Biological Psychology 127 (2017) 134–147

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Differentiation of subsequent memory effects between retrieval practice and elaborative study

Yi Liu[a], Timm Rosburg[d,e], Chuanji Gao[c], Christine Weber[c], Chunyan Guo[a,b,*]

[a] Beijing Key Laboratory of Learning and Cognition, Department of Psychology, Capital Normal University, Beijing, 100048, PR China
[b] Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, China
[c] Department of Psychology, Institute of Mind and Brain, University of South Carolina, Columbia, 29201, USA
[d] Department of Psychology, Experimental Neuropsychology Unit, Saarland University, Campus, D-66123 Saarbrücken, Germany
[e] Department of Forensic Psychiatry, University Psychiatric Clinics Basel, CH-4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Retrieval practice effect
Elaborative study
Event-related potential (ERP)
Subsequent memory effect (SME)

ABSTRACT

Retrieval practice enhances memory retention more than re-studying. The underlying mechanisms of this retrieval practice effect have remained widely unclear. According to the elaborative retrieval hypothesis, activation of elaborative information occurs to a larger extent during testing than re-studying. In contrast, the episodic context account has suggested that recollecting prior episodic information (especially the temporal context) contributes to memory retention. To adjudicate the distinction between these two accounts, the present study used the classical retrieval practice effect paradigm to compare retrieval practice and elaborative study. In an initial behavioral experiment, retrieval practice produced greater retention than elaboration and re-studying in a one-week delayed test. In a subsequent event-related potential (ERP) experiment, retrieval practice resulted in reliably superior accuracy in the delayed test compared to elaborative study. In the ERPs, a frontally distributed subsequent memory effect (SME), starting at 300 ms, occurred in the elaborative study condition, but not in the retrieval practice condition. A parietal SME emerged in the retrieval practice condition from 500 to 700 ms, but was absent in the elaborative study condition. After 700 ms, a late SME was present in the retrieval practice condition, but not in the elaborative study condition. Moreover, SMEs lasted longer in retrieval practice than in elaboration. The frontal SME in the elaborative study condition might be related to semantic processing or working memory-based elaboration, whereas the parietal and widespread SME in the retrieval practice condition might be associated with episodic recollection processes. These findings contradict the elaborative retrieval theory, and suggest that contextual recollection rather than activation of semantic information contributes to the retrieval practice effect, supporting the episodic context account.

1. Introduction

Previous research has shown that retrieval practice improves subsequent memory performance more than re-studying does (Roediger & Butler, 2011; Karpicke & Roediger, 2008). This phenomenon has been labeled the retrieval practice effect, and it has been repeatedly demonstrated using various laboratory and practical educational materials (e.g., Karpicke & Blunt, 2011; Carpenter, 2009; Lehman et al., 2014). However, the cognitive mechanism underlying the retrieval practice effect is still a matter of debate.

Carpenter (2009) proposed the elaborative retrieval hypothesis to explain the mechanisms underlying the retrieval practice effect. The core concept of this account is that retrieval practice activates more semantic information (words or concepts) related to the recalled targets than re-studying does (Carpenter & DeLosh, 2006; Carpenter, 2009, 2011). For example, in one study, Carpenter (2009) manipulated the strength of the cue-target relationship. Participants studied weakly related pairs (e.g., Basket: Bread) and strongly related pairs (e.g., Toast: Bread). Subsequently, they performed either a cued-recall task or a re-studying task. Five minutes later, they needed to recall all of the targets they could remember. The findings showed that weakly related cue-target pairs were retained better than strongly associated pairs for the retrieval practice condition in the final free recall task. However, the cue-to-target strength did not affect memory retention for previously restudied items. The author argued that for the weakly associated pairs, more information semantically related to the cues was generated and elaborated by retrieval practice than by re-studying. This mediating information was spontaneously activated in the later test and increased

* Corresponding author at: Department of Psychology, Capital Normal University, No.23 Baiduizijia, Fuwaidajie St Haidian District, Beijing, 100048, PR China.
E-mail address: guocy@cnu.edu.cn (C. Guo).

http://dx.doi.org/10.1016/j.biopsycho.2017.05.010
Received 7 July 2016; Received in revised form 2 May 2017; Accepted 17 May 2017
Available online 24 May 2017
0301-0511/

the likelihood of successful retrieval for targets (Carpenter, 2009).

However, much of the evidence supporting the elaborative retrieval hypothesis is indirect, based on inferences from behavioral measures (such as response times and accuracy), and did not stem from direct investigations of the underlying cognitive processes. Thus, these experimental results are correlational rather than explanatory evidence for the elaborative retrieval hypothesis (Lehman et al., 2014; Karpicke, Lehman, & Aue, 2014). Furthermore, the elaborative retrieval hypothesis violates the cue-overload principle, which states that additional retrieval cues decrease the efficiency of retrieval (Watkins & Watkins, 1976). The hypothesis also has difficulties in explaining retrieval practice effects beyond those of semantic words, such as visuospatial maps (Carpenter & Pashler, 2007; Kang, 2010).

Karpicke et al. (2014) argued that if the hypothesis that participants form an elaborative network with some activated semantic concepts during retrieval was reasonable, elaboration should produce mnemonic effects similar to those produced by retrieval practice. Karpicke and Blunt (2011) compared the impact of retrieval practice and of concept mapping (as an elaborative study condition) on memory retention in a test one week later. They found that participants in the retrieval practice group performed better in the final test than participants who participated in the concept mapping task, tentatively suggesting that the mechanisms behind the retrieval practice effect were different from the cognitive processes of elaboration. In addition, Karpicke and Smith (2012) observed consistent results when using other kinds of elaborative study strategies, such as an imagery-based keyword method.

Based on these experimental results, Karpicke et al. (2014) argued that activating semantically related information was not crucial for the retrieval practice effect. They put forth a new account for the retrieval practice effect, the episodic context account, which suggests that during active retrieval participants recall and reconstruct prior study episodes, particularly their temporal context. During active retrieval, participants update the episodic representation with reinstated temporal context information. This information can be used as a retrieval cue, making the target more retrievable in later memory tests. This account suggests that the amount of recalled detailed information adjusts the memory enhancement.

Nonetheless, Karpicke and colleagues' investigations of the effects of retrieval practice and elaborative study on memory performance were based on behavioral experiments, thus they could only infer the potential mechanisms behind the retrieval practice effect (Lehman et al., 2014). Therefore, whether elaborative processing or context reinstatement is the underlying mechanism of the retrieval practice effect still awaits more evidence. Only a few functional neuroimaging studies have aimed to reveal the neurocognitive processes underlying the retrieval practice effect (Eriksson, Kalpouzos, & Nyberg, 2011; Hashimoto, Usui, Taira, & Kojima, 2011; Keresztes, Kaiser, Kovács, & Racsmány, 2014; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013). For example, Wing et al. (2013) investigated subsequent memory effects (SMEs) in order to identify brain areas relevant for the retrieval practice effect. SMEs are differences in neural activity triggered by subsequently remembered items and by subsequently forgotten items. Wing et al. observed larger SMEs in the bilateral hippocampus, lateral temporal cortex and medial prefrontal cortex for the retrieval practice condition than for the re-studying condition. The increased activity in the hippocampus might be related to the reinstatement of previously formed associations, as well as the updating of representations via integration of disparate information. However, this study focused on comparing encoding processes during retrieval practice and re-studying without considering elaboration, so the authors were not able to differentiate between the elaborative retrieval hypothesis and the episodic context account.

Researchers using the event-related potential (ERP) technique, which possesses high temporal resolution, have consistently shown that episodic recollection and the quantity of retrieved information are both indexed by the late parietal component (LPC) at 500–800 ms latency (Vilberg, Moosavi, & Rugg, 2006; Friedman & Johnson, 2000; Mecklinger, 2006; Rugg & Curran, 2007). Recently, some researchers used the ERP method to examine the neural correlates of the retrieval practice effect (Rosburg, Johansson, Weigl, & Mecklinger, 2015; Gao et al., 2016). For instance, Rosburg et al. (2015) analyzed the electrophysiological consequences of active retrieval by comparing old/new effects of previously tested items and untested items in a source memory task. The results suggested that the left parietal old/new effect was significantly superior for previously tested items than for previously untested items (Rosburg et al., 2015). However, the study only compared ERP differences induced by tested and untested items, as opposed to differentiating neural activity between retrieval practice and elaborative study. In addition, the difference in exposure time between tested and untested items may have been a confounding factor in this study. Moreover, the authors examined the outcome of retrieval practice rather than the neural activity occurring at the time of active retrieval. In contrast, Bai, Bridger, Zimmer, and Mecklinger (2015) used a subsequent memory paradigm to elucidate ERP correlates of retrieval practice occurring at initial test. They observed that the scalp topography of the 500–700 ms SME for tested pairs resembled the parietal old/new effect. The findings of these two ERP studies indicated that retrieval-specific processes, such as detailed context recollection, might benefit the retrieval practice effect, and were consistent with the episodic context account.

To our knowledge, neurophysiological studies have so far never aimed to compare the elaborative retrieval hypothesis and the episodic context account. Thus, whether the underlying mechanism of the retrieval practice effect is semantic elaboration or recollection of episodic context remains unclear. In the present study, Experiment 1 compared the behavioral mnemonic effects of retrieval practice, elaborative study, and re-studying. Experiment 2 investigated the neural correlates of successful encoding for retrieval practice and elaborative study. This experiment explored SMEs at retrieval practice and elaborative study within a classical retrieval practice effect paradigm (comprising study phase, initial retrieval phase, and final test phase), see Fig. 1. During the study phase, participants memorized weakly related word pairs (such as Experiment-Chemical; Theory-Hypothesis; Grammar-Book). These items were then assigned to one of the initial retrieval conditions: re-studying (only in Experiment 1), elaborative study (participants had to generate semantic mediators to relate cues and targets), or retrieval practice (cued-recall task). One week later, participants completed a recognition test and had to identify presented word pairs as “old” (studied pairs, such as Grammar-Book), “recombined” (rearranged pairs, such as Experiment-Hypothesis), or “new” (unstudied pairs). Old, recombined, and new word pairs were all semantically weakly related in order to avoid reliance on semantic relationship strength when judging the items.

Our hypotheses were as follows: For Experiment 1, retrieval practice would outperform elaborative study and re-studying in the delayed test, even when restricting retrieval practice to a single test (Roediger & Butler, 2011; Rowland, 2014). For Experiment 2, retrieval practice and elaborative study would differ behaviorally and in their ERP correlates. Specifically, the subsequent memory performance of retrieval practice would be superior to that of elaborative study. For ERPs, spatiotemporally distinct SMEs would occur for retrieval practice and elaborative study. In the retrieval practice condition, SMEs would resemble the LPC, as reported by Bai et al., 2015. Superior memory performance and distinct SMEs for the retrieval practice condition in comparison to the elaborative study condition would both be considered counterevidence for the elaborative retrieval hypothesis.

