CrossMark

The role of retrieval mode and retrieval orientation in retrieval practice: insights from comparing recognition memory testing formats and restudying

Chuanji Gao[1,2] · Timm Rosburg[3,4] · Mingzhu Hou[5] · Bingbing Li[1] · Xin Xiao[6] ·
Chunyan Guo[1,7]

Abstract The effectiveness of retrieval practice for aiding long-term memory, referred to as the *testing effect*, has been widely demonstrated. However, the specific neurocognitive mechanisms underlying this phenomenon remain unclear. In the present study, we sought to explore the role of pre-retrieval processes at initial testing on later recognition performance by using event-related potentials (ERPs). Subjects studied two lists of words (Chinese characters) and then performed a recognition task or a source memory task, or restudied the word lists. At the end of the experiment, subjects received a final recognition test based on the remember–know paradigm. Behaviorally, initial testing (active retrieval) enhanced memory retention relative to restudying (passive retrieval). The retrieval mode at initial testing was indexed by more positive-going ERPs for unstudied items in the active-retrieval tasks than in passive retrieval from 300 to 900 ms. Follow-up analyses showed that the magnitude of the early ERP retrieval mode effect (300–500 ms) was predictive of the behavioral testing effect later on. In addition, the ERPs for correctly rejected new items during initial testing differed between the two active-retrieval tasks from 500 to 900 ms, and this ERP retrieval orientation effect predicted differential behavioral testing gains between the two active-retrieval conditions. Our findings confirm that initial testing promotes later retrieval relative to restudying, and they further suggest that adopting pre-retrieval processing in the forms of retrieval mode and retrieval orientation might contribute to these memory enhancements.

✉ Chunyan Guo
guocy@cnu.edu.cn

Chuanji Gao
chuanjigao@gmail.com

[1] Beijing Key Laboratory of Learning and Cognition, Department of Psychology, Capital Normal University, Beijing 100048, People's Republic of China

[2] Department of Psychology, Institute of Mind and Brain, University of South Carolina, Columbia, SC 29201, USA

[3] Department of Psychology, Experimental Neuropsychology Unit, Saarland University, Campus, D-66123 Saarbrücken, Germany

[4] Department of Forensic Psychiatry, University Psychiatric Clinics Basel, CH-4012 Basel, Switzerland

[5] Department of Psychology, University of Arizona, Tucson, AZ 85721, USA

[6] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[7] Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, People's Republic of China

It has been widely demonstrated by behavioral studies that taking an initial test can improve later memory, in contrast to mere restudying. This *active retrieval* not only provides feedback about what someone has learned, but also promotes long-term retention by the learner (for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006a).

Different ideas have been put forward to explain this testing effect phenomenon (for reviews, see Karpicke, Lehman, & Aue, 2014; Roediger & Butler, 2011), but, as was pointed out by Karpicke and colleagues (2014), the empirical evidence for many accounts is scarce or even conflicting. The transfer-

Springer

appropriate processing (TAP) theory, for example, purports that memory performance depends on how well the encoding operations match with those required at test, and processes associated with active retrieval under the testing context better match with processes in the final retrieval than do those associated with restudying (Morris, Bransford, & Franks, 1977; Roediger & Karpicke, 2006b; Thomas & McDaniel, 2007). However, as pointed out by Karpicke and colleagues (2014), the largest effects of retrieval practice are found after free recall or short test formats rather than after tests that match the final test. The elaborative retrieval hypothesis, as another example, suggests that recalling a cue can activate the semantically related information that will further faciliate later retrieval (Carpenter, 2009, 2011). However, this account is inconsistent with the cue overload principle (Watkins & Watkins, 1976), which claims that the likelihood for successful retrieval decreases when the number of associates to the cue increases (Karpicke et al., 2014). Moreover, Karpicke and colleagues (2014) argued that some accounts of the retrieval practice effect are correlational rather than explanatory. Bjork and Bjork's desirable-difficulties model (Bjork, 1994, 1999; Bjork & Bjork, 1992), for example, suggests that challenging learning conditions promote memory retention. Indeed, it has been shown that retrieval practice effects become larger with increasing retrieval difficulty (e.g., Pyc & Rawson, 2009; Roediger & Karpicke, 2006b), but the account itself does not provide an explanation how this effect actually occurs.

Karpicke and colleagues (2014) put forward a new, more comprehensive account for the retrieval practice effect. Their episodic context account suggests that retrieval practice improves later retrieval performances by providing more effective contextual cues. Specifically, at active retrieval, subjects would try to reinstate the prior study context and incorporate it with the current context, which would consequently make more distinct cues available for later retrieval (Karpicke et al., 2014; Lehman, Smith, & Karpicke, 2014).

The episodic context account makes the explicit prediction that intentional retrieval (being in an episodic retrieval mode) produces greater retrieval practice effects relative to incidental retrieval (Karpicke et al., 2014, p. 266). Intentional retrieval or being in an episodic retrieval mode is considered as a cognitive state wherein an individual consciously thinks of the past when he/she encounters a potential cue (e.g., a name or picture; Tulving, 1983). This state of "traveling back" to the study episode is assumed to be a prerequisite for successful episodic retrieval (Tulving, 1983, 2002). According to Rugg and Wilding (2000), retrieval mode constitutes a tonically maintained state that is entered when there is a need to engage in episodic retrieval, and, thus, it can be considered one form of pre-retrieval processing (see also Bridger & Mecklinger, 2012; Burgess & Shallice, 1996; Mecklinger, 2010). Experimentally, retrieval mode can be manipulated by varying the test requirements, for example by presenting a word cue in a direct, intentional memory task or in an indirect (incidental) task. The dichotomy between

intentional and incidental memory tasks is essentially parallel to the distinction between retrieval practice and repeated studying (Hornberger, Rugg, & Henson, 2006; Richardson-Klavehn & Bjork, 1988). Retrieval practice as a form of intentional retrieval requires retrieval mode, whereas restudying as a form of incidental retrieval does not. Behaviorally, it has already been found that intentional retrieval (being in retrieval mode) leads to better memory retention than incidental retrieval (Karpicke & Zaromb, 2010).

Retrieval orientation is another pre-retrieval process and is thought to determine the specific processing of a retrieval cue, depending on the to-be-retrieved episodic content or the task requirements (Rugg & Wilding, 2000). Thus, different processes are presumed to be involved when someone seeks to remember whether they have encountered a presented cue before in contrast to when they seek to remember the context of such encounter. Similarly, such cue processing is presumed to vary with the kind of to-be-retrieved episodic content (e.g., "Was the named object presented in red or green?" vs. "Did you perform task A or task B on this item?"). Retrieval orientation conceptualizes pre-retrieval processes, as it reflects processes that are present in retrieval attempts and that are not directly related to retrieval success (Mecklinger, 2010; Rugg & Wilding, 2000). Experimentally, retrieval orientation is therefore studied by investigating the processing of new items in memory tasks, which are by definition not subject to memory success. To our knowledge, the influence of adopting retrieval orientation on the retrieval practice effect has been investigated in only one behavioral study: Karpicke and coworkers compared memory accuracy after different initial tests (elaborative study task vs. standard yes/no recognition test vs. source constrained recognition test; quoting from Karpicke et al., 2014). Both tests led to better retrieval accuracy than with the elaborative study task. Moreover, the source constrained recognition test produced greater final recall than the standard yes/no recognition test. On the basis of the described findings, it appears reasonable to assume that the benefit of retrieval practice over restudy on later memory performance is modulated by adopting retrieval mode and retrieval orientation.

