Behavioral/Cognitive

Age-Related Decreases in the Retrieval Practice Effect Directly Relate to Changes in Alpha-Beta Oscillations

Catherine-Noémie Alexandrina Guran,[1] Nora Alicia Herweg,[2] and ⓘ Nico Bunzeck[1]

[1]Institute of Psychology I, University of Lübeck, 23562 Lübeck, Germany, and [2]Computational Memory Lab, Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

The retrieval (or testing) of information leads to better memory performance compared with reencoding. This phenomenon is known as "testing effect" or "retrieval practice effect" and has been primarily described in behavioral studies with healthy young subjects. However, possible age-related changes and their associated underlying neural processes, in particular neural oscillations, remain unclear. To address this issue, we used a previously established paradigm in healthy young ($N = 27$) and elderly ($N = 28$) male and female human adults while their brain activity was being recorded using EEG. Subjects viewed prefamiliarized scene images intermixed with new scenes and classified them as indoor versus outdoor (encoding task) or old versus new (retrieval task). Subsequently, subjects performed a recognition memory task 10 min and 24 h after encoding. Behaviorally, both age groups showed the testing effect at both time points but, importantly, it was less pronounced in the elderly. At the neural level, the retrieval compared with the encoding task was accompanied by power decreases in the alpha (9 – 12 Hz) and beta bands (13–30 Hz), possibly reflecting task demands, and this difference was more pronounced in the elderly. Finally, a correlation analysis revealed that those elderly who displayed a more pronounced testing effect exhibited a neural pattern that was more similar to the younger subjects. These findings provide evidence that the testing effect decreases across the life span, and they suggest that changes in alpha-beta oscillations play a direct role.

Key words: aging; alpha-beta oscillations; EEG; long-term memory; retrieval practice effect; testing effect

---

Significance Statement

Learning new and retrieving old information is part of everyday human life. Understanding how learning processes can be optimized therefore has direct applications in the realm of educational and rehabilitative contexts. Here, we show that retrieval practice is a strategy to optimize encoding into long-term memory in both young and elderly humans. Importantly, retrieval practice was significantly reduced in the elderly and closely related to changes in alpha (9 – 13 Hz) and beta band (13–30 Hz) oscillations. Our findings suggest that decreased retrieval practice effects across the life span contribute to, and may reflect, age-related declines in memory performance. They further provide new insights into the underlying neural mechanisms and point toward future avenues for neuro-modulatory interventions.

Introduction

A common approach to learning new material is to repeatedly encode the information at hand. However, retrieval practice (i.e., recall of previously encoded material from memory) is a more effective way to enhance long-term memory (Karpicke et al.,

2009). This so-called "retrieval practice effect" (RPE; or 'testing effect') has been shown for retention intervals of up to a week and various stimulus materials, such as vocabulary or scientific texts (Abbott, 1909; Spitzer, 1939; Karpicke and Roediger, 2008; Karpicke and Blunt, 2011). Importantly, while the RPE is well described in younger subjects, its age-related development and link to underlying neural oscillations remain unclear.

Learning and retrieval of declarative information gradually decline with age, which can be explained by specific structural brain changes to regions, such as the medial temporal lobe (i.e., hippocampus and surrounding cortex) and PFC (Hedden and Gabrieli, 2004). Therefore, age-related differences of the RPE should also be expected. Indeed, differences in size or dynamics of the RPE appear to differ between young and elderly, which may relate to changes in costs and benefits of repeated remembering
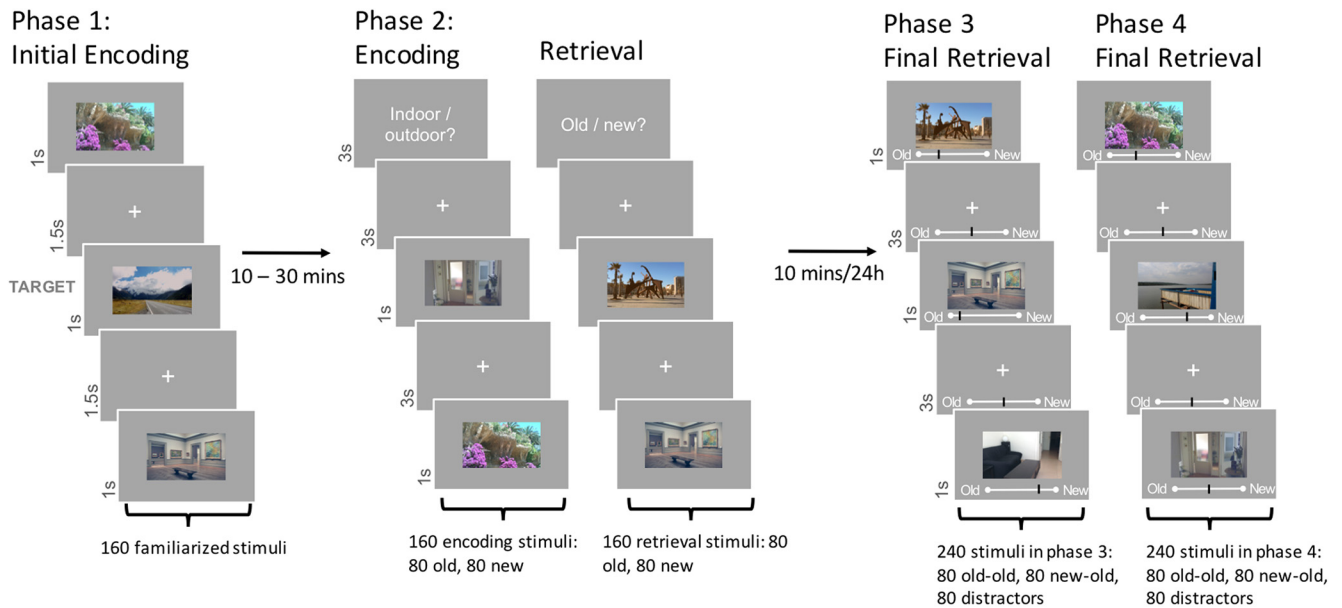
Figure 1. Experimental paradigm. Participants were familiarized with 160 images during the first phase (p1). In Phase 2, participants saw all 160 previously shown images (old) intermixed with 160 new images in either the encoding, or the retrieval task. In Phases 3 and 4, participants saw a randomized counterbalanced half of all previously shown images (old-old and new-old; 80 per category and phase) together with 80 new distractor images each (also 80 per category and phase). Phases 3 and 4 were performed 10 min or 24 h after Phase 2, respectively, and subjects gave recognition memory strength ratings. Modified with permission from Herweg et al. (2018).

(Henkel, 2007). Similarly, RPEs in the elderly are dependent on feedback availability during RP (Tse et al., 2010), and, as sometimes seen in younger subjects, they may only be observable after longer retention intervals (Roediger and Karpicke, 2006; Toppino and Cohen, 2009; Roediger and Butler, 2011; but for immediate RPE, see Meyer and Logan, 2013; Rowland and DeLosh, 2015; Herweg et al., 2018). In contrast, others indicate no age-related changes in RPEs (Rabinowitz and Craik, 1986; Balota et al., 1989; Meyer and Logan, 2013). For instance, the learning of unrelated word pairs benefits from retrieval relative to a study condition in both young and elderly (Coane, 2013). Together, RPEs can be observed in the elderly, yet they are task dependent and may be less pronounced compared with younger subjects.

