

# Retrieval Practice Can Insulate Items Against Intralist Interference: Evidence From the List-Length Effect, Output Interference, and Retrieval-Induced Forgetting

Oliver Kliegl and Karl-Heinz T. Bäuml  
Regensburg University

This study sought to determine whether nonselective retrieval practice after study can reduce memories' susceptibility to intralist interference, as it is observed in the list-length effect, output interference, and retrieval-induced forgetting. Across 3 experiments, we compared the effects of nonselective retrieval practice and restudy on previously studied material with regard to these 3 forms of episodic forgetting. When study of an item list was followed by a restudy cycle, recall from a longer list was worse than recall from a shorter list (list-length effect), preceding recall of studied nontarget items impaired recall of the list's target items (output interference), and repeated selective retrieval of some list items attenuated recall of other nonretrieved items at test (retrieval-induced forgetting). In contrast, none of these effects arose when study of the list was followed by a nonselective retrieval cycle. The findings are consistent with a combination of contextual variability theory and a variant of study-phase retrieval theory that assumes that retrieval can create more distinct context features for retrieved items than restudy does for restudied items, thus reducing items' susceptibility to interference relative to restudy cycles. The findings add to the view that nonselective retrieval practice can stabilize and consolidate memories.

*Keywords:* retrieval practice, interference, list-length effect, output interference, retrieval-induced forgetting

Within the last 10 years, there has been a surge of interest in the beneficial effects of retrieval practice on recall performance. One of the most prominent benefits of retrieval practice is the testing effect, which refers to the finding that intermediate retrieval practice between study and a final memory test can dramatically enhance final-test performance when compared with restudy trials (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006). But retrieval practice can induce further memorial benefits. For instance, retrieval practice in comparison to restudy can produce better transfer of the study material (Butler, 2010), improve organization of newly acquired knowledge (Zaromb & Roediger, 2010), or enhance the ability of the learner to benefit from a subsequent restudy opportunity (Arnold & McDermott, 2013; for a list of 10 benefits of retrieval practice, see Roediger, Putnam, & Smith, 2011). It is important to note that it can also reduce memories' susceptibility to retroactive and proactive interference.

## Effects of Retrieval Practice on Retroactive and Proactive Interference

Retroactive interference refers to the finding that memory for some previously studied material is typically impaired when ad-

ditional material is encoded between study and test (Müller & Pilzecker, 1900). Addressing the role of retrieval practice for retroactive interference, Halamish and Bjork (2011) asked subjects to study two lists of word pairs containing the same stimulus terms, and, between study of the two lists, to either retrieve or restudy the first list word pairs. Retrieval practice was conducted by presenting the stimulus words of each single paired associate and letting subjects generate the appropriate response word from the list. At test, the stimulus words were presented again and subjects were asked to recall the response words of the first list. Prior retrieval practice improved recall of the first-list items relative to prior restudy, suggesting that retrieval practice can reduce memories' susceptibility to retroactive interference (for related results, see Abel & Bäuml, 2014; Potts & Shanks, 2012).

Proactive interference refers to the finding that memory for recently studied material is typically impaired by the prior study of additional material (Underwood, 1957). Szpunar, McDermott, and Roediger (2008) applied a multiple-list paradigm, in which subjects successively studied four nontarget lists and a final target list in anticipation of a final cumulative recall test, to examine whether retrieval practice affects proactive interference. All subjects were tested immediately on the target list. One group of subjects was also tested immediately on the four nontarget lists, whereas another group restudied these lists, and a third group did a mathematical distractor task after each single list. Subjects who were tested immediately on the nontarget lists recalled more target items and showed fewer prior-list intrusions than subjects in the two no-testing groups, indicating that interpolated retrieval practice can insulate memories against proactive interference (for related results, see Bäuml & Kliegl, 2013; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Weinstein, McDermott, & Szpunar, 2011).

---

This article was published Online First August 24, 2015.

Oliver Kliegl and Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University.

Correspondence concerning this article should be addressed to Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany. E-mail: karl-heinz.baeuml@ur.de

The finding that retrieval practice can reduce memories' susceptibility to retroactive and proactive interference has been explained through enhanced list segregation processes, arguing that retrieval practice can help distinguishing retrieved information from other, nonretrieved information (Abel & Bäuml, 2014; Bäuml & Kliegl, 2013; Halamish & Bjork, 2011; Szpunar et al., 2008). On the basis of the view that, during study, people encode information not only about the study items but also about the temporal context in which the list is presented (e.g., Bower, 1972; Estes, 1955; Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981), enhanced list segregation may be mediated by the production of distinct context features that are created during retrieval practice and are encoded with the retrieved list items (see Karpicke, Lehman, & Aue, 2014). Arguably, such context features may enable better discrimination of studied lists at test and thus reduce items' susceptibility to retroactive and proactive interference.

### Possible Effects of Retrieval Practice on Interference in Single-List Learning

A common feature of the previous studies on the role of retrieval practice for retroactive and proactive interference is that the retrieved information and the other, nonretrieved information belong to different item lists and only the first studied information is retrieved, immediately after its study. Indeed, in the retroactive interference studies, subjects studied and retrieved the target list and subsequently studied the nontarget list before recalling the target list items (e.g., Abel & Bäuml, 2014; Halamish & Bjork, 2011); in the proactive interference studies, subjects studied and retrieved each single nontarget list and subsequently studied the target list before recalling the target list items (e.g., Pastötter et al., 2011; Szpunar et al., 2008). Although results show that retrieval practice can reduce interlist interference in these situations, it is unclear whether the results generalize to other types of interference situations. This study raises the question of whether retrieval practice can also reduce interference in single-list learning, that is, situations in which all studied items belong to a common single list, and retrieval practice occurs on all studied items and after study of all of the items. Because in this case no room is left for some intermittent retrieval practice to enhance segregation of a set of retrieved items from a set of other, nonretrieved items, the theoretical accounts of retrieval practice effects in retroactive and proactive interference do not speak directly to the possible role of retrieval practice for interference in single-list learning.