In addition to determining the neural basis of the retrieval practice effect, neuroimaging methods might be particularly useful for assessing the role of retrieval mode and retrieval orientation in retrieval practice. Functional magnetic resonance imaging (fMRI) studies on retrieval practice have shown more frontal activation (such as in the left inferior frontal gyrus, medial prefrontal cortex) for test than restudying in the initial test phase (van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013; see also Eriksson, Kalpouzos, & Nyberg, 2011; Hashimoto, Usui, Taira, & Kojima, 2011). Intriguingly, activation of similar prefrontal areas was found in studies on retrieval mode (Buckner et al., 1998; Herron & Wilding, 2006; Wagner, Desmond, Glover, & Gabrieli, 1998) and

retrieval orientation (Nolde, Johnson, & D'Esposito, 1998; Rugg, Fletcher, Chua, & Dolan, 1999). However, the activation of similar prefrontal areas by retrieval practice and by adopting retrieval mode/orientation represents only weak evidence for a relationship between retrieval mode, retrieval orientation, and the behavioral retrieval practice effect.

Given that lexical and memory processes take place within a few hundred milliseconds, the event-related potential (ERP) technique may be better suited to capture the neurocognitive processes underlying the testing effect due to its higher temporal resolution. Although the ERP technique has been used extensively in neuropsychological studies of recognition memory (for reviews, see Friedman & Johnson, 2000; Rugg & Curran, 2007), to our knowledge, only two ERP studies have focused on the testing effect. Bai, Bridger, Zimmer, and Mecklinger (2015) used ERPs to study subsequent memory effects at testing, but unfortunately the study revealed only a marginally significant behavioral effect of testing. Rosburg, Johansson, Weigl, and Mecklinger (2015) had subjects study perceived and imagined items in an initial encoding phase, and then make three-key responses (perceived, imagined, or new) in two consecutive source memory tests. Half of the studied items were shown in the first test and all items in the second test. The authors found that, as compared to previously untested items, previously tested items elicited a stronger late parietal complex (LPC) ERP component, which is believed to index recollection. However, in this study, the exposure time was unequal for previously tested and untested items. Therefore, it remains unclear whether the enhancement of the LPC for previously tested items was due to the additional exposure time or to the process of active retrieval. Moreover, the study focused on the outcome of testing rather than on the processes that actually take place at testing.

Of particular importance for the present study, the processes of retrieval mode and retrieval orientation have not been investigated in the context of the retrieval practice effect, even though these pre-retrieval processes have been the focus of numerous other ERP studies (e.g., Herron & Wilding, 2004). Some of these studies provided evidence that adopting a retrieval orientation is beneficial for ongoing retrieval (i.e., leads to better retrieval accuracy; Bridger, Herron, Elward, & Wilding, 2009; Bridger & Mecklinger, 2012; Rosburg, Johansson, Sprondel, & Mecklinger, 2014). However, it remains unclear whether this benefit also applies in the context of retrieval practice and whether adopting retrieval mode or orientation is also beneficial for later retrieval accuracy.

In the present study, we aimed to investigate retrieval mode and retrieval orientation and their role in retrieval practice. We used a typical three-phase testing-effect paradigm, including an encoding phase, a retrieval phase, and a final test (Roediger & Karpicke, 2006a; see Fig. 1). The task performed during the retrieval phase defined our experimental conditions: two active-retrieval tasks (a recognition test and a source memory
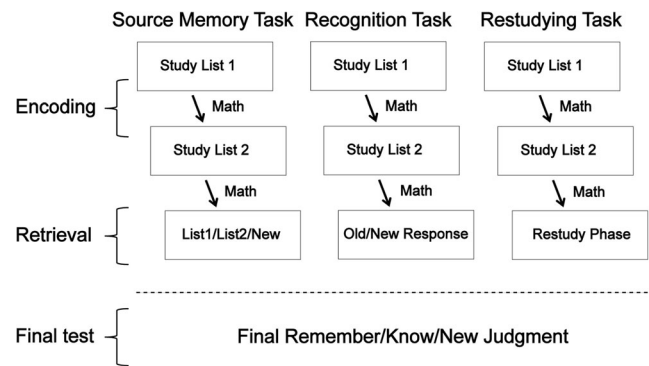

Fig. 1 General experimental procedure. Three within-subjects conditions were used: recognition, source monitoring, and restudying. In the encoding phase, subjects were asked to study two lists of characters. We did not tell the subjects a priori which task they would be expected to perform at subsequent retrieval. The subjects were exposed to one of the three conditions in the retrieval phase in each session, and each subject was tested in all three conditions. Then, the subjects performed a final test that encompassed both old items (items studied in encoding phases) and new items (ones not previously shown in the experiment), and they were requested to make a remember/know/new judgment

test) and one passive-retrieval (restudying) task. Subjects performed all three retrieval tasks for different study lists, with the order of the tasks counterbalanced across subjects. Variables related to stimulus exposure (total number of stimuli and amount of learning time) were the same for the three tasks in order to avoid additional exposure time as a confounding factor (Slamecka & Katsaiti, 1988; Thompson, Wenger, & Bartling, 1978). Furthermore, both old and new items were presented not only in the active-retrieval tasks, but also in the restudying condition. This approach was chosen in order to create a more realistic restudying scenario because students normally cannot selectively study only a tightly circumscribed set of facts for an upcoming test. Also, it enabled us to have ERP data for new items for all three retrieval conditions, such that retrieval orientation and retrieval mode could be investigated by contrasting the ERP responses to new items to avoid a contamination of these effects with effects of retrieval success. Previous studies on retrieval practice have also included new items in the initial intervening test (e.g., Carpenter & Delosh, 2006; Glover, 1989).

Neural correlates of retrieval mode were analyzed by contrasting the ERP responses to correctly rejected new items (CRs) in the two active-retrieval tasks to the ERP responses to new items in restudying condition. Neural correlates of retrieval orientation were investigated by comparing the ERPs to CRs between the recognition and the source memory task. For the final test, we chose a remember/know recognition test (Tulving, 1985) to determine whether the different forms of intermediate retrieval (testing vs. restudying) affect more familiarity-related or more recollection-related processes. Distinction between familiarity and recollection is the foundation of two-process models of episodic memory. According to these models, recognition can be

based on familiarity (i.e., identifying an event as previously encountered) without retrieving context information, whereas recollection is the slower and more effortful process of retrieving such context information (Rugg & Curran, 2007; Yonelinas, 2002).

For our study, we had the following three predictions. First, we expected that active retrieval (performing a recognition test or a source memory test during the retrieval phase) would lead to better retrieval accuracy than restudying. Second, on the basis of the desirable difficulties model (Bjork, 1994, 1999; Bjork & Bjork, 1992), we also hypothesized that the more recollection-based, and therefore effortful, source memory test would produce a larger behavioral retrieval practice effect than the recognition task. Third, we expected that the magnitude of the retrieval orientation and retrieval mode effects would modulate the behavioral testing effect. In other words, we expected to see larger behavioral retrieval practice effects in subjects who showed ERPs evidence of intentional retrieval and cue-specific processing at active retrieval.

Method

Subjects

Twenty-four right-handed volunteers (age range 20–27; 15 female, nine male) gave informed consent and were compensated ¥20 per hour. All subjects were students from Capital Normal University and spoke Chinese as their first language. All had normal or corrected-to-normal vision and reported good health. Two subjects were not included in the ERP analyses because of insufficient trial counts (<16). The study was approved by the Capital Normal University's research ethics committee.

