Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 2, 287–297

0278-7393/10/$12.00 DOI: 10.1037/a0021801

Retrieval During Learning Facilitates Subsequent Memory Encoding

Bernhard Pastötter, Sabine Schicker, Julia Niedernhuber, and Karl-Heinz T. Bäuml
Regensburg University

In multiple-list learning, retrieval during learning has been suggested to improve recall of the single lists by enhancing list discrimination and, at test, reducing interference. Using electrophysiological, oscillatory measures of brain activity, we examined to what extent retrieval during learning facilitates list encoding. Subjects studied 5 lists of items in anticipation of a final cumulative recall test and did either a retrieval or a no-retrieval task between study of the lists. Retrieval was from episodic memory (recall of the previous list), semantic memory (generation of exemplars from an unrelated category), or short-term memory (2-back task). Behaviorally, all 3 forms of retrieval enhanced recall of both previously and subsequently studied lists. Physiologically, the results showed an increase of alpha power (8–14 Hz) from List 1 to List 5 encoding when no retrieval activities were interpolated but no such increase when any of the 3 retrieval activities occurred. Brain–behavior correlations showed that alpha-power dynamics from List 1 to List 5 encoding predicted subsequent recall performance. The results suggest that, without intermittent retrieval, encoding becomes ineffective across lists. In contrast, with intermittent retrieval, there is a reset of the encoding process for each single list that makes encoding of later lists as effective as encoding of early lists.

Keywords: learning, encoding, retrieval, testing, brain oscillations

Retrieval is a powerful tool for effective learning. For instance, successful retrieval of previously studied material is known to improve its long-term retention more than restudy does, a prominent finding that has been documented in numerous research articles and has been referred to as the testing effect (see e.g., Hogan & Kintsch, 1971; Karpicke & Roediger, 2008). Failed retrieval of previously studied material also promotes effective learning, because it provides feedback to focus additional study time on the not-yet-learned material in the future (see e.g., Kornell, Hays, & Bjork, 2009; Pashler, Zarow, & Triplett, 2003). Finally, and of central interest for the present study, retrieval of previously studied material can improve retention of subsequently studied material (see e.g., Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974). This beneficial effect of retrieval is particularly striking, because it is on learning of new material that is not related to the previously retrieved information.

Behavioral Findings and the Context-Change Hypothesis of Retrieval

The beneficial effect of retrieval of previously studied material on learning of subsequently studied material has been reported in several research articles (see e.g., Darley & Murdock, 1971; Szpunar et al., 2008; Tulving & Watkins, 1974). For instance, Szpunar et al. (2008) let their subjects study five lists of items (Lists 1–5) in anticipation of a final cumulative recall test. All subjects were tested immediately on List 5. One group of subjects was also tested immediately on Lists 1–4, whereas another group restudied Lists 1–4, and a third group did a mathematical distractor task after each single list. Results showed that subjects who were tested immediately on Lists 1–4 recalled more List 5 items and showed fewer prior-list intrusions than did subjects in the two no-testing groups. Szpunar et al. proposed that immediate testing of Lists 1–4, but not restudy of the single lists, permits subjects to create list-specific context cues during encoding that enhance list discrimination and, at test, reduce interference. Consistently, recent work has shown that, when subjects have been tested immediately during the study of item lists, list discrimination concerning both items remembered and items not remembered on the immediate recall tests is enhanced on a final recognition test (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Chan & McDermott, 2007).

According to the context-change hypothesis of retrieval, which has been incorporated in recent computational and multinomial models (Criss & Shiffrin, 2004; Howard & Kahana, 2002; Jang & Huber, 2008; Klein, Shiffrin, & Criss, 2007), retrieval cycles between the study of lists drive internal context changes that promote list segregation. The proposal is that, during list encoding, the memory system binds each list item to the current representation of the subject's internal context. Retrieval between the study of lists alters the internal context, which leads to specific context cues for each single list. Then, at test, list-specific context cues can be used, enhancing list discrimination and reducing interference between lists. Prior work has shown that beneficial effects of retrieval can arise from (episodic) retrieval of previously studied material, (semantic) retrieval of information unrelated to the stud-

---

Bernhard Pastötter, Sabine Schicker, Julia Niedernhuber, and Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University, Regensburg, Germany.

We thank H. Backes, A. Schlichting, and T. Schwarzmüller for their help with data collection.

Correspondence concerning this article should be addressed to Bernhard Pastötter or Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany. E-mail: bernhard.pastoetter@psychologie.uni-regensburg.de or karl-heinz.baeuml@psychologie.uni-regensburg.de

ied material, and retrieval from autobiographical memory (Delaney, Sahakyan, Kelley, & Zimmerman, 2010; Jang & Huber, 2008; Pastötter, Bäuml, & Hanslmayr, 2008; Sahakyan & Kelley, 2002; Shiffrin, 1970; Szpunar et al., 2008; Tulving & Watkins, 1974). These findings suggest that various forms of retrieval can induce internal context changes, enhancing list discrimination and improving retention of the studied material.

Standard views on the memorial effects of internal context changes emphasize the important role of retrieval processes at test to account for context-dependent memory effects (Criss & Shiffrin, 2004; Dennis & Humphreys, 2001; Howard & Kahana, 2002; Jang & Huber, 2008; Mensink & Raaijmakers, 1988). Accordingly, whereas detrimental effects of a context change supposedly arise from a mismatch between study context and test context, the beneficial effects are explained by an improved interference situation at test, caused through the creation of list-specific context cues during encoding. Results from recent electrophysiological studies strengthen but also extend this view by indicating that internal context changes may even cause a reset of the encoding process, making the encoding of subsequent lists as effective as the encoding of previous lists.

Electrophysiological Findings and the Context-Change Hypothesis of Retrieval

Although traditional electroencephalogram (EEG) measures of brain activity have focused on event-related potentials (ERPs) to study the neural correlates of cognitive processes, more recent research has demonstrated that analysis of oscillatory brain activity can yield new insights into the functional role of EEG signals (Başar, Başar-Eroglu, Karakas, & Schürmann, 1999). Indeed, due to their critical role in establishing synchronized firing between cortical cell assemblies, which promotes proper communication within neural networks (Fries, 2005), brain oscillations have been investigated intensively in human cognition and, in particular, human memory (for reviews, see Klimesch, 1999, or Nyhus & Curran, 2010). Typically, brain oscillations are divided into different frequency bands, for example, ranging from theta (4–7 Hz), alpha (8–14 Hz), and beta (15–25 Hz) to gamma (<30 Hz). Oscillatory activity in these frequency bands has been shown to be related to successful memory encoding of both single and multiple item lists (see e.g., Bäuml, Hanslmayr, Pastötter, & Klimesch, 2008; Hanslmayr, Spitzer, & Bäuml, 2009; Klimesch et al., 1996; Osipova et al., 2006; Pastötter et al., 2008; Sederberg et al., 2006; Sederberg, Kahana, Howard, Donner, & Madsen, 2003).