That interference can play a critical role for recall in single-list learning has been demonstrated with numerous paradigms. One prominent example is the *list-length effect*, which refers to the finding that, as the number of words in a study list increases, the proportion of words recalled from the list decreases (Murdock, 1962; Postman & Phillips, 1965; Watkins & Watkins, 1975). The finding is generally attributed to retrieval competition at test, assuming that intralist interference for a target item increases with the number of the other list items, thus enhancing retrieval competition and reducing target recall (e.g., Raaijmakers & Shiffrin, 1981; Rundus, 1973). Another example is *output interference*, the demonstration that an item's recall probability can decline with the item's serial position in the testing sequence (Roediger, 1973; Smith, 1971; Tulving & Arbuckle, 1966). The finding is often

explained by blocking, assuming that preceding recall of some items at test blocks recall of the remaining items due to increased interference from the recalled items (e.g., Roediger & Neely, 1982; Rundus, 1973), although there may also be role for inhibition if the yet-to-be-recalled items interfere during preceding recall and are inhibited to reduce the interference (e.g., Bäuml, 1998; Verde, 2009). A third example comes from studies on *retrieval-induced forgetting*, which have demonstrated that repeated selective retrieval of some studied items can impair later recall of related nonretrieved list items (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995; Ciranni & Shimamura, 1999). Again, the effect has been attributed to inhibition and blocking. The inhibition account assumes that during selective retrieval of some items not-to-be retrieved items interfere and the interference is reduced by inhibition (Anderson, 2003), whereas blocking suggests that repeated retrieval just strengthens the retrieved items and such strengthening induces blocking of related nonretrieved items at test (Raaijmakers & Jakab, 2012; for a different view on retrieval-induced forgetting, see Jonker, Seli, & MacLeod, 2013).

Regarding the role of retrieval practice for interference in single-list learning, there is theoretical reason to predict that retrieval practice can reduce the interference. The prediction can be deduced from Karpicke et al.'s (2014) episodic context account of retrieval-based learning, which at its core is a combination of contextual variability theory (Bower, 1972; Estes, 1955) and a variant of study-phase retrieval theory (Greene, 1989; Thios & D'Agostino, 1976). These two theories have been successfully applied to a wide range of recall findings, including the recency effect, that is, the enhanced recall of end-of-list items in immediate free recall, the contiguity effect, that is, the tendency to successively recall neighboring list items, and the spacing effect, that is, the beneficial mnemonic effect of spaced over massed learning (e.g., Greene, 1989; Howard & Kahana, 2002; Kahana, 1996). Contextual variability theory assumes that temporal context - a current pattern of activity in the subject's mind that, among others, may be influenced by environmental factors, internal factors, as well as the experimental stimulus and task - drifts slowly during study and each studied item is associated with the temporal context in which the item is shown (Bower, 1972; Estes, 1955), an assumption that is included in many computational memory models (e.g., Howard & Kahana, 2002; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1988). Study-phase retrieval theory assumes that repetition of an item retrieves one's memories of the repeated item's earlier occurrences and their associated contexts (Greene, 1989; Thios & D'Agostino, 1976). For instance, when a previously studied item is retrieved, the context representation associated with that item may be updated such that it includes a composite of the features of both the (unique) study and the (unique) retrieval context of the particular item (e.g., Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009). Such an enriched set of (unique) context features may then provide a particularly distinctive retrieval cue that can reduce the item's susceptibility to interference, and thus attenuate forms of intralist interference, like the list-length effect, output interference, and retrieval-induced forgetting.

Critically, although, according to context retrieval theory, both retrieval and restudy may create such distinct context features, retrieval may be more effective in producing such an effect. Such proposal is at the heart of Karpicke et al.'s (2014) explanation of

the testing effect. According to this view, both retrieval and restudy may induce context retrieval, but without intentional retrieval instructions, context retrieval may not be obligatory during restudy cycles. Moreover, the degree of context updating supposed to occur during restudy may be reduced relative to retrieval practice, during which people deliberately search memory information about the prior occurrence of studied items. The suggested difference in context retrieval between retrieval practice and restudy conditions may induce a difference in the creation of unique context cues, and thus enhance long-term retention of retrieved items more than of restudied items. Recent results by [Bäuml and Dobler \(2015\)](#) support the view that retrieval may be more effective than restudy in context retrieval. These authors showed that, after a change in context between study and test, both selective retrieval and restudy of some of the studied items improve recall of the remaining list items, but this beneficial effect can be larger after selective retrieval than restudy. If these beneficial effects reflect context reactivation processes induced by item repetition, the finding indicates that repetition by virtue of retrieval can induce more context retrieval than repetition by virtue of restudy. If so, retrieval may also create more distinct context features, and forms of intralist interference, like the list-length effect, output interference, and retrieval-induced forgetting, may be less pronounced after retrieval than after restudy.<sup>1</sup>

### The Present Study

This study reports the results of three experiments designed to examine whether retrieval practice, in comparison to restudy, reduces interference in single-list learning. To achieve this goal, we compared the effects of retrieval practice and restudy on previously studied material with regard to the list-length effect, output interference, and retrieval-induced forgetting. In all three experiments, subjects studied a list of items and, on a later memory test, were asked to recall the items of the list. Critically, the study phase was always followed by a retrieval-practice or a restudy phase, in which subjects were asked to recall the studied material or the items were reexposed for additional learning. To compare the effects of retrieval practice and restudy on the previously studied items, Experiment 1 used the list-length effect, manipulating whether subjects at test had to recall from a longer or a shorter study list; Experiment 2 used output interference, examining how prior recall of some list items at test influenced the recall of the remaining list items; Experiment 3 used retrieval-induced forgetting, examining how the repeated selective retrieval of some list items influenced the recall of related nonretrieved items at test.

