Cogn Affect Behav Neurosci
DOI 10.3758/s13415-014-0310-y

How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval

Timm Rosburg · Mikael Johansson · Michael Weigl · Axel Mecklinger

Abstract The testing effect is conceptualized as the benefit for remembering items that were studied and tested rather than just studied. Thus far, little is known about the underlying neurocognitive mechanisms. In an event-related potential (ERP) study, we investigated the immediate consequences of testing on recollection processes. During an initial study phase, participants encountered object names together with pictures of the denoted objects ("perceived items") or with the instruction to mentally visualize them ("imagined items"). Directly afterward, they had to differentiate between perceived, imagined, and new items, in two consecutive source memory tests. Half of the studied items were presented in the first run, and all items in the second. Behaviorally, repeated testing led to improved item and source memory, as well as faster reaction times, relative to items that had been tested once. In accordance with these behavioral changes, the left parietal old–new effect (500–900 ms) as the putative correlate of recollection was strongly enhanced by previous testing. An enhancement after testing was also observed for the early portion of the late right frontal old–new effect (700–900 ms). In contrast, old–new effects after 900 ms were not modulated by previous testing. The finding of a stronger left parietal old–new effect for previously tested items suggests that testing leads to an elaboration of memory traces, whereas the faster reaction times are more likely explained in terms of transfer-appropriate processing. The combination of more elaborated memory traces and transfer-appropriate processing provides a tentative explanation for the effectiveness of testing in enhancing retrieval performance.

Keywords Episodic memory · Learning · Event-related potential (ERP) · Source memory · Late posterior negativity (LPN)

Already William James proposed that after studying, later memory might benefit more from subsequent active recall than from subsequent additional studying (James, 1890). The relevance of testing as one form of active retrieval for learning has more recently been shown by Karpicke and Roediger (2008). Although it is relatively undisputed that testing has a beneficial effect on later memory, the cognitive mechanisms behind the so-called *testing effect* have not been fully disclosed as yet. In their review, Roediger and Butler (2011) suggested that testing might lead to an elaboration of memory traces, which increases the likelihood that the stored information will be retrievable in the future. The increased effort required for active retrieval might contribute to such an elaboration. As another account, Roediger and Butler discussed transfer-appropriate processing (Morris, Bransford, & Franks, 1977). According to the concept of transfer-appropriate processing, retrieval accuracy is strongly determined by the match between processes engaged during encoding and processes engaged during testing. Finally, the testing effect might be explained in terms of storage strength and retrieval strength—that is, by an interaction of memory trace and retrieval process characteristics, as described in the model of disuse (Bjork & Bjork, 1992). Storage strength refers to how well an item was learned, whereas retrieval strength refers to how easy a memory trace can be assessed. One core assumption of the

Electronic supplementary material The online version of this article (doi:10.3758/s13415-014-0310-y) contains supplementary material, which is available to authorized users.

T. Rosburg (✉) · M. Weigl · A. Mecklinger
Department of Psychology, Experimental Neuropsychology Unit, Saarland University, Campus, Building A2.4, 66123 Saarbrücken, Germany
e-mail: trosburg@mx.uni-saarland.de

M. Johansson
Department of Psychology, Neuropsychology Division, Lund University, Lund, Sweden

Published online: 08 July 2014

model of disuse is that successful retrieval has larger benefits for subsequent retrieval the more difficult the act of retrieval.

Behavioral studies have been informative about factors influencing the size of the testing effect. Those factors included the difficulty of testing, with more difficult testing being more effective (Pyc & Rawson, 2009), the provision of feedback during testing (Butler & Roediger, 2008), and the retrievability of items in the initial test (Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012). In particular, the finding that the difficulty of testing has a positive effect on later memory might be taken as evidence for the idea that testing promotes the elaboration of memory traces, but might also be interpreted as increased retrieval strength (Bjork & Bjork, 1992). To date, psychophysiological methods, such as event-related potentials (ERPs), or neuroimaging methods, such as fMRI, have hardly been used to elucidate potential neurocognitive mechanisms behind the testing effect; relatively few fMRI studies (Eriksson, Kalpouzos, & Nyberg, 2011; Hashimoto, Usui, Taira, & Kojima, 2011; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013) and no ERP studies on the testing effect have been conducted.

The main reasons for the lack of such studies might be that experiments on the testing effect typically are rather complex and include repeated study–test runs. With psychophysiological experiments usually having a maximal recording time of 60 min, it is problematic to cover all phases of such experiments. In the present ERP study, we investigated the *immediate* neurobehavioral effects of a *single* test session. Thus, we focused on what we regarded as the central module of the testing effect, namely the test itself. Surprisingly, to our knowledge, immediate neurobehavioral effects of testing have been addressed in only a single ERP study (Spitzer, Hanslmayr, Opitz, Mecklinger, & Bäuml, 2009).

The lack of ERP studies for investigating the immediate effects of testing is astonishing, because ERP recordings have widely been used for determining the processes that underlie episodic memory. For example, dual-process models of recognition memory have been validated by ERP studies (Rugg & Curran, 2007; Yonelinas, 2002). In such studies, retrieval-related processes are presumed to be carved out when ERP responses to studied (old) and unstudied (new) items are contrasted (old–new effects). Dual-process models of recognition memory differentiate between familiarity and recollection, as independent processes forming the phenomenological experience of recognition memory: Familiarity conceptualizes early and fast occurring processes of retrieving an event as previously encountered, but without retrieving context information ("feeling of knowing").

In ERP recordings, familiarity is thought to be reflected in an early, midfrontal old–new effect between 300 and 500 ms (Bridger, Bader, Kriukova, Unger, & Mecklinger, 2012; Rugg & Curran, 2007; Stenberg, Hellman, Johansson, & Rosén, 2009; but see Paller, Voss, & Boehm, 2007). Its amplitude is modulated by the confidence of having encountered an event previously (Addante, Ranganath, & Yonelinas, 2012; Woodruff, Hayama, & Rugg, 2006; Yu & Rugg, 2010). In contrast to familiarity, *recollection* refers to a slower and more effortful process of retrieving contextual information that specify a previous episode in space and time ("remembering"). In ERP recordings, recollection is presumed to be reflected in the left parietal old–new effect, occurring between 500 and 800 ms. This old–new effect has been found to be larger after deep-encoding tasks than after shallow-encoding tasks (Rugg et al. 1998), larger for correct than for incorrect source judgments (Wilding, 2000), and larger for "remember" than for "know" responses (Düzel, Yonelinas, Mangun, Heinze, & Tulving, 1997).

Familiarity and recollection have been proposed to be supported by differential neural substrates (for reviews, see Diana, Yonelinas, & Ranganath, 2007; Eichenbaum, Yonelinas, & Ranganath, 2007; Montaldi & Mayes, 2010). Evidence for dual-process models of recognition memory, as well as for the involvement of different neural substrates in these two kinds of processes, has come from ERP studies on patients with intact implicit but deficient explicit memory (Addante, Ranganath, Olichney, & Yonelinas, 2012; Düzel, Vargha-Khadem, Heinze, & Mishkin, 2001; Mecklinger, von Cramon, & Matthes-von Cramon, 1998; Olichney et al., 2000). Further evidence for dual-process models, as opposed to single-process models, of recognition memory is discussed elsewhere (Diana, Reder, Arndt, & Park, 2006; Yonelinas, Aly, Wang, & Koen, 2010). However, it is important to note that the boundary between item and context information is not necessarily absolute and tests of source and associative recognition should not be taken as process-pure measures of recollection. Recognition of two (or more) associated pieces of information can be based on familiarity if they were encoded as a single configuration and form a unitized representation (e.g., Bader, Mecklinger, Hoppstädter, & Meyer, 2010; Diana, Van den Boom, Yonelinas, & Ranganath, 2011; Rhodes & Donaldson, 2008). Furthermore, a recent study showed that episodic context can sometimes be retrieved independently of recollection, probably via context familiarity (Addante, Ranganath, & Yonelinas, 2012).

In the present study, we were particularly interested in the modulation of the left parietal old–new effect by testing. The left parietal old–new effect varies with the amount of recoverable episodic information (Vilberg & Rugg, 2009; Wilding, 2000), which makes it an interesting candidate for validating the assumption that testing leads to more elaborated memory traces, as was suggested by Roediger and Butler (2011). Previous ERP research has so far not addressed this issue. Some ERP studies, however, have investigated how old–new effects are modulated by multiple encounters of study material in repeated study–test cycles (de Chastelaine, Friedman, Cycowicz, & Horton, 2009; R. J. Johnson, Kreiter, Russo, & Zhu, 1998; Nessler, Friedman, Johnson, & Bersick, 2007): These ERP studies showed that, alongside with improved memory accuracy, the magnitude of left parietal old–new effect increased with the repetition of study–test cycles. Yet, these studies cannot advocate whether the observed increase of the left parietal old–new effect stemmed from repeated studying, repeated testing, or both.

The ERP study of Spitzer et al. (2009) addressed the effects of retrieval practice more directly. Category–exemplar pairs were used as material in the study phase. In the subsequent practice phase, pairs with incomplete exemplar names were presented (e.g., FRUIT–AP…). This study revealed that retrieval practice led to better accuracy in the final recognition task and to a larger parietal old–new effect. However, since only strong exemplars of each category were used as study material, it remains open whether during the practice block category–exemplar pair associations were actually retrieved from episodic memory or whether the exemplars were retrieved from semantic knowledge (i.e., generated upon the presentation of the category name and initial letters). With other words, it is conceivable that participants sometimes came up with the correct responses during retrieval practice without remembering the initial study episode.