The underlying neural processes and associated age-related changes of the RPE are not well understood. Retrieval compared with restudy has been shown to engage a variety of brain areas, including the anterior hippocampus, lateral temporal cortices, mPFC, striatum and dopaminergic midbrain (van den Broek et al., 2013; Wing et al., 2013; Herweg et al., 2018), as well as to enhance connectivity between the ventrolateral PFC and midline regions (Wing et al., 2013). These effects have been interpreted as reflecting semantic processing, selective strengthening of associations, retrieval effort, cognitive control, and motivational processes, which fit to behavioral interpretations of the RPE (Roediger and Butler, 2011).

Electrophysiologically, a few EEG studies have focused on event-related potentials (ERPs). In a cue-target word pairs paradigm, subsequent memory effects at parietal electrodes (starting at ~500 ms after stimulus onset) were linked to RP conditions (Liu et al., 2017). Furthermore (in a different task), left parietal old-new effects (500–900 ms), indicating recollection, were stronger for tested than untested items (Rosburg et al., 2015). While both studies provide an EEG marker for the RPE, a more direct link comes from Gao et al. (2016), demonstrating that ERP

Table 1. Number of averaged correct and incorrect responses per condition for Phase 2 encoding (indoor vs outdoor) and retrieval task (old vs new)

Age group	Encoding		Retrieval			
			Old		New	
	Correct	Incorrect	Remembered	Forgotten	Correct rejections	False alarms
Young	151.04	6.78	49.48	28.96	66.44	12.63
Elderly	154.54	3.36	57.25	20.96	58.07	20.64

Table 2. d′ values for young and elderly participants in Phase 2[a]

Age	Task	Mean	SEM
Young	Encoding	3.55	0.14
	Retrieval	1.44	0.13
Elderly	Encoding	3.9	0.12
	Retrieval	1.36	0.14

[a]Performance was overall better in the encoding task. Older participants' performance difference between encoding and retrieval was larger.

magnitudes (300–500 ms after stimulus onset) in the retrieval condition correlate with RP benefits.

Apart from ERPs, neural oscillations, in particular within the theta (4–8 Hz), alpha (9–12 Hz), and beta (13–25 Hz) bands, can provide specific insights into the underlying neural mechanisms of learning and memory processes (e.g., Düzel et al., 2010; Hanslmayr and Staudigl, 2014). However, to our knowledge, no study has been published on how age-related changes in RPEs relate to specific neural oscillations. Therefore, we used a modified version of a previously established paradigm (Herweg et al., 2018) as well as EEG and time-frequency analysis. We expected RPEs in young and elderly at both time points (day of encoding [day 1] and subsequent day [day 2]), and reduced memory performance and reduced RPE in the elderly. Despite mixed evidence at the behavioral level (see above), these hypotheses are

Table 3. Number of averaged correct and incorrect responses per condition for Phase 3 and Phase 4

	Previously encoded stimuli				Previously retrieved stimuli			
	Old in Phase 2		New in Phase 2		Old in Phase 2		New in Phase 2	
Age group	Remembered	Forgotten	Remembered	Forgotten	Remembered	Forgotten	Remembered	Forgotten
Phase 3, final retrieval on day 1								
Young	32.85	6.89	23.04	16.56	36.26	3.56	28.3	11.3
Elderly	32.79	6.68	24.79	14.64	34.46	4.86	26.21	12.96
Phase 4, final retrieval on day 2								
Young	28.04	11.7	17.07	22.7	32	7.74	23.11	16.67
Elderly	30.82	8.82	20.6	19.29	32.5	7.36	22.6	16.93

primarily based on our own work (Herweg et al., 2018), showing an early RPE in the young, and on the proposed link between the RPE and brain regions, such as the medial temporal lobe and PFC (van den Broek et al., 2013; Wing et al., 2013; Herweg et al., 2018), which typically show age-related degeneration leading to impaired behavior (Hedden and Gabrieli, 2004). These behavioral effects should be mirrored in theta, alpha, and beta oscillations. While we had no clear hypothesis regarding the direction of theta and beta changes, alpha power decreases were expected in the retrieval condition due to higher task demands (e.g., Klimesch, 1999).

Materials and Methods

Participants. In this study, 33 young (18–30 years) and 31 elderly (from 50 years on) subjects were tested. All were right-handed and none had a history of neurological or psychiatric disorders. The elderly were screened with the Montreal Cognitive Assessment Scale (Freitas et al., 2013), and all scored above a threshold of 22 points (lowest score in our sample: 25 of 30). The study was approved by the local ethics committee and subjects gave written informed consent.

From the initial sample, 4 participants had to be excluded due to technical difficulties, and another 5 due to other problems discovered during screening in the laboratory (left-handedness and history of psychiatric diseases). This led to a total of 27 young subjects (14 male, age range: 18–29 years, mean ± SD age 21.4 ± 2.7 years), and 28 elderly subjects (13 male, age range: 53–80 years, 62.8 ± 7.6 years).

Experimental design and task. The experimental paradigm was based on Herweg et al. (2018). The procedure consisted of four phases (Fig. 1). In Phase 1, participants were familiarized with 160 outdoor and indoor images (80 each), by means of a target detection task. The target stimuli were presented initially (one indoor, one outdoor) for 12 s. Subsequently, the 160 images for familiarization were presented three times each, for 1 s in pseudorandom order intermixed with 9% of target trials (i.e., 48 target and 480 nontarget trials). Each image was followed by an interstimulus interval of 1.5 s (white fixation cross on gray background). Participants had 2 s to respond to the target stimuli with a button press, and had the opportunity to pause every 96 trials.

For Phase 2, taking place ~30 min after the first phase, EEG was recorded with a BrainVision 64-electrode active EEG system. Participants had to perform two different, randomly alternating tasks, while viewing 160 new stimuli randomly intermixed with the 160 familiarized stimuli. The tasks were designed to induce an encoding and a retrieval context for half of the new and familiarized stimuli each. The encoding task (ENC) was a simple indoor/outdoor categorization. In the retrieval task (RET), participants gave an old/new recognition judgment. The combination of the two factors task and stimulus novelty (ENC/RET and OLD/NEW) resulted in a 2 × 2 repeated-measures design with 80 stimuli per
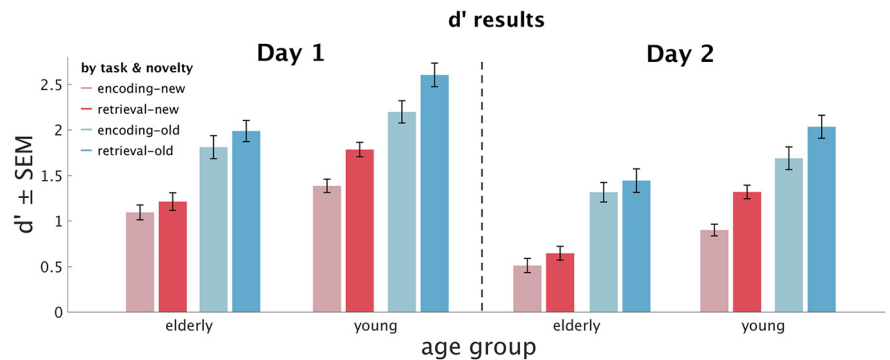

Figure 2. Descriptive results. Memory accuracy in Phase 3 (first experimental day) and 4 (second day), expressed in *d'* for elderly and young participants, encoded, retrieved, new, and old stimuli, respectively.

condition. Task blocks were 8 trials long, with each block containing 4 OLD and 4 NEW stimuli in random order.