Following [Karpicke et al. \(2014\)](#) and the suggested combination of contextual variability theory and study-phase retrieval theory, we expected that retrieval creates distinct context features for individual retrieved items and may thus reduce or eliminate intralist interference. In consequence, we expected retrieval to reduce or eliminate the list-length effect, output interference, and retrieval-induced forgetting. In contrast, on the basis of the view that restudy cycles can be less effective in creating distinct context features than retrieval-practice cycles are ([Bäuml & Dobler, 2015](#); [Karpicke et al., 2014](#)) and the results of previous studies indicating that restudy cycles do not eliminate items' susceptibility to retroactive and proactive interference ([Bäuml & Kliegl, 2013](#); [Halamish & Bjork, 2011](#); [Pastötter et al., 2011](#); [Szpunar et al., 2008](#)),

we expected the typical list-length effect, output interference, and retrieval-induced forgetting if items after study were not retrieved but reexposed for additional learning.

### Experiment 1

Experiment 1 examined whether retrieval practice reduces the list-length effect. The task consisted of a study phase, a practice phase, and a final test phase. In the study phase, subjects in the short-list condition studied a list of 10 target items, whereas subjects in the long-list condition studied a list of 10 target and 10 nontarget items. In the subsequent practice phase, all (10 or 20) study items were either retrieved or restudied. In the final test phase, subjects recalled the target items. We expected the typical list-length effect in the restudy condition, with target recall being impaired in the long-list compared with the short-list condition. In contrast, in the retrieval-practice condition, we expected this list-length effect to be reduced, or even to be eliminated, assuming that retrieval practice may create a more distinct set of context features for the retrieved items than restudy does for restudied items.

Expectations do also arise regarding recall levels in the retrieval practice condition relative to the restudy condition. For instance, on the basis of the typical testing effect finding (e.g., [Roediger & Butler, 2011](#)), one may expect recall in the retrieval practice conditions to be higher than in the restudy conditions. However, such finding is usually tied to longer retention intervals between retrieval practice and final test (e.g., [Roediger & Karpicke, 2006](#)) and demanding recall formats at test, like free recall (e.g., [Halamish & Bjork, 2011](#)). In contrast, when the retention interval is quite short and/or item-specific cues are provided at test, then restudy and retrieval practice often create similar recall levels at test (e.g., [Bäuml, Holterman, & Abel, 2014](#); [Halamish & Bjork, 2011](#); [Roediger & Karpicke, 2006](#)) or recall in the restudy condition even exceeds recall in the retrieval practice condition (e.g., [Abel & Bäuml, in press](#); [Bäuml et al., 2014](#); [Halamish & Bjork, 2011](#); [Kornell, Bjork, & Garcia, 2011](#)). Because, similar to the prior work on the role of retrieval practice for retroactive and proactive interference (e.g., [Abel & Bäuml, 2014](#); [Halamish & Bjork, 2011](#); [Pastötter et al., 2011](#); [Szpunar et al., 2008](#)), in Experiment 1, we used a retention interval of few minutes only between practice and final test and subjects were presented item-specific cues at test (see Method section), we expected recall levels at test to be lower in the retrieval practice condition than in the restudy condition, at least in the short-list condition of the experiment, that is, in the absence of the interfering nontarget items.

### Method

**Participants.** Thirty-two students (mean age = 22.6 years) at Regensburg University participated in Experiment 1. They took part in the experiment in return for either partial course credit or a

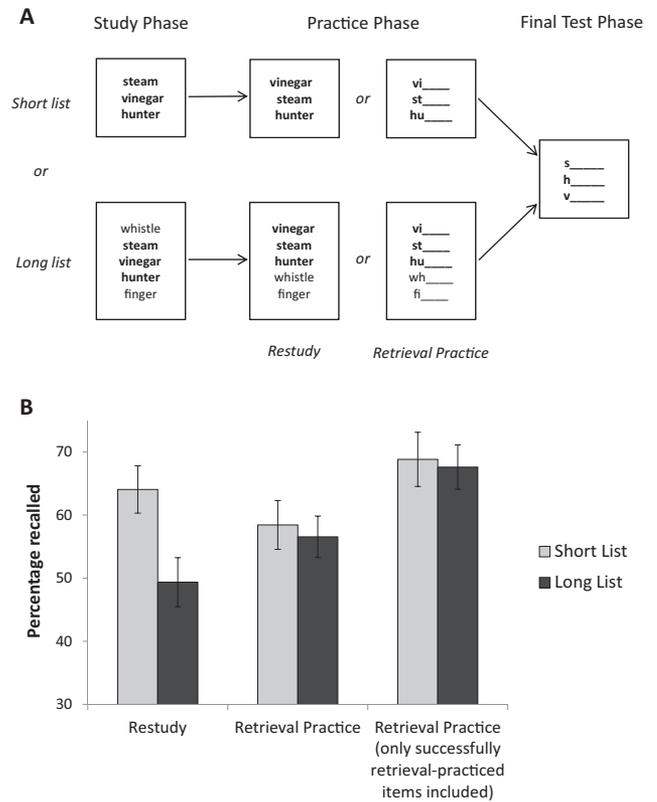
<sup>1</sup> This account of possible effects of retrieval practice is highly similar to recent accounts of the spacing effect (e.g., [Benjamin & Tullis, 2010](#); [Delaney, Verkoeijen, & Spiguel, 2010](#); [Siegel & Kahana, 2014](#); see also [Karpicke et al., 2014](#)). Indeed, the main difference is that here it is assumed that retrieval can induce an even higher level of context retrieval than restudy does.

compensatory amount of money. All of the subjects spoke German as their native language.