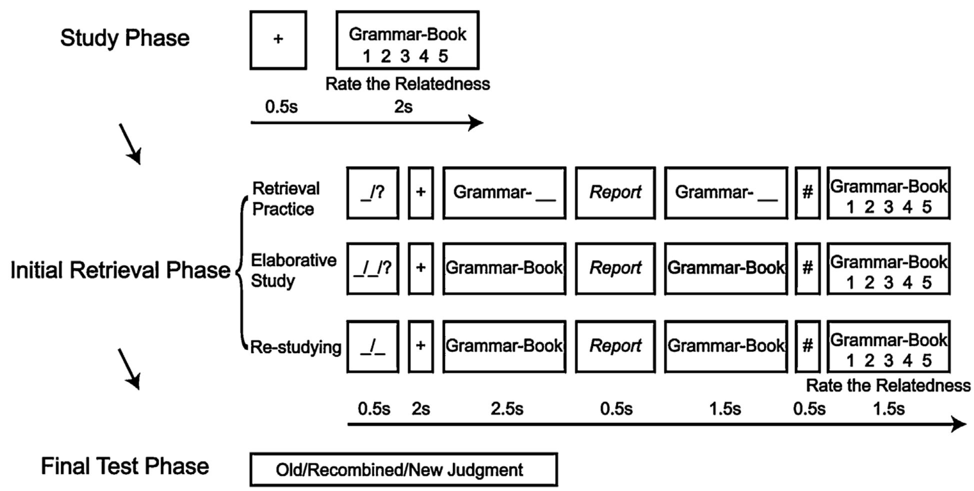

Fig. 1. Experimental procedure.
Experiment 1 comprised three within-subject conditions (Retrieval Practice, Elaborative Study, Re-studying), while Experiment 2 had two conditions (Retrieval Practice, Elaborative Study). In the study phase, participants needed to memorize weakly associated word pairs. Then, they were exposed to the initial retrieval phase and underwent the retrieval practice, elaborative study and re-studying conditions in Experiment 1. The latter condition was excluded in Experiment 2. A week after the initial retrieval phase, they performed an associative recognition task encompassing old word pairs (pairs that had been studied), recombined word pairs (rearranged pairs of studied cues and targets) and new word pairs (pairs of unstudied cues and unstudied targets).

2. Method

2.1. Participants

Thirty students (age, 19–25 years, 9 males) from Capital Normal University participated in Experiment 1. Twenty-one right-handed students (age, 19–26 years, 6 males) from Capital Normal University took part in Experiment 2. Five participants were excluded because of low ERP trial counts for the subsequently forgotten condition (less than 15 trials). The samples of Experiments 1 and 2 did not overlap. Participants had normal or corrected-to-normal vision and no neurological disorders. Participants gave written informed consent prior to participating and received ¥20 per hour in compensation. The Capital Normal University Institutional Review Board approved the study.

2.2. Materials

A total of 540 semantically weakly related cue-target word pairs were selected from the database of Nelson, McEvoy and Schreiber (1998). The average cue-to-target strength for these word pairs was 2.34% (range 1.0% to 4.8%), which represents the likelihood of producing the target upon presentation of a given cue. The cues and targets were translated into Chinese two-character pairs. To control for variations of the semantic relationship strength due to cultural differences, 13 students (7 male) from Capital Normal University rated the relatedness of the pairs, including an additional 180 weakly related recombined pairs, 100 unrelated pairs and 100 closely related pairs (mean cue-to-target strength: 32.6%). These students did not participate in Experiment 1 or Experiment 2. They rated the relatedness of each cue-target pair ("What is the possibility that you can produce the target depending on the given cue?") on a scale from 1 (very unlikely) to 5 (very likely). Through this procedure, rating scores for strongly related, weakly related, and unrelated word pairs were obtained. A one-way ANOVA revealed a significant effect of Relatedness (strongly related, weakly related, unrelated), $F(2, 24) = 209.71$, $p < 0.001$, $\varepsilon = 0.61$, $\eta_p^2 = 0.95$. Post-hoc contrasts (using Bonferroni correction) revealed that the scores of weakly related pairs [Mean(SD): 3.53(0.38)] were higher than the scores of unrelated pairs [1.27(0.33)] and lower than the scores of closely related pairs [4.20(0.43)], all p's < 0.001, which showed the selected stimuli were still weakly related word pairs in Chinese.

270 word pairs were selected for Experiment 1. Of these stimuli, 180 pairs were presented in the study phase, half of which were shown again as old pairs in the final test phase, while the other half were used as recombined pairs (90 recombined word pairs), and an additional 90 word pairs were used as new pairs. 540 word pairs were selected as materials for Experiment 2 in which participants needed to study 360 pairs. Half of studied word pairs acted as the same pairs (old word pairs) and the other half as recombined word pairs, and an additional 180 word pairs were used as new items in the final test.

2.3. Design and procedure

Both experiments used the classical paradigm for the retrieval practice effect (including three distinct phases: study phase, initial retrieval phase, and a final test phase after a delay of one week) (Fig. 1). All stimuli were presented using Presentation software (Neurobehavioral Systems, San Francisco, CA). Experiment 1 comprised three retrieval conditions (Retrieval Practice, Elaborative Study, Re-studying) and three blocks of 60 word pairs in study and initial retrieval. Experiment 2 comprised two conditions (Retrieval Practice, Elaborative Study) and six blocks of 60 word pairs in study and initial retrieval. Word pairs were always presented trial-by-trial. EEG recordings were collected during the initial retrieval and final test phases. The experimental procedures were largely identical across the two experiments, but, for practical reasons, the re-studying condition was excluded for Experiment 2.

In the study phase, participants were instructed to study and memorize sixty cue-target word pairs presented in the middle of the screen for 2 s each. They were asked to think about the relatedness of each pair and to indicate the likelihood of producing the target for the given cue by pressing buttons using a five-point scale ("1" representing "very unlikely" and "5" representing "very likely"). After studying all sixty pairs, the subjects completed a math task (counting continuously backwards by three from a given number) for 30 s.

After a resting phase for approximately two minutes, participants underwent the initial retrieval phase, including three conditions (Retrieval Practice, Elaborative Study and Re-studying) in Experiment 1. In each condition, they were exposed to one of three preparatory cues (_/?, _/_/?, _/_) for 500 ms, which reminded them to prepare for the following task ('_/?' Retrieval Practice; '_/_/?' Elaborative Study; '_/_' Re-studying). To reduce the influence of switching costs, the same task occurred twice in consecutive trials (Herron & Wilding, 2006), and there were also randomly added sets of three consecutive trials in order to lower participants' expectations. After the preparatory cue, the fixation cross appeared for 2 s, followed by the presentation of either an intact word pair (for the re-studying and elaborative study conditions) or a cue and an empty underline (for the retrieval practice condition) for 2.5 s in the center of the screen. Stimulus presentation order was pseudorandomized. The timing and number of stimuli during initial retrieval phase were identical for all three conditions.

In the retrieval practice condition, the symbol "_/?" prompted subjects to prepare for a cued-recall task of word pairs. One studied word and an empty underline appeared for 2.5 s, such as "Grammar-__", during which the participants needed to recall the corresponding target. Afterwards, the word "Report" was presented on the screen for 500 ms and then participants would be exposed to the cue again for 1.5 s, during which they had to report the target if possible. If they could not recover the corresponding target, they needed to say "I do not know". Immediately thereafter, the symbol "#" was presented for 500 ms, followed by the intact word pair for 1.5 s, during which participants rated the relationship between the cue and target again. The presentation of the intact pairs in this condition provided feedback to the participants, as well as an opportunity to memorize these pairs again, even when failing to retrieve them in the initial test.

The symbol "_/_/?" indicated that participants needed to prepare for the elaborative study task. Then an intact word pair appeared for 2.5 s, during which participants needed to covertly generate a word that was semantically associated to both the cue and target. The word pair was replaced by the word "Report" for 500 ms, and then participants were exposed to the intact word pair again for 1.5 s, during which they had to name the generated associate. If they could not think of an associated word, they were prompted to say "I do not know". After the participant's verbal response, the symbol "#" occurred for 500 ms and the intact word pair was presented for 1.5 s, during which they rated the relatedness of the word pair. Thus, participants needed to verbalize one word in both the retrieval practice and the elaborative study conditions. In the retrieval practice condition, they had to name the target; in the elaborative study condition, they had to name the generated associate.

The symbol "_/_" reminded participants to re-study the shown word pairs. Intact word pairs were presented for 2.5 s, during which participants could restudy them. Afterwards, the word "Report" was presented on the screen for 500 ms, and then participants would be exposed to the intact word pairs again for 1.5 s, during which they read the pairs overtly. After the participant's oral responses, the symbol "#" occurred for 500 ms and intact word pairs were presented for 1.5 s. As in the other two conditions, participants needed to rate the relationships of the pairs.

One week after studying the word pairs, the participants returned to the lab and underwent a recognition test. Each word pair was presented for 2 s, preceded by a fixation cross for 500 ms. The participants were told to judge via button press whether the presented word pairs were old (pairs that had been studied), recombined (pairs that consisted of cues and targets from different study pairs), or new (pairs that had not been studied).

2.4. EEG recording and data processing

EEG data were acquired from 62 Ag/AgCl electrodes positioned in an elastic nylon cap from the Neuroscan system according to the extended international 10–20 systems (Picton et al., 2000). Voltage from all electrodes during recording was referred to the left mastoid, and signals were re-referenced to the averaged recordings of mastoids. Four additional electrodes were attached above and below the left eye and on the outer canthi of both eyes in order to record the vertical and horizontal electrooculogram (VEOG and HEOG) respectively. Data were analyzed using the Neuroscan Scan 4.5 software. ERPs were time-locked to the word pairs presented for the first time in the initial retrieval phase and epochs were created from 200 ms prior to the occurrence of word pairs for 2000 ms. Signals from all channels (sampling rate 500 Hz) were filtered online with a 0.05–100 Hz bandpass (0.05–40 Hz filtered offline) and the impedance was maintained below 5 kΩ. Trials containing voltages exceeding the range ± 75 μV were excluded from the analyses after EOG blink artifacts were corrected using a linear regression estimate (Semlitsch, Anderer, Schuster, & Presslich, 1986).

2.5. Data analyses

2.5.1. Behavioral data

To measure performance in the initial retrieval phase, the accuracy of successful retrieval for the retrieval practice condition and the proportion of associates named for the elaborative study condition were calculated. Note that the proportion of "I don't know" responses in the elaborative study condition cannot be reported for Experiment 1 as the responses were unfortunately not protocolled. To measure performance in the final recognition test, the proportions of correctly recognized old pairs (labeled as "Hit_O"), old pairs falsely judged as recombined pairs (FA_O) and old pairs falsely judged as new pairs (Miss_O), correctly recognized recombined pairs (Hit_R), recombined pairs falsely judged as old pairs (FA_R), and recombined pairs falsely judged as new pairs (Miss_R) of different conditions (Experiment 1: Retrieval Practice, Elaborative Study, Re-studying; Experiment 2: Retrieval Practice, Elaborative Study) were calculated. Of note, there was only a single correct rejection rate for new pairs (CR_N), a single rate of new pairs falsely judged as recombined word pairs (FA_N_R) and a single rate of new word pairs falsely judged as old word pairs (FA_N_O) in each experiment, since the word pairs of the three conditions were presented in a mixed order in the final test.