To sum up these findings, previous studies suggested that combined study–test cycles, as well as retrieval practice, lead to enhanced parietal old–new effects. However, these studies are not informative as to whether this also holds true for testing alone, in particular for test situations when participants have to rely fully on their memories of the study phase (which represents the typical test situation). We hypothesized that such testing as a form of active retrieval of the initial study episode would lead to reencoding and more elaborated memory traces; in consequence, we expected that the left parietal old–new effect would be larger to previously tested items than to previously untested items. In contrast, we expected to find no modulation of the early midfrontal old–new effect, because combined study–test repetitions had no influence on it, either (de Chastelaine et al., 2009; R. J. Johnson et al., 1998; Nessler et al., 2007).

The possible effects of testing on two later ERP old–new effects—namely, the late right frontal old–new effect and the late posterior negativity (LPN)—were more difficult to forecast. The role of the late right frontal old–new effect in retrieval is yet disputed (Cruse & Wilding, 2009; Dobbins & Han, 2006; Hayama, Johnson, & Rugg, 2008). Some studies have reported that the right frontal effect was larger for old–new decisions with high confidence, conflicting with the assumption that it reflects monitoring processes (Cruse & Wilding, 2009; Woodruff et al., 2006). In some other ERP studies on episodic memory, the late frontal old–new effect was unexpectedly not apparent (e.g., Leynes, Cairns, & Crawford, 2005; Rosburg, Johansson, & Mecklinger, 2013). In view of these heterogeneous findings, we did not make any predictions as to how the late right frontal old–new effect would be modulated by testing.

The LPN, as the second major late old–new effect, is primarily observed in tasks that require source memory decisions and is much less pronounced in recognition tasks (Johansson, Stenberg, Lindgren, & Rosén, 2002). The LPN is presumed to reflect an attempt to reconstruct the prior study episode when task-relevant attribute conjunctions are not readily recovered or when they need continued evaluation (Johansson & Mecklinger, 2003). We expected that previous testing should increase the efficiency with which task-relevant attribute conjunctions can be recovered and should decrease the need for continued evaluation. In consequence, we hypothesized that the amplitude of the LPN would decrease when items were previously tested.

In our study, we investigated how ERP old–new effects were modulated by an immediately preceding source memory test. Participants studied object names that were presented together with a picture of the denoted object or with the instruction to mentally visualize the object (Fig. 1). In an initial test (1st test), half of the items were tested, whereas the other remained untested. In a subsequent test (2nd test), all items were tested. Possible effects of testing were analyzed by comparing the old–new effects between the 1st and 2nd tests, but also by comparing the old–new effects at the 2nd test between previously tested and untested items. Only when these two comparisons converged and showed a significantly larger or smaller old–new effect for previously tested items was an effect of testing assumed to be present.

Method

Participants

A group of 32 volunteers (16 female, 16 male), ranging in age from 20 to 31 years (mean age 25 years) took part in the experiment. The data from another five participants were discarded due to excessive ocular artifacts during electroencephalographic (EEG) recordings. Due to the nature of the task, only German native speakers were included. All of the participants but one were right-handed, and all had normal or corrected-to-normal vision. Participants were informed about the procedure of the experiment and gave written consent prior to participation. All participants were students at Saarland University. Participation was compensated with €8/h or course credit. The study has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Experimental procedure

The experimental setup was adopted from our previous study (Rosburg, Mecklinger, & Johansson, 2011a). The present experiment consisted of three phases: a study phase and two subsequent test phases. During the study phase (Fig. 1A), participants saw German object names followed by either a picture of the denoted object (perceive condition) or a blank

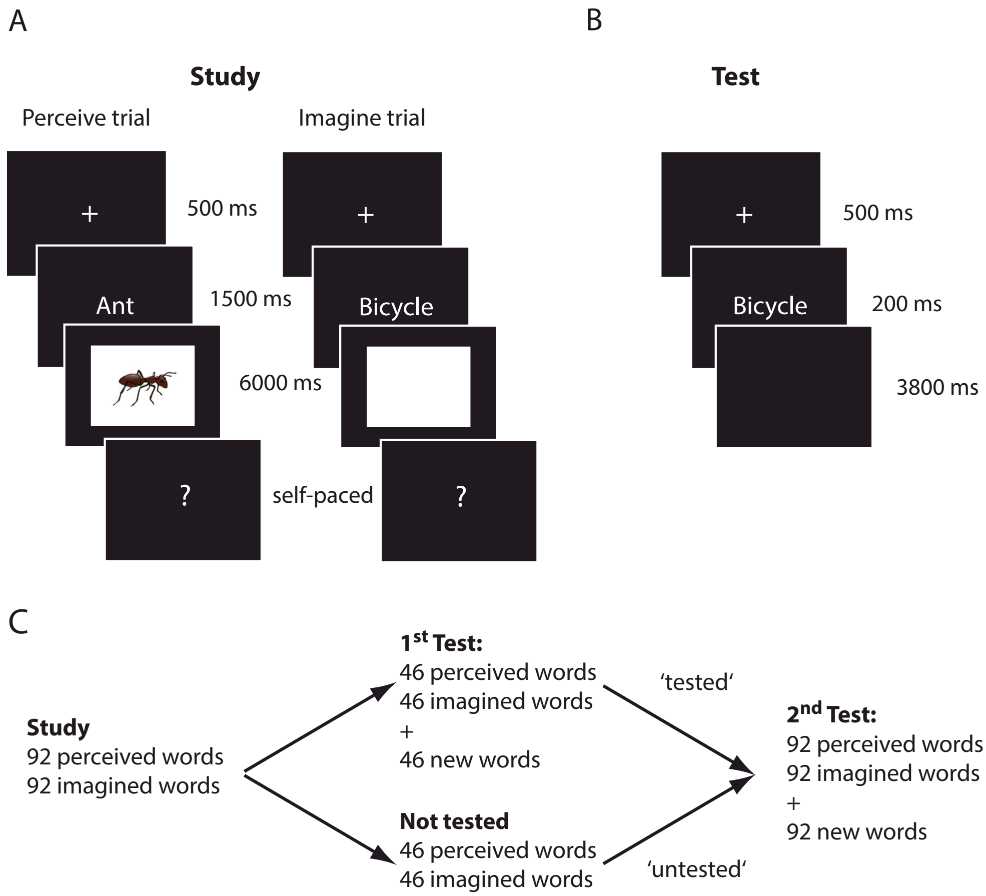

Fig. 1 Schematic view of the experiment: The experiment consisted of one study phase and two test phases. During the study phase, participants saw objects words, together with drawings of the denoted object (perceive trials) or followed by the instruction to imagine such drawings (imagine trials). In subsequent source memory tests, participants had to identify object words as being perceived, imagined, or new. Half of the studied items had been presented in the 1st test, and all items were presented at the 2nd test. The labels "tested" and "untested" at the 2nd test refer to whether study items had been presented at the 1st test ("tested") or not ("untested")

frame (imagine condition). Each trial started with the presentation of a fixation cross for 500 ms, followed by an object name that was presented for 1,500 ms. In the perceive condition, object names were followed by a colored picture of the object superimposed on a white rectangle; in the imagine condition, by a white rectangle without a picture. The pictures and empty rectangles were both shown for 6,000 ms. In the perceive condition, participants had to examine the graphic distinctiveness of the picture and to assess how well the picture illustrated the denoted object by pressing a button on the numerical part of the computer keyboard (with 1 for *well*, 2 for *fairly*, and 3 for *badly*). Artistic merit and clarity of representation were suggested as criteria for the judgment. In the imagine condition, participants were instructed to imagine a drawing of the named object and to project it on to the blank rectangle. Thereafter, participants had to rate the graphic distinctiveness of the imagined picture in the same way as for the perceived pictures. Identical study tasks had been used in the reality-monitoring studies of M.K. Johnson, Kounios, and Reeder (1994) and Johansson et al. (2002). The next trial started 500 ms after the rating. Perceived and imagined items were presented in a random order, with no more than three items of one condition occurring in succession.

In the two subsequent source memory tests (Fig. 1B), trials started with a 1,000-ms-long fixation cross, followed by an empty screen lasting for 500 ms. Next, object names were presented for 200 ms. Then a question mark appeared for 3,600 ms. Participants had to indicate by buttonpress on the keyboard whether a presented object had been perceived, imagined, or unstudied (new). The instructions emphasized responding as quickly and as accurately as possible. Participants pressed separate keys with the index and second finger of one hand in order to indicate the source of the recognized object names; they pressed a key with the index finger of the other hand when an object name was new. The mappings between fingers and response keys for source judgments were counterbalanced across participants. When participants did not press a response key within 3,600 ms, a missed response was recorded (0.4 % of the trials). Participants were informed at the beginning of the test phase that they would be tested twice, but not that some of the items would be presented at both tests.

Materials

Participants were seated 60–80 cm in front of a 17-in. monitor. All displays were centered and had a black background. Words were presented in white 18-point Courier New font. A white rectangle with a size of 75 % of the screen height and width displayed the pictures in the perceive condition or remained blank in the imagine condition. The colored drawings in the perceive condition originated from Rossion and Pourtois (2004) and were based on the picture set of Snodgrass and Vanderwart (1980). Object names had to meet the following criteria: (a) a word length between three and ten characters, and (b)a word frequency ranging from 1 to 475 occurrences per million (as verified with the CELEX linguistic database by Baayen, Piepenbrock, & van Rijn, 1993). Only names of highly concrete real-world objects were used.