**Material.** Four study lists were constructed, each containing 20 unrelated German nouns, 10 target and 10 nontarget items (Duyck, Desmet, Verbeke, & Brysbaert, 2004). Among all the items of a list, both each target and each nontarget item began with a unique initial letter. Two lists were used in the restudy and retrieval-practice conditions each.

**Design.** The experiment had a  $2 \times 2$  within-subjects design with the factors of practice (restudy vs. retrieval practice) and list length (long list vs. short list). Practice conditions differed in whether, following initial study, subjects were reexposed to the study list for a restudy cycle, or whether they were asked to retrieve the study list. List-length conditions differed in whether subjects studied a list of 10 (target) items or a list of 20 items, consisting of 10 target and 10 nontarget items; in the latter case, the distinction between target and nontarget items was unknown to the subjects. Order of conditions and assignment of lists to conditions were counterbalanced.

**Procedure.** The experiment consisted of four blocks, with each block consisting of three main phases: a study phase, a practice phase, and a final test phase (see Figure 1A). In the study phase, the study items were presented individually and in a blocked random order for 4 seconds each. In the short-list condition, subjects studied the 10 target items only. In the long-list condition, subjects additionally studied the 10 nontarget items, five nontargets before and five nontargets after presentation of the target items. After study, subjects in the long-list condition were asked to count backward from a three-digit number as a recency control for 30 s. This counting task was followed by the practice phase, during which subjects were either reexposed to all of the items from the study list for additional learning (restudy condition), or were asked to complete unique word stems of all the items from the study list. Subjects had 4 seconds to complete each word stem (retrieval-practice condition). Employing these item-specific cues guaranteed that subjects retrieved the 10 target items first and the 10 nontarget items last, which ensured comparable success rates for target items in the short-list and long-list conditions (see Results section). Afterward, subjects were asked to solve mathematical problems for 180 s. The procedure was analogous in the short-list condition, but the time that was spent on the counting and math (distractor) tasks was adjusted to ensure that the average time interval between study and test of the target items was identical in the two list-length conditions. In the final test phase, subjects in both conditions were asked to orally recall the 10 target items. In the long-list condition, subjects subsequently were asked to also recall the 10 nontarget items. Each of the target and nontarget items was cued with its unique initial letter. Both the target and nontarget cues were presented individually and in a blocked random order for 6 s each. One half of the subjects completed the retrieval-practice condition first and the restudy condition second, with a 5-min break between conditions; order of conditions was reversed for the other one half of the subjects. Within both the retrieval-practice and the restudy conditions, one half of the subjects completed the long-list condition first and the short-list condition second, with a 2-min break between conditions; order of conditions was reversed for the other one half of the subjects.



**Figure 1.** (A) Study, practice, and final test phases of Experiment 1. In the study phase, subjects studied either 10 target items (short-list condition) or 10 target items and 10 nontarget items (long-list condition). In the subsequent practice phase, subjects were either reexposed to all study items for additional learning (restudy condition), or were asked to recall all study items (retrieval-practice condition). In the final test phase, subjects recalled the target items. Predefined target items are depicted in bold letters. (B) Results of Experiment 1. Mean correct recall of target list items as a function of practice (restudy, retrieval practice) and list length (long list, short list). Error bars represent standard errors.

## Results

**Retrieval success in the practice phase.** In the retrieval-practice condition, mean retrieval success rates were 75.9% for target items in the short-list condition and 76.6% for target items in the long-list condition. The difference between conditions was not significant,  $t(31) < 1$ . The success rate for nontarget items in the long-list condition was 71.6%.

**Final test performance.** Figure 1B shows recall for target items in the restudy and retrieval-practice conditions, both for short and long lists. A  $2 \times 2$  analysis of variance (ANOVA) with the within-subjects factors of list length (short list vs. long list) and practice (restudy vs. retrieval practice) revealed a main effect of list length,  $F(1, 31) = 13.007$ ,  $MSE = .02$ ,  $p = .001$ , partial  $\eta^2 = .30$ , reflecting better recall in the short-list than the long-list condition, but no significant main effect of practice,  $F(1, 31) < 1$ . In addition, a significant interaction between the two factors arose,  $F(1, 31) = 11.272$ ,  $MSE = .01$ ,  $p = .002$ , partial  $\eta^2 = .27$ , suggesting that type of practice affected the list-length effect. Indeed, planned comparisons showed that there was a reliable

effect of list length in the restudy condition, with target recall being significantly worse in the long-list than the short-list condition (49.4% vs. 64.1%),  $t(31) = -5.110$ ,  $p < .001$ ,  $d = -.907$ , whereas in the retrieval-practice condition, there was no reliable difference between conditions (56.6% vs. 58.4%),  $t(31) < 1$ ,  $d = -.105$ . In the present experiment, both list length and practice were manipulated within subjects. It is important to note that none of the reported statistical effects interacted with the participants' testing order (all  $ps > .225$ ), and there was also no main effect of testing order (all  $ps > .598$ ).

**Additional analysis.** For the retrieval-practice condition, we also examined final test performance for target items when only those target items were taken into account that had been successfully retrieved during the practice phase. A  $2 \times 2$  ANOVA with the within subjects factors of retrieval success (all items vs. successfully retrieved items only) and list length (long list vs. short list) revealed a main effect of retrieval success,  $F(1, 31) = 52.199$ ,  $MSE = .03$ ,  $p < .001$ , partial  $\eta^2 = .63$ , but no main effect of list length,  $F(1, 31) < 1$ , and no interaction between the two factors,  $F(1, 31) < 1$ . The main effect of retrieval success reflects the expected superior target recall when only successfully retrieved items, rather than all items, were included in the analysis. Planned comparisons further showed that no list-length effect arose for successfully retrieved items (67.6% vs. 68.8%),  $t(31) < 1$ ,  $d = -.064$ . Again, none of the reported statistical effects interacted with participants' testing order (all  $ps > .563$ ), and there was also no main effect of testing order (all  $ps > .724$ ).<sup>2</sup>