To further examine the long-term memory retention for different conditions (Experiment 1: Retrieval Practice, Elaborative Study, Re-studying; Experiment 2: Retrieval Practice, Elaborative Study), the study distinguishes old/new discrimination and old/recombined discrimination. Old/new discrimination refers to participants' ability to distinguish old items from new items during the final test phase, whereas old/recombined discrimination refers to their ability to distinguish old items from recombined items. The study reports discriminability (d') and decision criterion (β) for both old/new discrimination and old/recombined discrimination.

Based on previous studies (Stanislaw & Todorov, 1999), d' for old/new discrimination (labeled as "$d'_{old/new}$") was computed as:

$$d'_{old/new} = Z_{Hit\,O} - Z_{FA\,N_O};$$

β for old/new discrimination ($\beta_{old/new}$) was computed as:

$$\beta_{old/new} = \frac{O_{Hit\,O}}{O_{FA\,N_O}};$$

d' for old/recombined discrimination ($d'_{old/rec}$) was calculated as:

$$d'_{old/rec} = Z_{Hit\,O} - Z_{FA\,R};$$

β for old/recombined discrimination ($\beta_{old/rec}$) was calculated as:

$$\beta_{old/rec} = \frac{O_{Hit\,O}}{O_{FA\,R}};$$

The value of the Y-axis (O) for the Z score according to the standard normal distribution was calculated as:

$$O = \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})}.$$

Two one-way repeated measures ANOVAs were respectively performed on d' and β with Condition (Experiment 1: Retrieval Practice, Elaborative Study, Re-studying; Experiment 2: Retrieval Practice, Elaborative Study) as a within-subject factor.

In addition, based on the performance in the initial retrieval phase (Retrieval Practice: successful retrieval vs. unsuccessful retrieval;

Table 1
Final recognition test (Experiment 1): Mean Hit, FA, and Miss rates for each type of word pair (Old, Recombined, and New pairs) (± SD).

Conditions (Experiment 1)	Old pairs			Recombined pairs			New pairs		
	Hit_O	FA_O	Miss_O	Hit_R	FA_R	Miss_R	CR_N	FA_N_o	FA_N_R
Retrieval Practice	0.66 ± 0.11	0.22 ± 0.09	0.11 ± 0.09	0.41 ± 0.13	0.41 ± 0.13	0.17 ± 0.09	0.71 ± 0.10	0.07 ± 0.06	0.22 ± 0.06
Elaborative Study	0.58 ± 0.19	0.29 ± 0.14	0.13 ± 0.10	0.39 ± 0.11	0.33 ± 0.15	0.28 ± 0.12			
Re-studying	0.57 ± 0.15	0.25 ± 0.11	0.17 ± 0.09	0.44 ± 0.11	0.26 ± 0.11	0.29 ± 0.13			

"Hit_O" means hit rates for Old pairs; "Hit_R" means hit rates for Recombined pairs; "CR_N" means correct rejection rates for New pairs; "FA_O" means false alarm rates of classifying Old pairs as Recombined pairs; "FA_R" means false alarm rates of classifying Recombined pairs as Old pairs; "FA_N_o " means false alarm rates of classifying New pairs as Old pairs; "Miss_O" means miss rates of classifying Old pairs as New pairs; "Miss_R" means miss rates of classifying Recombined as New pairs; "FA_N_R" means false alarm rates of classifying New pairs as Recombined pairs.

Elaborative Study: associate named vs. no associate named), the conditions in Experiment 2 were each differentiated into two further categories. To examine how the initial retrieval performance contributed to the retrieval accuracy in the final recognition test, the hit rates were analyzed separately for words pairs that were successfully retrieved in the retrieval practice condition vs. word pairs with no such retrieval success, and for word pairs for which an associate was named in the elaborative study condition vs. those for which no associate was named.

2.5.2. ERP data

Trial selection: The ERP data analysis excluded trials for which participants gave "I don't know" responses in the elaborative study condition (the mean proportion of not naming associates was 0.044), as no elaboration took place. In contrast, all trials of the retrieval practice condition were included (even when the initial retrieval was unsuccessful), as the participants presumably actively sought to retrieve targets in these trials and were provided with the correct word pairs at the end of each trial.

The selection of latency intervals and electrodes was based on previous studies examining memory retrieval or the retrieval practice effect (Rugg & Curran, 2007; Bai et al., 2015; Herron & Wilding, 2005; Evans, Williams, & Wilding, 2015) and on visual inspection of our waveforms. Four time-windows (300–500, 500–700, 700–1000, 1000–1500 ms) were used to examine the retrieval success effect and observe the subsequent memory effects of retrieval practice and elaboration. Three sets of electrodes along the anterior-posterior dimension were chosen for further analysis: frontal (F8/F4/FZ/F3/F7), central (T8/C4/CZ/C3/T7), and parietal (P8/P4/PZ/P3/P7), and the amplitudes across these electrodes were averaged for each cluster.

Retrieval success was defined as the difference between ERPs of word pairs that had been successfully retrieved and of word pairs that had been retrieved falsely or not at all during the initial retrieval phase. In order to examine the ERP correlates of retrieval success, a two-way repeated measures ANOVA with Retrieval Success (successful retrieval, unsuccessful retrieval) and Cluster (frontal, central, parietal) as within-subject factors was conducted for each of the selected time windows. To compute SMEs, we calculated the amplitude differences between ERPs for word pairs of initial retrieval phase that were Hit_O responses vs. Miss_O and FA_O responses in the final test. In order to examine SMEs for the retrieval practice and elaborative study conditions, a three-way repeated measures ANOVA with three factors, Condition (retrieval practice, elaborative study), Subsequent Memory (later remembered, later forgotten) and Cluster (frontal, central, parietal) as within-subject factors was performed on the mean amplitudes for each time-window. To examine if the scalp distributions of SMEs differed between retrieval practice and elaboration, additional topographic analyses on normalized mean amplitudes (Mccarthy & Wood, 1985) were conducted using all 62 recording electrodes ('Electrode Site'). A two-way repeated measures ANOVA with factors Condition (retrieval practice, elaborative study) and Electrode Site was performed for each time window. If this analysis showed a significant interaction (Condition × Electrode Site), different SME scalp topographies were presumed.

In addition, amplitudes of the SMEs were averaged for the electrode clusters and time windows, as used in the ANOVA analyses of the SMEs. Pearson's correlation coefficients were then computed between the SMEs and the behavioral measures in the final test (Hit_O, Hit_R, FA_O, FA_R, FA_N_O, FA_N_R), for the retrieval practice and elaborative study conditions separately.

The mean trial counts (and ranges) for the retrieval practice condition in the initial retrieval phase were: successful retrieval: 75(39–114); unsuccessful retrieval: 84(51–104); subsequently remembered: 51(30–69); subsequently forgotten: 29(15–43). For the elaborative study condition, the mean trial counts in the initial retrieval phase were: subsequently remembered: 44(25–57); subsequently forgotten: 33(19–48). All statistical analyses were performed using SPSS 19.0 and the alpha level was 0.05. The Greenhouse-Geisser correction was used when the sphericity assumption was violated. In such cases, corrected p values are reported with the indication of ε values. Bonferroni-correction was used for all *post-hoc* pairwise comparisons. Effect sizes (partial eta squared: η_p^2) and 95% confidence intervals (95% CIs) were reported alongside each inferential statistic.

3. Results

3.1. Experiment 1

For the initial retrieval phase, the mean proportion of successful target recall in the active retrieval condition was 0.45 (± 0.14). For the final test phase, mean Hit, FA and Miss rates for old pairs, recombined pairs and new pairs (Hit_O, FA_O, Miss_O, Hit_R, FA_R, Miss_R, CR_N, FA_N_R, FA_N_O) are specified in Table 1 and Fig. 2. Statistical comparisons of these performance measures can be found in Supplementary Materials.

3.1.1. Old/new discrimination

The ANOVA revealed $d'_{old/new}$ was significantly different across conditions, $F(2, 58) = 7.75$, $p = 0.001$, $\eta_p^2 = 0.21$. $d'_{old/new}$ of the retrieval practice condition [Mean(SD): 2.02(0.53)] was significantly higher than that of both the elaborative study [1.79(0.67)], $t(29) = 3.30$, $p = 0.008$, 95% CI [0.05, 0.41], $\eta_p^2 = 0.27$ and re-studying conditions [1.76(0.56)], $t(29) = 3.71$, $p = 0.003$, 95% CI [0.08, 0.43], $\eta_p^2 = 0.32$. $d'_{old/new}$ of the elaborative study and re-studying conditions did not differ, $t(29) = 0.32$, $p > 0.05$, 95% CI [−0.17, 0.22], $\eta_p^2 = 0.003$, see Fig. 3.

The ANOVA revealed that there were no significant differences between $\beta_{old/new}$ of the retrieval practice condition [4.05(2.84)], elaborative study condition [4.11(3.09)], and re-studying condition [4.45(3.61)], $F(2, 58) = 3.02$, $p = 0.077$, $\varepsilon = 0.70$, $\eta_p^2 = 0.09$, see Fig. 3.

3.1.2. Old/recombined discrimination

There were no significant differences between $d'_{old/rec}$ of the retrieval practice condition [0.68(0.42)], elaborative study condition