An instruction screen informed participants about the upcoming task before the start of each block. Images were presented for 1 s with an interstimulus interval of 3 s (fixation cross). Participants gave their response within 2.8 s using their right index and middle finger. Response-button mappings were counterbalanced across participants. Participants could make a self-paced pause every 64 trials.

In Phase 3, subjects performed a surprise recognition task 5–10 min after the end of the second phase. A total of 160 stimuli, counterbalanced for location (indoor/outdoor), novelty, and task of the 320 images from Phase 2, were intermixed with 80 unseen distractor images (i.e., 80 stimuli from the encoding context, 80 stimuli from the retrieval context, and 80 unseen distractors). Each image was presented for 1 s with a visual analog scale below the image (ranging from "certainly old" to "certainly new"). The interstimulus interval was 3 s (fixation cross and analog scale) during which participants could still give their response using a mouse device. Participants could make a self-paced pause every 60 trials.

The task in Phase 4 was identical to Phase 3. However, it was conducted on the following day. Participants were shown the remaining 160 stimuli from Phase 2, which had not been presented in Phase 3, as well as another 80 previously unseen (distractor) stimuli (i.e., 80 stimuli from the encoding context, 80 stimuli from the retrieval context, and 80 unseen distractors).

Images were randomly assigned to the different phases and conditions for each subject. To control for effects of illumination, mean luminance on each color channel (R,G,B) was set to 127 (scale from 0 to 288), and images were presented on a gray background of equal luminance. Before each phase of the experiment, participants completed a brief training session. Images used during the training phase were different from those used during testing.

Statistical analyses of behavioral data. Accuracy of behavioral responses was assessed on the basis of signal detection theory (Stanislaw and Todorov, 1999). Specifically, *d'* was calculated by subtracting the inverse φ (conversion of probabilities into *z* scores according to the normal cumulative distribution function) of the hit rate from the inverse φ of the false alarm rate for each subject and condition. For target detection during
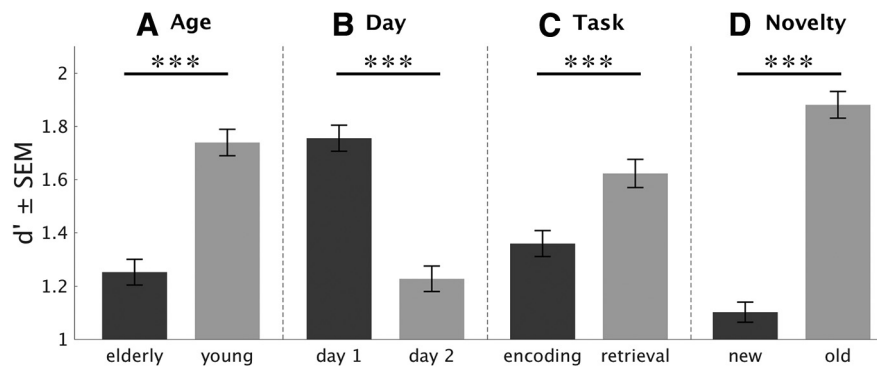
Figure 3. Main effects of recognition memory performance in Phases 3 and 4. Recognition memory performance (d') is depicted for age (***A***), task (***C***), and novelty (***D***) across Phases 3 and 4 (i.e., day 1 and 2, respectively), and for retrieval day (***B***). All factors (age, day, task, novelty) revealed significant main effects. ***$p < 0.001$.

Phase 1, hits were defined as correctly detected targets. For old/new categorization during Phase 2, 3, and 4, hits were defined as old stimuli correctly classified as old. For indoor/outdoor categorization during Phase 2, hits were defined as indoor images correctly classified as indoor. Because the inverse ϕ of 0 and 1 is $-\infty$ and ∞, respectively, 0.5 was added to the number of hits and false alarms and 1 was added to the number of signal and no signal trials (Stanislaw and Todorov, 1999).

Memory performance in Phases 3 and 4 was measured on a continuous visual analog scale. Here, we not only report d' (based on dichotomized responses) but also raw memory strength rating (measured as memory confidence ratings) per condition. Data were statistically analyzed using a 2 (Novelty: OLD/NEW) × 2 (Task: ENC/RET) × 2 (Retrieval Day: Phase 3/Phase 4) × 2 (Age: young/elderly) repeated-measures ANOVA, and t tests where appropriate.

EEG data analysis. EEG activity was acquired with an EasyCap system (BrainProducts) with 64 standard active electrodes. For detecting vertical and horizontal eye movement, electrodes with indices 29 through 32 were used (VEOG/EOG). Impedances were maintained <10 kΩ. An electrode on location TP8 served as reference, and an additional electrode placed behind the right ear as ground electrode.

The sampling rate was at 500 Hz with online high-pass (0.1 Hz) and low-pass (140 Hz) filters. EEGLAB (version 13) (Delorme and Makeig, 2004) and customized MATLAB (The MathWorks; RRID:SCR_001622) tools were used for preprocessing the EEG data offline. First, data were high-pass (1 Hz), low-pass (60 Hz), and notch-filtered (50 Hz). Second, data were inspected visually for paroxysmal artifacts. All trials were epoched (5 s trials, from 2 s before stimulus onset to 3 s after) and downsampled to 250 Hz. The ICA was performed and components to remove were selected with help of the ADJUST toolbox (Mognon et al., 2011), which suggests components for removal, as well as visual inspection of all components. A second visual inspection for, and removal of, remaining artifacts was performed. Finally, data were rereferenced to the average reference.

Baseline correction from 1.5 to 1.4 s before stimulus onset and data analysis were performed using the FieldTrip toolbox (Oostenveld et al., 2011) (http://www.ru.nl/neuroimaging/fieldtrip). Time-frequency decompositions were conducted from 4 to 30 Hz, across each epoch from 2000 ms before to 3000 ms after stimulus onset, using convolution on the single-trial time series with complex Morlet wavelets (width = 4), with steps of 4 ms in the time and 0.5 Hz in the frequency domain. For each condition, power was averaged across trials.

Statistical analysis of EEG data. To test for differences between the power bands in the different tasks and novelty conditions, two-tailed nonparametric cluster-based permutation tests (CBPs) (Maris and Oostenveld, 2007), with a family-wise error level of 0.05, were performed. All scalp electrodes were included in the test. t tests were performed for each sample: Within-subject factors were analyzed with repeated measures t tests, whereas the between-subjects factor was analyzed in an independent t test. Clusters were formed based on adjacent significant samples ($p < 0.05$), considering only effects with three or more significant neigh-

boring channels. To control for multiple comparisons, a Monte Carlo estimate of the permutation p value was computed by randomly permuting condition labels ($N = 1000$). The proportion of surrogate clusters whose cluster statistics (sum of t values) exceeded that of clusters observed with original condition labels constitutes the corrected p value. Nonparametric CBP tests, as described above, were conducted on the time-frequency data from 0 to 2000 ms after stimulus onset, and 4 to 30 Hz. While CBP tests are a very robust method to analyze EEG data, they are unsuitable for ANOVA analyses. To assess interactions, we used t tests on the averages over conditions.