In single-list learning, Sederberg et al. (2006) examined oscillatory power as a function of studied items' serial list position. Their analyses showed that theta and alpha power increase from early to late serial list positions at widespread electrodes over the scalp. Because increases of both theta power and alpha power have been related to high memory load (for a review, see Jensen, 2006) and increases of alpha power have also been linked to inattention (for a review, see Palva & Palva, 2007), Sederberg et al. suggested that, with cumulative list length, increases of theta and alpha power reflect a breakdown of encoding processes due to increased memory load and inattention. Generalizing the item-specific results to list learning, Pastötter et al. (2008) recently showed that stimulus-induced theta and alpha power also increase from a first to a second to-be-studied list. Following Sederberg et al., it was argued that the study of a second list is accompanied by a shift from focused to less-attentive encoding when compared with the study of the first list.

Pastötter et al. (2008) varied subjects' internal context between the study of two lists by interpolating autobiographical memory retrieval. Either subjects were cued to change their internal context through a simple imagination task, in which they were asked to mentally walk through their parents' house and tell their imaginations to the experimenter (see e.g., Pastötter & Bäuml, 2007; Sahakyan & Kelley, 2002), or they were not given such cue. Consistent with the prior work, the behavioral results showed that the change in context led to impaired recall of the first list and to memory enhancement of the second. Going beyond the prior work, the context change was found to affect oscillatory brain activity during List 2 encoding. Whereas an increase of theta and alpha power from List 1 to List 2 encoding was found when the context was left unchanged, there was no such increase in oscillatory activity when the context was changed between lists. In particular, brain–behavior correlations revealed that stimulus-induced alpha power during List 2 encoding predicted the beneficial effect of context change on List 2 recall. These results indicate that an internal context change between the study of two lists leads to a reset of the encoding process for the second list. The reset likely reduces memory load and inattention during List 2 encoding that would build up when the encoding context was left unchanged. In particular, the reset makes encoding of items from the second list as effective as encoding of items from the first list.

The Present Experiment

Changing internal context between the study of lists affects alpha-power dynamics during item encoding, which predicts memory performance for material studied after the internal context change (Pastötter et al., 2008). This finding holds for autobiographical memory retrieval, as it occurs in the imagination task (see Sahakyan & Kelley, 2002). Because not only intermittent retrieval from autobiographical memory but also intermittent retrieval from episodic memory and semantic memory have been suggested to induce internal context changes (Jang & Huber, 2008; Szpunar et al., 2008), in this study we tested the hypothesis that multiple forms of retrieval during learning can affect subsequent memory encoding and later recall.

Following the experimental design of Szpunar et al. (2008), subjects studied five lists of 20 items (Lists 1–5), which they were asked to remember for a final cumulative recall test. All subjects were tested immediately for List 5. There were three retrieval groups and two no-retrieval groups. Regarding the two no-retrieval groups, one group of subjects restudied each of Lists 1–4 and one group did a backward-counting distractor task after each single list. Regarding the three retrieval groups, one group of subjects was tested immediately on Lists 1–4, one group did a semantic generation task between the study of lists, and one group did a 2-back short-term memory task after each single list. On the basis of the context-change hypothesis of retrieval, according to which retrieval activities enhance list segregation, we expected to find a List 5 recall benefit in the three retrieval groups compared with the two no-retrieval groups. If retrieval enhances list segregation regardless of whether retrieval is from episodic memory, semantic memory, or short-term memory, then the List 5 recall benefit should be observed in all three retrieval groups. Following Szpunar et al. (2008), we did not expect a difference in recall performance between the two no-retrieval groups.

In order to examine to what extent intermittent retrieval affects subsequent encoding, we recorded EEGs during study of the five lists and compared stimulus-induced power changes from List 1 to List 5 encoding between experimental conditions. Extending the findings from two-list learning (Pastötter et al., 2008), and on the basis of the context-change hypothesis of retrieval, we expected an increase in alpha power from List 1 to List 5 encoding in the two no-retrieval groups, in which lists cannot be segregated by interlist retrieval. In contrast, no such increase of oscillatory brain activity from List 1 to List 5 encoding should arise in the three retrieval groups, at least if all forms of intermittent retrieval induced list segregation and caused a reset of encoding processes for each single list. A comparison of electrophysiological brain activity between conditions should bring results that indicate (a) whether retrieval during learning makes encoding of later lists as effective as encoding of early lists (i.e., induces a reset of encoding processes) and (b) whether all forms of intermittent retrieval promote such beneficial encoding.

Method

Subjects

Ninety students (59 women and 31 men) at Regensburg University in Germany participated in the study. Their mean age was 23.0 years, with a range of 20 to 33 years ($SD = 2.1$). No subject reported any history of neurological disease. All subjects gave written informed consent, reported normal or corrected-to-normal vision, spoke German as native language, and received 15 euros (US$20) for participation.

Materials

One hundred unrelated German nouns of medium frequency were drawn from the CELEX database using the Wordgen v1.0 software toolbox (Duyck, Desmet, Verbeke, & Brysbaert, 2004). For each subject, five lists of 20 items (Lists 1–5) were prepared. The assignment of items to lists was random for all subjects.

Design

The experiment had a 5 × 5 design with the within-subjects factor of list (List 1 to List 5) and the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back). In all experimental groups, subjects encoded five lists of items. Experimental groups differed in interlist activity. In the distractor condition, subjects counted backward in steps of three after Lists 1–4; in the restudy condition, subjects restudied Lists 1–4; in the testing condition, subjects were tested immediately for Lists 1–4; in the generation condition, subjects generated exemplars from a semantic category after Lists 1–4; and in the 2-back condition, subjects did a 2-back task after the encoding of Lists 1–4. All subjects were tested immediately for List 5. After a distractor, all subjects were asked to recall in any order they wished all items of all previously studied lists. EEGs were recorded during the encoding of Lists 1–5.

Procedure

All experimental conditions consisted of a learning phase, a distractor phase, and a final test phase. Prior to the learning phase, subjects were informed about the general nature of the experiment. We told all subjects that several item lists would have to be encoded, and we encouraged them to pay close attention to the presented words for an upcoming final free recall test, in which all of the previously learned items would be tested. They were also told to expect various activities that may follow the presentation of each single list: backward counting in steps of three; restudy of words from a list that had just been studied in a new random order; an immediate free-recall test of words that had just been studied, semantic generation of exemplars from a semantic category unrelated to the words that had just been studied; and a 2-back task in which numbers from 0 to 9 should be matched. All subjects were acquainted with backward counting and the 2-back task. We pretended that activities following each list were determined randomly. In fact, interlist activities differed between conditions, and subjects passed through the same activities after the encoding of Lists 1–4 within each experimental condition (for similar procedure, see Szpunar et al., 2008).