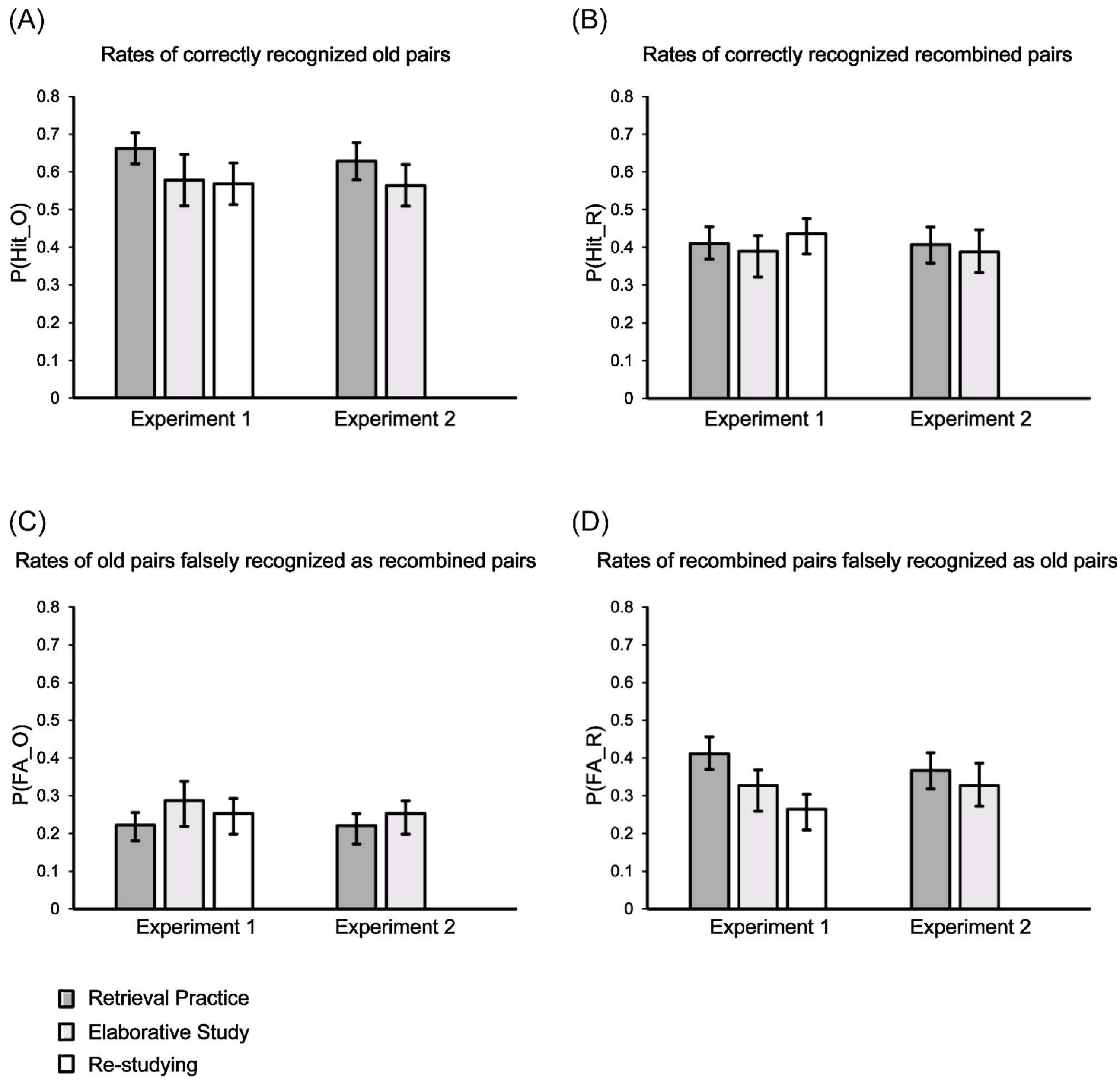

Fig. 2. Behavioral results of Experiments 1 and 2.
Word pairs that appeared in the study phase were presented again in the recognition test a week later. Participants needed to distinguish old, recombined and new pairs. (A) Proportions of correctly recognized old pairs in the final recognition test, for each condition and experiment; (B) Proportions of correctly recognized recombined pairs in the final recognition test, for each condition and experiment; (C) Proportions of old pairs falsely considered as recombined pairs in the final recognition test, for each condition and experiment; (D) Proportions of recombined pairs falsely recognized as old pairs in the final recognition test, for each condition and experiment. Error bars index 95% confidence intervals (CIs).

[0.69(0.45)] and re-studying condition [0.87(0.36)], $F(2, 58) = 2.85$, $p = 0.066$, $\eta_p^2 = 0.09$, see Fig. 3.

There were significant differences between $\beta_{old/rec}$ of the three conditions, $F(2, 58) = 7.18$, $p = 0.008$, $\varepsilon = 0.60$, $\eta_p^2 = 0.20$. $\beta_{old/rec}$ of the retrieval practice condition [0.95(0.18)] was lower than that of both the elaborative study [1.10(0.29)], $t(29) = -3.10$, $p = 0.013$, 95% CI [−0.27, −0.03], $\eta_p^2 = 0.25$ and re-studying conditions [1.36(0.76)], $t(29) = -1.81$, $p = 0.013$, 95% CI [−0.75, −0.08], $\eta_p^2 = 0.25$. $\beta_{old/rec}$ of the elaborative study and re-studying conditions did not differ, $t(29) = -2.83$, $p = 0.143$, 95% CI [−0.60, 0.06], $\eta_p^2 = 0.13$, see Fig. 3.

In addition, the one-way ANOVA indicated that, within the retrieval practice condition, the hit rate in the final test was higher for word pairs that had been retrieved in the initial test [Mean(SD) = 0.74(0.11)] than for word pairs that had not been retrieved [0.53(0.13)], $F(1, 15) = 51.77$, $p < 0.001$, 95% CI [0.15, 0.28], $\eta_p^2 = 0.78$. Within the elaborative study condition, the one-way ANOVA showed no significant difference between the final recognition performance for word pairs that had been elaborated in the initial retrieval phase [0.57(0.11)] and word pairs that had not been elaborated [0.41(0.36)], $F(1, 15) = 3.43$, $p = 0.084$, 95% CI [−0.02, 0.33], $\eta_p^2 = 0.19$.

3.2. Experiment 2

3.2.1. Behavioral data

The mean proportion of successful target recall in the retrieval practice condition was 0.47 (± 0.10) and the mean proportion of successful associate generation in the elaborative study condition was 0.96 (± 0.05). Mean Hit, FA and Miss rates for old pairs, recombined pairs, and new pairs (the Hit_O, FA_O, Miss_O, Hit_R, FA_R, Miss_R,

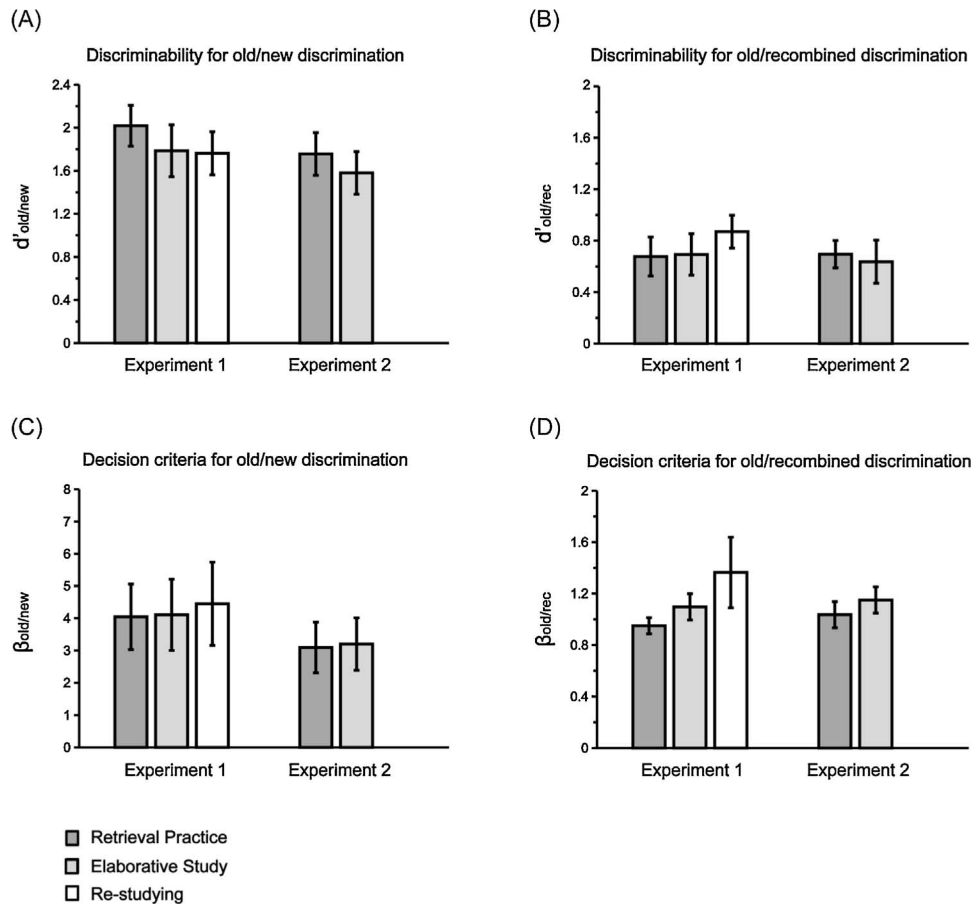

Fig. 3. Discriminability (d') and decision criteria (β) for old/new discrimination and old/recombined discrimination during the final test phase for Experiments 1 and 2. (A) Discriminability when distinguishing old pairs from new pairs in the final recognition test, for each condition and experiment; (B) Discriminability when distinguishing old pairs from recombined pairs in the final recognition test, for each condition and experiment; (C) Decision criteria when distinguishing old pairs from new pairs in the final recognition test, for each condition and experiment; (D) Decision criteria when distinguishing old pairs from recombined pairs in the final recognition test, for each condition and experiment. Error bars index 95% confidence intervals (CIs).

Table 2
Final recognition test (Experiment 2): Mean Hit, FA, and Miss rates for each type of word pair (Old, Recombined, and New pairs) (± SD).

Conditions (Experiment 2)	Old pairs			Recombined pairs			New pairs		
	Hit_O	FA_O	Miss_O	Hit_R	FA_R	Miss_R	CR_N	FA_N_o	FA_N_R
Retrieval Practice	0.63 ± 0.10	0.22 ± 0.07	0.15 ± 0.08	0.41 ± 0.10	0.37 ± 0.12	0.22 ± 0.07	0.67 ± 0.11	0.10 ± 0.07	0.23 ± 0.07
Elaborative Study	0.56 ± 0.11	0.25 ± 0.07	0.18 ± 0.08	0.39 ± 0.12	0.33 ± 0.14	0.28 ± 0.09			

For abbreviations, see Table 1.

CR_N, FA_N_R, FA_N_O) are specified in Table 2 and Fig. 2. Supplementary Materials show the statistical comparisons for these performance measures.

3.2.1.1. Old/new discrimination. The ANOVA revealed $d'_{old/new}$ of the retrieval practice condition [1.76(0.40)] was significantly higher than that of the elaborative study condition [1.58(0.40)], $F(1, 15) = 12.48$, $p = 0.003$, 95% CI [0.07, 0.28], $\eta_p^2 = 0.45$. There was no significant difference between $\beta_{old/new}$ of the retrieval practice condition [3.09(1.61)] and that of the elaborative study condition [3.20(1.66)], $F(1, 15) = 1.68$, $p = 0.21$, 95% CI [−0.27, 0.07], $\eta_p^2 = 0.10$, see Fig. 3.

3.2.1.2. Old/recombined discrimination. There was no significant difference between $d'_{old/rec}$ of the retrieval practice condition [0.70(0.22)] and that of the elaborative study condition [0.64(0.34)], $p > 0.01$. $\beta_{old/rec}$ of the retrieval practice condition [1.04(0.21)] was significantly lower than that of the elaborative study condition [1.15(0.21)], $F(1, 15) = 6.12$, $p = 0.026$, 95% CI [−0.21, −0.02], $\eta_p^2 = 0.29$, see Fig. 3.

3.2.2. ERP data

3.2.2.1. Retrieval success effect. The Retrieval Success × Cluster ANOVA revealed no significant effects involving Retrieval Success from 300 to 500 ms (all F's < 1), see Fig. 4. From 500–700 ms, the ANOVA revealed a significant Retrieval Success main effect [$F(1, 15) = 15.66$, $p = 0.001$, $\eta_p^2 = 0.51$], but there was no Retrieval Success × Cluster interaction ($F < 1$). Successfully retrieved items induced ERPs that were more positive [Mean (SD): 1.23(2.42)] than the ERPs for not successfully retrieved items [0.45(2.56)], 95% CI [0.36, 1.20].