## Discussion

In the restudy condition, recall of target items was impaired in the long-list condition relative to the short-list condition, thus replicating the typical list-length effect (e.g., Murdock, 1962; Postman & Phillips, 1965; Watkins & Watkins, 1975). In contrast, in the retrieval-practice condition, recall of target items was equivalent in the two list conditions and no list-length effect arose. The findings are consistent with the suggested combination of contextual variability theory and study-phase retrieval theory and the theoretical view that retrieval creates more distinct context features for individual retrieved items than restudy does for restudied items, thus reducing retrieval competition between items and attenuating the list-length effect. Recall levels at test were numerically lower in the retrieval practice condition than in the restudy condition, at least in the short-list condition of the experiment when the interfering nontarget items were absent, and when the analyses included both successfully and unsuccessfully retrieved items. This finding replicates results from previous retrieval practice studies that, like Experiment 1, used a retention interval of few minutes only between practice and final test or provided item-specific cues at test (e.g., Bäuml et al., 2014; Halamish & Bjork, 2011; Kornell et al., 2011). When the analyses included successfully retrieved items only, recall levels at test were no longer lower in the retrieval practice condition than in the restudy condition. Importantly, however, the list-length effect remained absent. The goal of Experiments 2 and 3 was to examine whether retrieval-induced reductions in intralist interference are restricted to the list-length effect or generalize to other forms of episodic forgetting, like output interference and retrieval-induced forgetting.

## Experiment 2

Experiment 2 examined whether retrieval practice reduces output interference. The task consisted of a study phase, a practice phase, and a final test phase. In the study phase, subjects studied a list of predefined target and nontarget items. In the subsequent practice phase, all study items were either retrieved or restudied. In the final test phase, subjects were asked to recall the target items either with or without prior recall of the nontarget items. Analogous to Experiment 1, we expected that in the restudy condition, typical output interference would arise and target recall be impaired in the presence compared with the absence of prior nontarget recall (e.g., Bäuml & Sameniéh, 2010, 2012). In contrast, in the retrieval-practice condition, we expected output interference to be reduced, or even to be eliminated, compared with the restudy condition. Analogous to Experiment 1, the expectation was based on the view that retrieval practice reduces intralist interference, thus attenuating possible effects of blocking and inhibition and reducing output interference. Because Experiment 2, like Experiment 1, used a retention interval of few minutes only between practice and test and subjects were presented item-specific cues at test (see Method section), we again expected recall levels at test to be lower in the retrieval practice condition than in the restudy condition, at least when target items were recalled first and there was no preceding recall of the nontarget items.

## Method

**Participants.** Thirty-two students (mean age = 23.7 years) at Regensburg University participated in Experiment 2. They took part in the experiment in return for either partial course credit or a compensatory amount of money. All of the subjects spoke German as their native language.

**Material.** Four study lists were constructed, each containing 15 unrelated German nouns, 5 target and 10 nontarget items each (Duyck et al., 2004). Target items were determined by the experimenter but were unknown to the subjects. Among all items, each target item began with a unique initial letter and each nontarget item had a unique word stem. Two lists were used in the restudy and retrieval-practice conditions each.

**Design.** The experiment had a  $2 \times 2$  within-subjects design with the factors of practice (restudy vs. retrieval practice) and prior selective retrieval (targets first vs. targets last). Practice conditions differed in whether, following initial study, subjects were reexposed to the study list for additional learning, or whether they were asked to retrieve the items of the list. Prior selective retrieval

<sup>2</sup> For the long-list condition, we also analyzed subjects' final-test performance for nontarget items. A  $2 \times 2$  ANOVA with the within subjects factors of practice (restudy vs. retrieval practice) and item type (target vs. nontarget) revealed no main effect of practice,  $F(1, 31) < 1$ , and no main effect of item type,  $F(1, 31) < 1$ , but an interaction of the two factors,  $F(1, 31) = 5.386$ ,  $MSE = .01$ ,  $p = .027$ , partial  $\eta^2 = .15$ , which reflects a marginally increased recall performance for target items compared with nontarget items in the retrieval practice condition (56.6% vs. 50.0%),  $t(31) = 1.862$ ,  $p = .072$ , but no difference between target and nontarget items in the restudy condition (49.4% vs. 52.2%),  $t(31) < 1$ . The difference between target and nontarget items in the retrieval-practice condition is not surprising, because nontarget items had lower success rates during retrieval practice (see main text) and at test were recalled after the target items, thus likely suffering from output interference.

conditions differed in whether participants were asked to retrieve the 10 nontarget items prior to retrieving the five target items, or were asked to retrieve the five target items first. Order of conditions and assignment of lists to conditions was counterbalanced.

**Procedure.** The experiment consisted of four blocks, with each block consisting of three main phases: a study phase, a practice phase, and a final test phase (see Figure 2A). In the study phase, the study items were presented individually and in a random order for 4 s each. After a 30-s distractor task, in which subjects were asked to count backward from a three-digit number as a recency control, subjects were either reexposed to the items from the study list for additional learning (restudy condition) or were given 60 s to orally recall in any order as many words as possible from the study list (retrieval-practice condition). In contrast to Experiment 1, we did not use item-specific cues to control subjects' output order in the retrieval-practice condition because here exactly the same items were to be practiced in the targets-first and targets-last conditions. After solving mathematical problems for 180 s, the final test phase followed, in which subjects were asked to orally recall the five target items, either with (targets-last condition) or without (targets-first condition) prior retrieval of the nontarget items. Each of the nontarget items was cued with its word stem to increase recall chances, whereas each of the target items was cued with its unique initial letter (see Bäuml & Sameni, 2010, 2012). Both the target and nontarget cues were presented indi-

