3-fold cross validation. Both training dataset and testing dataset were normalized using the mean and standard deviation of training dataset. The decoding process was conducted on every 20 ms time point and for 10 iterations (results remained the same for up to 100 iterations, see supplementary Figure S3G). For Think and No-Think conditions, the chance levels were 1/12 (8.33%) given that there were 12 unique stimuli in each of these two conditions. For *PERCEPTUAL BASELINE* trials, the chance level was 1/6 (16.67%).

Given we have different item numbers in Perceptual Baseline (6 items) and Think/No-Think (12 items), in order to directly compare the decoding accuracy in Think or No-Think with Perceptual Baseline, we conducted a resampled decoding in Think and No-Think, respectively. The resampled decoding is similar to the normal decoding, except that during each iteration we randomly selected 6 out of all 12 items before dividing and averaging into 3

sets. Considering the randomization used only half of the items, we increased iterations to 20 times. An item-level decoding with 20-iterations was also rerun in Perceptual Baseline, to be compared with the resampled decoding.

Condition-/Item-level Decoding with Time-Frequency Domain EEG: Time domain EEG was wavelet transformed into time-frequency domain data in Fieldtrip Toolbox [29] before decoding. Frequencies of interest increased logarithmically from 2.8 Hz to 30 Hz, resulting in 22 frequency bins. Wavelet cycles increased linearly along with frequencies from 3 to 7. Then the decoding was conducted for each frequency bin data across time in the same procedure as described in **Channel-/Item-level Decoding with Time Domain EEG** (as if treating each frequency bin data as a time domain data).

Channel Searchlight Decoding: Both condition- and item-level decoding used EEGs from all 61 channels as features. To examine which electrodes contributed the most to the decoding accuracy, we conducted a channel searchlight decoding using subsets of the 61 channels as features [30].

Specifically, we first divided all channels into 61 neighbourhoods, centering each channel according to its location (conducted in Fieldtrip Toolbox [29] via ft_prepare_neighbours() function using 'triangulation' method). Immediately neighbouring channels were clustered together, resulting in 6.39±1.50 channel neighbours for each channel (with overlaps). Then the time domain EEG was averaged on time windows of interest, i.e., averaged on 0-500 ms, 500-3,000 ms, etc., to inspect the decoding topographical distribution on different time windows. The rest of the procedure was the same as time domain EEG decoding: we divided data into 3 sets and averaged within each set before splitting training and testing datasets; then we normalized them using mean and standard deviation of training sets. Finally, the decoding was conducted with a 3-fold cross validation and 10 iterations. Theta/alpha searchlight was conducted in the same way as time-domain searchlight, after averaging time-frequency power on respective oscillation range (theta: 4-8 Hz; alpha: 9-12 Hz).

Time Frequency Analyses: Six electrode clusters were selected for Time Frequency analyses: left parietal (CP3/5, P3/5), parietal (Pz, CP1/2, P1/2), right parietal (CP2/4, P2/4), frontocentral (FC1/2, C1/2, FCz, Cz), left prefrontal (AF3, F3/5) and right prefrontal (AF4, F4/6).

Time frequency transformation was performed using the same parameters as in decoding analyses in Fieldtrip [29], with additional decibel baseline normalization using power on -500 to -200 ms. We focus on the early theta power change on 200-400 ms which is indicator of inhibitory control [12, 13], and theta and alpha power change on a post hoc late time window (500-3000 ms) following condition level decoding results.

Correlation Analyses: We calculated Spearman's Rho for all correlations. In condition-level decoding, memory of Think and No-think was normalized by subtracting and then divided by Baseline memory, then correlated with time domain condition-level decoding accuracy on 500-3000 ms. To investigate the time course of these correlations, Spearman's Rho was calculated at each time point. For condition-level alpha decoding, we investigated correlation between memory and decoding accuracy on 1,000-2,000 ms considering the findings from Fellner, et al. [17].

In item-level time-domain decoding, we investigated the correlations between decoding accuracy and absolute memory score of the same condition, on 0-500 ms and 500-3000 ms, respectively. To link item-level decoding with condition level inhibitory control theta power change, we calculated correlation between decoding accuracy difference between Think and No-Think on 420-600 ms, and theta power difference between Think and No-Think on 200-400 ms.

In the High- vs. Low-Suppression Grouping correlation, we calculated correlation between decoding accuracy on 280-420 ms and No-Think minus Baseline Detail memory score, to be consistent with the grouping measure.

High- vs. Low-Suppression Grouping: We divided 40 participants into High- vs. Low-Suppression Groups based on their No-Think-minus-Baseline *Detail* scores ranking, and median split into 20 participants in each group. *Detail* measure was used because it captured both variability and suppression effect compared to *Identification* (limited variability since it was a dichotomous measure) and *Gist* (did not show suppression effect). The pre-TNT learning was not different between Think and No-Think in neither group (*p*s > .116).