From 700–1000 ms, the Retrieval Success main effect was significant [$F(1, 15) = 7.54$, $p = 0.015$, $\eta_p^2 = 0.34$], but the Retrieval Success × Cluster interaction was not significant [$F(2, 30) = 1.48$,

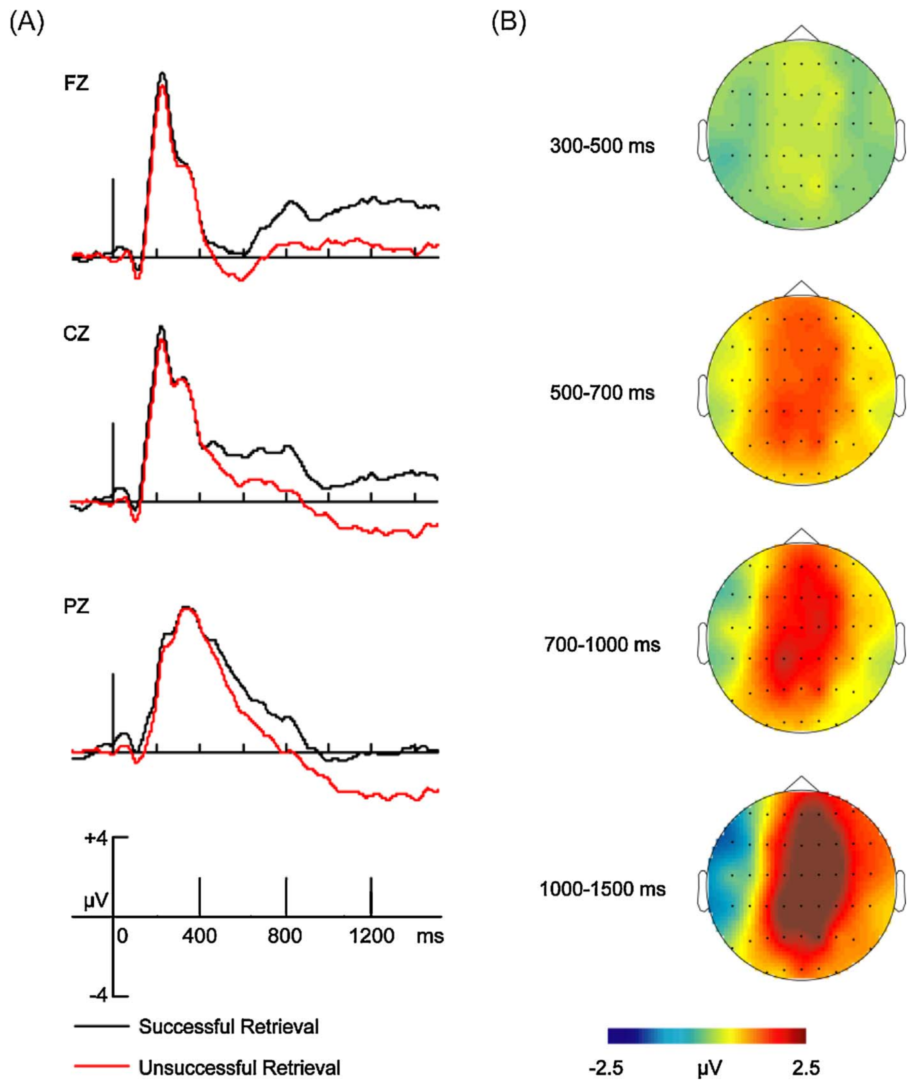

Fig. 4. Initial retrieval phase (retrieval practice condition): ERPs for items that were successfully retrieved and that could not be retrieved.
(A) The average ERP waveforms at electrodes Fz, Cz, and Pz for word pairs that were successfully retrieved and that could not be retrieved (unsuccessful retrieval); (B) Topographic maps of the ERP differences between items that were successfully and unsuccessfully retrieved, displayed for four time windows (300–500, 500–700, 700–1000, 1000–1500 ms).

$p = 0.244$, $\eta_p^2 = 0.09$]. Successfully retrieved items induced ERPs that were more positive [1.38(2.14)] than the ERPs for unsuccessfully retrieved items [0.61(2.32)], 95% CI [0.17,1.38].

Similarly, in the later time window (1000–1500 ms), there was a significant Retrieval Success main effect [$F(1, 15) = 22.98$, $p < 0.001$, $\eta_p^2 = 0.61$], but the Retrieval Success × Cluster interaction was not significant [$F(2, 30) = 1.70$, $p = 0.211$, $\eta_p^2 = 0.10$]. Successfully retrieved items induced ERPs that were more positive [1.26(2.56)] than the ERPs for unsuccessfully retrieved items [−0.09(2.51)], 95% CI [0.75,1.95].

3.2.2.2. SMEs comparisons. In the 300–500 ms time interval, the Condition × Subsequent Memory × Cluster ANOVA exhibited a significant Condition main effect, a significant Subsequent Memory main effect, and a significant Condition × Subsequent Memory × Cluster interaction (see Table 3). The Condition × Subsequent Memory and Condition × Cluster interactions were not significant (all *F*'s < 1). Subsidiary ANOVAs analyzing the SMEs in the elaborative study and retrieval practice conditions separately revealed additional effects (see Table 4). For the elaborative study condition, there was a Subsequent Memory main effect and a significant Subsequent Memory × Cluster interaction. Follow-up analyses for each electrode cluster using pairwise *t*-tests showed later remembered items elicited significantly more positive ERPs than later forgotten items in the frontal cluster [subsequently remembered minus forgotten = 1.11(1.42), $t(15) = 3.12$, $p = 0.007$, 95% CI [0.35, 1.86], $\eta_p^2 = 0.39$], whereas no significant SMEs were observed in the other two clusters [central: 0.65(1.33), $t(15) = 1.95$, $p = 0.07$, 95% CI [−0.06, 1.36], $\eta_p^2 = 0.20$; parietal: 0.09(1.36), $t(15) = 0.28$, $p = 0.79$, 95% CI [−0.63, 0.82], $\eta_p^2 = 0.005$]. For the retrieval practice condition, there were no SMEs in this latency range [Subsequent Memory main effect: $F(1, 15) = 1.62$, $p = 0.22$, $\eta_p^2 = 0.10$; Subsequent Memory × Cluster interaction: $F < 1$], see Fig. 5.

In the 500–700 ms time window, the omnibus repeated measures ANOVA revealed significant main effects of Condition and Subsequent Memory, as well as a significant Condition × Cluster interaction. There was also a significant Condition × Subsequent Memory × Cluster interaction (Table 3). The subsidiary ANOVA (Table 4) for the elaborative study condition revealed no Subsequent Memory main effect [$F(1, 15) = 3.88$, $p = 0.07$, $\eta_p^2 = 0.21$], although the Subsequent Memory × Cluster interaction was significant. Follow-up pairwise comparisons for each electrode cluster showed that subsequently remembered items induced ERPs that were more positive than those for subsequently forgotten items in the frontal cluster [subsequently remembered minus forgotten = 0.99(1.70), $t(15) = 2.33$, $p = 0.034$,

Table 3
The Condition, Subsequent Memory, and Cluster effects for each time window; Follow-up results are shown in Table 4 for analyses with significant Condition × Subsequent Memory × Cluster interactions.

Window (ms)	Condition Main Effect	Subsequent Memory Main Effect	Condition × Subsequent Memory Interaction	Condition × Cluster Interaction	Condition × Subsequent Memory × Cluster Interaction
300–500	$F(1, 15) = 5.80$ $p = 0.029$ $\eta_p^2 = 0.28$	$F(1, 15) = 5.36$ $p = 0.035$ $\eta_p^2 = 0.26$	*n.s.*	*n.s.*	$F(2, 30) = 4.40$ $p = 0.044$ $\varepsilon = 0.60$ $\eta_p^2 = 0.23$
500–700	$F(1, 15) = 25.67$ $p < 0.001$ $\eta_p^2 = 0.63$	$F(1, 15) = 7.35$ $p = 0.016$ $\eta_p^2 = 0.33$	*n.s.*	$F(2, 30) = 14.15$ $p = 0.001$ $\varepsilon = 0.61$ $\eta_p^2 = 0.49$	$F(2, 30) = 17.68$ $p < 0.001$ $\varepsilon = 0.69$ $\eta_p^2 = 0.54$
700–1000	$F(1, 15) = 19.35$ $p = 0.001$ $\eta_p^2 = 0.56$	$F(1, 15) = 5.56$ $p = 0.032$ $\eta_p^2 = 0.27$	$F(1, 15) = 4.29$ $p = 0.056$ $\eta_p^2 = 0.22$	$F(2, 30) = 9.26$ $p = 0.001$ $\varepsilon = 0.71$ $\eta_p^2 = 0.38$	$F(2, 30) = 15.16$ $p < 0.001$ $\varepsilon = 0.64$ $\eta_p^2 = 0.50$
1000–1500	*n.s.*	*n.s.*	*n.s.*	*n.s.*	$F(2, 30) = 4.37$ $p = 0.036$ $\varepsilon = 0.73$ $\eta_p^2 = 0.23$

Non-significant is abbreviated as n.s.

95% CI [0.08, 1.90], $\eta_p^2 = 0.27$], whereas there were no significant differences for ERPs induced by later remembered and forgotten items in the central and parietal clusters [central: 0.66(1.34), $t(15) = 1.96$, $p = 0.07$, 95% CI [−0.06, 1.38], $\eta_p^2 = 0.20$; parietal: 0.31(1.28), $t(15) = 0.97$, $p = 0.346$, 95% CI [−0.37, 0.99], $\eta_p^2 = 0.06$; Fig. 5]. Subsidiary ANOVAs (Table 4) for the retrieval practice condition showed no Subsequent Memory main effect [$F(1, 15) = 2.54$, $p = 0.13$, $\eta_p^2 = 0.15$], but a significant Subsequent Memory × Cluster interaction was present. Pairwise *t*-tests showed subsequently remembered items produced significantly more positive ERPs than subsequently forgotten items in the parietal cluster [subsequently remembered minus forgotten = 1.14(1.57), $t(15) = 2.90$, $p = 0.01$, 95% CI [0.30, 1.97], $\eta_p^2 = 0.36$], whereas no SMEs occurred in the more anterior two clusters [frontal: 0.17(1.69), $t(15) = 0.41$, $p = 0.69$, 95% CI [−0.73, 1.08], $\eta_p^2 = 0.01$; central: 0.52(1.62), $t(15) = 1.28$, $p = 0.22$, 95% CI [−0.35, 1.35], $\eta_p^2 = 0.10$; Fig. 5].