A set of 184 object names served as the materials for the study phase, with 92 items each being presented in the two conditions (Fig. 1C). In the 1st test, 46 perceived and 46 imagined items were presented, together with 46 new object names. In the 2nd test, all 184 studied object names (half of them tested during the 1st test) were presented, together with 92 new object names. The study phase and the two tests were separated by short breaks of a couple of minutes. Lists of items presented in the two study conditions, as well as lists of tested and untested items during the 1st test, were counterbalanced across participants. Due to the limited number of colored drawings, the lists of study items and lures were not counterbalanced across participants. However, we made sure that these word lists did not vary in word length and word frequency. Different sets of new object names were used for the 1st and 2nd test—that is, new items at the 1st test were not used as lures at the 2nd test.

EEG recording

Prior to the study phase, an elastic cap (Easycap, Herrsching, Germany) with 58 embedded silver/silver-chloride EEG electrodes was attached to a participant's head. The electrode locations in these caps are based on the extended 10–20 system. EEG was continuously recorded, referenced to the left mastoid. In addition, electro-ocular activity was recorded by a pair of electrodes affixed to the outer canthi and by a pair of electrodes placed below and above the right eye. Data were sampled at 500 Hz and filtered online from 0.016 Hz (time constant 10 s) to 250 Hz. Electrode impedances were kept below 5 kΩ.

ERP data processing

Offline, data were digitally filtered from 0.1 to 40 Hz (48 dB), with an additional notch filter at 50 Hz to suppress line activity, and re-referenced to linked mastoids. Data were sampled down to 200 Hz and exported to EEGLab (Swartz Center for Computational Neuroscience, University of California San Diego, USA). In EEGLab, an independent component analysis (ICA) was run in order to identify and to eliminate the impact of ocular, muscular, electrocardiographic, and technical artifacts. Data were segmented into epochs of 3,000-ms duration, including a 500-ms baseline. Data were baseline-corrected and screened for artifacts that had remained undetected by the ICA procedure. Trials with EEG activity exceeding ±100 μV, exhibiting abnormal trends (R^2 limit = 0.3), or that were abnormally distributed (±5 *SD*s from the mean) were excluded. For the 1st test, ERPs were calculated for correctly rejected new items (CR_New) and for correct source judgments of old items (Hit_hits_Imagined, Hit_hits_Perceived). For the 2nd test, separate ERPs were calculated for correct rejections of new items and for correct source judgments of old items that had previously been tested ("tested") and that had not been tested before ("untested"), again for imagined and perceived items separately. Individual ERPs were based on at least 19 trials. Exact trial numbers are provided in Supplementary Table S1. The ERPs to false alarms could not be analyzed due to the limited number of such trials.

Statistics

Response accuracy for studied object names was quantified as the discrimination indices for item and source memory. In detail, we differentiated between P_Hit_hits as the hit rate for studied items with correct source judgments and P_Hit_misses as the hit rate for studied items with incorrect source judgments. Subsequently, the false alarm rate to new items (item memory) or to nontargets (source memory) was subtracted from the hit rates. Thus, the discrimination index for item memory was defined as Pr_Item = (P_Hit_hits + P_Hit_misses) – P_FA (false alarms) to new items; the discrimination index for source memory as Pr_Source = P_Hits_hits – P_FA to nontargets (Snodgrass & Corwin, 1988). These accuracy measures were calculated for the two sources (perceived and imagined items) separately. The response accuracies and reaction times (RTs) were compared by a repeated measures analysis of variance (ANOVA) with Source (imagined vs. perceived) and Test Session (1st vs. 2nd test) as within-subjects factors. Test Session effects were analyzed separately for "tested" and "untested" items. Behavioral testing effects were further analyzed by comparing the response accuracies and RTs within the 2nd test in a repeated measures ANOVA with Source (imagined vs. perceived) and Test Status ("tested" vs. "untested" items) as within-subjects factors. The effects of test session on the behavioral responses to new items were tested by paired *t* tests.

For the analyses of the ERP effects, we first assessed whether significant old–new effects occurred at each test.

For these analyses, the mean ERP amplitudes between 300 and 500, 500 and 700, 700 and 900, 900 and 1,200, 1,200 and 1,500 ms, and 1,500 and 1,800 ms were calculated. The individual old–new effects were analyzed at single electrodes that were representative for them, by comparing the ERPs to old and new items in a repeated measures ANOVA with Item (Hit_hit_imagined vs. Hit_hit_perceived vs. CR_new) as within-subjects factor. The selection of the time windows and electrodes was based on previous studies (e.g., Johansson et al., 2002; Rosburg et al., 2013; Rosburg, Mecklinger, & Johansson, 2011b; Rugg et al., 1998), but the topography and latencies of old–new effects at the 1st test were also taken into account, in particular for the early and parietal old–new effects, as we document below. ANOVAs were run separately for the 1st and 2nd tests, and, within the 2nd test, separately for previously tested and previously untested items. Second—and this was our primary interest—we assessed whether old–new effects were modulated by repeated testing. For that, we calculated the old–new effects by subtracting the ERP responses to new items from the ERP responses to old items, for each category of items separately. Then, we compared the old–new effects between the 1st and 2nd tests in two repeated measures ANOVAs with Source (imagined vs. perceived) and Test Session (1st vs. 2nd test) as within-subjects factors, again separately for the "tested" and "untested" items. In addition, we compared the old–new effects within the 2nd test by an ANOVA with Source (imagined vs. perceived) and Test Status ("tested" vs. "untested" items) as within-subjects factors.

For all repeated measures ANOVAs, a Greenhouse–Geisser correction for nonsphericity was performed when necessary, and corrected p values are reported, as is indicated by the citation of ε values. Exploratory correlation analyses of significant testing effects were performed by calculating Pearson product-moment correlation coefficients between them. The α criterion was set to $p = .05$ for all analyses. Statistical tests were computed with the SPSS software package 19.0 (IBM, USA) and GPower 3.1.9 (University of Düsseldorf, Germany; Faul, Erdfelder, Lang, & Buchner, 2007).

Results

Behavioral data

Study phase The ratings during the study phase did not vary between the two conditions (mean rating for perceived items = 1.39, $SD = 0.27$, vs. mean rating for imagined items = 1.47, $SD = 0.30$), $t(30) = 1.516$, $p = .140$, $d = 0.272$, with the data of one participant excluded as statistical outlier.

Test Session effects (1st vs. 2nd test) The rate of correct rejections decreased from the 1st to the 2nd test [Table 1; $t(31) = 2.325$, $p = .027$, $d = 0.413$], whereas the RTs to these items remained constant [$t(31) = 0.646$, $p = .523$, $d = 0.114$; Table 2]. Item and source memory improved for previously tested object names (i.e., 2nd vs. 1st test) [$F(1, 31) = 10.531$, $p = .003$, $\eta^2 = .254$, and $F(1, 31) = 7.697$, $p = .009$, $\eta^2 = .199$, respectively; Table 1]. Study condition did not modulate these accuracy measures (source and Source × Test Session effects: all Fs < 1.044, ps > .314, η^2s < .033). For on average about 10.0 items (range 2 to 23 items), participants provided a more accurate response at the 2nd test, whereas their responses were less accurate at the 2nd test for on average about 5.4 items (range 0 to 17 items) (for details, see Table S2). Participants were faster to correctly identify previously tested object names [$F(1, 31) = 91.104$, $p < .001$, $\eta^2 = .746$; Table 2]. The RT decrease was not modulated by the study condition [Source × Test Session: $F(1, 31) = 0.349$, $p = .559$, $\eta^2 = .011$], but RTs were generally shorter for perceived than for imagined items [source: $F(1, 31) = 12.456$, $p = .001$, $\eta^2 = .287$]. For "tested" items, RTs to correctly attributed old items at the 2nd test (Hit_hits) were separately calculated across sources for items that were attributed to the correct source at the 1st test as well as for items that were subject to an incorrect response at the 1st test. The comparison of these RTs in a post-hoc analysis with the RTs in the 1st test showed that the RTs sped up only for items that were correctly identified in both tests (1st test, 1,168.8 ± 217.9 ms, vs. 2nd test, 959.0 ± 194.6 ms) [$F(1, 31) = 110.577$, $p < .001$, $\eta^2 = .781$], but not for items for which the source attribution was corrected (2nd test, 1,168.0 ± 291.4 ms) [$F(1, 31) < 0.001$, $p = .984$, $\eta^2 < .001$].

Item memory for previously untested object words at the 2nd test was poorer than item memory at the 1st test [$F(1, 31) = 22.698$, $p < .001$, $\eta^2 = .423$]. The drop in item memory performance was modulated by source [Source × Test Session interaction: $F(1, 31) = 7.183$, $p = .012$, $\eta^2 = .188$], being more pronounced for imagined [$t(31) = 4.664$, $p < .001$, $d = 0.824$] than for perceived [$t(31) = 2.130$, $p = .041$, $d = 0.376$] items. For source memory, we found no significant main effect of test session [$F(1, 31) = 1.546$, $p = .223$, $\eta^2 = .048$], but again, a significant Source × Test Session interaction [$F(1, 31) = 6.742$, $p = .014$, $\eta^2 = .179$]. Pairwise comparisons revealed marginally significant poorer source memory for "untested" imagined items at the 2nd test than for imagined items at the 1st test [$t(31) = 1.997$, $p = .055$, $d = 0.354$], but no such difference for perceived items. The RTs did not vary between old items of the 1st test and "untested" items of the 2nd test [test session, $F(1, 31) = 2.785$, $p = .105$, $\eta^2 = .082$; Source × Test Session interaction, $F(1, 31) = 2.291$, $p = .140$, $\eta^2 = .069$].