To correlate EEG findings with behavioral data, we extracted power values from peak electrodes within a cluster (visual inspection), which were then Pearson correlated with d' values. To compare correlations between age groups, we used the CoCor software (Diedenhofen and Musch, 2015). All figures were created in MATLAB, with the help of the EEGLAB toolbox (Delorme and Makeig, 2004), and in case of the bar plots, the gramm toolbox (Morel, 2018).

Results
Behavioral results
Phase 1
Participants' accuracy in the first phase, which consisted of a target detection task, was similarly high in both age groups (d'_{young} (mean ± SD) = 5.3 ± 0.36, $d'_{elderly}$ = 5.47 ± 0.3), and a t test yielded no significant difference between the groups ($p > 0.2$).

Phase 2
The number of correct and incorrect responses per condition can be seen in Table 1. In Phase 2, data were analyzed in a 2 × 2 ANOVA, with task (ENC vs RET) as a within-subjects factor and age group as between-subjects factor. There was a main effect of task for response accuracy (d', ENC/RET; $F_{(1,53)} = 541.19$, $p < 0.001$, $\eta_p^2 = 0.911$; Table 2), as participants responded more accurately in the encoding than the retrieval task. There was a significant interaction of task and age group (ENC/RET vs young/elderly: $F_{(1,53)} = 4.7$, $p = 0.036$, partial $\eta_p^2 = 0.081$; Table 2). The difference in accuracy between retrieval and encoding task was significantly smaller in young participants than in elderly ($t_{(1,53)} = 2.16$, $p = 0.036$). There was no main effect of age group ($p > 0.39$). As d' for the retrieval condition was calculated based on responses for old and new images, these data cannot be analyzed in a 2 × 2 ANOVA.

Phase 3 and 4
The number of correct and incorrect response per condition can be seen in Table 3 as well as Figure 2. We conducted 2 × 2 × 2 × 2 ANOVAs, with age group as a between-subjects factor, and retrieval day (Phase 3/Phase 4), task (ENC/RET), and novelty (OLD/NEW) as within-subject factors for each dependent variable, namely, d' as well as memory strength ratings.

There were main effects (Fig. 3) of age group (young/elderly, $F_{(1,53)} = 16.16$, $p < 0.001$, $\eta_p^2 = 0.23$), retrieval day (Phase 3/Phase 4, $F_{(1,53)} = 181.71$, $p < 0.001$, $\eta_p^2 = 0.77$), task (ENC/RET, $F_{(1,53)} = 83.3$, $p < 0.001$, $\eta_p^2 = 0.61$), and stimulus novelty (OLD/NEW, $F_{(1,53)} = 270.4$, $p < 0.001$, $\eta_p^2 = 0.84$) on memory accuracy (d'). Young participants were more accurate than the elderly, participants gave more accurate memory judgments on the first experimental day, they were more accurate for old stimuli (i.e., which were already present in Phase 1, therefore old in
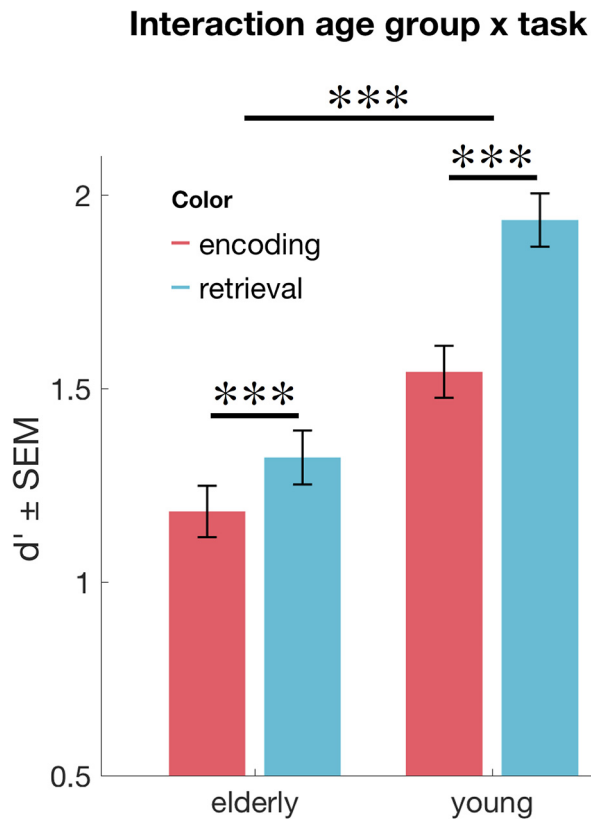
Interaction age group x task

Figure 4. Age-dependent RPEs. Across Phases 3 and 4 (i.e., day 1 and 2, respectively), recognition memory performance was lower in the elderly. Importantly, elderly subjects showed a significant RPE, which was significantly lower compared with the younger subjects. ***$p <$ 0.001.

Table 4. Results of the time-frequency analysis

Cluster	Frequency range (Hz)	Duration (ms)	Significant electrodes	p
Retrieval versus encoding	4 – 30	456 – 2000	All channels	Negative, 0.004
Old versus new	4 – 28	580 – 2000	All but TP8	Negative, 0.012
Young versus elderly	12 – 30	200 – 1630	All but T8	Positive, 0.04
Retrieval versus encoding: young versus elderly	5 – 30	960 – 2000	All channels	Positive, 0.024
Old versus New: Young versus elderly	4 – 26	660 – 2000	All but P8 and TP8	Negative, 0.044

showed a more pronounced decrease in memory strength ratings ($t_{(53)} = 2.02$, $p = 0.049$), while for the latter, memory strength decreased more strongly for novel stimuli than for old ones ($t_{(54)} = 2.85$, $p < 0.01$). No other effects were significant ($p > 0.05$).

EEG results

In a first step, Monte Carlo cluster-based permutation tests were run on the time-frequency data of young and elderly participants grouped together, from 0 to 2000 ms after picture onset, and from 4 to 30 Hz. Most effects were widespread in terms of time, significant electrodes and frequency (Table 4): therefore, we describe power peaks after visual inspection of the significant differences between conditions. We found a main effect of task (RET vs ENC), with decreased frontocentral alpha-beta power (10 – 20 Hz) in retrieval from 1.0 to 1.4 s, a main effect of novelty (OLD vs NEW), with decreased occipital and left frontal alpha power (8 – 10 Hz) for old stimuli from 0.8 to 1.1 s, a main effect of age group (young vs elderly), with stronger frontocentral beta power (16 – 23 Hz) for young participants from 1.0 to 1.4 s (Fig. 5, *A*, *B*, and *C* respectively). Importantly, there was an interaction between task and age group, with stronger frontocentral alpha-beta power (10 – 20 Hz) differences for retrieval versus encoding in elderly participants compared with young ones, from 1.0 to 1.4 s (Fig. 6). To further assess these effects, we performed separate CBP tests for distinct frequency bands: alpha (9 – 12 Hz), lower beta (13 – 19 Hz), and higher beta (20 – 30 Hz). Again, it revealed statistically significant effects in all three frequency bands (with $p < 0.01$, $p < 0.05$, and $p < 0.01$, respectively) that were similar in terms of electrode location and temporal dynamics compared with the initial analysis (4 – 30 Hz). Results for the analyses of separate frequency bands are not shown but will be made available upon request.