Stimuli

The stimuli were 420 words (Chinese characters; word frequency: mean = 7 occurrences/million, range = 1–13 occurrences/million; stroke numbers: mean = 11.0, range = 4–21). Words were distributed randomly into 12 lists with matched word frequency (mean for each list ranged from 6.3 to 7.0 occurrences/million; Beijing Language College, 1986) and numbers of strokes (mean for each list range from 10.8 to 11.2). The lists were distributed pseudorandomly into the different experimental conditions and counterbalanced between subjects. Words were presented in white color on a black background on a 17-in. computer CRT monitor, covering a visual angle of approximately 2.5° × 2.5°. A fixation cross appeared at the central location during each interstimulus interval (ISI).

Procedure

The experimental design is presented in Fig. 1. Briefly, three within-subjects conditions were employed. In each condition, subjects underwent an encoding phase and retrieval phase, with the condition being defined by the kind of task at the retrieval phase. These included two active-retrieval tasks (performing a recognition task and performing a source memory task) and one passive-retrieval task (restudying). Each subject was tested in all three conditions in separate study–retrieval cycles. The timing and number of stimuli in both the encoding and retrieval phases were exactly the same across the three conditions.

In the encoding phases, the subjects were asked to study and memorize two lists (different words) of 35 characters each (that only differed in the order in which they were presented). Subjects were informed that there would be a memory test later in the experiment. In the initial study phase, subjects were not provided with any further instruction regarding how to memorize the words. Each character was presented at the center of the monitor for 1,000 ms, followed by a variable 800- to 1,200-ms ISI. After studying a list, subjects were instructed to count backward by three from a specified number for 90 s. The encoding phases were identical for all three conditions, but different study items were presented in each condition.

In the retrieval phase, the item lists consisted of 70 studied characters and 35 new characters. The stimuli were shown for 1,000 ms, with a variable 1,500- to 2,500-ms ISI. In the recognition task, the subjects had to make an old–new judgment for each item. In the source memory task, the subjects performed a modified source recognition test (Drosopoulos, Wagner, & Born, 2005). In this test, they were instructed to press one of three keys to indicate whether an item appeared in the first study list, appeared in the second study list, or was new. For the restudying task, subjects studied the two word lists a second time, together with 35 items that had not been presented before. They were instructed to study and memorize the items regardless of whether they were familiar or not familiar. The order of the three conditions within a session was pseudorandomized and counterbalanced across the 24 subjects.

After the last retrieval phase, all subjects took off the electrode caps and washed their hair (approximately 5–10 min), and then played a video game (Tetris) on the computer for 10–15 min. The retention interval for each subject was about 20 min. Then, all subjects performed the final test in which only items from the initial encoding phase and completely new items were presented. Subjects were exposed to all of the 210 initially studied items, and 105 completely new items. Thus, new items presented during the retrieval phases were not again presented. All stimuli appeared for 1,000 ms, with an ISI ranging from 1,500 to 2,500 ms. Subjects were asked to indicate for each item whether it was remembered (R), known

(K), or new (N). Specifically, they were instructed to press the "R" key if they could recollect specific details associated with the item's presentation, to press "K" key if an item was familiar but their memory lacked specific details, and to press the "N" key if they believed an item was new. The keys designated for each response type were counterbalanced across subjects.

Electroencephalographic (EEG) recordings

The EEG data (range 0.05–100 Hz, sampling rate 500 Hz) were recorded from 62 Ag–AgCl scalp electrodes embedded in an elastic cap with a NeuroScan SynAmps system (NeuroScan Inc. Sterling, Virginia, USA). The electrode locations in the cap were based on the extended international 10–20 system (Picton et al., 2000). Voltage was referenced to the left mastoid online and re-referenced offline to the average of the left and right mastoids. Eye movements were monitored by a pair of electrodes placed outside the outer canthi and a pair of electrodes placed below and above the left eye. Impedance was kept below 5 kΩ. Recordings were digitally filtered with a bandpass of 0.05–40 Hz, and epochs were created beginning 200 ms prior to stimulus onset, with a length of 1,400 ms. Waveforms were corrected relative to the 200-ms prestimulus baseline period. EOG blink artifacts were corrected using a linear regression estimate (Semlitsch, Anderer, Schuster, & Presslich, 1986). Trials containing EEG activity exceeding ±75 μV were rejected before averaging. The minimum number of trials per condition was set at $n = 16$. The exact cutoff value was chosen on the basis of those used in our previous studies (e.g., Rosburg, Mecklinger, & Johansson, 2011a, 2011b).

Data analysis

In all analyses, the Greenhouse–Geisser correction for sphericity violation was used when appropriate, and the corrected degrees of freedom are given in the text. Bonferroni correction was applied for post-hoc pairwise comparisons. An alpha level of .05 was used for all statistical tests.

Behavioral data The discrimination score (Pr) was defined as the difference between the hit rate and the false alarm rate (Snodgrass & Corwin, 1988). Response bias was measured by index Br (false alarm rate/[1 – (hit rate – false alarm rate)]) (Snodgrass & Corwin, 1988). Behavioral responses were compared between conditions by means of paired t tests and repeated measures analysis of variance (ANOVA). To determine the probability in the final remember–know test that an item was familiar, we adopted the independent remember/know (IRK) procedure proposed by Yonelinas and Jacoby (1995; proportion of K responses/[1 – proportion of R responses]).

ERP data For the analysis of the retrieval mode effect, ERPs for new items were contrasted between the active-retrieval tasks (CRs to new items) and restudying (all new items). We considered a retrieval mode effect to be present when condition effects were found in both contrasts (recognition vs. restudying and source vs. restudying). The ERP retrieval mode effect was quantified as difference potentials (active retrieval – passive retrieval). Correlation analyses were then conducted to investigate the relationship between the ERP retrieval mode effect and behavioral testing effect. The mean (range) numbers of trials for these ERPs were 27.8 (17–34) for the recognition task, 26.0 (17–34) for the source memory task, and 32.5 (27–35) for restudying.

For the analysis of the retrieval orientation effect, we contrasted the ERPs to new items in the active-retrieval tasks (recognition vs. source), again just using the ERPs to CRs. The ERP retrieval orientation effect was quantified as the difference between the recognition and source potentials. Subsequently, Pearson's correlation coefficients between behavioral testing gains and retrieval orientation effects were calculated in order to examine whether differential retrieval cue processing modulated the behavioral testing effect.

For the analysis of the retrieval mode and retrieval orientation effects, ERP amplitudes were averaged over four electrode clusters along the anterior–posterior axis: frontopolar (FP1, FPZ, FP2), frontal (F5, FZ, F6, FC5, FCZ, FC6), central (C5, CZ, C6), and parietal electrodes (CP5, CPZ, CP6, P5, PZ, P6). The selection of these electrode sites and the analyzed latency ranges was based on inspection of the waveforms and previous research (Dzulkifli & Wilding, 2005; Voss & Federmeier, 2011; Werkle-Bergner, Mecklinger, Kray, Meyer, & Düzel, 2005). The latency intervals of 300–500, 500–700, 700–900, and 900–1,100 ms were selected for the analyses, as presented below.

Results

Behavioral results

Initial testing The retrieval accuracy (hits rates and false alarm rates), as well as the reaction times (RTs) of the two active-retrieval tasks (recognition vs. source) are shown in Table 1. To examine the subjects' memory performance in these initial tests, t tests were conducted. For the source task, hits with and without a correct source were collapsed into an overall hit rate. For both tasks, the hit rates were reliably greater than the false alarm rates [recognition task, $t(23) = 13.86$, $p < .001$; source memory task, $t(23) = 11.98$, $p < .001$]. Moreover, in the source memory task, subjects made significantly more accurate than inaccurate source judgments [$t(23) = 5.33$, $p < .001$]. Retrieval accuracy for List 1 and List

Table 1 Mean accuracy and reaction times (RTs) in the initial tests (± *SD*)

	Recognition		Source			
	Hit	CR	Overall Hit	Hit_Correct	Hit_Incorrect	CR
Accuracy	.79 ± .12	.80 ± .16	.77 ± .14	.48 ± .13	.29 ± .09	.76 ± .16
RT	777.8 ± 149.1	831.1 ± 140.7	1,070.1 ± 273.8	1,072.7 ± 276.6	1,069.0 ± 290.9	953.1 ± 200.6

CR means a correct rejection. Hit_Correct means correct source judgments. Hit_Incorrect means incorrect source judgments but correct old/new judgments (pressed "List 2" button for "List 1" items, or vice versa). Overall Hit = Hit_Correct + Hit_Incorrect.