Test Status effects ("tested" vs. "untested" items at the 2nd test) Direct comparison of discrimination indices for items of the 2nd session showed that item and source memory was better for "tested" items than for "untested" items [$F(1, 31) = 42.778$, $p < .001$, $\eta^2 = .580$, and $F(1, 31) = 19.739$, $p < .001$, η^2

Table 1 Behavioral response accuracy in the two source memory tests

Accuracy	Perceived		Imagined		New
1st Test	P_Hit_hits	.82 (.10)	P_Hit_hits	.81 (.12)	CR .97 (.04)
	Pr_Item	.88 (.08)	Pr_Item	.88 (.10)	
	Pr_Source	.75 (.14)	Pr_Source	.73 (.17)	
2nd Test (“tested”)	P_Hit_hits	.86 (.08)	P_Hit_hits	.86 (.13)	CR .95 (.05)
	Pr_Item	.92 (.05)	Pr_Item	.90 (.09)	
	Pr_Source	.79 (.12)	Pr_Source	.78 (.15)	
2nd Test (“untested”)	P_Hit_hits	.79 (.12)	P_Hit_hits	.79 (.14)	
	Pr_Item	.86 (.09)	Pr_Item	.82 (.11)	
	Pr_Source	.74 (.14)	Pr_Source	.71 (.16)	

Response accuracy was quantified as hit rate for studied items with a correct source judgment (P_Hits_hits), as discrimination index for correctly identifying studied items as old (Pr_Item = P_Hit_hits+ P_Hits_misses– P_False alarms new items), as discrimination index for correctly remembering the source of items that were identified as old (Pr_Source = P_Hit_hits– P_False alarms nontargets), and finally as rate for correctly rejecting new items (CR_New). Response accuracies for retested (“tested”) and previously untested (“untested”) items at the 2nd test are provided separately. Numbers in brackets indicate standard deviations (*SD*s).

= .389, respectively]. Furthermore, participants responded faster to “tested” than to “untested” items [$F(1,31) = 189.546$, $p < .001$, $\eta^2 = .859$].

Taken together, item memory and source memory for previously tested items were better, and RTs were shorter, as compared both to previously untested items and to the initial testing. Memory source did not modulate these testing effects.

ERP old–new effects

The ERPs at selected electrodes are displayed in Fig. 2, for the 1st test, as well as for “tested” and “untested” items at the 2nd test, separately. The topographies of the corresponding old–new effects are shown in Fig. 3. The ANOVA results are summarized in Table 3. The table focuses on the main effects of Item, Test Session, Test Status, and Source, since significant Source × Test Session and Source × Test Status interactions were, with one exception detailed below, not observed.

Early old–new effect (Fz, Cz, Pz) In the present study, the early old–new effect had its maximum at posterior electrode sites, whereas it usually has a midfrontal distribution (Fig. 3). In order to take this unexpected distribution into account, old–new effects were calculated for the 300- to 500-ms time window at three electrodes (Fz, Cz, Pz). Across all old–new contrasts, the early old–new effect was most reliably observed at electrode Pz (Table 3). At electrode Fz, a significant Test Session effect was observed for “tested” items, but the Test Status effect did not reach significance. In contrast, at electrodes Cz and Pz, the early old–new effect was larger for “tested” than for “untested” items, as indicated by significant Test Status effects, but the Test Session effects for “tested” items did not reach significance. In sum, the analysis provided no converging evidence for an effect of testing on the early old–new effect.

Left parietal old–new effect (P5) At the 1st test, the left parietal old–new effects for hits were relatively short-lasting and had already ended at about 700 ms (Fig. 2). This might have been due to an early onset of the LPN. In order to take the short duration of the left parietal old–new effect at the 1st test into account, the left parietal old–new effect was analyzed for two consecutive time windows, from 500 to 700 ms and 700 to 900 ms. From 500 to 700 ms, the left parietal old–new effect was present for hits at 1st and 2nd test. The effect was larger for hits to “tested” items than for hits at the 1st test and for hits to “untested” items at the 2nd test, whereas the magnitudes did not vary between the latter two (Fig. 2, Table 3). From 700 to 900 ms, the effect was only present at the 2nd test. Consequently, the test session effect was significant for both for ”tested“ than for “untested” items. However, within the 2nd test, the parietal old–new effect from 700 to 900 ms was again larger for “tested” than for “untested” items (Table 3). Taken together, the left parietal old–new effect was increased for previously tested items from 500 to 900 ms, as indicated by the significant Test Session *and* Test Status effects.

Table 2 Mean reaction times (RTs, with ±*SD*s) for correct source judgments and correct rejections of new items in the two tests

RTs	Hit_hits_Perceived	Hit_hits_Imagined	CR_New
1st test	1,125.7 (200.1)	1,215.8 (259.0)	884.8 (186.5)
2nd test (“tested”)	947.2 (172.5)	1,027.0 (257.5)	894.2 (203.7)
2nd test (“untested”)	1,116.1 (184.8)	1,167.1 (239.4)	

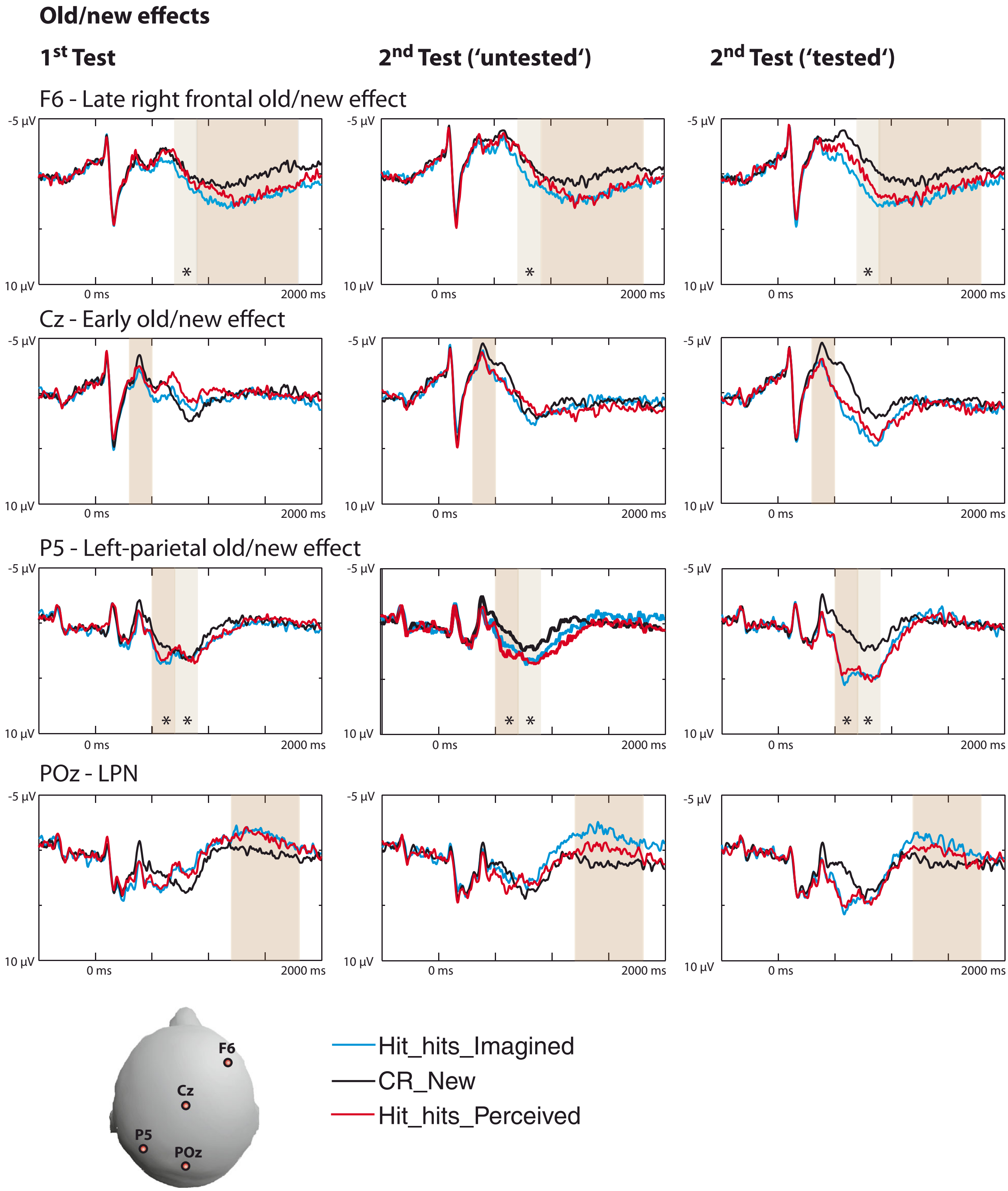

Fig. 2 Event-related potentials (ERPs) at electrodes selected for the analysis of old–new effects: In each column, ERPs to correctly rejected new items (black lines), to hits for perceived items (red lines), and to hits for imagined items (blue lines) are depicted. In the left column, the ERP data of the 1st test are shown. The columns in the middle and on the right depict the ERP data of the 2nd test for previously untested ("untested") and previously tested ("tested") items, respectively. Each row contains ERP data from a single electrode, as indicated by the subheadings. The electrode positions are provided by the head view at the bottom. Latencies with significant old–new effects in at least one comparison are marked by shading. Different shadings are used to highlight differential condition effects; latencies with converging Test Status and Test Session effects, indicating significant testing effects, are marked by asterisks. Negative values are plotted upward