In the 700–1000 ms time window, there were significant main effects of Condition and Subsequent Memory (see Table 3). The ANOVA revealed a marginally significant Condition × Subsequent Memory interaction as well as a significant Condition × Cluster interaction. The Condition × Subsequent Memory × Cluster 3-way interaction was also significant. Subsidiary ANOVAs for the elaborative study condition revealed no Subsequent Memory main effect ($F < 1$), but a significant Subsequent Memory × Cluster interaction. Follow-up pairwise comparisons for each electrode cluster indicated that subsequently remembered items gave rise to more negative ERPs than subsequently forgotten items in the parietal cluster [subsequently remembered minus forgotten = −0.65(1.17), $t(15) = -2.21$, $p = 0.043$, 95% CI [−1.27, −0.02], $\eta_p^2 = 0.25$], i.e. there was a polarity reversal of the SME for the elaborative study condition over the parietal region in this latency

Table 4
Subsequent Memory and Cluster effects are shown for the retrieval practice and elaborative study conditions for each time window for which significant Condition × Subsequent Memory × Cluster interactions were revealed in the initial ANOVA (Table 3).

Window (ms)	Elaborative Study condition			Retrieval Practice condition		
	Subsequent Memory Main Effect	Subsequent Memory × Cluster Interaction	Pairwise Comparisons	Subsequent Memory Main Effect	Subsequent Memory × Cluster Interaction	Pairwise Comparisons
300–500	$F(1, 15) = 4.59$ $p = 0.049$ $\eta_p^2 = 0.23$	$F(2, 30) = 4.89$ $p = 0.029$ $\varepsilon = 0.68$ $\eta_p^2 = 0.25$	Frontal: SR > SF Central: *n.s.* Parietal: *n.s.*	*n.s.*	*n.s.*	–
500–700	*n.s.*	$F(2, 30) = 3.51$ $p = 0.043$ $\eta_p^2 = 0.19$	Frontal: SR > SF Central: *n.s.* Parietal: *n.s.*	*n.s.*	$F(2, 30) = 8.14$ $p = 0.006$ $\varepsilon = 0.68$ $\eta_p^2 = 0.35$	Frontal: *n.s.* Central: *n.s.* Parietal: SR > SF
700–1000	*n.s.*	$F(2, 30) = 8.47$ $p = 0.005$ $\varepsilon = 0.69$ $\eta_p^2 = 0.36$	Frontal: *n.s.* Central: *n.s.* Parietal: SR < SF	$F(1, 15) = 6.81$ $p = 0.02$ $\eta_p^2 = 0.31$	*n.s.*	–
1000–1500	*n.s.*	*n.s.*	–	*n.s.*	$F(2, 30) = 3.36$ $p = 0.048$ $\eta_p^2 = 0.18$	Frontal: *n.s.* Central: *n.s.* Parietal: SR > SF

The Symbol " > " shown in the table, for example, "SR > SF" in pairwisecomparisons column means that amplitudes induced by subsequently remembered items are significantly more positive than those induced by subsequently forgotten items. η_p^2 (partial eta squared) is used as index of effect size. "-" represents that pairwise comparisons are not conducted when the interaction is not significant. Non-significant is abbreviated as n.s.

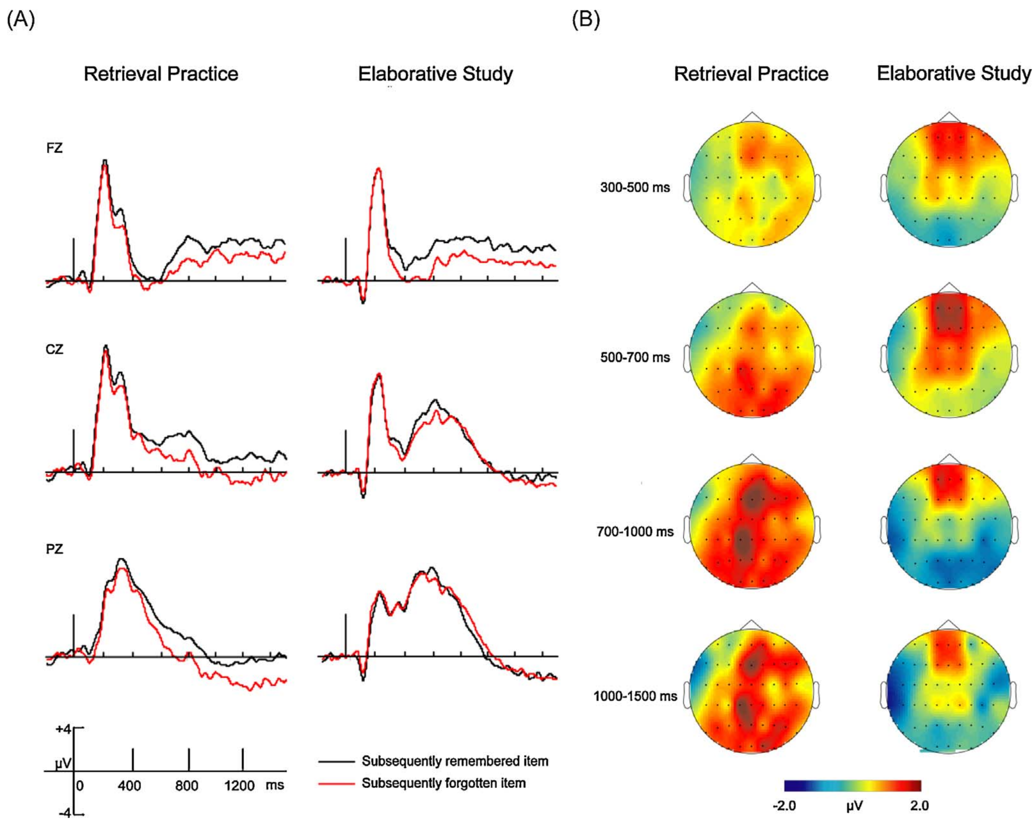

Fig. 5. Initial retrieval phase: SMEs for the retrieval practice and elaborative study conditions.
(A) The average ERP waveforms for subsequently remembered items and forgotten items in the retrieval practice and elaborative study conditions in the initial retrieval phase, and (B) topographic maps of the SMEs (differences between subsequently remembered and forgotten items) for the two conditions in four time windows (300–500, 500–700, 700–1000, 1000–1500 ms).

range. There were no significant differences over the other two clusters [frontal: 0.70(1.63), $t(15) = 1.71$, $p = 0.108$, 95% CI [−0.17, 1.57], $\eta_p^2 = 0.163$; central: −0.08(0.82), $t(15) = -0.41$, $p = 0.691$, 95% CI [−0.52, 0.36], $\eta_p^2 = 0.01$], see Fig. 5 and Table 4.

Subsidiary ANOVAs comparing ERP differences between subsequently remembered and forgotten items of the retrieval practice condition revealed a significant Subsequent Memory main effect. There was no significant Subsequent Memory × Cluster interaction [$F(2, 30) = 2.21$, $p = 0.145$, $\varepsilon = 0.72$, $\eta_p^2 = 0.13$]. The results indicated that subsequently remembered items [1.72(2.03)] induced more positive ERPs than subsequently forgotten items [0.70(2.56)], 95% CI [0.19, 1.85]. The SME of the retrieval practice condition was widely distributed, as can be seen in Fig. 5, and was not modulated by Cluster.

In the 1000–1500 ms time window, there were no significant main effects of Condition or Subsequent Memory, and no 2-way interactions involving Condition (all p's > 0.1). However, there was a significant Condition × Subsequent Memory × Cluster 3-way interaction (Table 3). Subsidiary ANOVAs comparing ERP differences between subsequently remembered and forgotten items in the elaborative study condition showed no Subsequent Memory main effect or significant Subsequent Memory × Cluster interaction (all F's < 1), see Table 4. Subsidiary ANOVAs comparing ERP differences between subsequently remembered and forgotten items in the retrieval practice condition revealed no Subsequent Memory main effect [$F(1, 15) = 2.55$, $p = 0.131$, $\eta_p^2 = 0.15$]. However, a significant Subsequent Memory × Cluster interaction was observed. Pairwise t-tests showed later remembered items induced more positive ERPs than later forgotten items in the parietal area [later remembered-forgotten: 1.21(1.97), $t(15) = 2.45$, $p = 0.027$, 95% CI [0.16, 2.26], $\eta_p^2 = 0.29$]. No SMEs were found in the other two clusters [frontal: 0.44(1.92), $t(15) = 0.91$, $p = 0.377$, 95% CI [−0.59, 1.46], $\eta_p^2 = 0.05$; central: 0.65(2.22), $t(15) = 1.16$, $p = 0.263$, 95% CI [−0.54, 1.83], $\eta_p^2 = 0.08$]. The results revealed that a SME with a parietal distribution occurred in the retrieval practice condition, whereas no SME was present in the elaboration condition at this latency (Fig. 5; Table 4).

3.2.2.3. SMEs topographic analyses. In the 300–500 ms time window, the Condition × Electrode Site ANOVA revealed there was no Condition main effect [$F(1, 15) = 0.05$, $p = 0.83$, $\eta_p^2 = 0.003$], whereas the Condition × Electrode Site interaction was significant [$F(61, 915) = 1.61$, $p = 0.003$, $\eta_p^2 = 0.10$]. The results suggested the scalp distributions of SMEs were significantly different between the retrieval practice and elaborative study conditions from 300 to 500 ms. The ANOVAs for SMEs in these two conditions and the visual inspection of the topographic map in the 300–500 ms time interval indicated a frontally distributed SME occurred in the elaborative study condition, but not in the retrieval practice condition.

In the 500–700 ms time window, the ANOVA revealed no Condition main effect ($F < 1$), but a significant Condition × Electrode Site interaction [$F(61, 915) = 2.77$, $p < 0.001$, $\eta_p^2 = 0.16$]. The results suggested SMEs for the retrieval practice and elaborative study conditions had different scalp distributions from 500 to 700 ms. The ANOVAs for SMEs in these two conditions and the visual inspection of the topographic map in the 500–700 ms time interval showed that a frontally distributed SME was present in the elaborative study condition, whereas a parietally distributed SME occurred in the retrieval practice condition.

In the 700–1000 ms time window, there was a significant Condition main effect [$F(1, 15) = 5.58$, $p = 0.03$, $\eta_p^2 = 0.27$] and the interaction between Condition and Electrode Site was also significant [$F(61, 915) = 2.48$, $p < 0.001$, $\eta_p^2 = 0.14$]. The results indicated that the scalp distributions of SMEs differed between the retrieval practice and elaborative study conditions. Converging results of ANOVAs for SMEs in these two conditions and the visual inspection of the topographic map in the 700–1000 ms time interval indicated that a parietal SME with reversed (negative) polarity occurred in the elaborative study condition, whereas the SME in the retrieval practice condition was widely distributed and had a positive polarity.

In the later time window (1000–1500 ms), there was neither a Condition main effect nor a Condition × Electrode Site interaction (all F's < 1). Thus, the SMEs scalp topographies of the retrieval practice and elaborative study conditions did not differ in this latency range.

3.2.2.4. SME-behavior correlations. Pearson's correlation coefficients investigating the associations between SMEs and final test performance revealed no significant correlations between the SMEs and final test performance in any condition (Hit_O, Hit_R, FA_O, FA_R, FA_N_O, FA_N_R, data not shown).