2 items did not differ in either the recognition task or the source memory task (all *p*s > .1, data not shown). To examine any differences in memory performance between the two active-retrieval tasks, additional *t* tests were conducted. The recognition task was characterized by higher CR ratios [*t*(23) = 2.18, *p* < .05] than was the source task. Response bias between the two active-retrieval tasks did not differ [recognition = .44 ± .24 vs. source = .49 ± .24; *t*(23) = 1.20, n.s.], and we found no difference in either the overall hit rates [recognition = .79 ± .12 vs. source = .77 ± .14; *t*(23) = 0.69, n.s.] or discrimination scores [recognition = .59 ± .21 vs. source = .54 ± .22; *t*(23) = 1.56, n.s.].

Final testing Memory performance in the final test is shown in Table 2 as proportions of K and R responses for each of the three conditions separately. However, as we mentioned in the Method section above, the data analyses were based on familiarity estimates obtained by the IRK procedure in order to examine the effect of retrieval practice on familiarity-based and recollection-based recognition more accurately. To measure the testing effect, a 3 × 2 repeated measures ANOVA was run on correct responses to studied items in the final test, with Condition (recognition, source, restudying) and Processing Type (recollection, familiarity) as within-subjects factors. The ANOVA revealed a significant condition main effect [*F*(2, 46) = 13.08, *p* < .001] but did not support a dissociation between recollection and familiarity [Condition × Processing Type interaction: *F*(1.60, 36.75) = 1.25, n.s.]. Additional multiple comparisons showed higher hit rates in the final test for

the recognition task condition (*p* = .001) and the source memory task condition (*p* = .01) than for the restudying condition, but no significant difference emerged between the two active-retrieval conditions (*p* = .08). We also conducted an analysis on the effects of block order—that is, 1st versus 2nd versus 3rd encoding block—on retrieval accuracy, to see whether memory performance declined over time (note that conditions were balanced across encoding blocks). This analysis revealed no significant distinctions between encoding blocks for combined hit rates ("remember" hit rates plus "know" hit rates), recollection, and familiarity ("independence" K scores) in the final testing (all *p*s > .1).

ERP results

ERP correlates of retrieval mode A Condition (recognition, restudying) × Latency (300–500, 500–700, 700–900, 900–1,100 ms) × Electrode Cluster (frontopolar, frontal, central, parietal) repeated measures ANOVA was conducted. It exhibited a significant Condition × Latency × Electrode Cluster interaction [*F*(3.93, 82.62) = 8.77, *p* < .001]. Likewise, a Condition (source, restudying) × Latency × Electrode Cluster repeated measures ANOVA was also conducted. It exhibited a significant condition main effect [*F*(1, 21) = 8.53, *p* = .008] and a marginally significant Condition × Latency × Electrode Cluster interaction [*F*(2.38, 49.88) = 2.53, *p* = .08]. Separate analyses were run within each 200-ms time window to analyze the temporal extent of the retrieval

Table 2 Mean accuracy and reaction times (RTs) in the final test (± *SD*)

	Old						New
	Restudy		Recognition		Source		
	Hit_R	Hit_K	Hit_R	Hit_K	Hit_R	Hit_K	CR
Accuracy	.41 ± .21	.35 ± .15	.52 ± .18	.34 ± .14	.46 ± .22	.37 ± .15	.67 ± .23
RT	917.5 ± 146.6	964.2 ± 293.2	884.3 ± 121.1	1,043.4 ± 218.1	904.5 ± 155.3	1,039.9 ± 195.6	971.6 ± 158.2

Raw proportions of R and K responses are listed in this table. However, the data analyses were based on familiarity estimates obtained by the IRK procedure. Hit_R means correct R responses for old items. Hit_K means correct K responses for old items. CR means correct rejections.

mode effect for each of the two contrasts in greater detail. The condition main effects were significant from 300 to 1,100 ms (for recognition vs. restudying) and 300 to 900 ms (for source vs. restudying), as is shown in Table 3. More positive-going ERPs were found for both conditions of active retrieval than for restudying. Topographically, the retrieval mode effect was relatively widespread (Fig. 2).

The relationship between behavioral memory enhancements at the final test and the ERP correlates of retrieval mode at initial retrieval was addressed by correlation analyses. The purpose of these analyses was to evaluate whether the size of the ERP retrieval mode effect could predict the behavioral benefit of active retrieval. Behavioral difference scores were obtained by subtracting the hit ratios in the final test for items from the restudying condition from the hit ratios for items from the recognition or source memory conditions. ERP correlates of retrieval mode at the initial test were obtained by subtracting the ERPs for new items in the restudying condition from the ERPs to CRs of new items in the recognition condition or source memory condition. Given that our previous analyses revealed no regional differences, we quantified the ERP retrieval mode effect for this purpose as the average effect across all analyzed electrodes, for each of the three latency intervals that showed significant retrieval mode effects (300–500, 500–700, and 700–900 ms) and for both active-retrieval conditions separately. The results of this correlation analysis are shown in Table 4. For the recognition condition, a positive correlation between the behavioral testing effect and the retrieval mode effect was observed for the 300- to 500-ms latency interval ($n = 22$, $r = .49$, $p < .05$), but not for the later time windows. No significant correlations were observed for the source condition versus restudying. Thus, the early retrieval mode effect (300–500 ms) predicted the later behavioral testing effect only for the recognition condition (Fig. 3a). Notably, the findings on retrieval mode effects for both the ANOVA and correlation analyses were not altered when the ERPs to all new items (rather than just CRs) were analyzed for the two active-retrieval tasks.

ERP correlates of retrieval orientation A $2 \times 4 \times 4$ repeated measures ANOVA was conducted between two test types (recognition, source) for four electrode clusters (frontopolar, frontal, central, parietal), and the 300- to 500-ms, 500- to 700-ms, 700- to 900-ms, and 900- to 1,100-ms latency intervals. The analysis revealed a significant Latency × Test Type × Electrode Cluster interaction [$F(2.25, 47.29) = 3.47$, $p < .05$]. Separate analyses were run to analyze the test type effects within each time window. The ANOVA results are summarized in Table 5. The retrieval orientation effect was present from 500 to 900 ms, and this effect was pronounced at posterior electrode sites. The ERPs to CRs in the recognition task were more positive-going than the ERPs to CRs in the source memory task (Fig. 4).

To elucidate the role of retrieval orientation on the behavioral testing effect, correlations between behavioral testing gains and retrieval orientation ERP effects were calculated. Behavioral testing gains were obtained by subtracting the hit ratios in the final test for the recognition task condition from the hit ratios in the final test for the source memory task condition. For the retrieval orientation effect, ERPs to correctly rejected new items were contrasted between the two active-retrieval tasks. On the basis of the ANOVA analyses, the retrieval orientation effect was largest at central and parietal electrodes and was largely absent for more frontal electrodes. Therefore, the mean retrieval orientation effect averaged across central and parietal electrodes was used for calculating correlations with the behavioral testing gains. For this analysis, we only considered data from the two time windows in which we had revealed significant retrieval orientation effects (500–700 and 700–900 ms). A significant positive correlation was observed in the 700- to 900-ms latency interval ($n = 22$, $r = .52$, $p < .05$), whereas the effect was only marginally significant in the earlier time window ($r = .37$, $p = .09$). The larger retrieval orientation effects from 700 to 900 ms were surprisingly associated with greater behavioral testing gains in the recognition condition, as is illustrated in Fig. 3b.