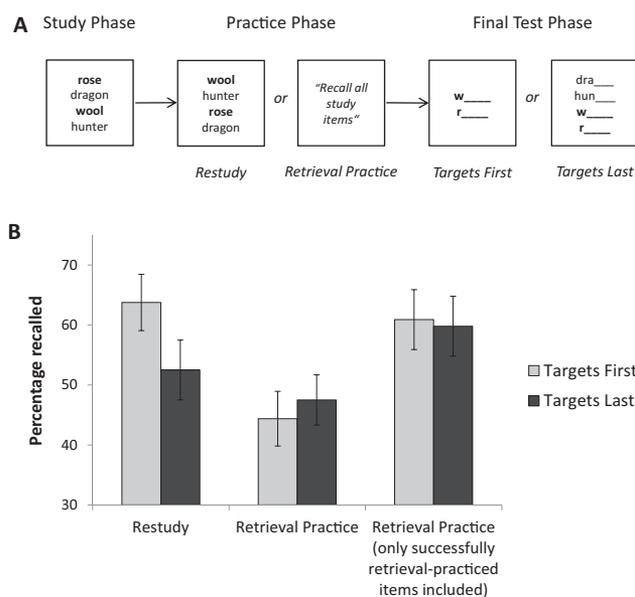
vidually and in a random order for 6 s each. One half of the subjects completed the retrieval-practice condition first and the restudy condition second, with a 5-min break between conditions; for the other half of the subjects, the order of the two conditions was reversed. Within both the retrieval-practice and restudy conditions, one half of the subjects completed the targets-first condition first and the targets-last condition second, with a 2-min break between conditions; for the other half of the subjects, order of conditions was reversed.

## Results

**Retrieval success in the practice phase.** In the retrieval-practice condition, mean retrieval success rates for target items were 67.5% in the target-first condition and 71.3% in the target-last condition. Mean success rates for nontarget items were 71.9% in the target-first condition and 70.0% in the target-last condition. The numerical differences between conditions were not significant,  $t(31) < 1$ .

**Final test performance.** Figure 2B shows recall for target items in the restudy and retrieval-practice conditions, in both the target first and the target last conditions. A  $2 \times 2$  ANOVA with the within-subjects factors of prior selective retrieval (targets first vs. targets last) and practice (restudy vs. retrieval practice) revealed no main effect of prior selective retrieval,  $F(2, 31) < 1$ , but a significant main effect of practice,  $F(1, 31) = 9.847$ ,  $MSE = .04$ ,  $p = .004$ , partial  $\eta^2 = .24$ , which reflects better recall in the restudy than the retrieval-practice condition. In addition, a significant interaction between the two factors arose,  $F(1, 31) = 4.429$ ,  $MSE = .04$ ,  $p = .044$ , partial  $\eta^2 = .13$ , suggesting that type of practice affected prior selective retrieval differently. Indeed, planned comparisons showed that there was a reliable effect of prior selective retrieval in the restudy condition, with recall of target items being significantly worse in the presence than in the absence of prior nontarget recall (51.3% vs. 62.5%),  $t(31) = -2.150$ ,  $p = .039$ ,  $d = -.379$ , whereas in the retrieval-practice condition, recall rates were not affected by the presence or absence of prior nontarget recall (47.5% vs. 43.8%),  $t(31) < 1$ ,  $d = .116$ . Both prior selective retrieval and practice were manipulated within participants in this experiment. Importantly, none of the reported statistical effects interacted with testing order (all  $p$ s  $> .231$ ), and there was also no main effect of testing order on recall performance (all  $p$ s  $> .302$ ).

**Additional analysis.** Again, we also examined final-test performance for target items in the retrieval-practice condition when for each single subject only those items were taken into account that had been successfully retrieved during the practice phase. A  $2 \times 2$  ANOVA with the within subjects factors of retrieval success (all items vs. successfully retrieved items only) and prior selective retrieval (targets first vs. targets last) revealed a main effect of retrieval success,  $F(1, 31) = 26.149$ ,  $MSE = .06$ ,  $p < .001$ , partial  $\eta^2 = .46$ , reflecting the expected superior recall performance when only successfully retrieved items, rather than all items, were included in the analysis. Besides, there was no main effect of prior selective retrieval,  $F(1, 31) < 1$ , and no significant interaction between the two factors,  $F(1, 31) < 1$ . Planned comparisons further showed that no output-interference effect arose for successfully retrieved items, (59.7% vs. 61.2%),  $t(31) < 1$ ,  $d = -.040$ . Importantly, none of the reported statistical effects interacted with



**Figure 2.** (A) Study, practice, and final test phases of Experiment 2. In the study phase, subjects studied a list of five target items and 10 nontarget items. In the subsequent practice phase, they were reexposed to the study items for additional learning (restudy condition), or were asked to recall the study items (retrieval-practice condition). In the final test phase, subjects were asked to recall the target items, in either the absence (targets-first condition) or the presence (targets-last condition) of prior retrieval of the nontarget items. Predefined target items are depicted in bold letters. (B) Results of Experiment 2. Mean correct recall of target list items as a function of practice (restudy, retrieval practice) and prior selective retrieval (targets first, targets last). Error bars represent standard errors.

testing order (all  $p$ s > .490), and there was also no main effect of testing order (all  $p$ s > .301).<sup>3</sup>

## Discussion

In the restudy condition, prior recall of nontarget items impaired recall of the target items compared with when no nontarget items were recalled before, thus replicating the typical output-interference effect (e.g., Bäuml & Sameniéh, 2010; Roediger, 1973; Smith, 1971). In contrast, in the retrieval-practice condition, prior recall of the nontarget items left recall of the target items unaffected and no output-interference effect arose. These findings are consistent with the view that retrieval practice of studied items can create more distinct context features than restudy of the same items, thus reducing possible effects of blocking and inhibition and attenuating output interference. In line with Experiment 1, recall levels at test were again lower in the retrieval practice condition than in the restudy condition, at least when target items were recalled first and when the analyses included both successfully and unsuccessfully retrieved items. When the analyses included successfully retrieved items only, recall levels at test were no longer lower in the retrieval practice condition than in the restudy condition, but the output interference effect remained absent.