4. Discussion

This study examined the behavioral retrieval practice effect, and sought to dissociate the underlying mechanisms of retrieval practice and elaborative study. Behaviorally, the results from Experiment 1 showed that the old/new discrimination was superior in the retrieval practice condition as compared to that in the elaborative study and re-studying conditions. But the old/new discrimination of the elaborative study and re-studying conditions was almost identical. Experiment 2 results replicated the behavioral retrieval practice effect, as the old/new discrimination of the retrieval practice condition was again better than that of the elaborative study condition (the effect size was large according to Cohen's criteria; Cohen, 1973). The old/recombined discrimination did not, however, vary between conditions, likely because only a single study-test cycle was used. The behavioral results are widely in line with previous behavioral studies on the retrieval practice effect (Karpicke & Blunt, 2011; Karpicke & Smith, 2012; Lehman et al., 2014). Participants are in retrieval mode when they actively retrieve targets (Tulving, 1983; Rugg & Wilding, 2000). Conceptually, retrieval mode refers to the cognitive state in which items are processed as retrieval cues to initiate the recall of prior spatiotemporal context information. As proposed by the episodic context account, the diagnostic value and effectiveness of retrieval cues presumably increase by recollecting the previous episode and restricting the search set, which improves the participants' ability to distinguish a particular target from other potential candidates (Gao et al., 2016; Karpicke et al., 2014; Karpicke & Blunt, 2011).

In addition, Karpicke and Zaromb (2010) have argued that retrieval practice enhances item-specific processing. Enhanced item-specific characteristics increase the distinctiveness of retrieved items and enable participants to distinguish these items from nontargets more easily (Peterson & Mulligan, 2013; Mulligan & Peterson, 2015). However, previous studies indicate that distinctiveness is related to recollection-based retrieval (Park, Arndt, & Reder, 2006). The observed old/new effects at the final test phase do not provide evidence that participants relied on recollection when making the old/recombined/new decisions, making it unlikely that actively retrieved items were indeed highly distinct. One might speculate that a single study-test run is insufficient to increase the distinctiveness of actively retrieved items in a one-week delayed test.

The behavioral results showed that semantic elaboration was not associated with the same levels of learning as those produced by retrieval practice and did not lead to better retrieval accuracy than re-studying, which contradicts the elaborative retrieval hypothesis for the retrieval practice effect. For elaborative study, the participants need to encode and conceptualize more semantic features about the materials on the basis of prior formed memory traces (Craik, 2002). However, such an elaboration of memory traces may produce cue overload and decrease the likelihood of recovering a particular target (Watkins & Watkins, 1976). Moreover, for the elaborative study task, participants increase the number of encoded features to generate semantic mediators, which may in turn decrease the distinctiveness of a particular pair (Park et al., 2006). Cue overload and reduced distinctiveness represent tentative explanations for why the hit rate in the delayed test for the elaborative study condition was not superior to re-studying.

In both experiments, aside from an increased hit rate, the recombined word pairs were more likely to be wrongly considered as old word pairs in the retrieval practice condition than in other conditions. Thus, the increased hit rate (Hit_O) in the retrieval practice condition was accompanied by an increased false alarm rate to recombined word pairs (FA_R), see Supplementary Materials. This reveals that participants were more liberal in giving old responses in the retrieval practice condition. We speculate that this increased false alarm rate indicates that the retrieval-based learning is not finalized after a single test. In previous studies on the retrieval practice effect (Karpicke & Roediger, 2007, 2008), participants usually have to retrieve the materials repeatedly in order to ensure an adequate initial learning. In contrast, in our study, each word pair was only tested once during the initial retrieval phase. By using just a single initial retrieval phase, we sought to isolate the retrieval practice effect and avoid the influence of repetition (as Wing et al., 2013). However, this procedure may be insufficient to avoid such high false alarm rates for previously tested but subsequently recombined word pairs in the delayed test.

4.1. Retrieval success effect

Based on previous studies about the retrieval success effect that have used a cued-recall task, retrieval success may be associated with various brain regions, such as the inferior prefrontal cortex, parietal cortex and middle temporal gyrus (Hayama, Vilberg, & Rugg, 2012). Tibon and Levy (2013) examined the ERP retrieval success effect using a cued-recall task with pictorial materials. They found that the retrieval success effect was characterized by markedly early scalp potential differences between 350 and 600 ms over frontal areas. In contrast, the present study showed widespread and long-lasting retrieval success effects, which more resembled the retrieval success effects which have been observed in word-stem cued-recall tasks (Allan, Doyle, & Rugg, 1996; Allan, Wolf, Rosenthal, & Rugg, 2001). The present study did not observe a distinct LPC, perhaps because successful recall in the cued-recall task is related to retrieval of contextual information, and also related to the evaluation of whether the information represents an appropriate prior episode (Hayama et al., 2012).

4.2. Frontal SMEs

A frontal SME, starting at 300 ms and extending to 700 ms, occurred in the elaborative study condition, but such a frontal SME was absent in the retrieval practice condition. There could be two potential interpretations for the frontal SME in the elaborative study condition: One possibility is that this early SME is associated with semantic processing and linked to the N400. Previous electrophysiological studies have suggested that the N400 reflects automatic and implicit semantic processing for a variety of meaningful materials, such as words, pictures, and sounds (Kutas & Federmeier, 2011; Voss & Federmeier, 2011). However, such an interpretation of the observed frontal SME must be considered with caution, bearing in mind the typically centro-parietal distribution of the N400. According to the proposed semantic interpretation, for elaborative study, participants need to encode semantic attributes of cues and targets and process the semantic relationship between them in order to generate associates. The effectiveness of processing this semantic relationship could influence later memory success. In contrast, the cued-recall task in the retrieval practice condition does not require participants to process the semantic relationship.

The second possibility is that the early frontal SME indicates working memory-based elaboration (Kamp & Zimmer, 2015). Kamp and Zimmer (2015) observed a similar frontal SME, which has been interpreted as active elaboration of representations in working memory. By this interpretation, this effect may be related to varying working memory load (Bosch, Mecklinger, & Friederici, 2001; Lehnert & Zimmer, 2008; Ruchkin, Johnson, Canoune, & Ritter, 1990) and deep encoding (Schott, Richardson-Klavehn, Heinze, & Düzel, 2002; Guo, Zhu, Ding, Fan, & Paller, 2004). Thus, the early frontal SME in the elaborative study condition may indicate the working memory load with which cues and targets were elaborated during this task, and increased working memory load may have resulted in better retention.

Future studies may address the issue of the functional significance of the frontal SME by contrasting the SME for Hit_O and FA_O, which could not be done in the current study due to insufficient trial numbers, or by manipulating the working memory demands during elaboration.

4.3. Parietal SMEs

A parietally distributed SME occurred in the retrieval practice condition from 500 to 700 ms, but not in the elaborative study condition. A topographic map contrast revealed that the SMEs for the retrieval practice and elaborative study conditions had different scalp distributions from 500 to 700 ms.

Our results replicate the recent finding of Bai et al. (2015), who observed a SME for retrieval practice. The authors assumed this SME to be functionally similar to the parietal old/new effect, indexing recollection (Mecklinger, 2006; Rugg & Curran, 2007). Notably, Bai and colleagues (2015) showed a widespread distributed effect rather than a localized parietal effect as in our study. As proposed by the episodic context account, during retrieval practice, participants need to reinstate the contextual features from when an event occurred (e.g., temporal order information) and then incorporate the retrieved context with the current context representation. In contrast, voluntary recollection can be presumed absent during elaborative study. Our study showed a parietal SME for the retrieval practice condition, which resembled the LPC in its temporal and spatial characteristics, but showed no such effect for the elaborative study condition. Previous studies have suggested the LPC is related to episodic recollection (e.g. Rugg & Curran, 2007). Considering the spatiotemporal characteristics of the effect and the cued-recall task that participants completed in the retrieval practice condition, we speculate that the parietal SME in the 500–700 ms time window during retrieval practice may be associated with successful recollection processes that benefit later retrieval. In line with this assumption, retrieval accuracy in the final test phase was much better when the initial retrieval was successful than when it was not. Unfortunately, we could not run an analogue ERP analysis (SME for initially successful retrieval trials vs. SME for initially unsuccessful trials), due to insufficient trial counts.

The topographic comparison of SMEs for the retrieval practice and elaborative study conditions in the 500–700 ms time interval also showed different spatial distributions of SMEs for these two conditions. The ANOVAs of SMEs for these two conditions and the visual inspection of the topographic map in the 500–700 ms time interval indicated that a SME with frontal distribution was present in the elaborative study condition, whereas a SME with parietal distribution occurred in the retrieval practice condition. The dissociated SMEs underline that different encoding processes are involved in retrieval practice and semantic elaboration.

4.4. Late SMEs

For the retrieval practice condition, a topographically widespread late SME was present from 700 to 1000 ms, but this positive SME was absent in the elaborative study condition. The topographic comparison revealed that the SMEs for the retrieval practice condition and the ERP retrieval success effect were topographically indistinct in this interval [the Condition × Electrode Site interaction was not significant; $F(61, 915) = 1.71$, $p = 0.12$, $\eta_p^2 = 0.10$]. The SME for the retrieval practice condition showed similarities to the retrieval success effect, as observed in the initial retrieval phase. This suggests that these processes, related to retrieval of contextual information and to operations on information retrieved in response to a cue, may promote later memory performance. Furthermore, the retrieval practice task used in our study requires subjects to recollect the associative targets, which suggests that successful retrieval is mainly (but not exclusively) due to recruiting recollection processes (Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1997). The late SME findings support the episodic context account, which assumes that the retrieval practice effect may be driven by the engagement of episodic reinstatement. The SME for the retrieval practice condition extended to 1500 ms, whereas no such late SME occurred in the elaborative study condition, indicating that the SME for the former condition outlasted the SME for the latter condition.

In contrast, the SME of the elaborative study condition showed some polarity reversal over parietal areas from 700 to 1000 ms. The subsequently remembered items induced ERPs that were more negative than the ERPs induced by the subsequently forgotten items over parietal areas (the effect size was large), which was the opposite of the SME in the retrieval practice condition. We assume that a different set of neural structures may have been engaged in generating such negative SMEs in the elaborative study condition, as compared to the positive SMEs in the active retrieval condition (Otten & Rugg, 2001). During retrieval practice, participants completed a cued-recall task, where they needed to recall the targets that were shown in the previous study episode, whereas during elaboration the participants had to retrieve stored knowledge from semantic memory to generate related words. Episodic memory is subserved by widely distributed brain regions, which may overlap with and extend beyond areas involved in semantic memory (Tulving, 2002).

Two caveats of the present study should be pointed out: For the analysis of SMEs in the retrieval practice condition, we needed to merge trials of initially successful retrieval attempts and initially unsuccessful retrieval attempts. Previous studies have indicated that both successful retrieval and unsuccessful retrieval that are followed by feedback can benefit later memory performance (Kornell, Klein, & Rawson, 2015; Kang, McDermott, & Roediger, 2007). However, retrieval success and retrieval failure produce different neural correlates (McDermott, Jones, Petersen, Lageman, & Roediger, 2000; Konishi, Wheeler, Donaldson, & Buckner, 2000). The current study could not differentiate the SMEs for items that were initially successfully retrieved and the SMEs for items that were initially not successfully retrieved, due to insufficient trial numbers for these sub-categories. Future studies focusing on the active retrieval condition may allow such differentiation and provide additional insights about the neurocognitive mechanisms underlying retrieval practice. Another limitation is that the present study included only a single study-test cycle (instead of multiple cycles). Findings during the initial (and here only) study-test cycle might not fully reveal the neurocognitive effects of the retrieval practice effect. Based on the observed old/new effects in the final test phase, participants apparently depended on familiarity rather than on recollection to make their old/recombined/new judgments in the one-week delayed test, which strongly suggests that the retrieval-based learning was not finalized after a single study-test cycle. We presume that the degree to which subjects rely on familiarity would likely decrease after repeated study-test cycles.