In the learning phase, the items of the five lists were visually presented in the center of a computer screen. Following a prestimulus interval of variable duration (800–1,200 ms), in which a fixation cross was shown, the 20 words of each list were exposed individually for 2,000 ms. Subsequent to the presentation of each list (approximately 1 min), all subjects counted backward from a random three-digit number for 30 s. Experimental conditions differed in interlist activity that followed this backward counting after each single list. In the distractor condition, backward counting was prolonged for 1 min. In the restudy condition, a list's words were reexposed in a new random order (approximately 1 min). In the testing condition, subjects were given 1 min to write down in any order they wished as many words as possible from the list they had just studied; different sheets of paper were used for the different lists. In the generation condition, subjects were given 1 min to generate as many words as possible from one of four semantic categories (vegetables, sports, four-legged animals, professions). The assignment of the categories to interlist activities following Lists 1–4 encoding was random; none of the to-be-learned items belonged to one of the four semantic categories. In the 2-back task, subjects were asked to decide whether stimuli in a sequence of numbers matched the ones that had appeared two trials ago (see Owen, McMillan, Laird, & Bullmore, 2005). Pseudorandom sequences of 60 numbers from 0 to 9 were presented in the center of the screen at a 1-s interval, and subjects were instructed to respond to each target stimulus by pressing the space bar. No response should be made to nontargets (i.e., stimuli that were different from the stimulus presented two trials ago). The number of targets (20) and nontargets (40) was fixed.

Between the learning phase and the final test phase, subjects did a mathematical distractor for 5 min in which they added pairs of three-digit numbers as fast and correctly as possible. In the final-test phase, subjects were given 7 min to recall in any order they wished as many words as possible from all five lists of words they had studied. They wrote down the words on a sheet of paper. It was emphasized to subjects that they should use the 7 min efficiently in their attempt to recall study materials. The experiment was completed in approximately 45 min by all subjects, at which point they were thanked for their participation, fully debriefed, and paid.

Recordings of EEG Data

During list encoding, EEGs were recorded from 61 equidistant active electrodes mounted in an elastic cap (ActiCAP, Montage 10, Brain Products, Gilching, Germany). ActiCAP, with its active electrode system, enables fast electrode placement and low electrode–skin impedances due to amplification circuitry built into the electrodes boosting the signal and reducing the noise. Electrode–skin impedance was kept below 20 kΩ. Vertical and horizontal eye movements were recorded from two additional channels. Electrode Cz served as the common reference. Signals were digitalized with a sampling rate of 500 Hz and amplified between 0.15 and 100 Hz with a notch filter at 50 Hz to remove line noise (BrainAmpMR plus, BrainVision Recorder, Brain Products, Gilching, Germany).

Processing of EEG Data

EEG recordings were offline-rereferenced against average reference and were EOG-corrected using calibration data to generate individual artifact coefficients (Ille, Berg, & Scherg, 2002) as implemented in the software package BESA (Brain Electrical Source Analysis, MEGIS Software v5.1.8, Gräfelfing, Germany). Remaining artifacts were marked by careful visual inspection.

Power spectral density was computed for each item presentation and was then averaged across artifact-free trials for each subject in each item list and interlist-activity condition. As implemented in BESA (see Hoechstetter et al., 2004), each subject's EEG was transformed into the time–frequency domain using a complex demodulation algorithm (Papp & Ktonas, 1977). This algorithm consists of a multiplication of the time-domain signal with a complex periodic exponential function, with a frequency equal to the frequency under analysis, and a subsequent low-pass filter. The low-pass filter is a finite impulse-response filter of Gaussian shape in the time domain, which is related to the envelope of the moving window in wavelet analysis. The data were filtered in a frequency range from 2 to 30 Hz. Time resolution was set to 78.8 ms (full width at half maximum, or FWHM), and frequency resolution was set to 1.42 Hz (FWHM). Time–frequency data were exported in bins of 50 ms and 1 Hz.

We examined stimulus-induced power changes by calculating the percentage of power decrease or power increase in relation to a prestimulus baseline interval (Pfurtscheller & Aranibar, 1977), which was set from 750 ms to 250 ms prior to stimulus onset. A priori, data were analyzed with and without subtracting the evoked signal from single trials prior to time–frequency transformation. As it turned out, the conclusions were the same for the two analysis methods. The present results are based on analyses without subtraction. For statistical analysis, power data were collapsed in order to obtain three frequency bands (theta: 4–7 Hz, alpha: 8–14 Hz, beta: 15–25 Hz) and eight time windows of 250 ms each from word onset (0 s) to word offset (2 s). Actually, we also analyzed stimulus-induced power in the delta (2–4 Hz), lower gamma (30–45 Hz), and upper gamma (60–90 Hz) frequency bands. However, for these bands, we did not find any effects of multiple-list encoding or interlist activity. Therefore, in the Results section, we do not report on the detailed findings from these analyses.

Statistical Analysis of EEG Data

First, to examine electrophysiological effects of multiple-list encoding, we compared stimulus-induced power changes between List 1 and List 5 encoding for each frequency band (3), time window (8), and electrode site (61), irrespective of interlist activity. To account for multiple testing, a two-stage randomization procedure was carried out. At first, Wilcoxon sign-rank tests were calculated for each electrode in order to investigate which electrodes differ between conditions ($p < .05$). Thereafter, a randomization test using 2,000 permutations was run in which we shuffled the order of lists consistently across electrodes (see Blair & Karniski, 1993). Similar procedures had already been applied in several other studies from our lab (see e.g., Bäuml et al., 2008; Hanslmayr et al., 2009; Pastötter et al., 2008). The procedure evaluates whether a given number of electrodes, exhibiting a significant difference between List 1 and List 5 encoding, is expected by chance. If the p value (p_{corr}) of this randomization test was below .05, then less than 5% of the permutation runs exhibited equal or more electrode sites with a significant difference between conditions. If we found an effect to persist for two (or more) consecutive time windows (e.g., from 250 ms to 500 ms and from 500 ms to 750 ms), then we ran a new randomization test for the expanded time window of interest (e.g., from 250 ms to 750 ms). Across significant electrode sites, we then compared the timing of power changes between List 1 and List 5 encoding.

Only if the randomization test exhibited a significant difference between conditions was power data averaged across all significant electrodes for the specific frequency band and (expanded) time window and entered into analyses of variance (ANOVAs) with the within-subjects factor of list (List 1 to List 5) and the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back). Greenhouse–Geisser correction was applied in all ANOVAs. Finally, we examined the relationship between power effects and behavior by calculating correlational analyses between power dynamics from List 1 to List 5 encoding and immediate List 5 recall performance.