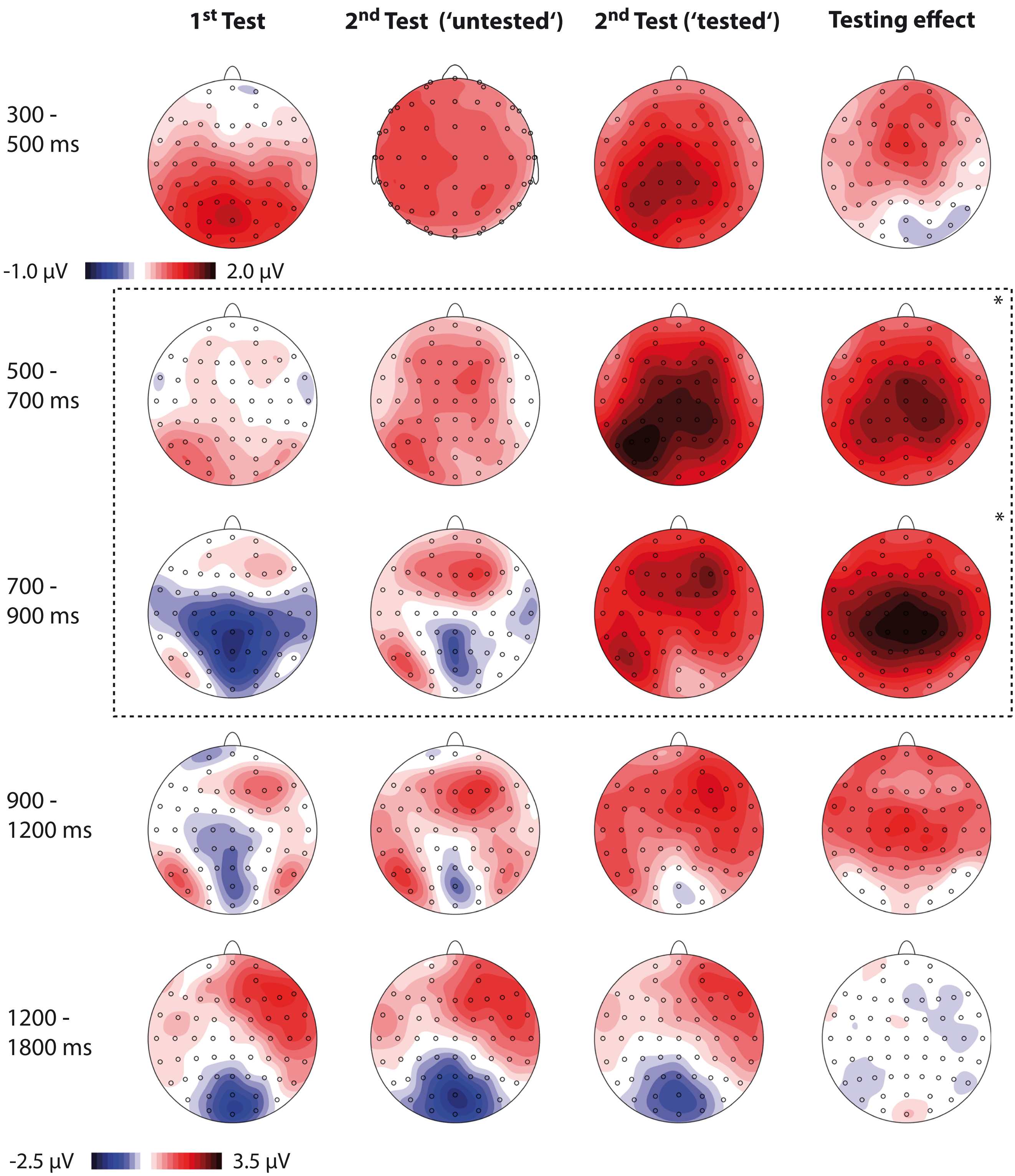

Fig. 3 Old–new effects for the 1st and 2nd tests, and ERP testing effects for the latency ranges from 300 to 500, 500 to 700, 700 to 900, 900 to 1,200, and 1,200 to 1,800 ms: Old–new effects are collapsed across item sources in order to highlight the variation of the old–new effects by repeated testing. The first three columns depict the old–new effects of the 1st test (left), as well as of the 2nd test for "untested" items (left middle) and "tested" items (right middle) separately. The right column illustrates the effects of testing on the ERP old–new effects ([ERP to "tested" items – ERP to new items at the 2nd test] – [ERP to old items – ERP to new items at the 1st test]). The magnitude of the old–new effects is provided by color scaling. Latencies with converging Test Status and Test Session effects, indicating significant testing effects, are marked by asterisks and highlighted by the dotted box

Table 3 Summary of the conducted repeated measures analyses of variance on the event-related potential data

	ITEM 1st test	ITEM 2nd test 'Tested'	ITEM 2nd test 'Untested'	TEST SESSION 'Tested'	TEST SESSION 'Untested'	TEST STATUS 2nd test	SOURCE 1st test	SOURCE 2nd test
Early frontal old/new effect (300-500 ms)								
300-500 ms Fz	F = 0.455 P = 0.600 η^2 = 0.014	F = 4.297 P = 0.018 η^2 = 0.122	F = 3.146 P = 0.050* η^2 = 0.092	F = 5.473 P = 0.026 η^2 = 0.150	F = 2.495 P = 0.124 η^2 = 0.074	F = 0.597 P = 0.446 η^2 = 0.019	t = 0.905 P = 0.373 d = 0.160	F = 1.836 P = 0.185 η^2 = 0.056
300-500 ms Cz	F = 3.777 P = 0.028* η^2 = 0.109	F = 9.683 P < 0.001 η^2 = 0.238	F = 3.946 P = 0.034*,** η^2 = 0.113	F = 3.197 P = 0.084 η^2 = 0.093	F = 0.094 P = 0.761 η^2 = 0.003	F = 4.713 P = 0.038 η^2 = 0.132	t = 1.773 P = 0.086 d = 0.314	F = 0.007 P = 0.932 η^2 = 0.000
300-500 ms Pz	F = 8.602 P = 0.001 η^2 = 0.217	F = 9.902 P < 0.001 η^2 = 0.242	F = 4.562 P = 0.020*** η^2 = 0.128	F = 0.086 P = 0.771 η^2 = 0.003	F = 1.858 P = 0.183 η^2 = 0.057	F = 6.384 P = 0.017 η^2 = 0.171	t = 0.858 P = 0.398 d = 0.200	F = 1.822 P = 0.187 η^2 = 0.056
Left-parietal old/new effect (500-900 ms)								
500-700 ms P5	F = 7.218 P = 0.002 η^2 = 0.189	F = 45.665 P < 0.001 η^2 = 0.596	F = 12.744 P < 0.001 η^2 = 0.291	F = 27.382 P < 0.001 η^2 = 0.469	F = 1.746 P = 0.234 η^2 = 0.045	F = 47.713 P < 0.001 η^2 = 0.606	t = 1.283 P = 0.209 d = 0.227	F = 0.172 P = 0.681 η^2 = 0.006
700-900 ms P5	F = 0.063 P = 0.939 η^2 = 0.003	F = 20.079 P < 0.001 η^2 = 0.393	F = 5.683 P = 0.005 η^2 = 0.155	F = 34.616 P < 0.001 η^2 = 0.528	F = 11.723 P = 0.002 η^2 = 0.274	F = 15.271 P < 0.001 η^2 = 0.330	t = 0.284 P = 0.778 d = 0.050	F = 0.100 P = 0.754 η^2 = 0.003
LPN (700-1800 ms)								
700-900 ms POz	F = 6.507 P = 0.003 η^2 = 0.173	F = 1.350 P = 0.267 η^2 = 0.042	F = 2.315 P = 0.107 η^2 = 0.069	F = 21.485 P < 0.001 η^2 = 0.489	F = 2.219 P = 0.146 η^2 = 0.067	F = 11.961 P = 0.002 η^2 = 0.278	t = 0.534 P = 0.597 d = 0.094	F = 0.044 P = 0.835 η^2 = 0.001
900-1200 ms POz	F = 2.342 P = 0.105 η^2 = 0.070	F = 2.115 P = 0.129 η^2 = 0.064	F = 3.787 P = 0.028 η^2 = 0.109	F = 1.961 P = 0.171 η^2 = 0.060	F = 0.025 P = 0.877 η^2 = 0.001	F = 2.062 P = 0.161 η^2 = 0.062	t = 0.157 P = 0.876 d = 0.028	F = 7.573 P = 0.010 η^2 = 0.196
1200-1500 ms POz	F = 5.084 P = 0.009 η^2 = 0.141	F = 9.970 P < 0.001 η^2 = 0.243	F = 15.258 P < 0.001 η^2 = 0.330	F = 0.033 P = 0.858 η^2 = 0.001	F = 1.164 P = 0.289 η^2 = 0.036	F = 2.251 P = 0.144 η^2 = 0.068	t = 0.510 P = 0.613 d = 0.090	F = 8.767 P = 0.006 η^2 = 0.220
1500-1800 ms POz	F = 3.671 P = 0.031 η^2 = 0.106	F = 5.363 P = 0.007* η^2 = 0.147	F = 10.738 P = 0.001**** η^2 = 0.257	F = 0.580 P = 0.452 η^2 = 0.018	F = 0.730 P = 0.399 η^2 = 0.023	F = 9.304 P = 0.005 η^2 = 0.231	t = 0.516 P = 0.610 d = 0.091	F = 6.572 P = 0.015 η^2 = 0.175
Late right frontal old/new effect (700-1800 ms)								
700-900 ms	F = 3.795	F = 18.161	F = 7.508	F = 12.207	F = 0.147	F = 16.197	t = 2.526	F = 14.459
F6	P = 0.028* η^2 = 0.109	P < 0.001 η^2 = 0.369	P = 0.001* η^2 = 0.195	P = 0.001 η^2 = 0.283	P = 0.704 η^2 = 0.005	P < 0.001 η^2 = 0.343	P = 0.017 d = 0.446	P = 0.001 η^2 = 0.318
900-1200 ms F6	F = 7.680 P = 0.001* η^2 = 0.199	F = 10.625 P < 0.001 η^2 = 0.255	F = 3.869 P = 0.026* η^2 = 0.111	F = 3.076 P = 0.089 η^2 = 0.090	F = 0.002 P = 0.968 η^2 = 0.000	F = 3.541 P = 0.069 η^2 = 0.103	t = 2.777 P = 0.009 d = 0.491	F = 1.835 P = 0.185 η^2 = 0.056
1200-1500 ms F6	F = 13.078 P < 0.001 η^2 = 0.297	F = 6.803 P = 0.002 η^2 = 0.180	F = 13.267 P < 0.001 η^2 = 0.300	F = 1.017 P = 0.321 η^2 = 0.032	F = 0.005 P = 0.942 η^2 = 0.000	F = 1.676 P = 0.205 η^2 = 0.051	t = 0.447 P = 0.658 d = 0.079	F = 0.659 P = 0.423 η^2 = 0.021
1500-1800 ms F6	F = 20.152 P < 0.001 η^2 = 0.394	F = 7.704 P = 0.001 η^2 = 0.199	F = 10.752 P < 0.001 η^2 = 0.258	F = 1.666 P = 0.206 η^2 = 0.051	F = 0.732 P = 0.399 η^2 = 0.023	F = 0.211 P = 0.649 η^2 = 0.007	t = 0.849 P = 0.402 d = 0.150	F = 0.663 P = 0.422 η^2 = 0.021