5. Conclusion

The current study dissociated the mnemonic effects of retrieval practice and elaboration for both the behavioral and ERP findings. Elaboration did not lead to superior retrieval accuracy in the delayed test, and the SMEs during encoding varied between elaboration and active retrieval. Both findings provide evidence against the elaborative retrieval hypothesis for the retrieval practice effect. Our findings suggest that context reinstatement, rather than semantic elaborative processing, is of importance for this effect. On the basis of the observed improved concurrent behavioral memory retention and parietally distributed SME for the retrieval practice condition, we conclude that episodic recollection occurring in the retrieval practice condition is crucial for the retrieval practice effect, providing new evidence for the episodic context account. To our knowledge, this is the first electrophysiological study dissociating mnemonic effects of retrieval practice and elaboration.

Acknowledgments

The present study was supported by National Natural Science Foundation of China (31671127), Beijing Municipal Commission of Education Key Program of Science and Technology (KZ201410028034) and Beijing Advanced Innovation Center for Imaging Technology (BAICIT-2016018). T.R. was financially supported by the German Research Foundation (DFG Ro 3533/1-1). We greatly appreciate Erik A. Wing for providing his experimental materials. We thank Xinrui Mao and Bingbing Li for helpful comments and data analysis assistance. We also thank Mengtong Cai and Meng Han for critical reading of the manuscript and research assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biopsycho.2017.05.010.

References

Allan, K., Doyle, M. C., & Rugg, M. D. (1996). An event-related potential study of word-stem cued recall. *Cognitive Brain Research*, *4*(4), 251–262. http://dx.doi.org/10.1016/S0926-6410(96)00061-4.

Allan, K., Wolf, H. A., Rosenthal, C. R., & Rugg, M. D. (2001). The effect of retrieval cues on post-retrieval monitoring in episodic memory: an electrophysiological study. *Cognitive Brain Research*, *12*(2), 289–299. http://dx.doi.org/10.1016/S0926-6410(01)00061-1.

Bai, C. H., Bridger, E. K., Zimmer, H. D., & Mecklinger, A. (2015). The beneficial effect of testing: An event-related potential study. *Frontiers in Behavioral Neuroscience*, *9*, 248–259. http://dx.doi.org/10.3389/fnbeh.2015.00248.

Bosch, V., Mecklinger, A., & Friederici, A. D. (2001). Slow cortical potentials during retention of object, spatial, and verbal information. *Cognitive Brain Research*, *10*(3), 219–237. http://dx.doi.org/10.1016/S0926-6410(00)00040-9.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268–276. http://dx.doi.org/10.3758/BF03193405.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*(3), 474–478. http://dx.doi.org/10.3758/BF03194092.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1563–1569. http://dx.doi.org/10.1037/a0017021.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(6), 1547–1552. http://dx.doi.org/10.1037/a0024140.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement*, *33*(1), 107–112. http://dx.doi.org/10.1177/001316447303300111.

Craik, F. I. M. (2002). Human memory and aging. In L. Bäckman, & C. von Hofsten (Eds.), *Psychology at the turn of the Millennium* (pp. 261–280). Hove, UK: Psychology Press.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters*, *505*(1), 36–40. http://dx.doi.org/10.1016/j.neulet.2011.08.061.

Evans, L. H., Williams, A. N., & Wilding, E. L. (2015). Electrophysiological evidence for retrieval mode immediately after a task switch. *Neuroimage*, *108*, 435–440. http://dx.doi.org/10.1016/j.neuroimage.2014.12.068.

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, *51*(1), 6–28. http://dx.doi.org/10.1002/1097-0029(20001001)51:1<6:AID-JEMT2>3.3.CO;2-I.

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 re-studying. *Cognitive Affective & Behavioral Neuroscience*, *16*(6), 977–990. http://dx.doi.org/10.3758/s13415-016-0446-z.

Guo, C., Zhu, Y., Ding, J., Fan, S., & Paller, K. A. (2004). An electrophysiological investigation of memory encoding, depth of processing, and word frequency in humans. *Neuroscience Letters*, *356*(2), 79–82. http://dx.doi.org/10.1016/j.neulet.2003.09.049.

Hashimoto, T., Usui, N., Taira, M., & Kojima, S. (2011). Neural enhancement and attenuation induced by repetitive recall. *Neurobiology of Learning and Memory*, *96*(2), 143–149. http://dx.doi.org/10.1016/j.nlm.2011.03.008.

Hayama, H. R., Vilberg, K. L., & Rugg, M. D. (2012). Overlap between the neural correlates of cued recall and source memory: Evidence for a generic recollection network? *Journal of Cognitive Neuroscience*, *24*(5), 1127–1137. http://dx.doi.org/10.1162/jocn_a_00202.

Herron, J. E., & Wilding, E. L. (2005). An electrophysiological investigation of factors facilitating strategic recollection. *Journal of Cognitive Neuroscience*, *17*(5), 777–787. http://dx.doi.org/10.1162/0898929053747649.

Herron, J. E., & Wilding, E. L. (2006). Brain and behavioral indices of retrieval mode. *Neuroimage*, *32*(2), 863–870. http://dx.doi.org/10.1016/j.neuroimage.2006.03.046.

Kamp, S. M., & Zimmer, H. D. (2015). Contributions of attention and elaboration to associative encoding in young and older adults. *Neuropsychologia*, *75*, 252–264. http://dx.doi.org/10.1016/j.neuropsychologia.2015.06.026.

Kang, S. H., McDermott, K. B., & Roediger, H. L., III (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. http://dx.doi.org/10.1080/09541440601056620.

Kang, S. H. K. (2010). Enhancing visuo-spatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*(8), 1009–1017. http://dx.doi.org/10.3758/MC.38.8.1009.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. http://dx.doi.org/10.1126/science.1199327.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. http://dx.doi.org/10.1016/j.jml.2006.09.004.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. http://dx.doi.org/10.1016/j.jml.2006.09.004.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17–29. http://dx.doi.org/10.1016/j.jml.2012.02.004.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*(3), 227–239. http://dx.doi.org/10.1016/j.jml.2009.11.010.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, *61*, 237–284.

Keresztes, A., Kaiser, D., Kovács, G., & Racsmány, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex*, *24*(11), 3025–3035. http://dx.doi.org/10.1093/cercor/bht158.

Konishi, S., Wheeler, M. E., Donaldson, D. I., & Buckner, R. L. (2000). Neural correlates of episodic retrieval success. *Neuroimage*, *12*(3), 276–286. http://dx.doi.org/10.1006/nimg.2000.0614.

Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(1), 283–294. http://dx.doi.org/10.1037/a0037850.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. http://dx.doi.org/10.1146/annurev.psych.093008.131123.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 1–8. http://dx.doi.org/10.1037/xlm0000012.

Lehnert, G., & Zimmer, H. D. (2008). Common coding of auditory and visual spatial information in working memory. *Brain Research*, *1230*(14), 158–167. http://dx.doi.org/10.1016/j.brainres.2008.07.005.

McDermott, K., Jones, T., Petersen, S., Lageman, S., & Roediger, H. (2000). Retrieval success is accompanied by enhanced activation in anterior prefrontal cortex during recognition memory: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *12*(6), 965–976. http://dx.doi.org/10.1162/08989290051137503.

Mccarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*(3), 203–208. http://dx.doi.org/10.1016/0168-5597(85)90015-2.

Mecklinger, A. (2006). Electrophysiological measures of familiarity memory. *Clinical EEG and Neuroscience*, *37*(4), 292–299. http://dx.doi.org/10.1177/155005940603700406.

Mulligan, N. W., & Peterson, D. J. (2015). Negative and positive testing effects in terms of item-specific and relational information. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(3), 859–871. http://dx.doi.org/10.1037/xlm0000056.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://www.usf.edu/FreeAssociation/.

Otten, L. J., & Rugg, M. D. (2001). Electrophysiological correlates of memory encoding are task-dependent. *Cognitive Brain Research*, *12*(1), 11–18. http://dx.doi.org/10.1016/S0926-6410(01)00015-5.

Park, H., Arndt, J., & Reder, L. M. (2006). A contextual interference account of distinctiveness effects in recognition. *Memory & Cognition*, *34*(4), 743–751. http://dx.doi.org/10.3758/BF03193422.

Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(4), 1287–1293. http://dx.doi.org/10.1037/a0031337.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. http://dx.doi.org/10.1017/s0048577200000305.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003.

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. *Cognitive Affective & Behavioral Neuroscience*, *15*(1), 195–210. http://dx.doi.org/10.3758/s13415-014-0310-y.

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

Ruchkin, D. S., Johnson, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, *76*(5), 419–439. http://dx.doi.org/10.1016/0013-4694(90)90096-3.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. http://dx.doi.org/10.1016/j.tics.2007.04.004.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences*, *4*(3), 108–115. http://dx.doi.org/10.1016/S1364-6613(00)01445-5.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1997). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*(6), 2073–2083. http://dx.doi.org/10.1093/brain/119.6.2073.

Schott, B., Richardson-Klavehn, A., Heinze, H. J., & Düzel, E. (2002). Perceptual priming versus explicit memory: Dissociable neural correlates at encoding. *Journal of Cognitive Neuroscience*, *14*(4), 578–592. http://dx.doi.org/10.1162/08989290260045828.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*(6), 695–703. http://dx.doi.org/10.1111/j.1469-8986.1986.tb00696.x.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods*, *31*(1), 137–149. http://dx.doi.org/10.3758/BF03207704.

Tibon, R., & Levy, D. A. (2013). The time course of episodic associative retrieval: Electrophysiological correlates of cued recall of unimodal and crossmodal pair-associate learning. *Cognitive Affective & Behavioral Neuroscience*, *14*(1), 220. http://dx.doi.org/10.3758/s13415-013-0199-x.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. http://dx.doi.org/10.1146/annurev.psych.53.100901.135114.

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, *1122*(1), 161–170. http://dx.doi.org/10.1016/j.brainres.2006.09.023.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, *48*(4), 532–546. http://dx.doi.org/10.1111/j.1469-8986.2010.01085.x.

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society*, *7*(3), 289–291. http://dx.doi.org/10.3758/bf03337192.

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: An fMRI study of the testing effect. *Neuropsychologia*, *51*(12), 2360–2370. http://dx.doi.org/10.1016/j.neuropsychologia.2013.04.004.

van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, *78*, 94–102. http://dx.doi.org/10.1016/j.neuroimage.2013.03.071.