Results

Behavioral Results

Immediate recall. Immediate List 5 recall rates are shown on the left side of Figure 1A and prior-list intrusions during immediate List 5 recall are shown in Figure 1B. ANOVAs with the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back) revealed significant main effects for both correct List 5 recall, $F(4, 85) = 3.8$, $MSE = .039$, $p < .01$, $\eta_p^2 = .15$, and Lists 1–4 intrusions, $F(4, 85) = 3.0$, $MSE = .625$, $p < .05$, $\eta_p^2 = .12$. Pairwise comparisons showed that subjects in the three retrieval conditions (testing, generation, 2-back) correctly recalled more List 5 items than did subjects in the two no-retrieval conditions (distractor, restudy), all ts$(34) > 2.2$, all ps $< .05$, all ds $> .88$; all other differences were nonsignificant, all ts$(34) < 1$. With regard to prior-list intrusion, pairwise comparisons revealed that subjects in the testing condition and the generation condition

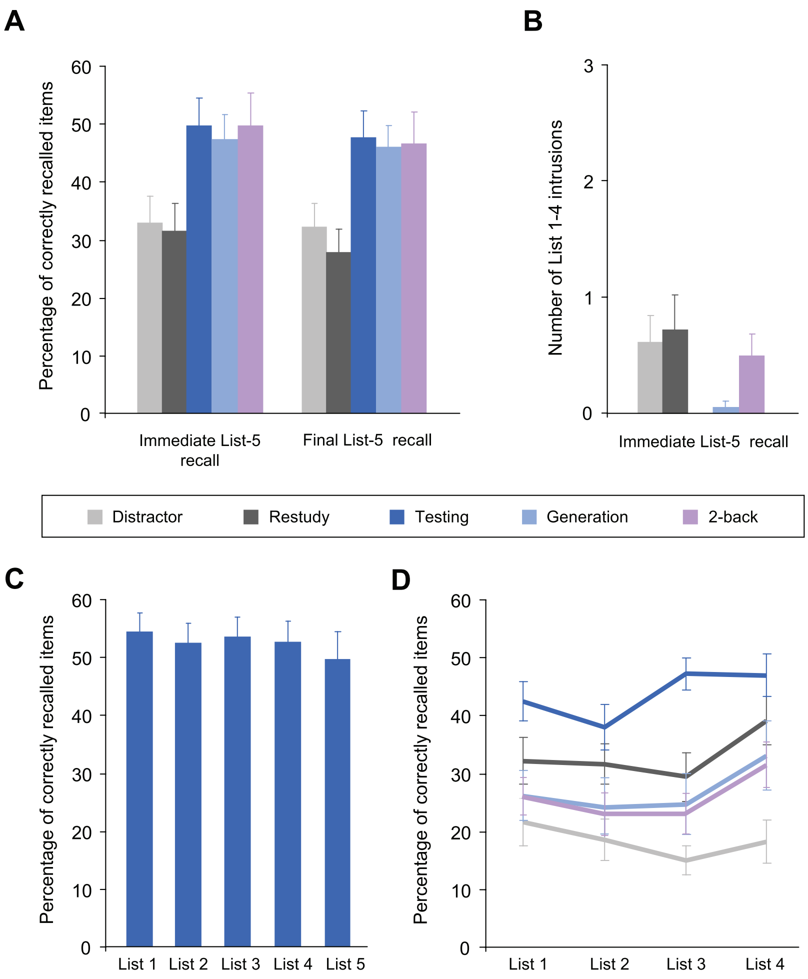

Figure 1. Behavioral results. A: List 5 recall rates as a function of interlist activity (distractor, restudy, testing, generation, 2-back) in the immediate List 5 recall test and the final recall test. B: Number of prior-list intrusions as a function of interlist activity in the immediate List 5 recall test. C: Immediate recall rates for each single list in the testing condition. D: List 1–4 final recall rates as a function of interlist activity. All error bars indicate standard errors.

produced fewer intrusions than did subjects in the distractor condition, the restudy condition, and the 2-back condition, all *t*s(34) > 2.1, all *p*s < .05, all *d*s > .72; all other differences were nonsignificant, all *t*s(34) < 1; conclusions were the same for nonparametric testing.

Subjects in the testing condition remembered approximately half of a list's items in immediate recall of Lists 1–5 (see Figure 1C) and produced hardly any prior-list intrusions in recall of Lists 2–5 (0.07 intrusions per list). Both immediate-recall rates and prior-list intrusions did not differ between lists (*F*s < 1). Thus, in the testing condition, previous study of other lists did not influence retention of subsequently studied material.

Final recall. Final List 5 recall rates are shown on the right side of Figure 1A. They mimic immediate-recall results for this list. Thus, one-way ANOVA revealed a significant effect of interlist activity on List 5 recall, $F(4, 85) = 4.5$, $MSE = .036$, $p < .005$, $\eta_p^2 = .17$, and pairwise comparisons showed that subjects in the three retrieval conditions (testing, generation, 2-back) recalled more List 5 items than did subjects in the two no-retrieval conditions (distractor, restudy), all *t*s(34) > 2.1, all *p*s < .05, all *d*s > .70; all other differences were nonsignificant, all *t*s(34) < 1.

Final Lists 1–4 recall performance is shown in Figure 1D. A two-way ANOVA with the within-subjects factor of list (Lists 1–4) and the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back) revealed main effects of list, $F(3, 255) = 3.8$, $MSE = .020$, $p < .01$, $\eta_p^2 = .04$, and interlist activity, $F(4, 85) = 11.4$, $MSE = .057$, $p < .001$, $\eta_p^2 = .35$, but no interaction between the two factors, $F(12, 255) < 1$. Subjects recalled more List 4 items than List 3 and List 2 items, both *t*s(89) > 2.6, both *p*s < .01, both *d*s > .30; all other differences between lists were nonsignificant, all *t*s(89) < 1.6. In particular, post hoc comparisons of mean Lists 1–4 recall between groups revealed that subjects in the testing condition recalled more Lists 1–4 items than did subjects in any other condition, all *t*s(34) > 2.2, all *p*s < .05, all *d*s > .74. In contrast, subjects in the distractor condition recalled fewer Lists 1–4 items than did subjects in any other condition, all *t*s(34) > 2.2, all *p*s < .05, all *d*s > .66. In the restudy condition, Lists 1–4 recall tended to be higher than in the generation condition, $t(34) = 1.5$, $p = .14$, $d = .49$, and the 2-back condition, $t(34) = 1.6$, $p = .11$, $d = .53$, although the differences were nonsignificant. There was no difference in Lists 1–4 recall between the generation condition and the 2-back condition, $t(34) < 1$. Thus, in conditions in which Lists 1–4 items were reprocessed after study (testing, restudy), recall performance for Lists 1–4 was best. In the two retrieval conditions in which no reprocessing of List 1–4 took place (generation, 2-back), performance was better than in the distractor condition.