Table 3 Summary of the repeated measures ANOVA conducted on the ERP correlates of retrieval mode

Comparison	Effect	Time windows (F value)			
		300–500	500–700	700–900	900–1,100
ANOVA (RM, O vs. R)	RM main effect	21.94***	29.00***	35.40***	16.59**
	RM × Cluster	2.83	1.00	2.13	1.95
ANOVA (RM, S vs. R)	RM main effect	7.92*	9.92**	9.07**	2.59
	RM × Cluster	6.43**	1.10	0.12	0.35

RM is the abbreviation for the factor Retrieval Mode. Cluster included four levels (frontopolar, frontal, central, and parietal). The letter "R" is the abbreviation for "Restudy," "O" for "Recognition," and "S" for "Source." Follow-up analyses for the significant RM × Cluster interaction in the source versus restudy comparison indicated that significant RM main effects were found for frontopolar and frontal clusters but not for central and parietal clusters. * $p < .05$, ** $p < .01$, *** $p < .001$
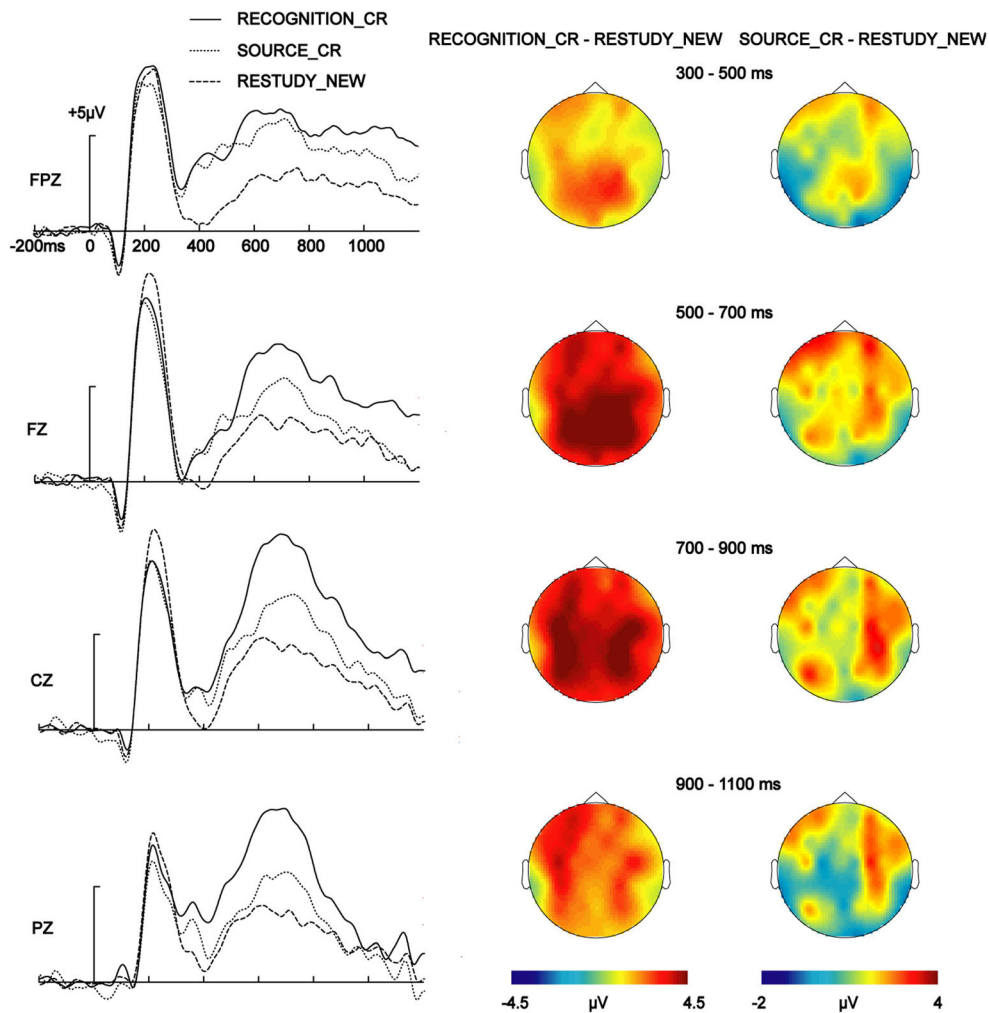
Fig. 2 Event-related potential (ERP) correlates of retrieval mode. Time–voltage plots are shown for the RECOGNITION_CR (correct rejections of new items), SOURCE_CR (CRs of new items), and RESTUDY_NEW (new items) for each of four electrode locations (*left column*); topographies of the ERP differences between recognition (CRs of new items) and restudying (new items) for the four latency intervals (middle column); and topographic maps of the source (CRs of new items) versus restudying (new items) ERP differences in the four time windows (*right column*)

On the basis of previous studies indicating that retrieval orientation might generally contribute to retrieval accuracy, and not just for one kind of targeted information (Bridger et al., 2009; Bridger & Mecklinger, 2012; Rosburg et al., 2014), correlations between the general behavioral testing

Table 4 Values for Pearson's *r* relating ERP amplitude differences of retrieval mode effect across all analyzed electrode clusters with behavioral difference scores for "Recognition versus Restudy" and "Source versus Restudy," respectively

	Time windows (*r* value)		
Condition	300–500	500–700	700–900
Recognition–restudy	.49[*]	.31	.16
Source–restudy	.27	.23	−.03

All tests were two tailed. [*]*p* < .05

effect and the retrieval orientation ERP effects were calculated for two time windows, as above. The general behavioral testing effect (active-retrieval conditions vs. restudying) was obtained by averaging the behavioral testing effects in the recognition task and the source memory task. The calculation of the retrieval orientation ERP effects was the same as above. Positive correlations were observed in the 500- to 700-ms latency interval (*n* = 22, *r* = .50, *p* < .05) and the 700- to 900-ms latency interval (*n* = 22, *r* = .60, *p* < .01; Fig. 3c).