The columns comprehend the main effects of Item (old vs. new), Test Session (1st vs. 2nd test), Test Status ("tested" vs. "untested"), and source (perceived vs. imagined). For a better overview, source effects are constrained to the findings within the 1st and 2nd tests, and reference to the few significant interactions is found in the text. Converging Test Session and Test Status effects are marked by black edging. * Post-hoc tests showed a significant old–new effect for imagined items, but not for perceived items; ** ε = .749; *** ε = .827, **** ε = .795.

LPN (POz) At the 1st test, an early-occurring LPN was observed from 700 to 900 ms, whereas no such early LPN occurred at the 2nd test (Figs. 2 and 3, Table 3). For the 900- to 1,200-ms time window, a significant LPN was observed only in response to "untested" items at the 2nd test. From 1,200 to 1,800 ms, a LPN was present for hits at both tests. The LPN to "tested" items from 900 to 1,800 ms was not modulated by test session (Table 3). For the LPN to "untested" items, no main effect of test session was found, either, but a significant Source × Test Session interaction was observed in the 1,200- to 1,500-ms latency window [$F(1, 31) = 5.097$, $p = .031$, $\eta^2 = .141$]; subsequent pairwise comparisons revealed a marginally significant increase of LPN amplitude for "untested" imagined items [$t(31) = 1.927$, $p = .063$, $d = 0.341$], but not for "untested" perceived items [$t(31) = 0.147$, $p = .884$, $d = 0.026$]. Within the 2nd test, the LPN was more negative for "untested" than for “tested” items, but only from 1,500 to 1,800 ms; furthermore, the LPN was more negative for imagined than for perceived items from 900 to 1,800 ms. Taken together, no converging effects of Test Session and Test Status on the LPN of previously tested items were apparent, except at the early latency range (700–900 ms).

Right frontal old–new effect (F6) The late right frontal old–new effect started already at about 700 ms. Its early portion (700–900 ms) was influenced by testing: It was found to be larger for ”tested” items at the 2nd test than for both “untested” items at the 2nd test and items at the 1st test (Fig. 2). In the subsequent time windows (900–1,800 ms), the old–new effect was present for hits at both tests and was not modulated by previous testing (Figs. 2 and 3, Table 3). Effects were larger for imagined items than for perceived items between 700 and 1,200 ms at the 1st test, and between 700 and 900 ms at the 2nd test. Taken together, only the early portion of the right frontal old–new effect (700–900 ms) was increased after previous testing.

Correlation analysis

Exploratory correlation analyses were run across variables that showed a significant effect of previous testing. The mean changes from the 1st to 2nd test for “tested” items across the conditions were calculated for these variables (e.g., ΔP_Hit_hit = P_Hit_hits 2nd test – P_Hit_hits 1st test). The variables included the left parietal old–new effect at P5 from 500 to 700 ms and from 700 to 900 ms, the frontal old–new effect at F6 from 700 to 900 ms, the Pr_source and Pr_item scores, and the RT to Hit_hits. For the increased old–new effects, we observed no correlations with the speed-up of RTs ($.053 < r\text{s} < .196$, $p > .285$) and with the measures of improved retrieval accuracy ($-.325 <$ all $r\text{s} < .080$, $p > .068$; see Supplementary Table S4).

Discussion

The main findings can be summarized as follows: For previously tested items, retrieval accuracy was better and reaction times were shorter than for the initial testing and the previously untested items. In the ERPs, converging effects of testing were observed for the left parietal old–new effect (500–900 ms) and for the early portion of the right frontal old–new effect (700–900 ms), with larger old–new effects for “tested” items than for items at the 1st test and for “untested” items. The effects of testing did not differ between imagined and perceived items. In contrast to the previously mentioned old–new effects, the early old–new effect (300–500 ms) was not modulated by previous testing. Substantial effects of testing on the later old–new effects (900–1,800 ms) were not observed, either. In the following, we will discuss, the major findings in detail.

Behavioral effects

In the present study, we investigated the immediate neurobehavioral effects of a single test session, by contrasting the retrieval performance of items at their 1st and 2nd test (Test Session effect) and by contrasting the retrieval performance for previously tested and untested items at the 2nd test (Test Status effect). Using two baselines is important, because for ”untested” items the 1st test represents a retention interval, during which some items might possibly be forgotten or inhibited by the retrieval of other items at the 1st test (*retrieval induced forgetting*; Anderson, Bjork, & Bjork, 1994). Indeed, the present behavioral data showed that item memory accuracy was lower for ”untested” items at the 2nd test than for items at the 1st test. In contrast, item and source memory accuracy for ”tested” items improved from the 1st to the 2nd test even though the participants did not receive any feedback during the tests. Presumably, at the 1st test, participants became aware of some of their incorrect responses and adjusted them in the 2nd test. Such an adjustment was about two times more likely than a deterioration of a correct source decision. Increased retrieval accuracy after testing without feedback was also reported by Hashimoto et al. (2011).

In addition to the improved retrieval accuracy, reaction times to correctly identified ”tested” items strongly decreased from the 1st to the 2nd test. These shorter reaction times cannot be regarded as the consequence of a general practicing effect because reaction times to new items did not speed up. The improved retrieval accuracy and speed-up of reaction times for previously tested items underline that even single testing has a beneficial effect on later memory performance. However, the speed-up reaction times for ”tested” items at the 2nd test were only found when these items were attributed to the correct source at the 1st test as well.

Early old–new effect (300–500 ms) We hypothesized that the early old–new effect would not be modulated by testing, because it was reported not to be influenced by repeating study–test cycles (de Chastelaine et al., 2009; Johnson et al., 1998; Nessler et al., 2007). In our analysis, we did not reveal clear evidence that the early old–new effect was modulated by testing. However, strikingly, the early old–new effect had a posterior scalp distribution. Such early, posteriorly distributed old–new effects have been functionally dissociated from the early midfrontal old–new effect and have been associated with perceptual fluency and implicit memory processes (Bridger et al., 2012; Leynes & Zish, 2012; Rugg et al., 1998; Yu & Rugg, 2010; but see Voss & Federmeier, 2011). Thus, on the basis of the topography of the currently observed early old–new effect, presumably perceptual fluency or implicit memory processes were elicited by the test cues, rather than familiarity-related processes.

Two factors might have contributed to why familiarity-related processes were not observed: First, the analyzed ERP contrast (Hit_hit vs. CR_new) highlighted recollection (accurate source judgments) rather than familiarity-related processes (Evans & Wilding, 2012). Second, the encoding task emphasized pictorial information, whereas verbal cues were presented at test. Consequently, format change from study to test might have played some role, and such a format change from study to test is known to attenuate familiarity-related processes (e.g., Schloerscheidt & Rugg, 2004; Stenberg, Johansson, & Rosén, 2006). Taken together, the present study is inconclusive with regard to the modulation of familiarity-related processes by testing.

Parietal old–new effect (500–900 ms) We hypothesized that testing is associated with more elaborated memory traces and, therefore, should result in a larger left parietal old–new effect (Vilberg & Rugg, 2009; Wilding, 2000). This hypothesis was confirmed: In both tested time intervals (500–700 ms and 700–900 ms), the left parietal old–new effect was strongly increased after previous testing.

In the later time interval (700–900 ms), the left parietal old–new effect to ”untested” items was increased as compared to the 1st test, as well. We interpreted this increase for ”untested” items as a general practice effect. At the 1st test, the left parietal old–new effect from 700 to 900 ms was absent, whereas an LPN was present in this latency range (Table 3). It has previously been shown that the early portion of the LPN decreases with repetition of a retrieval task (Herron, 2007): When participants performed several study–test blocks with *different* study materials, the early LPN decreased across blocks. In our study, a similar decrease of the early LPN was observed: It was present at the 1st test, but absent at the 2nd test. Since the LPN and the left parietal old–new effect have opposite polarities, we presume that the parietal old–new effect between 700 and 900 ms at the 1st test was widely diminished by the early LPN, as a result of component overlap. With increasing test practice and decreasing LPN, the left parietal old–new effect in the later time window increased from the 1st to the 2nd test, not just for ”tested” but also for ”untested” items.