Finally, the interaction between novelty and age group was significant ($p = 0.044$, data not shown). *Post hoc* analysis revealed that the contrasts old versus new stimuli calculated for elderly participants did not reach significance, whereas the contrast for young participants only reached trend level ($p = 0.052$), suggesting a bigger power difference between old and new stimuli for young participants.

The cluster-based permutation analysis revealed no significant results for the comparisons of novelty and task, or novelty, task, and age. Contrasts for each cluster are indicated below. The significance level, duration in milliseconds, and the list of electrodes that were detected in the cluster are shown for each contrast (Table 4). Figures 5 and 6 represent the main effects and interaction between task and age group, respectively.
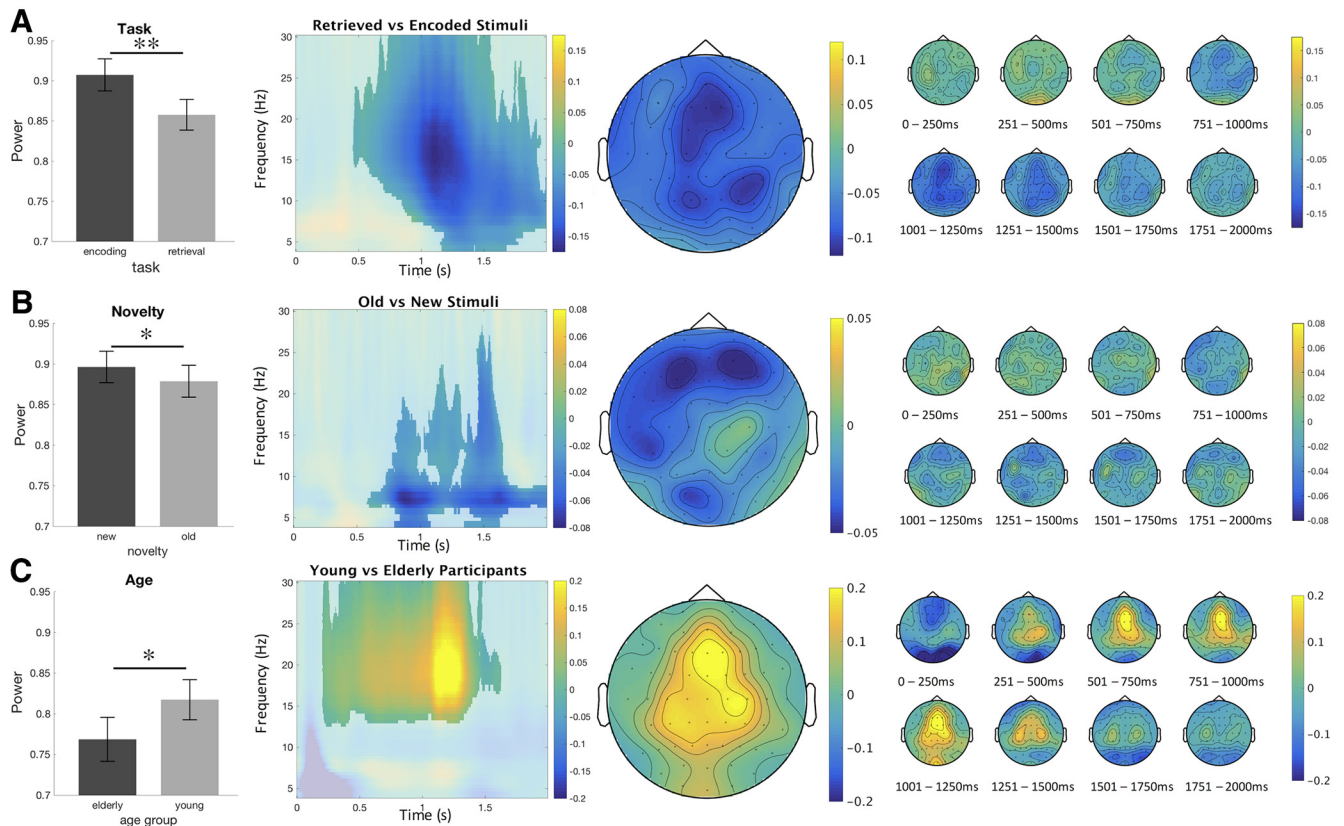
Phase 2) than for new images, and they were more accurate for stimuli presented in the retrieval task than for those presented in the encoding task.

Importantly, there was a significant interaction between task and age group (ENC/RET × young/elderly: $F_{(1,53)} = 18.82$, $p < 0.001$, $\eta_p^2 = 0.26$): this interaction was driven by a stronger memory benefit in the retrieval task compared with the encoding task for young versus elderly participants ($t_{(1,53)} = 4.338$, $p < 0.001$; Fig. 4). In other words, the RPE was more pronounced in the young compared with the elderly subjects. There were no other significant interactions ($p > 0.13$).

The findings from the memory strength ratings replicated these results in the main effects, with lower memory strength on the second retrieval day (Phase 3/Phase 4: $F_{(1,53)} = 68.94$, $p < 0.001$, $\eta_p^2 = 0.57$), higher memory strength for stimuli from the retrieval task (ENC/RET: $F_{(1,53)} = 87.14$, $p < 0.001$, $\eta_p^2 = 0.62$), and higher memory strength for previously old stimuli (OLD/NEW: $F_{(1,53)} = 255.56$, $p < 0.001$, $\eta_p^2 = 0.83$) with one exception: age group did not have a significant influence on the memory strength ratings ($F_{(1,53)} = 0.42$, $p > 0.8$). The interaction between task and age ($F_{(1,53)} = 17.52$, $p < 0.001$, $\eta_p^2 = 0.25$) was compatible with the d' analysis and showed a significantly smaller RPE for memory strength ratings for elderly than for young participants ($t_{(1,53)} = 4.19$, $p < 0.001$). In addition to the findings in the d' ratings, there were two significant interactions in memory strength ratings: day × age group ($F_{(1,53)} = 4.06$, $p = 0.049$, $\eta_p^2 = 0.071$), and day × novelty ($F_{(1,53)} = 7.95$, $p < 0.01$, $\eta_p^2 = 0.13$). For the former, memory strength ratings dropped less from one day to another for the elderly participants, while the young ones