Discussion

The goal of our present experiment was to examine the cognitive and neural mechanisms underlying the testing effect. First, we found that retrieval practice (active retrieval) indeed produced stronger memory than restudying (passive retrieval).
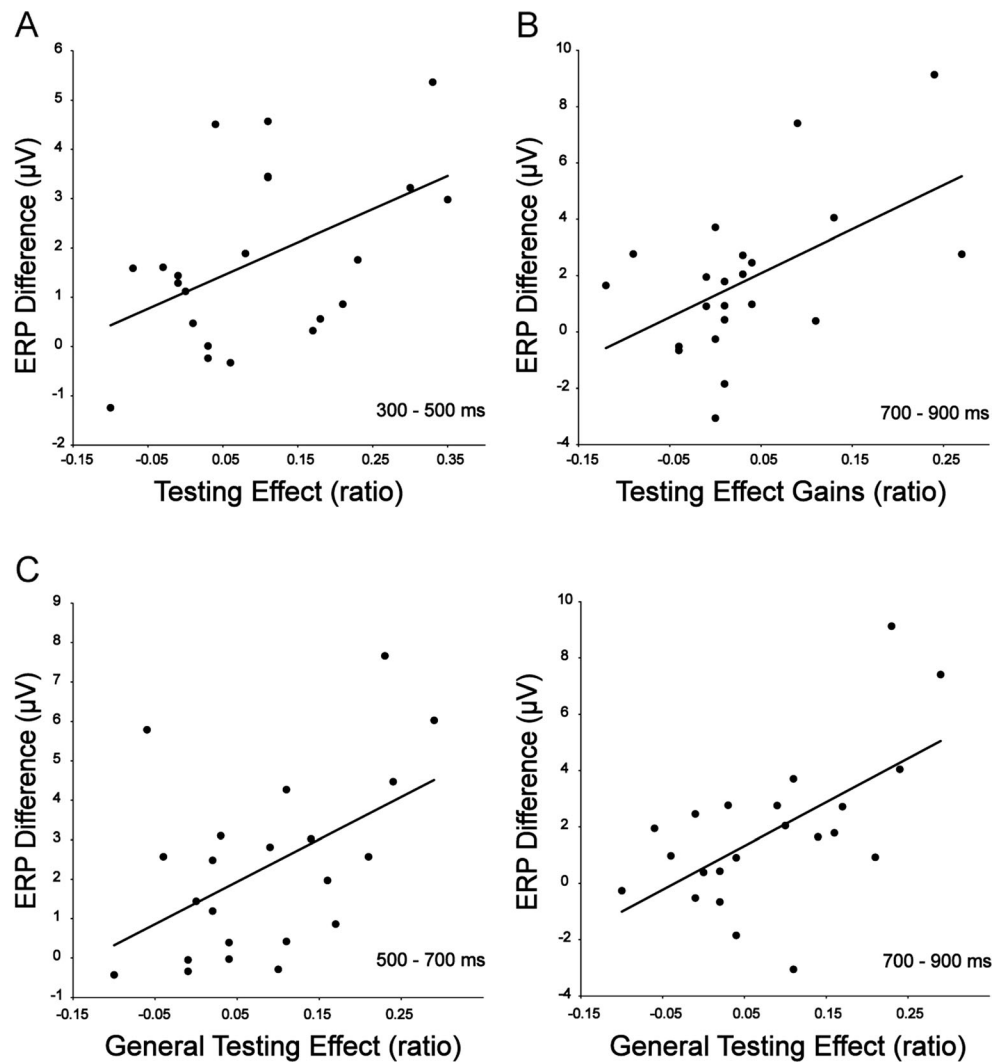
Fig. 3 Correlations between behavioral estimates of the testing effect (testing effect and testing gains) and event-related potential (ERP) correlates. (**a**) Scatterplot of significant correlations between the behavioral testing effects and retrieval mode ERP effects in the 300- to 500-ms latency interval. Behavioral testing effects for the recognition task were calculated by subtracting the hit ratios in the final test between the recognition task condition and the restudy condition. The retrieval mode effect was averaged across all analyzed electrodes. (**b**) Scatterplot of the significant correlations between the behavioral testing gains and retrieval orientation ERP effects in the 700- to 900-ms latency interval. Behavioral testing gains were calculated by subtracting the hit ratios in the final test between the recognition task condition and the source memory task condition. The retrieval orientation effect was averaged across central and parietal electrodes. (**c**) Scatterplots showing significant correlations between the general behavioral testing effect and the retrieval orientation ERP effects in the 500- to 700-ms (*left panel*) and 700- to 900-ms (*right panel*) latency intervals. General behavioral testing effects were calculated as the mean values for the behavioral testing effects in the recognition and source memory tasks

However, contrary to the predictions of the desirable difficulties model (Bjork, 1994, 1999; Bjork & Bjork, 1992), the retrieval practice effect did not vary between the recognition task and the source memory task. Second, for the correlates of retrieval mode, we observed that both forms of active retrieval had more positive-going ERPs for new items from 300 to 900 ms than in the restudying condition. These retrieval mode effects were widespread over the scalp. Follow-up correlation analyses revealed an association between the early (300 to 500 ms) retrieval mode effect and the behavioral testing effect for the recognition versus restudying contrast. Third, neural correlates of retrieval orientation were found between 500 and 900 ms, with a maximum over centro-parietal electrode sites. For the 700- to 900-ms time window, follow-up analyses showed an association between the retrieval orientation effect and the behavioral retrieval practice difference between the recognition task and the source memory task. Taken together, these findings suggest that retrieval practice (active retrieval) promotes later memory retention relative to restudying (passive retrieval), and that retrieval mode and retrieval orientation modulate this behavioral testing effect.

Behaviorally, we found that retrieval accuracy in the final test was better after active retrieval (i.e., performing a recognition task or source memory task) than after restudying the

Table 5 Summary of the repeated measures ANOVA on the ERP correlates of retrieval orientation

Effect	Time windows (*F* value)			
	300–500	500–700	700–900	900–1,100
RO main effect	0.95	10.47**	11.29**	3.19
RO × Cluster	2.52	3.39 (*p* < .06)	1.00	0.3
Frontopolar	–	0.98	(1.54)	–
Frontal	–	4.46	(6.77)	–
Central	–	18.91***	(14.96**)	–
Parietal	–	20.52***	(5.16*)	–

Main effects of retrieval orientation (RO) were found between 500 and 900 ms. The Cluster factor included four levels (frontopolar, frontal, central, parietal). Even though the RO × Cluster interaction was not significant for 700–900 ms, the results of follow-up analyses for each cluster suggest that the RO effect in this latency window still appeared to be more pronounced at central and parietal sites, similar to the 500- to 700-ms time window. The results of these post-hoc tests are, however, displayed in parentheses, as they were not warranted by a significant RO × Cluster interaction. * $p < .05$, ** $p < .01$, *** $p < .001$

words. As expected, our results show that active-retrieval conditions had a beneficial effect on later retrieval accuracy. This is important as it enabled us to delineate the role of pre-retrieval processes for the behavioral retrieval practice effect. Surprisingly, the behavioral effect was not larger after the source memory task than after the recognition task. Based on the desirable difficulties model (Bjork, 1994, 1999; Bjork & Bjork, 1992), we expected a larger effect for the source memory condition than for the recognition condition, since the source memory task is considered to be more effortful than the recognition task, and recollection of context-specifying information is usually required in the source memory task but not in the recognition task. Indeed, when analyzing the old/new effects for the two tasks, we found a late parietal complex (LPC) only in the source memory task (see the supplemental material). Also, on the basis of the episodic context account (Karpicke et al., 2014), one would predict larger retrieval practice effects for conditions that require recollection of contextual information.

This discrepancy between such theory-based predictions and the obtained results might result from the experimental set-up. Our subjects were rather good at discriminating old and new items (in both tasks), but were not very accurate in discriminating the two sources (List 1 vs. List 2). The high rate of incorrect source judgments for studied items indicates that subjects were often unable to retrieve the task-relevant prior study context. Overly difficult tasks are known to hinder successful learning task (Bjork, 1994, 1999). Here, increased difficulty was likely due to the relatively long word lists and the short duration of item presentation at encoding. As a consequence of this difficulty retrieving the prior study context, subjects would not have been able to incorporate the prior

study context with the current context, as claimed by the episodic context account (Karpicke et al., 2014; Lehman et al., 2014). Furthermore, the choice of the source memory task may have been suboptimal because for temporal source judgments subjects might also rely on factors other than recollection. For instance, subjects may infer the list membership of recognized items from their memory strengths, as suggested by research on recency judgments (Hintzman, 2005), and one might further argue that subjects would particularly rely on such factors when they had difficulties retrieving the prior temporal context.