Quantification and Statistical Analysis

Behavioral Analyses: We conducted separate one-way ANOVAs with three within-subject conditions (Think vs. No-Think vs. Baseline) on the percentage of Identification, percentage of correctly recalled Gist, and number of correctly recalled Details. We then examined the suppression-induced forgetting effect by conducting planned pairwise t test between No-think and Baseline, with a negative difference (i.e., when subtracting Baseline scores from No-think scores) being indicative of forgetting due to retrieval suppression, below the baseline level.

We report findings with $p < .05$ as significant. Within-subject analyses of variance (ANOVAs) are reported with Greenhouse-Geisser corrected p-values whenever the assumption of sphericity was violated. In terms of effect sizes, we report Cohen's *dz* given our within-subjects design [31].

Condition-/Item-level Decoding with Time Domain EEG: Following the statistical analysis procedure reported by [9], decoding accuracy at each time point (on 0-3000 ms) was compared to chance level by one-tailed paired t test. Multiple comparisons were controlled by non-parametric cluster-based Monte-Carlo procedure. Specifically, the null distribution was constructed by assigning trial level classification results to random classes (as if the classifier has no knowledge of actual information), and then timepoint-by-timepoint t-tests were performed to obtain a maximum summed t-value of continuous significant time cluster, which then repeated for 1,000 times. The resulting null distribution contained 1,000 summed t-values, which would be the distribution of the cluster summed ts when there is no true difference between decoding results and chance level. Both the cluster alpha and the alpha to obtain critical values from the permutation null distribution were set at 0.05 (on the positive tail, one-tail against chance).

The between-condition comparison of decoding accuracy along time were similar, except that the null distribution was constructed by randomly assigning condition labels to trial level classification results with two-tail repeated measure t-test and clusters were obtained on positive/negative tails, respectively. Thus, the critical values from the permutation null

distribution were at 2.5% on the negative clusters null distribution and 97.5% on the positive clusters null distribution.

Channel Searchlight Decoding: We compared channel searchlight topographies between item-level decoding in Think and No-think with a two-tailed paired-sample t test at each channel. The multiple comparisons were controlled by cluster correction of channel neighbour clusters in Fieldtrip [29]. The neighborhood was defined in the exact same way as the channel searchlight analysis. Cluster alpha was set at 0.05. Observed clusters were compared to null distribution on positive/negative tails respectively.

Channel-/Item-level Decoding with Time-Frequency Domain EEG: The statistical analyses for time-frequency domain decoding were similar to those of time domain decoding, except that here clusters were calculated in a 2-D matrix instead of on a 1-D time axis, and the cluster alpha was set at 0.05. Also, observed clusters were compared to the null distribution clusters of the same rankings. The statistical comparison of a single time-frequency decoding was performed against chance level (one-tailed), and that of the difference between two time-frequency decoding was performed against 0 (two-tailed). Theta (4-8 Hz) and alpha (9-12 Hz) oscillations decoding were assessed after averaging across the corresponding frequency bin.

Time Frequency Analyses: The early theta power at each electrode was compared between No-Think and Perceptual Baseline after averaging on 200-400 ms across 4 to 8 Hz, and then cluster corrected according to electrode positions in Fieldtrip [29]. The suppression-associated reduction of theta and alpha power on later time window was examined by averaging on 500-3000 ms across 4-8 Hz (theta) and 9-12 Hz (alpha), and then compared between No-Think and Think/Perceptual Baseline with neighbour cluster correction in Fieldtrip. The channel neighbours were defined in the same way as in channel searchlight analysis.

Correlation Analyses: The cluster correction for correlation time course was performed in this way: we first transformed Spearman's Rho back to *t*-values to obtain the observed time-course clustered *t*-values and a null distribution. The null distribution was obtained by randomizing labels of the two variables of interest before calculating the Spearman's Rho and corresponding t value. The cluster alpha was set as 0.05, and the observed clusters were calculated for positive and negative clusters respectively. The critical values of null distribution were at the 2.5% on both tails. The comparison between 2 correlation coefficients was conducted through a two-sided z test controlling for dependence [32].

High- vs. Low-Suppression Groups Comparison: Decoding accuracy at each time point on 0-3000 ms was compared between High- and Low suppression groups using two-tail independent t-test. The null distribution was constructed by randomly assigning group labels to each subject before by-timepoint t-test, to obtain the max summed-t of continuous significant time cluster when group labels are randomized, which repeated for 10,000 times. The resulting 10,000 summed-t values would be the null distribution when no true difference exists between the two groups. Critical values from the permutation null distribution were at 2.5% on the negative clusters null distribution and 97.5% on the positive clusters null distribution (two-tail, $\alpha s = 0.05$).