Physiological Results

Alpha power. In the analysis of electrophysiological data, the first step was to compare stimulus-induced alpha power (8–14 Hz) between List 1 and List 5 encoding, irrespective of interlist activity. Randomization tests, which evaluate whether alpha-power differences between List 1 and List 5 encoding are expected by chance (see Method section), revealed a significant increase in alpha power from List 1 to List 5 encoding at widespread electrodes over the scalp (see Figure 2A), beginning at stimulus onset and lasting for 1,000 ms ($p_{\text{corr}} < .005$; see Figure 2B).

On the basis of those electrodes that exhibited significant differences between List 1 and List 5 encoding in the randomization test, alpha power from 0 ms to 1,000 ms entered into a two-way ANOVA with the within-subjects factor of list (List 1 to List 5) and the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back; see Figure 2C). This analysis

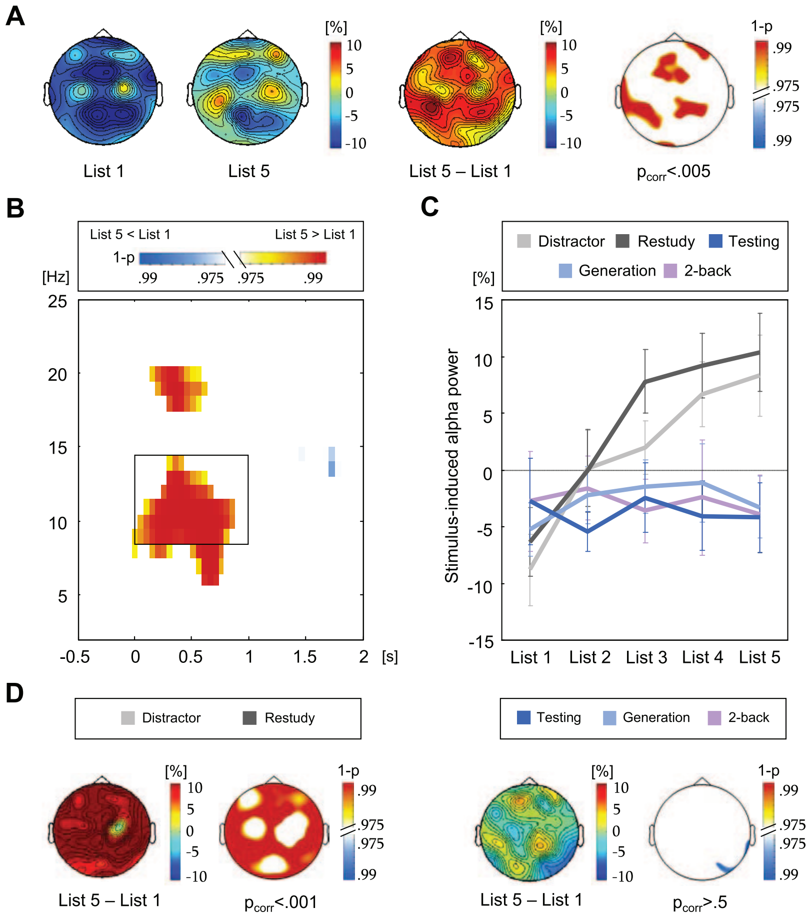

Figure 2. Alpha-power results (8–14 Hz). A: Topographical maps of stimulus-induced alpha power for List 1 and List 5 encoding from 0 ms to 1,000 ms following stimulus onset, irrespective of interlist activity (distractor, restudy, testing, generation, 2-back), with cold colors indicating a stimulus-induced decrease of power and warm colors indicating an increase of power. The difference between List 1 and List 5 encoding is expressed by topographical plotting of the p levels obtained by nonparametric Wilcoxon sign-rank tests. A randomization test (see Method section) revealed that differences in alpha power were reliable ($p_{\text{corr}} < .005$). B: Time–frequency spectrogram shows the timing of significant power differences between List 1 and List 5 encoding averaged across electrodes that exhibited a significant difference in the alpha-power randomization test, irrespective of interlist activity. C: Alpha-power dynamics at corresponding electrodes from List 1 to List 5 encoding as a factor of interlist activity. All error bars indicate standard errors. D: Difference maps of stimulus-induced alpha power between List 1 and List 5 encoding from 0 ms to 1,000 ms following stimulus onset, with randomization tests for alpha data pooled across subjects in the two no-retrieval conditions (distractor, restudy) and pooled across subjects in the three retrieval conditions (testing, generation, 2-back).

revealed main effects of list, $F(4, 340) = 4.9$, $MSE = .015$, $p < .001$, $\eta_p^2 = .06$, and interlist activity, $F(4, 85) = 3.1$, $MSE = .035$, $p < .05$, $\eta_p^2 = .13$, and, crucially, a significant List × Interlist Activity interaction, $F(16, 340) = 1.8$, $MSE = .015$, $p < .05$, $\eta_p^2 = .08$. Post hoc analyses showed that the interaction arose from an increase of alpha power from List 1 to List 5 encoding in the two no-retrieval conditions, that is, the distractor condition, $F(2.8, 47.9) = 8.0$, $MSE = .010$, $p < .001$, $\eta_p^2 = .32$, and the restudy condition, $F(3.3, 55.9) = 4.5$, $MSE = .021$, $p < .005$, $\eta_p^2 = .21$, whereas alpha power did not differ between lists in the three retrieval conditions, that is, the testing condition, the generation condition, and the 2-back condition (all $F\text{s} < 1$). This held while alpha power for List 1 encoding did not differ across conditions, $F(4, 85) < 1$. Consistently, alpha power pooled across subjects in the two no-retrieval conditions (distractor, restudy) increased from List 1 to List 5 encoding over the whole scalp ($p_{corr} < .001$; see Figure 2D), whereas no such increase was found for alpha power pooled across subjects in the three retrieval conditions (testing, generation, 2-back; $p_{corr} > .5$; see Figure 2D).