Moreover, behaviorally, we found that recollection and familiarity were not differentially influenced by the two kinds of active retrieval. On the basis of the TAP account (Morris et al., 1977) and on the episodic context account (Karpicke et al., 2014), we expected that the source memory task would enhance recollection-based ("remember") responses relative to the recognition task, as the latter task does not specifically require recollection. This lack of differentiation between the two active-retrieval tasks might be similarly explained by the fact that subjects did not (or could not) exclusively rely on recollection when performing the source memory task.

Both active-retrieval conditions were characterized by more positive-going ERPs for new items than in the restudy condition. We presumed that this ERP effect reflects retrieval mode (i.e., as the cognitive state in which an individual consciously thinks of the past when he/she encounters a potential cue; Tulving, 1983). We have to acknowledge, however, that with such a qualitative change of state other processes are presumably also modulated (such as the focus of attention, effort, or arousal), even though we avoided the influence of retrieval success ("ecphory") by analyzing the ERPs to new items. The observed ERP effect was relatively widespread, which indicates that more than one brain region generated this effect. In line with predictions made on the basis of the episodic context account (Karpicke et al., 2014), we observed that increased levels of retrieval mode (quantified as the ERP difference between active and passive retrieval) were associated with better retrieval accuracy in the final test. However, this association was only found for the early time window and just for the recognition versus restudying contrast. This might be due to the low signal-to-noise ratio of the ERP data and small sample size, but also to the possibility that other psychological and physiological factors, as described above, modulated the magnitude of ERP effect.

Retrieval orientation was evaluated from the comparison between the recognition task and the source memory task. ERP correlates of retrieval orientation effects were found in the 500- to 900-ms time window, with a centro-parietal maximum. Moreover, we found that the magnitude of the retrieval orientation effect correlated with the behavioral retrieval
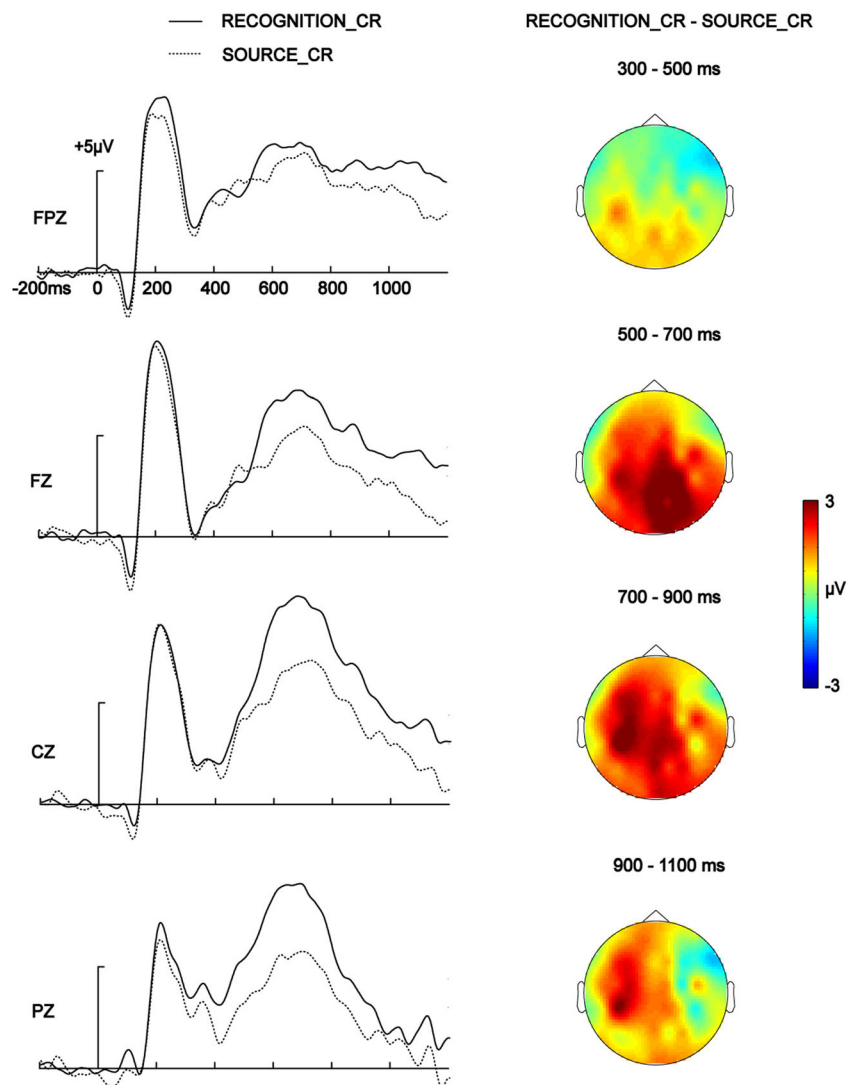
Fig. 4 Event-related potential (ERP) correlates of retrieval orientation effects: ERP waveforms for the RECOGNITION_CR (correct rejections of new items) and SOURCE_CR (CRs of new items) conditions for each of four electrode sites (*left column*), and average ERP differences between RECOGNITION_CR and SOURCE_CR for the four time intervals, plotted topographically (*right column*)

practice difference between the recognition task and the source memory task. Surprisingly, more positive ERP retrieval orientation effects were associated with larger behavioral gains in the recognition condition than in the source memory condition. Thus, as predicted, adopting retrieval orientation was beneficial for later retrieval accuracy, but, unexpectedly, the recognition condition (and not the source memory condition) showed this benefit.

Although the episodic context account does not directly refer to the concept of retrieval orientation, Karpicke and colleagues (2014) claim that reinstating the initial study context during retrieval practice enhances retention. We would argue that retrieval orientation, as investigated by contrasting the ERP responses to new items between the source memory and recognition tasks, reflects attempts to reinstate the previous episodic context, which further satisfies the demands of a given retrieval task. Given this, one might expect a larger behavioral retrieval practice effect the more a subject adopts such a retrieval orientation. Previous research has indeed shown that retrieval orientation can modulate retrieval accuracy at the time point of testing (Bridger et al., 2009; Bridger & Mecklinger, 2012; Rosburg et al., 2014). The present findings extend these results and show that it can also influence future retrieval accuracy. To our knowledge, this is the first ERP evidence linking retrieval orientation to memory performances after an interval.

There are, however, two caveats. First, the ERP retrieval orientation effect does not have a true baseline, as it is the contrast of two retrieval conditions. Thus, cue processing that is common to both retrieval conditions is not reflected.

Moreover, unspecific factors such as effort and arousal might vary between the conditions and partly be reflected in the ERP retrieval orientation effect. In other words, the ERP retrieval orientation effect cannot be presumed to exclusively reflect qualitatively distinct forms of cue processing. Second, in our study the ERP retrieval orientation effect was associated with better retrieval accuracy for the task condition that did not require recollection and for which we found no empirical evidence of recollection. It is usually suggested that recollection is involved in the reinstatement of processes or representations that were active when the episode was encoded (e.g., Johnson & Rugg, 2007). Although this does not imply that reinstatement processes are completely absent when no recollection takes place, it appears reasonable to presume that reinstatement processes are fragmentary in the absence of recollection. Thus, the relative behavioral retrieval practice effect (recognition vs. source memory condition) cannot be explained by terms of reinstatement, except we presume that under specific circumstances less reinstatement is beneficial for later retrieval, contrary to the suggestions of the episodic context account.

Indeed, it could be argued that the final test did not require subjects to retrieve the list membership. Instead, the subjects were asked to evaluate whether they remembered an item or not. This subjective measure of recollection is more inclusive and the endorsement of "remember" may or may not require the recollection of relevant source information (Gao, Hermiller, Voss, & Guo, 2015; Wang, Li, Gao, Xu, & Guo, 2015). Thus, a superficial reinstatement of the study episode might have been sufficient for enhancing the association between an item and its (list-unspecific) encoding context, which later on leads to an increase of 'remember' (and 'know') responses. In contrast, more extensive reinstatement might be less efficient when it is accompanied by frequent failures, which could then be experienced as such and provide the negative feedback of not-remembering.