Therefore, it is important to stress that the left parietal old–new effect to ”tested” items was increased even when contrasted with the old–new effect to ”untested” items. As mentioned before, we interpret this finding as evidence that previous testing leads to more elaborated memory traces (Vilberg & Rugg, 2009; Wilding, 2000). In a related account, it has been suggested that larger left parietal old–new effects for “remember” than for “know” responses reflect recollection of more details or more vivid details (Leynes & Phillips, 2008). The account points out that with quantitative changes (retrieving more details) the rememberer might also experience qualitative changes (higher vividness of memorized information). The increased left parietal old–new effect to ”tested” items might also reflect other factors secondary to the elaboration of the memory traces. It has been reported that the left parietal old–new effect is also modulated by the salience of information (e.g., Langeslag & van Strien, 2008) or by attention (e.g., Curran, 2004). However, it is important to note that cues of ”tested” items were perceptually not more salient than cues of ”untested” items, and could not per se have attracted more attention, since all of the test cues had the same form. Rather, the salience of cues or the levels of attention paid to them could only have varied on the basis of their memory representations.

However, it might be argued that the effects of testing observed here are the consequence of repeated cue presentation on these memory representations. Previous studies have indeed shown that repeated cue presentation at study leads to better item recognition (Ferrari, Bradley, Codispoti, Karlsson, & Lang, 2013; Finnigan, Humphreys, Dennis, & Geffen, 2002). These findings suggest that even sole cue repetition is to some extent effective for learning. In the present experiment, participants might have endorsed items they remembered as cues from the 1st test (but not from study) by chance to one of the two study conditions. However, behaviorally this strategy would lead to higher estimates for item memory but not for source memory, because the false alarm rates for nontargets (P_Hit_misses) would increase in parallel with the correct source attributions (P_Hit_hits). Instead, we found improved item memory and improved source memory for previously tested items at the 2nd test, as well as no variation of the false alarm rates for nontargets between the tests (Supplementary Table S3). Thus, no evidence was apparent that participants relied on the described strategy. Moreover, the analysis of the behavioral data showed that the speed-up of reaction times for correctly identified ”tested” items crucially depended on whether or not these items were successfully remembered at the 1st test. The RTs to items that were successfully remembered at the 2nd test but not at the 1st test were not accelerated. Thus, cue repetition alone did not determine the accelerated reaction times.

Previous ERP studies have shown an enhancement of the left parietal old–new effect after repeated cue presentation at study (Ferrari et al., 2013; Finnigan et al., 2002). However, even though items in these studies were presented more often (three to four times), the increase of the left parietal old–new effect after multiple presentations was clearly smaller than the currently observed increase after testing. In consequence, we would argue that the observed increase of the left parietal old–new effect after testing is unlikely to be the consequence of the cue repetition alone. At this point, it might be important to reiterate that it was not the purpose of our study to show the superior effectiveness of testing for learning over restudying, but to explore whether and how ERP old–new effects are modulated by previous testing. The effectiveness of testing is usually shown by contrasting conditions in which items were studied and tested to conditions in which items were repeatedly studied. However, the effectiveness of studying and, consequently, also of restudying depends on the study task. Even though our participants studied items only once, their retrieval accuracy was relatively high, probably due to a combination of the picture superiority effect (Paivio, Rogers, & Smythe, 1968) and generation effect (Slamecka & Graf, 1978) at encoding. Under such highly effective study conditions, testing without feedback might be even less efficient than restudying items.

As we described in the introduction, there are several theoretical accounts for the testing effect (Roediger & Butler, 2011). The three described accounts (the elaborated-memory-trace account, the transfer-appropriate processing account, and the model-of-disuse account) do not necessarily rule out each other, but emphasize different aspects of how testing presumably affects memory traces and retrieval processes. One major difference might be seen in the extents to which the testing effect is the consequence of an altered memory trace, an altered retrieval process, or both.

In recent studies using experimental manipulations leading to increased memory strengths or more elaborated memory traces, such as deeper encoding (Herron & Rugg, 2003; Rugg et al., 1998) or longer encoding times (Paller, Hutson, Miller, & Boehm, 2003; Vilberg & Rugg, 2009), retrieval accuracy increased, but reaction times were often widely unaffected. This contrasts to the strong speed-up of reaction times in the present study, which was found in addition to increased retrieval accuracy. Furthermore, we found no evidence in the form of a significant or close-to-significant correlation that the increase of the left parietal old–new effect after previous testing was associated with the decrease in reaction times for previously tested items. In light of these findings, we would argue that the currently observed effect of testing on reaction times is unlikely to be explained by increased memory strength or by more elaborated memory traces alone.

Transfer-appropriate processing has been considered as another mechanism underlying the testing effect (Roediger & Butler, 2011). The transfer-appropriate processing account proposes that memory is enhanced when similar cognitive processes are engaged at study and test (Morris et al., 1977). Following this account, testing has to be considered a situation in which reencoding takes place. On the basis of this presupposition, processes engaged at a previous test provide a better match with processes at a later test than do the processes involved in just (re)studying the material. In contrast to the elaborated-memory-trace account, the transfer-appropriate processing account stresses the quality of the encoding rather than its quantity. To our knowledge, no previous study has directly tested the influence of transfer-appropriate processing on the left parietal old effect. However, it has been shown that transfer-appropriate processing leads to considerably shorter reaction times (McBride & Abney, 2012). Therefore, as one putative explanation, we suggest that the shorter reaction times for ”tested” items in the present study might be regarded primarily as an effect of transfer-appropriate processing. Indeed, the speed-up of RTs was only observed when ”tested” items were correctly identified at the 1st and 2nd tests—thus, for items for which the reencoding situation (1st test) matched with the later (2nd) test—whereas no speed-up was observed for items for which the response was corrected from the 1st to the 2nd test. Future studies that, for example, manipulate the response key assignments between the 1st and 2nd tests might shed light on the contribution of other factors, such as response priming (Rosenbaum & Kornblum, 1982), on the shorter reactions times for ”tested” items. Future studies are also warranted whether the effects of more elaborated memory traces and transfer-appropriate processing can be dissociated experimentally, since the lack of a significant correlation between the increase of the left parietal old–new effect and the speed-up of reaction times could, in principle, also be due to a lack of statistical power.

As a third account for the testing effect, the model of disuse claims that the probability for successful recall is completely determined by retrieval strength and independent of storage strength (Bjork & Bjork, 1992). The model also proposes that retrieval results in increases of storage strength *and* retrieval strength. The present experiment was not designed to modulate storage strength and retrieval strength independently, and in consequence, does not allow for an evaluation of this model as an underlying mechanism of the testing effect.

LPN (900–1,800 ms) The amplitudes of the late LPN (900–1,800 ms) to ”tested” items did not vary between the 1st and 2nd tests. Thus, we could not confirm our hypothesis that the LPN amplitude would decrease for previously tested items. The speeded reaction times for ”tested” items clearly indicate that the efficiency with which task-relevant attribute conjunctions could be recovered was increased after previous testing. However, this apparently had no impact on the LPN. There is also evidence from other studies that task difficulty (or the efficiency with which task-relevant attribute conjunctions can be recovered) does not necessarily affect the LPN amplitude (Sprondel, Kipp, & Mecklinger, 2012).

At the 2nd test, the LPN amplitude was larger for imagined than for perceived items. Such a difference was also observed in our previous study (Rosburg et al., 2011b). It has been proposed that the amount of contextual information potentially available for the reconstruction of the study episode represents one factor influencing the LPN amplitude (Johansson & Mecklinger, 2003; Mecklinger, Johansson, Parra, & Hanslmayr, 2007). This possibly explains why the LPN was modulated by memory source: The amount of available contextual information can be considered as being larger for imagined than for perceived items, because items from both study conditions were associated with visual information, but for imagined items this visual information was complimented by information about the cognitive operations associated with generating these items (M.K. Johnson, Hashtroudi, & Lindsay, 1993). The increased amount of available contextual information for “tested” items (the original encoding task and the 1st testing) might also provide an explanation for the missing modulation of the LPN by previous testing: It might have offset the effects of an increased efficiency with which task-relevant attribute conjunctions were recovered.

As an alternative account, we have recently suggested that the amount of contextual information *actually used* for the reconstruction of the study episode, rather than the amount of contextual information *potentially available* for it, might influence the LPN magnitude (Rosburg et al., 2013). In consequence, the LPN might vary with the to-be-remembered source information, but also with the test instructions. Indeed, Curran, DeBuse, and Leynes (2007) showed that the LPN was larger when a conservative response bias was instructed: Such an instruction might encourage the retrieval of more comprehensive episodic information and continued evaluation of these task-relevant context features. In our study, the test instructions did not vary between the two tests; therefore, the subjectively experienced need for evaluating the retrieved source information might still have been high at the 2nd test, no matter that some of the items had been tested before. This, in turn, might have led to similar LPN effects in the two test runs.

Frontal old–new effect Only the early portion of the late frontal effect (700–900 ms) was modulated by testing and found to be increased for previously tested items. This increase could have, however, been the consequence of the increased parietal old–new effect, since the difference map in this latency range showed a parietal distribution (testing effect; Fig. 3, 700–900 ms). On the basis of the present data, we cannot rule out this possibility, and therefore, we refrain from a functional interpretation of the finding.