Figure 5. Time frequency results. Main effects of (**A**) task (retrieval vs encoding). Participants have a higher deactivation in the retrieval task, across high alpha and low beta. **B**, Novelty (old vs new). Participants have a higher deactivation for old stimuli. **C**, Age (young vs elderly). Young participants have stronger beta power at ∼1–1.4 s than the elderly. Raw power values are plotted in bar plots. Single topoplot across the frequency peak of the cluster. Time-frequency samples not significant on any channel within the cluster are displayed opaque. Multiple topoplots represent the time course of the cluster. Color bars represent the relative change in power from baseline. *$p < 0.05$, **$p < 0.01$.
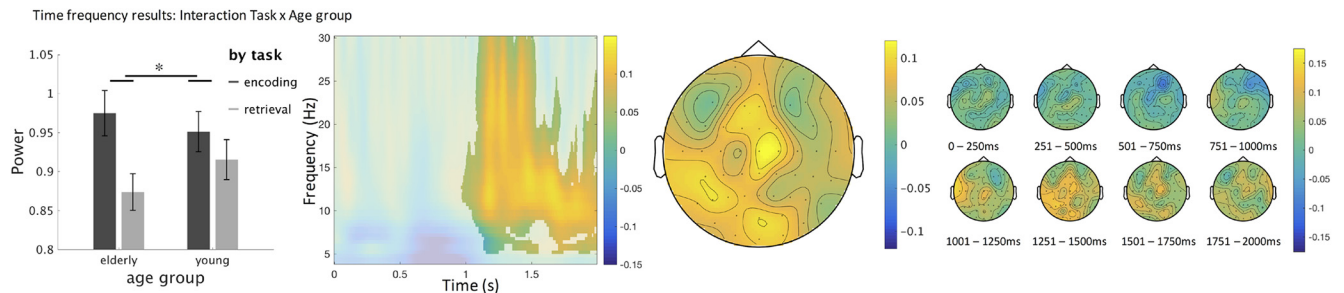

Figure 6. Interaction of task and age, calculated by subtracting the difference between encoding and retrieval in the young participants from the difference in the elderly. Elderly have a stronger deactivation in retrieval than young participants. *$p < 0.05$. Raw power values are plotted in a bar graph to illustrate the interaction. Single topoplot across the visual frequency peak (8–30 Hz) of the cluster (see singleplot). Time-frequency samples not significant on any channel within the cluster are displayed opaque. Multiple topoplots represent the time course of the cluster. Color bars represent the relative change in power from baseline.

Link between behavioral and EEG results: correlation analysis
To further assess the relationship between task-related behavior (i.e., RPE) and neural oscillations (i.e., power changes), we extracted power values for all electrodes within the clusters, and calculated Pearson correlations between (1) power difference between encoding and retrieval task in Phase 2 and (2) memory accuracy (d') difference between previously encoded versus retrieved stimuli (across both recall Phases 3 and 4). Specifically, visual inspection revealed slight differences between young and elderly in terms of peak electrodes (Fig. 7A,B, topoplots). For younger subjects, the average power (8–25 Hz) was strongest on electrodes F2 and F4 (concording with visual inspection, see Fig. 7B), whereas for elderly participants, the average power (8–25 Hz) was strongest on electrodes FC1, Cz, and F1. To have the same number of electrodes and comparable

locations between both age groups, we chose FC1 and F1 for the analysis in the elderly. For elderly participants, we found a significant positive correlation between the memory accuracy difference between encoding and retrieval ($d'_{ENC} - d'_{RET}$) at final recall in Phases 3 and 4 on one hand, and the activation difference between ENC and RET on the other hand ($r_{ENC/RET} = 0.41$, $p = 0.033$; Fig. 7A). There was no such correlations for young participants ($p > 0.3$; Fig. 7B). A direct comparison between the correlations of the elderly and young participants showed a significant difference (Fisher's $Z = -2.12$, $p = 0.034$).

Finally, we did not perform a subsequent memory analysis (i.e., difference due to later memory; DM analysis) because there were not enough trials in some of the four possible conditions (in particular for forgotten stimuli, see Table 3).

Discussion

Retrieval practice (RP) is a well-described strategy to improve long-term memory, but the underlying neural processes and possible age-related changes remain unclear. In this study, we can show that both young and elderly subjects benefit from RP but the effect on subsequent long-term memory is reduced in the elderly. At the electrophysiological level, retrieval (compared with encoding) was, across age groups, associated with decreased alpha/beta power, which indicates increased attentional demands and possibly deeper semantic processing (Klimesch et al., 1997; Klimesch, 2012). Importantly, the retrieval related alpha/beta decrease was more pronounced in the elderly, suggesting a close relationship to age-related changes in the RPE. Indeed, correlation analyses revealed that elderly subjects with larger RPEs show a more youth-like alpha/beta response, whereas those with small RPEs show a larger alpha/beta drop during retrieval. As such, our findings indicate a functional role and age-related changes of alpha/beta oscillations in the RPE across the life-span.

As expected, long-term memory in both young and elderly subjects was improved by RP (Fig. 3C). This is in line with long-standing evidence that testing of information leads to better memory compared with repeated study (Abbott, 1909; Spitzer, 1939; Karpicke and Roediger, 2008; Karpicke and Blunt, 2011). Although this is true for various stimulus material and retention intervals, the literature is less conclusive regarding age-related changes. While some report no age effects (Rabinowitz and Craik, 1986; Balota et al., 1989; Meyer and Logan, 2013), our findings are compatible with others (e.g., Tse et al., 2010; Rogalski et al., 2014) by showing that elderly subjects benefit from retrieval but, importantly, their RPE was less pronounced. We speculate that differences in study designs (i.e., stimulus material, retention interval, and type of retrieval) might help to explain divergent findings of age-related changes in RPE, which need to be addressed in future studies.

Overall, recognition memory was (1) lower for the elderly subjects (Fig. 3A), and (2) it declined from the early (day of encoding) to the late (1 d after encoding) time point (Fig. 3B). This is both compatible with previous literature on long-term memory in the elderly (e.g., Hedden and Gabrieli, 2004; Nyberg et al., 2012), and the notion of how we forget over time (e.g., Wixted, 2004; Oberauer and Lewandowsky, 2008). More importantly, however, the RPE was, in both groups, observed at both time points independent from retention interval (i.e., no interaction between task × day; Fig. 2). Specifically, an RPE on day 1 replicates our previous observation in younger subjects (Herweg et al., 2018) and is compatible with the hypothesis of a "fast-route to consolidation": accordingly, retrieval leads to fast consolidation by enhancing the integration of new information into pre-existing neocortical networks (Antony et al., 2017). Specifically, this mechanism is supposed to rely on fast modifications of neocortical representations, which have been shown in schema-based learning paradigms (Tse et al., 2007).
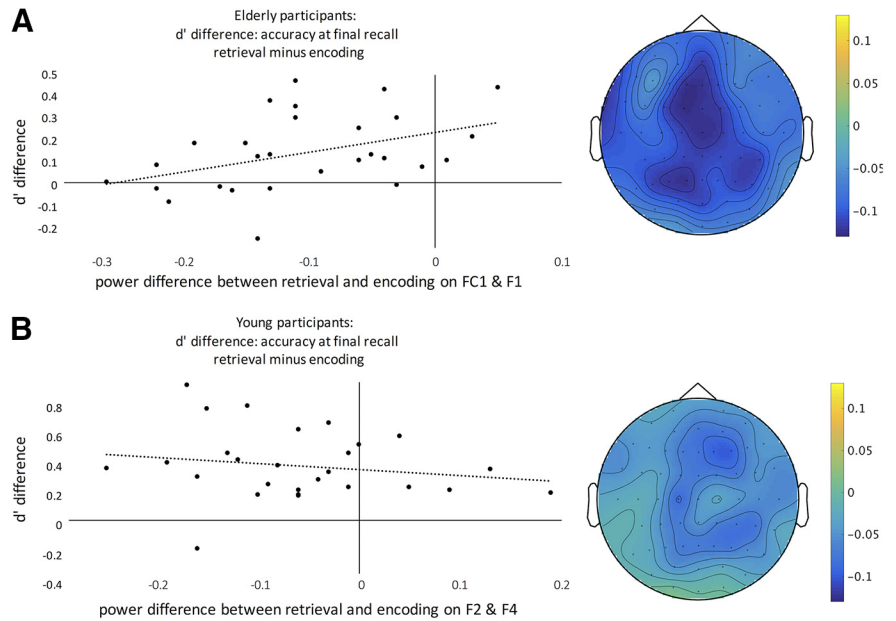
Figure 7. Correlation of power difference between retrieval and encoding in the central electrodes, and memory accuracy. This correlation was significant in the elderly, $r = 0.405$, (**A**) but not in the young participants, (**B**). Topoplots of retrieval condition minus encoding for elderly and young subjects, respectively, over peak frequencies (8 – 25 Hz). The effects are broadly centrally located, whereas the peak for the young subjects is shifted frontally.