However, retrieval orientation should not solely be conceptualized as the attempt to reinstate a study episode. Johnson, Kounios, and Nolde (1997) suggested that differences between ERPs evoked by classes of unstudied words reflect how memory traces are probed for different kinds of information. Such a differential probing might be achieved by processes that enhance the interaction between internal representation of the retrieval cues and memory traces (cue bias) or by processes that directly act on memory representations and modulate their accessibility (target bias) (Anderson & Bjork, 1994; Dzulkifli & Wilding, 2005; Mecklinger, 2010; Rosburg et al., 2013; Rosburg et al., 2011a). Moreover, it has also been proposed that such operations increase overlap between cue and target processing (Robb & Rugg, 2002). As such, retrieval

orientation should be considered as an additional factor besides retrieval mode that contributes to the retrieval practice effect. When analyzing the association between the general behavioral retrieval practice effect (active-retrieval conditions vs. restudying) and the ERP retrieval orientation effect, we unexpectedly found a relatively strong association between the two. This finding resembles previous findings that adopting retrieval orientation is generally, and not just for one kind of targeted information, beneficial for retrieval accuracy at the time point of testing (Bridger et al., 2009; Bridger & Mecklinger, 2012; Rosburg et al., 2014).

One limitation of our present study is that we did not record ERPs at the final testing, which would have enabled the exploration of the neurobehavioral consequences of active retrieval, such as investigated by Rosburg et al. (2015). Moreover, we did not test objective memory (i.e., source memory) in the final test, which would have had some greater practical relevance.

Taken together, our study provides evidence for the episodic context theory, by showing that retrieval mode modulates the retrieval practice effect. Our study failed to show a larger retrieval practice effect for the more difficult source memory condition, as predicted by the desirable difficulties model. This failure was in all likelihood due to characteristics of the initial encoding, source memory task, and final testing. Aside from retrieval mode, we showed for the first time that adopting a retrieval orientation can modulate retrieval accuracy not just at the time point of testing, but also at a later test. This observation does not conflict with the episodic-retrieval account, but suggests that retrieval orientation should be considered by this account as well.

References

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265–325). San Diego: Academic Press.

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.

Beijing Language College. (1986). *Modern Chinese word frequency dictionary.* Beijing: Beijing Language College Press.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimanura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitve regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus disfluctuation. In A. Healy, S. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale: Erlbaum.

Bridger, E. K., Herron, J. E., Elward, R. L., & Wilding, E. L. (2009). Neural correlates of individual differences in strategic retrieval processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1175–1186.

Bridger, E. K., & Mecklinger, A. (2012). Electrophysiologically dissociating episodic preretrieval processing. *Journal of Cognitive Neuroscience, 24,* 1476–1491.

Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional anatomic study of episodic memory retrieval II: Selective averaging of event-related FMRI trials to test the retrieval success hypothesis. *NeuroImage, 7,* 163–175.

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory, 4,* 359–411.

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,* 1563–1569.

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,* 1547–1552.

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,* 268–276.

Drosopoulos, S., Wagner, U., & Born, J. (2005). Sleep enhances explicit recollection in recognition memory. *Learning and Memory, 12,* 44–51.

Dzulkifli, M. A., & Wilding, E. L. (2005). Electrophysiological indices of strategic episodic retrieval processing. *Neuropsychologia, 43,* 1152–1162.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters, 505,* 36–40.

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique, 51,* 6–28.

Gao, C., Hermiller, M. S., Voss, J. L., & Guo, C. (2015). Basic perceptual changes that alter meaning and neural correlates of recognition memory. *Frontiers in Human Neuroscience, 9*(49), 1–10.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399.

Hashimoto, T., Usui, N., Taira, M., & Kojima, S. (2011). Neural enhancement and attenuation induced by repetitive recall. *Neurobiology of Learning and Memory, 96,* 143–149.

Herron, J. E., & Wilding, E. L. (2004). An electrophysiological dissociation of retrieval mode and retrieval orientation. *NeuroImage, 22,* 1554–1562.

Herron, J. E., & Wilding, E. L. (2006). Brain and behavioral indices of retrieval mode. *NeuroImage, 32,* 863–870.

Hintzman, D. L. (2005). Memory strength and recency judgments. *Psychonomic Bulletin & Review, 12,* 858–864.

Hornberger, M., Rugg, M. D., & Henson, R. N. A. (2006). ERP correlates of retrieval orientation: Direct versus indirect memory tasks. *Brain Research, 1071,* 124–136.

Johnson, M. K., Kounios, J., & Nolde, S. F. (1997). Electrophysiological brain activity and memory source monitoring. *NeuroReport, 8,* 1317–1320.

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex, 17,* 2507–2515.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 61, pp. 237–284). San Diego: Elsevier Academic Press.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62,* 227–239.

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,* 1787–1794.

Mecklinger, A. (2010). The control of long-term memory: Brain systems and cognitive processes. *Neuroscience & Biobehavioral Reviews, 34,* 1055–1065.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Nolde, S. F., Johnson, M. K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: an event-related fMRI study. *NeuroReport, 9,* 3509–3514.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37,* 127–152.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475–543.

Robb, W. G. K., & Rugg, M. D. (2002). Electrophysiological dissociation of retrieval orientation and retrieval effort. *Psychonomic Bulletin & Review, 9,* 583–589.

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27.

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210.

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Rosburg, T., Johansson, M., Sprondel, V., & Mecklinger, A. (2014). Retrieving self-vocalized information: An event-related potential (ERP) study on the effect of retrieval orientation. *Brain and Cognition, 92C,* 123–132.

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,* 195–210.

Rosburg, T., Johansson, M., & Mecklinger, A. (2013). Strategic retrieval and retrieval orientation in reality monitoring studied

by event-related potentials (ERPs). *Neuropsychologia, 51,* 557–571.

Rosburg, T., Mecklinger, A., & Johansson, M. (2011a). Electrophysiological correlates of retrieval orientation in reality monitoring. *NeuroImage, 54,* 3076–3084.

Rosburg, T., Mecklinger, A., & Johansson, M. (2011b). Strategic retrieval in a reality monitoring task. *Neuropsychologia, 49,* 2957–2969.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11,* 251–257.

Rugg, M. D., Fletcher, P. C., Chua, P. M., & Dolan, R. J. (1999). The role of the prefrontal cortex in recognition memory and memory for source: An fMRI study. *NeuroImage, 10,* 520–529.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4,* 108–115.

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,* 695–703.

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 716–727.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study–test processing in memory and metacomprehension. *Memory & Cognition, 35,* 668–678.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53,* 1–25.

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.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology, 48,* 532–546.

Wagner, A. D., Desmond, J. E., Glover, G., & Gabrieli, J. D. E. (1998). Prefrontal cortex and recognition memory: FMRI evidence for context-dependent retireval processes. *Brain, 121,* 1985–2002.

Wang, W., Li, B., Gao, C., Xu, H., & Guo, C. (2015). Conceptual fluency increases recollection: Behavioral and electrophysiological evidence. *Frontiers in Human Neuroscience, 9*(377), 1–11.

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bullletin of the Psychonomic Society, 7,* 289–291.

Werkle-Bergner, M., Mecklinger, A., Kray, J., Meyer, P., & Düzel, E. (2005). The control of memory retrieval: Insights from event-related potentials. *Cognitive Brain Research, 24,* 599–614.

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,* 2360–2370.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622–643.