Analysis of group data already indicates that the difference in alpha activity between Lists 1 and 5 is related to List 5 recall performance (see Figures 1A and 2C). In fact, groups showing an increase in alpha activity across lists show low List 5 recall levels, and groups showing no such increase show high List 5 recall levels. Individual difference analysis complements this finding. Correlational analysis on all subjects showed a negative relationship between immediate List 5 recall and the increase of alpha power from List 1 to List 5 ($r = -.52$, $p < .001$; see Figure 3A). This negative relationship did not differ significantly between experimental conditions, $\chi^2(4) < 1$. Consistently, single-group analyses largely confirmed the picture by showing that in four of the five conditions the correlation was (marginally) significant (distractor: $r = -.47$, $p = .05$; restudy: $r = -.49$, $p < .05$; testing: $r = -.43$, $p = .07$; generation: $r = -.25$, $p = .32$; 2-back: $r = -.47$, $p = .05$). Thus, alpha-power dynamics from List 1 to List 5 encoding depended on interlist activity and predicted List 5 recall performance.

Theta and beta power. We also compared stimulus-induced theta power (4–7 Hz) and beta power (15–25 Hz) between List 1 and List 5 encoding, irrespective of interlist activity. Randomization tests revealed a significant increase of beta power from List 1 to List 5 encoding, mostly at central electrode sites (see Figure 4A) from 250 ms to 750 ms following stimulus onset ($p_{corr} < .01$; see Figure 4B). In contrast, there was no significant difference in theta power in any time window, neither when analyzing data of all subjects nor when analyzing data sets pooled across subjects in the two no-retrieval conditions (distractor, restudy) or the three retrieval conditions (testing, generation, 2-back) separately (all values of $p_{corr} > .05$).

Beta power from 250 ms to 750 ms averaged across those electrodes that exhibited significant differences in the randomization test entered into a two-way ANOVA with the within-subjects factor of list (List 1 to List 5) and the between-subjects factor of interlist activity (distractor, restudy, testing, generation, 2-back). This analysis revealed a main effect of list, $F(8.6, 320.1) = 8.6$, $MSE = .008$, $p < .001$, $\eta_p^2 = .09$, but neither a main effect of interlist activity nor a List × Interlist Activity interaction (both $F\text{s} < 1$). Thus, beta power increased from List 1 to List 5 encoding irrespective of interlist activity (see Figure 4C). Consistently, beta power also increased from List 1 to List 5 encoding when pooled across subjects in the two no-retrieval conditions (distractor, restudy; $p_{corr} < .001$) and when pooled across subjects in the three

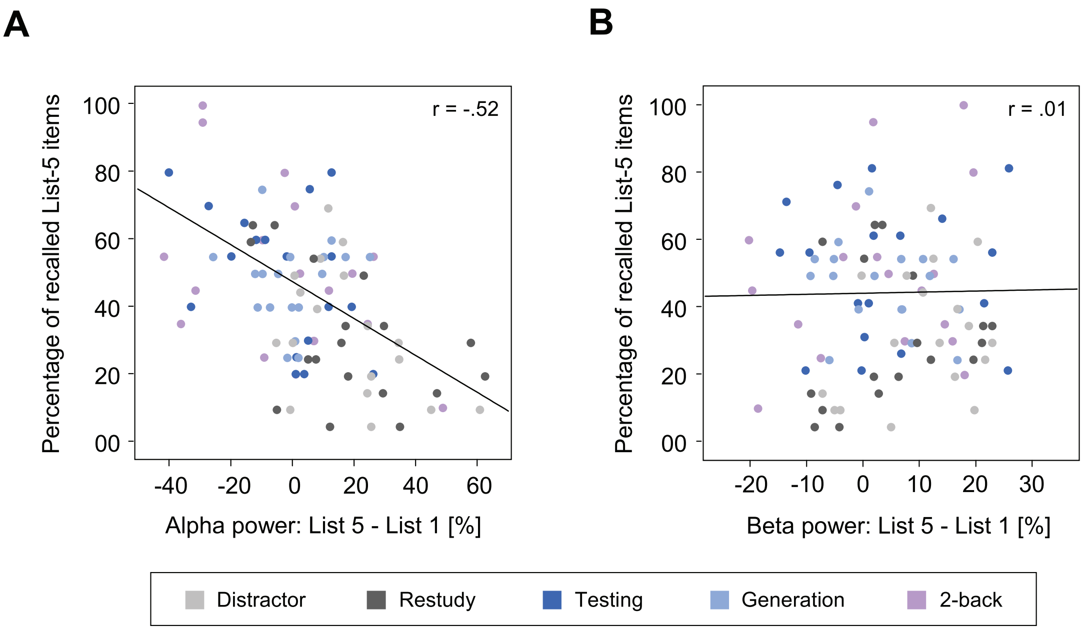

Figure 3. Brain–behavior correlations. A: Scatter plot showing a negative relationship between immediate List 5 recall and the increase of alpha power from List 1 to List 5 encoding ($r = -.52$). B: No brain–behavior relationship was found between immediate List 5 recall and the increase of beta power from List 1 to List 5 encoding ($r = .01$).