## Experiment 3

The goal of Experiment 3 was to examine whether retrieval practice also reduces retrieval-induced forgetting. The task consisted of a study phase, a nonselective practice phase, a selective retrieval phase, and a final test phase. In the study phase, subjects studied a list consisting of exemplars from several semantic categories. In the subsequent nonselective practice phase, all study items were either retrieved or restudied, before, in the selective retrieval phase, subjects retrieved some of the exemplars from some of the categories presented during study (e.g., Anderson et al., 1994; Anderson & Spellman, 1995). At test, all items from the study list were tested. Analogous to Experiments 1 and 2, we expected typical retrieval-induced forgetting in the restudy condition, with impaired recall for the nonretrieved items belonging to the same categories as the selectively retrieved items, relative to the control items belonging to those categories from which no item was selectively retrieved. In contrast, in the nonselective retrieval-practice condition, we expected retrieval-induced forgetting to be reduced or even to be eliminated, assuming that nonselective retrieval practice after study may attenuate intralist interference and thus reduce possible detrimental effects of inhibition and blocking. Because Experiment 3, like Experiments 1 and 2, used a relatively short retention interval between practice and test and subjects were presented item-specific cues at test (see Method section), we again expected recall levels at test to be lower in the retrieval practice condition than in the restudy condition, at least for the control items, for which the preceding selective retrieval of other category exemplars was absent.

## Method

**Participants.** Thirty-six students (mean age = 20.6 years) at Regensburg University participated in Experiment 3. They took part in the experiment in return for either partial course credit or a

compensatory amount of money. All of the subjects spoke German as their native language.

**Material.** Two sets of 36 items were compiled, each consisting of six exemplars from six semantic categories (Van Overschelde, Rawson, & Dunlosky, 2004). Within a category, all items had a unique initial letter. Each item was equally often used as a selectively retrieved item, a nonretrieved item belonging to the same category as a selectively retrieved item, and a control item belonging to those categories from which no item was retrieved in the selective retrieval phase. For each subject, one item set was used in the restudy condition and the other item set in the retrieval-practice condition.

**Design.** The experiment had a  $2 \times 3$  within-subjects design with the factors of practice (restudy vs. retrieval practice) and item type (SR+ items, SR- items, C items). Like in Experiments 1 and 2, practice conditions differed in whether, following initial study, subjects were represented the items from the study list for additional learning, or were asked to retrieve all the list items. In the selective retrieval phase, three items from four of the six categories were repeatedly retrieved, resulting in three item types: selectively retrieved items (SR+ items), nonretrieved items belonging to the same categories as the selectively retrieved items (SR- items), and control items belonging to those categories from which no item was selectively retrieved (C items).

**Procedure.** The experiment consisted of two blocks, with each block consisting of four main phases: a study phase, a practice phase, a selective retrieval phase, and a final test phase (see Figure 3A). In the study phase, the 36 category-exemplar pairs were presented individually and in a random order for 4 s each. Subsequently, subjects were asked to count backward from a three-digit number for 30 s as a recency control. In the practice phase, subjects were either reexposed to all category-item pairs from the study list for additional learning (restudy condition), or they were presented with the six category cues and were asked to recall as many exemplars from the study list as possible. The category cues were presented individually and in a random order for 24 s each (retrieval-practice condition). Similar to Experiment 2, we did not use item-specific cues to control subjects' output order in the retrieval-practice condition. After another 60 s of backward counting, the selective retrieval phase followed, in which subjects were asked to recall half of the exemplars from four of the six categories in two successive retrieval cycles. Following typical retrieval-induced forgetting experiments, the category name and the first two or three letters of the exemplar were presented individually for 4 s each as retrieval cues. After solving mathematical problems for 5 min, the final test phase followed, in which subjects were asked to orally recall all 36 exemplars. Each exemplar was cued with its category name and its unique initial

<sup>3</sup> Regarding nontarget recall in the final test, a  $2 \times 2$  ANOVA with the within subjects factors of prior selective retrieval (targets first vs. targets last) and practice (restudy vs. retrieval practice) revealed no main effect of prior selective retrieval,  $F(1, 31) < 1$ , a significant main effect of practice,  $F(1, 31) = 9.146$ ,  $MSE = .02$ ,  $p = .005$ , partial  $\eta^2 = .23$ , reflecting better recall in the restudy than the retrieval practice condition, and no interaction between the two factors,  $F(1, 31) < 1$ . The lack of a main effect of prior selective retrieval for nontarget items is not surprising given that (a) each list consisted of five target items only, and (b) in the final test, each nontarget item was cued with its word stem (whereas target items were cued with their initial letters).



with C items) arose only in the restudy condition (71.1% vs. 86.8%),  $t(35) = -7.162$ ,  $p < .001$ ,  $d = -1.203$ , but not in the nonselective retrieval-practice condition (75.0% vs. 72.5%),  $t(35) = 1.405$ ,  $p = .169$ ,  $d = .235$ . Both item type and practice were manipulated within participants in the present experiment. Importantly, none of the reported statistical effects interacted with participants' testing order (all  $ps > .163$ ), and there was also no main effect of testing order (all  $ps > .165$ ).