Of note, the effect of memory source (imagined vs. perceived) on the early portion of the late frontal old–new effect cannot be explained in such terms, because memory source did not modulate any other old–new effect in this latency range. The early portion of the late frontal old–new effect has been regarded as functionally distinct from the later portions, as has been indicated by the studies of Woodruff et al. (2006) and Cruse and Wilding (2009, 2011). It was found to be larger for remembered items than for highly familiar old items (Woodruff et al., 2006). The effect was absent when participants had to remember spoken materials (male vs. female voice; Cruse & Wilding, 2011), but present when they had to remember visually presented verbal material (Cruse & Wilding, 2009), suggesting a link between this effect and the recovery of modality-specific information. Alternatively, as was discussed by Cruse and Wilding (2011), the effect might depend on the distinctiveness of the source information.

Conclusion

Previous testing was associated with various behavioral consequences. First, even in the absence of external feedback, previous testing allowed the participants to adjust their response at the 2nd test for some studied items that had initially been missed or attributed to the incorrect source. Second, the reaction times to items with such adjusted responses did not speed up at the 2nd test. In contrast, reaction times to correctly identified ”tested” items sped up considerably at the 2nd test, when those items had been correctly identified in the initial test as well. We consider this finding as evidence that transfer-appropriate processing underlies the observed shorter reaction times for previously tested items. Third, previous testing was associated with increased left parietal old–new effects at the 2nd test, providing support for the hypothesis that testing is associated with an elaboration of memory traces. In sum, we suggest that the testing effect might be mediated by multiple rather than singular neurocognitive processes. Further functional studies are warranted to further elucidate these processes and to test alternative theoretical accounts for the testing effect, such as the model of disuse (Bjork & Bjork, 1992).

Author note The research was supported by the German Research Foundation (DFG; Grant No. Ro 3533/1-1). The authors greatly appreciate the technical assistance of Maxie Siebert, Hannah Heindrihoff, and Patrizia Schneider.

References

Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia, 50,* 3004–3014. doi:10.1016/j.neuropsychologia.2012.07.038

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *NeuroImage, 62,* 439–450.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087. doi:10.1037/0278-7393.20.5.1063

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bader, R., Mecklinger, A., Hoppstädter, M., & Meyer, P. (2010). Recognition memory for one-trial-unitized word pairs: Evidence from event-related potentials. *NeuroImage, 50,* 772–781. doi:10.1016/j.neuroimage.2009.12.100

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. 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, NJ: Erlbaum.

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage, 63,* 1334–1342.

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. doi:10.3758/MC.36.3.604

Cruse, D., & Wilding, E. L. (2009). Prefrontal cortex contributions to episodic retrieval monitoring and evaluation. *Neuropsychologia, 47,* 2779–2789. doi:10.1016/j.neuropsychologia.2009.06.003

Cruse, D., & Wilding, E. L. (2011). Temporally and functionally dissociable retrieval processing operations revealed by event-related potentials. *Neuropsychologia, 49,* 1751–1760.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42,* 1088–1106.

Curran, T., DeBuse, C., & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 2–17.

de Chastelaine, M., Friedman, D., Cycowicz, Y. M., & Horton, C. (2009). Effects of multiple study–test repetition on the neural correlates of recognition memory: ERPs dissociate remembering and knowing. *Psychophysiology, 46,* 86–99.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13,* 1–21. doi:10.3758/BF03193807

Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2011). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. *Brain Research, 1367,* 278–286. doi:10.1016/j.brainres.2010.10.030

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11,* 379–386. doi:10.1016/j.tics.2007.08.001

Dobbins, I. G., & Han, S. (2006). Isolating rule- versus evidence-based prefrontal activity during episodic and lexical discrimination: a functional magnetic resonance imaging investigation of detection theory distinctions. *Cerebral Cortex, 16,* 1614–1622.

Düzel, E., Vargha-Khadem, F., Heinze, H. J., & Mishkin, M. (2001). Brain activity evidence for recognition without recollection after early hippocampal damage. *Proceedings of the National Academy of Sciences, 98,* 8101–8106.

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, 94,* 5973–5978.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30,* 123–152. doi:10.1146/annurev.neuro.30.051606.094328

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.

Evans, L. H., & Wilding, E. L. (2012). Recollection and familiarity make independent contributions to memory judgments. *Journal of Neuroscience, 32,* 7253–7257.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. doi:10.3758/BF03193146

Ferrari, V., Bradley, M. M., Codispoti, M., Karlsson, M., & Lang, P. J. (2013). Repetition and brain potentials when recognizing natural scenes: Task and emotion differences. *Social Cognitive and Affective Neuroscience, 8,* 847–854.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP "old/new" effects: Memory strength and decisional factor(s). *Neuropsychologia, 40,* 2288–2304.

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.

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: Specific or non-specific? *Neuropsychologia, 46,* 1211–1223.

Herron, J. E. (2007). Decomposition of the ERP late posterior negativity: Effects of retrieval and response fluency. *Psychophysiology, 44,* 233–244.

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review, 10,* 703–710.

James, W. (1890). *The principles of psychology.* New York, NY: Henry Holt.

Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *Quarterly Journal of Experimental Psychology, 65,* 37–41.

Johansson, M., & Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: Action monitoring and retrieval of attribute conjunctions. *Biological Psychology, 64,* 91–117.

Johansson, M., Stenberg, G., Lindgren, M., & Rosén, I. (2002). Memory for perceived and imagined pictures: An event-related potential study. *Neuropsychologia, 40,* 986–1002.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi:10.1037/0033-2909.114.1.3

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1409–1419. doi:10.1037/0278-7393.20.6.1409

Johnson, R. J., Kreiter, K., Russo, B., & Zhu, J. (1998). A spatio-temporal analysis of recognition-related event-related brain potentials. *International Journal of Psychophysiology, 29,* 83–104.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Langeslag, S. J. E., & van Strien, J. W. (2008). Age differences in the emotional modulation of ERP old/new effect. *International Journal of Psychophysiology, 70,* 105–114.

Leynes, P. A., Cairns, A., & Crawford, J. T. (2005). Event-related potentials indicate that reality monitoring differs from external source monitoring. *American Journal of Psychology, 118,* 497–524.

Leynes, P. A., & Phillips, M. C. (2008). Event-related potential (ERP) evidence for varied recollection during source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 741–751.

Leynes, P. A., & Zish, K. (2012). Event-related potential (ERP) evidence for fluency-based recognition memory. *Neuropsychologia, 50,* 3240–3249.

McBride, D. M., & Abney, D. H. (2012). A comparison of transfer-appropriate processing and multi-process frameworks for prospective memory performance. *Experimental Psychology, 59,* 190–198.

Mecklinger, A., Johansson, M., Parra, M., & Hanslmayr, S. (2007). Source-retrieval requirements influence late ERP and EEG memory effects. *Brain Research, 1172,* 110–123.

Mecklinger, A., von Cramon, D. Y., & Matthes-von Cramon, G. (1998). Event-related potential evidence for a specific recognition memory deficit in adult survivors of cerebral hypoxia. *Brain, 121,* 1919–1935. doi:10.1093/brain/121.10.1919

Montaldi, D., & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus, 20,* 1291–1314. doi:10.1002/hipo.20853

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. doi:10.1016/S0022-5371(77)80016-9

Nessler, D., Friedman, D., Johnson, R. J., & Bersick, M. (2007). Does repetition engender the same retrieval processes in young and older adults? *NeuroReport, 18,* 1837–1840.

Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia: Electrophysiological measures of impaired and spared memory. *Brain, 123,* 1948–1963. doi:10.1093/brain/123.9.1948

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science, 11,* 1–2.

Paller, K. A., Hutson, C. A., Miller, B. B., & Boehm, S. G. (2003). Neural manifestations of memory with and without awareness. *Neuron, 38,* 507–516.

Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Sciences, 11,* 243–250.

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. doi:10.1016/j.jml.2009.01.004

Rhodes, S. M., & Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *NeuroImage, 39,* 873–884. doi:10.1016/j.neuroimage.2007.08.041

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. doi:10.1016/j.tics.2010.09.003

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.

Rosenbaum, D., & Kornblum, S. (1982). A priming method for investigating the selection of motor responses. *Acta Psychologica, 1,* 223–243.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33,* 217–236. doi:10.1068/p5117

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

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392,* 595–598. doi:10.1038/33396

Schloerscheidt, A. M., & Rugg, M. D. (2004). The impact of change in stimulus format on the electrophysiological indices of recognition. *Neuropsychologia, 42,* 451–466.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. doi:10.1037/0278-7393.4.6.592

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215. doi:10.1037/0278-7393.6.2.174

Spitzer, B., Hanslmayr, S., Opitz, B., Mecklinger, A., & Bäuml, K.-H. (2009). Oscillatory correlates of retrieval-induced forgetting in recognition memory. *Journal of Cognitive Neuroscience, 21,* 976–990.

Sprondel, V., Kipp, K. H., & Mecklinger, A. (2012). Electrophysiological evidence for late maturation of strategic episodic retrieval processes. *Developmental Science, 15,* 330–344.

Stenberg, G., Hellman, J., Johansson, M., & Rosén, I. (2009). Familiarity or conceptual priming: Event-related potentials in name recognition. *Journal of Cognitive Neuroscience, 21,* 447–460. doi:10.1162/jocn.2009.21045

Stenberg, G., Johansson, M., & Rosén, I. (2006). Conceptual and perceptual memory: Retrieval orientations reflected in event-related potentials. *Acta Psychologica, 122,* 174–205.

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.

Vilberg, K. L., & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping, 30,* 1490–1501.

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.

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology, 35,* 81–87.

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.

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research, 1100,* 125–135.

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. doi:10.1006/jmla.2002.2864

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20,* 1178–1194. doi:10.1002/hipo.20864

Yu, S. S., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research, 1320,* 74–84.