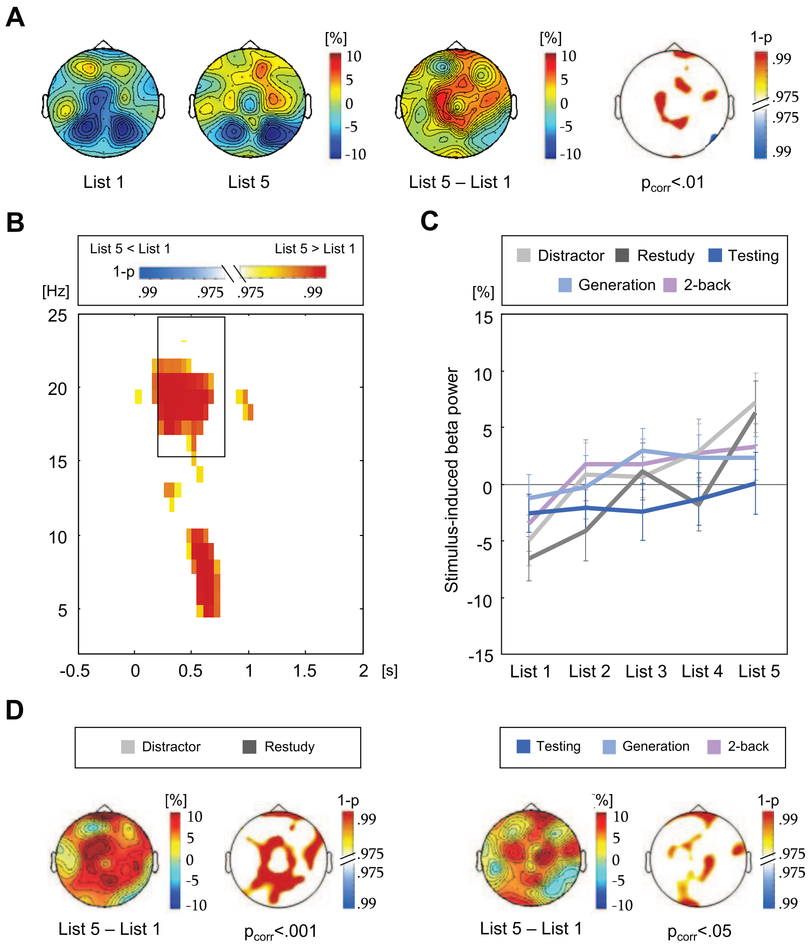

Figure 4. Beta-power results (15–25 Hz). A: Topographical maps of stimulus-induced beta power for List 1 and List 5 encoding from 250 ms to 750 ms following stimulus onset, irrespective of interlist activity (distractor, restudy, testing, generation, 2-back), with cold colors indicating a stimulus-induced decrease of power and warm colors indicating an increase of power. The difference between List 1 and List 5 encoding is expressed by topographical plotting of the *p* levels obtained by nonparametric Wilcoxon sign-rank tests. A randomization test (see Method section) revealed that differences in beta power were reliable ($p_{\text{corr}} < .01$). B: Time–frequency spectrogram shows the timing of significant power differences between List 1 and List 5 encoding averaged across electrodes that exhibited a significant difference in the beta-power randomization test, irrespective of interlist activity. C: Beta-power dynamics at corresponding electrodes from List 1 to List 5 encoding as a factor of interlist activity. All error bars indicate standard errors. D: Difference maps of stimulus-induced beta power between List 1 and List 5 encoding from 250 ms to 750 ms following stimulus onset, with randomization tests for beta data pooled across subjects in the two no-retrieval conditions (distractor, restudy) and pooled across subjects in the three retrieval conditions (testing, generation, 2-back).

retrieval conditions (testing, generation, 2-back; $p_{\text{corr}} < .05$; see Figure 4D).

Analysis of group data suggests that the difference in beta activity between Lists 1 and 5 is not related to List 5 recall performance (see Figures 1A and 4C). Although recall levels depended on subjects' interlist activity, the increase of beta power from List 1 to List 5 encoding did not. Consistently, correlational analysis showed no relationship between immediate List 5 recall and the increase of beta power from List 1 to List 5 ($r = .01$, $p = .90$; see Figure 3B). Thus, beta-power dynamics from List 1 to List 5 encoding did not depend on interlist activity and did not predict List 5 recall performance.

Discussion

Behavioral Results

We showed that, consistent with prior work, in multiple-list learning, testing during learning enhances recall of subsequently studied lists, whereas restudy of the same material does not (Szpunar et al., 2008; Tulving & Watkins, 1974). We extend prior work by showing that semantic retrieval and short-term memory retrieval mimic the effects of testing on subsequently studied lists and show equivalent beneficial effects on recall performance. The behavioral results also agree with prior work by showing that testing during learning improves recall of previously studied lists and improves recall to a larger extent than does restudy of the lists (Karpicke & Roediger, 2008; Szpunar et al., 2008). Again, semantic retrieval and short-term memory retrieval mimic the effects of testing, although the natural absence of reprocessing of previously studied items in semantic and short-term memory retrieval reduces the beneficial effect for this material. In contrast, the natural presence of reprocessing of previously studied items in the restudy condition creates some beneficial effect for this material.

The results are consistent with the view that retrieval—be it from episodic, semantic, or short-term memory—enhances list discrimination and thus improves recall of the single lists (Szpunar et al., 2008). In particular, the results agree with the context-change hypothesis of retrieval (Criss & Shiffrin, 2004; Howard & Kahana, 2002; Jang & Huber, 2008; Klein et al., 2007), according to which retrieval during learning triggers internal context changes between the single lists. The changes in internal context supposedly lead to specific context cues for each single list, enhancing list discrimination and, at test, reducing interference. Retrieval, therefore, should improve recall performance for both previously and subsequently studied lists. This is exactly what the present results show.

Physiological Results

The EEG results demonstrate that retrieval during learning facilitates encoding of the subsequently studied lists. We found that alpha power increased across lists when there was no retrieval between lists, whereas alpha power was unaffected when retrieval activities between lists occurred. This finding held regardless of whether retrieval was from episodic, semantic, or short-term memory. On both a group level and an individual level, the increase in alpha power predicted later recall of List 5: A high amount of increase led to poor recall, and a low amount of increase to good recall. In previous work, high levels of alpha power have been attributed to inattentional encoding of items and lists (Pastötter et al., 2008; Sederberg et al., 2006). On the basis of this view, the present results suggest that retrieval during learning abolishes inattentional encoding that would build up when no retrieval during learning occurs. Retrieval thus seems to induce a reset of encoding for each single list, making the encoding of later lists as effective as the encoding of early lists.

The results fit well with previous work in which it was shown that changes in internal context, as induced by autobiographical memory retrieval, influence encoding of the subsequently studied material. Using a two-list paradigm, Pastötter et al. (2008) showed that alpha power increased from List 1 to List 2 encoding when the internal context was left unchanged, whereas no such increase was observed when the internal context was changed between lists. Standard views on context-dependent memory, including computational models (Criss & Shiffrin, 2004; Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009), typically attribute retrieval effects on subsequently studied material to context changes and improved list discrimination, effectively caused by the generation of new context cues during encoding. Although the present behavioral results are consistent with this view, the present EEG results extend this view by demonstrating that internal context changes even cause a reset of encoding processes, making the encoding of later lists as effective as the encoding of early lists.

In the present study, retrieval between lists affected alpha activity, which was accompanied by beneficial effects on later recall. In contrast, retrieval between lists did not affect beta activity, which was also not accompanied by effects on recall. This held although beta power increased across lists. The localization of the beta effect at central electrode sites may hint at an involvement of the motor cortex, for which beta dynamics have been shown to be related to movement execution in general and to silent reading and vocalization in particular (for a review, see Salmelin, 2007). The increase of stimulus-induced beta power from List 1 to List 5 encoding thus may reflect a strategic shift from rote rehearsal to nonvocalization forms of encoding that some subjects may have performed independent of interlist activity. In contrast to previous two-list learning (Pastötter et al., 2008), in the present five-list paradigm we did not observe an increase of theta power in the no-retrieval conditions. Thus, in multiple-list learning, alpha seems to be the dominant frequency in which oscillatory activity is linked to beneficial encoding after an internal context change.

Short-Term Memory Retrieval Versus Long-Term Memory Retrieval

The present results for short-term memory retrieval mimic those from episodic retrieval (testing) and semantic generation in many respects. As a first parallel, the three forms of memory retrieval show equivalent behavioral effects on recall of subsequently encoded items. As a second parallel, the three forms of memory retrieval show equivalent physiological effects during encoding of the single lists. Finally, as a third parallel, short-term and semantic memory retrieval show the same effects on previously studied items and differ from episodic memory retrieval only because testing includes the reprocessing of previously studied material and thus leads to enhanced recall levels for this material.