References

1. Staresina, B.P., and Wimber, M. (2019). A neural chronometry of memory recall. Trends in cognitive sciences *23*, 1071-1085.
2. Staresina, B.P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher, N., and Fell, J. (2016). Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. Elife *5*, e17397.
3. Staresina, B.P., Reber, T.P., Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2019). Recollection in the human hippocampal-entorhinal cell circuitry. Nature communications *10*, 1-11.
4. Anderson, M.C., Ochsner, K.N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S.W., Glover, G.H., and Gabrieli, J.D. (2004). Neural systems underlying the suppression of unwanted memories. Science *303*, 232-235.
5. Depue, B.E., Curran, T., and Banich, M.T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. science *317*, 215-219.
6. Gagnepain, P., Henson, R.N., and Anderson, M.C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. Proceedings of the National Academy of Sciences *111*, E1310-E1319.
7. Hu, X., Bergström, Z.M., Gagnepain, P., and Anderson, M.C. (2017). Suppressing unwanted memories reduces their unintended influences. Current Directions in Psychological Science *26*, 197-206.
8. Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., Vallée, T., Klein-Peschanski, C., Viader, F., and De la Sayette, V. (2020). Resilience after trauma: The role of memory suppression. Science *367*.
9. Bae, G.-Y., and Luck, S.J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. Journal of Neuroscience *38*, 409-422.
10. Anderson, M.C., and Green, C. (2001). Suppressing unwanted memories by executive control. Nature *410*, 366-369.
11. Küpper, C.S., Benoit, R.G., Dalgleish, T., and Anderson, M.C. (2014). Direct suppression as a mechanism for controlling unpleasant memories in daily life. Journal of Experimental Psychology: General *143*, 1443.
12. Cavanagh, J.F., and Frank, M.J. (2014). Frontal theta as a mechanism for cognitive control. Trends in cognitive sciences *18*, 414-421.
13. Nigbur, R., Ivanova, G., and Stürmer, B. (2011). Theta power as a marker for cognitive interference. Clinical Neurophysiology *122*, 2185-2194.
14. Anderson, M.C., and Hulbert, J.C. (2020). Active forgetting: Adaptation of memory by prefrontal control. Annual Review of Psychology *72*.
15. Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron *85*, 390-401.
16. Colgin, L.L. (2013). Mechanisms and functions of theta rhythms. Annual review of neuroscience *36*, 295-312.
17. Fellner, M.-C., Waldhauser, G.T., and Axmacher, N. (2020). Tracking selective rehearsal and active inhibition of memory traces in directed forgetting. Current Biology *30*, 2638-2644. e2634.
18. Jensen, O., Gelfand, J., Kounios, J., and Lisman, J.E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cerebral cortex *12*, 877-882.

19. Anderson, M.C., and Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. Trends in cognitive sciences *18*, 279-292.
20. Colgin, L.L. (2016). Rhythms of the hippocampal network. Nature Reviews Neuroscience *17*, 239-249.
21. Lavenex, P., and Amaral, D.G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. Hippocampus *10*, 420-430.
22. van Schie, K., and Anderson, M.C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. Memory *25*, 1201-1216.
23. Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., Hartmann, T., Raabe, M., Greenlee, M.W., and Bäuml, K.-H.T. (2012). Prefrontally driven downregulation of neural synchrony mediates goal-directed forgetting. Journal of Neuroscience *32*, 14742-14751.
24. Bäuml, K.-H., Hanslmayr, S., Pastötter, B., and Klimesch, W. (2008). Oscillatory correlates of intentional updating in episodic memory. NeuroImage *41*, 596-604.
25. Xie, S., Kaiser, D., and Cichy, R.M. (2020). Visual imagery and perception share neural representations in the alpha frequency band. Current Biology *30*, 2621-2627. e2625.
26. Waldhauser, G.T., Bäuml, K.-H.T., and Hanslmayr, S. (2015). Brain oscillations mediate successful suppression of unwanted memories. Cerebral Cortex *25*, 4180-4190.
27. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods *134*, 9-21.
28. Lopez-Calderon, J., and Luck, S.J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. Frontiers in human neuroscience *8*, 213.
29. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience *2011*.
30. Treder, M.S. (2020). MVPA-Light: a classification and regression toolbox for multi-dimensional data. Frontiers in Neuroscience *14*, 289.
31. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in psychology *4*, 863.
32. Lenhard, W., and Lenhard, A. (2014). Hypothesis tests for comparing correlations. Bibergau, Germany: Psychometrica.