Another important replication of our previous study (Herweg et al., 2018) is an RPE for those images that were novel in Phase 2 (for a similar effect, see Chan et al., 2006; Cho et al., 2017). That means that retrieval mode not only drives long-term memory for retrieved old stimuli but also other (i.e., new stimuli) that are presented in the same context. While this cannot easily be explained by the "fast-route to consolidation" hypothesis, it might be attributable to enhanced difficulty (or effort) associated with retrieval (Bjork, 1999; Rowland, 2014), or a broader search-strategy and activation of related semantic concepts (Chan et al., 2006). Accordingly, more effort may lead to deeper information processing, which is associated with slower response times (as in our study) and increased activity in brain regions that drive long-term memory performance, such as the mesolimbic system (Boehler et al., 2011; van den Broek et al., 2013; Herweg et al., 2018).

The "fast-route to consolidation" hypothesis is mainly supported by our behavioral findings: the EEG activity, on the other hand, was not acquired during the final retrieval, which would be most informative regarding the "fast-route to consolidation" framework. Finally, it should be noted that an RPE on day 1 is partly incompatible with other studies (Roediger and Karpicke, 2006), which further suggests that the RPE depends on factors, such as stimulus material (verbal vs pictures) and retrieval mode (free recall vs recognition) (e.g., Roediger and Butler, 2011).

At the neural level, the retrieval of information (compared with repeated encoding) led to power decreases in the alpha/beta band (Fig. 5A). This effect was most pronounced at frontal, but also central and parietal, electrodes with a peak at 1000 – 1250 ms after stimulus onset. In general terms, alpha oscillations have been associated with attentional processes, and, more specifically, alpha power decreases with attentional and semantic memory demands (see, e.g., Klimesch, 1999), which nicely fit with the notion of enhanced task difficulty/effort during retrieval outlined above. This is further supported by the interaction between task and age (Fig. 6), which was driven by stronger alpha/beta sup-

pression in the elderly and may reflect increased cognitive effort during retrieval due to age related impairments (see below for further discussion of the age effect).

A common effect of retrieval on both alpha and beta is compatible with previous studies, which have linked both frequency bands with memory processes. In particular, alpha/beta power decreases have been observed during the encoding of verbal information, which might reflect semantic processing and drive subsequent memory performance (for review, see Hanslmayr et al., 2012). Interestingly, decreased alpha/beta power also relate to memory retrieval, indicating material specific memory reactivation (Hanslmayr et al., 2012). In our study, a direct comparison between encoding and retrieval task revealed lower alpha/beta power during the retrieval of information (Fig. 5A). Importantly, this effect was more pronounced in the elderly, suggesting age-related changes with a link to memory performance. Indeed, the correlation analysis between RPE and alpha/beta power (i.e., difference in alpha/beta between encoding vs retrieval mode; Fig. 7) demonstrates that elderly with a more pronounced RPE show a more youth-like alpha/beta pattern (i.e., a less pronounced power difference), whereas those with a smaller RPE show a more pronounced alpha/beta difference. While this age-related effect can be explained by compensation or dedifferentiation (see below), the lack of a correlation in the young participants is unlikely to be due to differences in variance and/or overall higher performance in this group (see, e.g., Table 1; Fig. 4).

From a more mechanistic point of view, both alpha and beta oscillations have been associated with several functions that are mostly complementary rather than contradictory. For instance, alpha/beta predominantly represent top-down-directed influences (Fries, 2015), whereas beta oscillations appear to represent long-range communication between brain regions, preservation of the current brain state, and endogenous (re)activation of cortical representations (Kopell et al., 2000; Engel and Fries, 2010; Fries, 2015; Spitzer and Haegens, 2017).

How can age-related changes in neural activity be interpreted? The more pronounced difference in alpha/beta activity for retrieval versus encoding in the elderly (Fig. 6) is compatible with previous studies. Specifically, changes in neural activity in the elderly, often expressed as increased fMRI activity, have been reported previously (Cabeza et al., 2002; Dolcos et al., 2002). This may either reflect compensatory mechanisms to preserve cognitive functioning (Barulli and Stern, 2013), or it may reflect an age-related loss of regional specificity (i.e., dedifferentiation hypothesis) (Reuter-Lorenz and Park, 2010). While the compensation hypothesis might be more parsimonious in explaining our group differences, the correlation analysis is in favor of the dedifferentiation hypothesis. Indeed, the compensation hypothesis would predict that high performing elderly show more compensatory activity to reach the behavioral level of the younger ones. Instead, the high performing elderly showed a more youth-like EEG response. While our study was not designed to specifically disentangle both hypotheses, recent methodological advances (i.e., multivariate tests) may help to further address this issue (Morcom and Henson, 2018) as well as the relationship to possible underlying anatomical brain changes (Hedden and Gabrieli, 2004).

Together, young and elderly subjects' long-term memory benefits from retrieval. In both groups, retrieval was associated with decreases in alpha/beta power, which indicates a direct link and might reflect attentional demands associated with increased effort. This was further supported by reduced RPEs in the elderly, which were associated with a more pronounced alpha/beta drop.

As such, our findings provide evidence that the elderly benefit less from retrieval practice, and they suggest that changes in alpha-beta oscillations play a direct role.

References

Abbott EE (1909) On the analysis of the factor of recall in the learning process. Psychol Rev Monogr Suppl 11:159.

Antony JW, Ferreira CS, Norman KA, Wimber M (2017) Retrieval as a fast route to memory consolidation. Trends Cogn Sci 21:573–576.

Balota DA, Duchek JM, Paullin R (1989) Age-related differences in the impact of spacing, lag, and retention interval. Psychol Aging 4:3–9.