Concluding from all this that retrieval from short-term memory is perfectly equivalent in effect to (episodic and semantic) retrieval from long-term memory might be premature, however. Indeed, despite the many similarities between the three forms of memory retrieval, one difference arose between short-term and long-term memory retrieval: Whereas, relative to the two no-retrieval conditions, both episodic and semantic memory retrieval effectively eliminated List 1–4 intrusions during immediate List 5 recall, short-term memory retrieval led to intrusion rates that were equivalent in amount to the no-retrieval conditions. This difference in intrusion errors may suggest that the effects of short-term and long-term memory retrieval are not perfectly identical.

Using the “list before the last” free-recall paradigm (Shiffrin, 1970), Jang and Huber (2008) drew a similar conclusion. They showed that, without interlist activity, increasing list length of an intervening List 2 reduced recall of a previously studied List 1. Crucially, testing and semantic generation, but not retrieval from short-term memory (2-back task), between study of the two lists abolished this negative effect, suggesting that retrieval from long-term memory, but not retrieval from short-term memory, drives internal context change and enhances list discrimination. The present results differ from this previous finding by showing beneficial effects of both short-term and long-term memory retrieval on recall of previously studied lists, and they extend prior work by showing beneficial effects of both short-term and long-term memory retrieval on encoding and recall of subsequently studied lists.

Still, because of the present difference in intrusion rates between short-term and long-term memory retrieval, the common suggestion arises that the effects of short-term memory retrieval may differ in detail from those of long-term memory retrieval. Discovering exactly how the effects of short-term memory retrieval differ from those of long-term memory retrieval is a high priority for future work.

Conclusions

Consistent with prior work, this study shows that retrieval during learning enhances list discrimination and improves recall of both previously and subsequently studied lists. Going beyond prior work, this study shows that retrieval during learning facilitates subsequent encoding, making the encoding of later lists as effective as the encoding of early lists. All of these results hold regardless of whether retrieval is from episodic, semantic, or short-term memory. The findings suggest that retrieval activities during learning induce a change in internal context, which creates a new context cue and induces a reset of the encoding process.

References

Başar, E., Başar-Eroglu, C., Karakas, S., & Schürmann, M. (1999). Oscillatory brain theory: A new trend in neuroscience. *IEEE Engineering in Medicine and Biology, 18,* 56–66.

Bäuml, K.-H., Hanslmayr, S., Pastötter, B., & Klimesch, W. (2008). Oscillatory correlates of intentional updating in episodic memory. *NeuroImage, 41,* 596–604.

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology, 30,* 518–524.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory, 18,* 385–393.

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 431–437.

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review, 111,* 800–807.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 91,* 66–73.

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. A. (2010). Remembering to forget: The amnesic effect of daydreaming. *Psychological Science, 21,* 1036–1042.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). Wordgen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments and Computers, 36,* 488–499.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9,* 474–480.

Hanslmayr, S., Spitzer, B., & Bäuml, K.-H. (2009). Brain oscillations dissociate between semantic and non-semantic encoding of episodic memories. *Cerebral Cortex, 19,* 1631–1640.

Hoechstetter, K., Bornfleth, H., Weckesser, D., Ille, N., Berg, P., & Scherg, M. (2004). BESA source coherence: A new method to study cortical oscillatory coupling. *Brain Topography, 16,* 233–238.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology, 19,* 113–124.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127.

Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. *Neuroscience, 139,* 237–249.

Karpicke, J. D., Roediger, H. L., III. (2008, February 15). The critical importance of retrieval for learning. *Science, 319,* 966–968.

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 171–189). New York, NY: Psychology Press.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29,* 169–195.

Klimesch, W., Schimke, H., Doppelmayr, M., Ripper, B., Schwaiger, J., & Pfurtscheller, G. (1996). Event-related desynchronization (ERD) and the DM effect: Does alpha desynchronization during encoding predict later recall performance? *International Journal of Psychophysiology, 24,* 47–60.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews, 34,* 1023–1035.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience, 26,* 7523–7531.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25,* 46–59.

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. *Trends in Neurosciences, 30,* 150–158.

Papp, N., & Ktonas, P. (1977). Critical evaluation of complex demodulation techniques for the quantification of bioelectrical activity. *Biomedical Sciences Instrumentation, 13,* 135–145.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1051–1057.

Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 977–982.

Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change: Evidence for a reset of encoding processes. *NeuroImage, 43,* 173–181.

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology, 42,* 817–826.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072.

Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach. *Clinical Neurophysiology, 118,* 237–254.

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage, 32,* 1422–1431.

Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience, 23,* 10809–10814.

Shiffrin, R. M. (1970, June 26). Forgetting, trace erosion or retrieval failure? *Science, 168,* 1601–1603.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399.

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on recall of another. *Journal of Verbal Learning and Verbal Behavior, 13,* 181–193.

Received June 18, 2010
Revision received September 9, 2010
Accepted September 11, 2010 ■

Call for Papers: Special Section on *Theory and Data in Categorization: Integrating Computational, Behavioral, and Cognitive Neuroscience Approaches*

The Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP:LMC) invites manuscripts for a special section on approaches to categorization, to be compiled by guest editors Stephan Lewandowsky and Thomas Palmeri working together with journal Associate Editor Michael Waldmann.

The goal of the special section is to showcase high-quality research that brings together behavioral, computational, mathematical, neuropsychological, and neuroimaging approaches to understanding the processes underlying category learning. There has been some divergence between approaches recently, with computational-mathematical models emphasizing the unity of category-learning processes while neuropsychological models emphasize the distinction between multiple underlying memory systems. We are seeking articles that integrate cognitive neuroscience findings in designing models or interpreting results, and behavioral studies and modeling results that constrain neuroscientific theories of categorization. In addition to empirical papers, focused review articles that highlight the significance of cognitive neuroscience approaches to cognitive theory—and/or the importance of behavioral data and computational models on constraining neuroscience approaches—are also appropriate.

The submission deadline is **June 1st, 2011**. The main text of each manuscript, exclusive of figures, tables, references, or appendixes, should not exceed 35 double-spaced pages (approximately 7,500 words). Initial inquiries regarding the special section may be sent to Stephan Lewandowsky (stephan.lewandowsky@uwa.edu.au), Tom Palmeri (thomas.j.palmeri@Vanderbilt.Edu), or Michael Waldmann (michael.waldmann@bio.uni-goettingen.de).

Papers should be submitted through the regular submission portal for *JEP: LMC* (http://www.apa.org/pubs/journals/xlm/submission.html) with a cover letter indicating that the paper is to be considered for the special section. For instructions to authors and other detailed submission information, see the journal Web site at http://www.apa.org/pubs/journals/xlm.