Barulli D, Stern Y (2013) Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. Trends Cogn Sci 17:502–509.

Bjork RA (1999) Assessing our own competence: heuristics and illusions. In: Attention and performance, Vol XVII: Cognitive regulation of performance: interaction of theory and application (Gopher D, Koriat A, eds), pp 435–459. Cambridge, MA: Massachusetts Institute of Technology.

Boehler CN, Hopf JM, Krebs RM, Stoppel CM, Schoenfeld MA, Heinze HJ, Noesselt T (2011) Task-load-dependent activation of dopaminergic midbrain areas in the absence of reward. J Neurosci 31:4955–4961.

Cabeza R, Anderson ND, Locantore JK, McIntosh AR (2002) Aging gracefully: compensatory brain activity in high-performing older adults. Neuroimage 17:1394–1402.

Chan JC, McDermott KB, Roediger HL 3rd (2006) Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. J Exp Psychol Gen 135:553–571.

Cho KW, Neely JH, Crocco S, Vitrano D (2017) Testing enhances both encoding and retrieval for both tested and untested items. Q J Exp Psychol 70:1211–1235.

Coane JH (2013) Retrieval practice and elaborative encoding benefit memory in younger and older adults. J Appl Res Mem Cogn 2:95–100.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Diedenhofen B, Musch J (2015) cocor: a comprehensive solution for the statistical comparison of correlations. PLoS One 10:e0121945.

Dolcos F, Rice HJ, Cabeza R (2002) Hemispheric asymmetry and aging: right hemisphere decline or asymmetry reduction. Neurosci Biobehav Rev 26:819–825.

Düzel E, Penny WD, Burgess N (2010) Brain oscillations and memory. Curr Opin Neurobiol 20:143–149.

Engel AK, Fries P (2010) Beta-band oscillations: signalling the status quo? Curr Opin Neurobiol 20:156–165.

Freitas S, Simões MR, Alves L, Santana I (2013) Montreal Cognitive Assessment: validation study for mild cognitive impairment and Alzheimer disease. Alzheimer Dis Assoc Disord 27:37–43.

Fries P (2015) Rhythms for cognition: communication through coherence. Neuron 88:220–235.

Gao C, Rosburg T, Hou M, Li B, Xiao X, Guo C (2016) The role of retrieval mode and retrieval orientation in retrieval practice: insights from comparing recognition memory testing formats and restudying. Cogn Affect Behav Neurosci 16:977–990.

Hanslmayr S, Staudigl T (2014) How brain oscillations form memories: a processing based perspective on oscillatory subsequent memory effects. Neuroimage 85:648–655.

Hanslmayr S, Staudigl T, Fellner MC (2012) Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. Front Hum Neurosci 6:74.

Hedden T, Gabrieli JD (2004) Insights into the ageing mind: a view from cognitive neuroscience. Nat Rev Neurosci 5:87–96.

Henkel LA (2007) The benefits and costs of repeated memory tests for young and older adults. Psychol Aging 22:580–595.

Herweg NA, Sommer T, Bunzeck N (2018) Retrieval demands adaptively change striatal old/new signals and boost subsequent long-term memory. J Neurosci 38:745–754.

Karpicke JD, Blunt JR (2011) Retrieval practice produces more learning than elaborative studying with concept mapping. Science 331:772–775.

Karpicke JD, Roediger HL 3rd (2008) The critical importance of retrieval for learning. Science 319:966–968.

Karpicke JD, Butler AC, Roediger HL 3rd (2009) Metacognitive strategies in student learning: do students practise retrieval when they study on their own? Memory 17:471–479.

Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev 29:169–195.

Klimesch W (2012) Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn Sci 16:606–617.

Klimesch W, Doppelmayr M, Pachinger T, Russegger H (1997) Event-related desynchronization in the alpha band and the processing of semantic information. Cogn Brain Res 6:83–94.

Kopell N, Ermentrout GB, Whittington MA, Traub RD (2000) Gamma rhythms and beta rhythms have different synchronization properties. PNAS 97:1867–1872.

Liu Y, Rosburg T, Gao C, Weber C, Guo C (2017) Differentiation of subsequent memory effects between retrieval practice and elaborative study. Biol Psychol 127:134–147.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods 164:177–190.

Meyer AN, Logan JM (2013) Taking the testing effect beyond the college freshman: benefits for lifelong learning. Psychol Aging 28:142–147.

Mognon A, Jovicich J, Bruzzone L, Buiatti M (2011) ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48:229–240.

Morcom AM, Henson RN (2018) Increased prefrontal activity with aging reflects nonspecific neural responses rather than compensation. J Neurosci 38:7303–7313.

Morel P (2018) Gramm: grammar of graphics plotting in Matlab. JOSS 3:568.

Nyberg L, Lövdén M, Riklund K, Lindenberger U, Bäckman L (2012) Memory aging and brain maintenance. Trends Cogn Sci 16:292–305.

Oberauer K, Lewandowsky S (2008) Forgetting in immediate serial recall: decay, temporal distinctiveness, or interference? Psychol Rev 115:544–576.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:256869.

Rabinowitz JC, Craik FI (1986) Prior retrieval effects in young and old adults. J Gerontol 41:368–375.

Reuter-Lorenz PA, Park DC (2010) Human neuroscience and the aging mind: a new look at old problems. J Gerontol B Psychol Sci Soc Sci 65:405–415.

Roediger HL 3rd, Butler AC (2011) The critical role of retrieval practice in long-term retention. Trends Cogn Sci 15:20–27.

Roediger HL 3rd, Karpicke JD (2006) Test-enhanced learning: taking memory tests improves long-term retention. Psychol Sci 17:249–255.

Rogalski Y, Altmann LJ, Rosenbek JC (2014) Retrieval practice and testing improve memory in older adults. Aphasiology 28:381–400.

Rosburg T, Johansson M, Weigl M, Mecklinger A (2015) How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. Cogn Affect Behav Neurosci 15:195–210.

Rowland CA (2014) The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. Washington, DC: American Psychological Association.

Rowland CA, DeLosh EL (2015) Mnemonic benefits of retrieval practice at short retention intervals. Memory 23:403–419.

Spitzer B, Haegens S (2017) Beyond the status quo: a role for beta oscillations in endogenous content (re-) activation. eNeuro 4:ENEURO-0170-17.2017.

Spitzer HF (1939) Studies in retention. J Educ Psychol 30:641.

Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures. Behav Res Methods Instrum Comput 31:137–149.

Toppino TC, Cohen MS (2009) The testing effect and the retention interval. Exp Psychol 56:252–257.

Tse CS, Balota DA, Roediger HL 3rd (2010) The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. Psychol Aging 25:833–845.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG (2007) Schemas and memory consolidation. Science 316:76–82.

van den Broek GS, Takashima A, Segers E, Fernández G, Verhoeven L (2013) Neural correlates of testing effects in vocabulary learning. Neuroimage 78:94–102.

Wing EA, Marsh EJ, Cabeza R (2013) Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. Neuropsychologia 51:2360–2370.

Wixted JT (2004) The psychology and neuroscience of forgetting. Annu Rev Psychol 55:235–269.