**Additional analysis.** For the nonselective retrieval-practice condition, we also examined final-test performance when only those target items were taken into account that had been successfully retrieved during the nonselective practice phase. A  $2 \times 2$  ANOVA with the within-subjects factors of retrieval success (all items vs. successfully retrieved items only) and item type (SR– vs. C) revealed a main effect of retrieval success,  $F(1, 35) = 121.142$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .78$ , no main effect of item type,  $F(1, 35) = 1.840$ ,  $MSE = .01$ ,  $p = .184$ , and no interaction between the two factors,  $F(1, 35) < 1$ . The main effect of retrieval success reflects superior recall performance when only successfully retrieved items, rather than all items, were included. Planned comparisons showed similar recall performance for the SR– and C items (94.5% vs. 92.8%),  $t(35) < 1$ ,  $d = .134$ , suggesting that no retrieval-induced forgetting arose for successfully retrieval-practiced items. Importantly, none of the reported statistical effects interacted with testing order (all  $ps > .490$ ), and there was no main effect of testing order (all  $ps > .541$ ).

## Discussion

In the restudy condition, selective retrieval of some category exemplars impaired recall of the categories' nonretrieved exemplars, relative to baseline items belonging to categories from which no item was selectively retrieved, thus demonstrating typical retrieval-induced forgetting (e.g., Anderson et al., 1994; Anderson & Spellman, 1995). In contrast, in the nonselective retrieval-practice condition, no retrieval-induced forgetting arose and the subsequent selective retrieval of some category exemplars left recall of the same categories' nonretrieved items unaffected. The findings are consistent with the view that nonselective retrieval practice can create distinct context features for the retrieved items that can reduce possible effects of inhibition and blocking and thus insulate items against the detrimental effects of selective memory retrieval. In line with Experiments 1 and 2, recall levels at test were again lower in the nonselective retrieval practice condition than in the restudy condition, at least for the control items, for which no preceding selective retrieval of other category exemplars occurred, and when the analyses included both items that were successfully retrieved and items that were unsuccessfully retrieved in the nonselective retrieval practice phase. When the analyses included successfully retrieved items only, recall levels at test were no longer lower in the nonselective retrieval practice condition than in the restudy condition. Importantly, however, retrieval-induced forgetting remained absent.<sup>4</sup>

## General Discussion

The results of the present study replicate prior work by providing another demonstration of the list-length effect, output interference, and retrieval-induced forgetting. Indeed, when study of an

item list was followed by a restudy cycle, recall from a longer list was worse than recall from a shorter list (list-length effect; Watkins & Watkins, 1975), preceding recall of studied nontarget items impaired recall of studied target items (output interference; Roediger, 1973), and repeated selective retrieval of some list items attenuated recall of related nonretrieved items at test (retrieval-induced forgetting; Anderson et al., 1994). Going beyond the prior work, the results of the present study also showed that nonselective retrieval practice can eliminate all these effects. Consistently, there were no list-length effect, no output interference, and no retrieval-induced forgetting if after study of an item list subjects were asked to recall (all) the studied items, indicating that retrieval practice compared with restudy cycles can attenuate episodic forgetting in these paradigms.

## The Distinct Context Features View on Retrieval Practice Effects

The present findings are consistent with the theoretical view that nonselective retrieval practice can create more distinct context features for retrieved items than restudy does for restudied items, and the presence of these context features reduces the items' susceptibility to interference. The view follows Karpicke et al.'s (2014) episodic context account and at its core includes contextual variability theory (Bower, 1972; Estes, 1955) and a variant of study-phase retrieval theory (Greene, 1989; Thios & D'Agostino, 1976) that assumes more context retrieval after retrieval practice than restudy. According to this view, list items differ in their contextual representations after study due to gradual contextual drift during encoding (e.g., Howard & Kahana, 2002; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1988). When it comes to retrieval of an item, the context representation of the item is then updated such that the new representation includes a composite of the features of both the (unique) study context and the (unique) retrieval context of the item (e.g., Howard & Kahana, 2002; Polyn et al., 2009), which can then provide a particularly effective retrieval cue for the item reducing its susceptibility to interference. Critically, in this approach, repetition by virtue of restudy is assumed to be less effective in creating unique context features

<sup>4</sup> Regarding final recall of those items that were unsuccessfully retrieved during nonselective retrieval practice, recall levels in the single experiments were fairly low, which is consistent with prior work (e.g., Halamish & Bjork, 2011). In Experiment 1, subjects recalled 26.8% of the items in the short-list condition, and 22.9% of the items in the long-list condition; in Experiment 2, subjects recalled 7.7% of the items in the targets-first condition and 17.2% of the items in the targets-last condition; in Experiment 3, subjects recalled 33.3% of the C items and 32.8% of the SR– items. Planned comparisons showed no significant list-length, output-interference, or retrieval-induced forgetting effects for unsuccessfully retrieved items when each experiment was analyzed separately (all  $ps > .29$ ). Similarly, when pooling the data of the three experiments' "high-interference conditions" (i.e., the long-list condition, the targets-last condition, SR– items) and the three experiments' "low-interference conditions" (i.e., the short-list condition, the targets-first condition, C items), no reliable difference between the two conditions was found,  $t(99) < 1$ . On the basis of the distinct context features view on retrieval practice effects, these findings may not be surprising. If the successfully retrieved items become more distinct list items, this higher distinctiveness level may also reduce interference for the unsuccessfully retrieved items, thus making also the latter items relatively immune against increases in list length, output interference, and retrieval-induced forgetting.

than is repetition by virtue of retrieval, possibly because restudy is less effective than retrieval in reactivating the study context (Bäuml & Döbler, 2015). Thus, though interference effects may still emerge after restudy cycles, they may be reduced, or even be absent, after nonselective retrieval cycles, which is exactly what the present results suggest.

The present indication that nonselective retrieval practice creates more distinct context features for retrieved items than restudy does for restudied items and thus reduces intralist interference fits with theoretical accounts of the list-length effect, output interference, and retrieval-induced forgetting. The list-length effect is generally attributed to retrieval competition at test, assuming that intralist interference for a target item increases with the number of other list items, thus enhancing retrieval competition and reducing target recall (e.g., Raaijmakers & Shiffrin, 1981; Rundus, 1973); the blocking explanation of retrieval-induced forgetting and output interference assumes that prior retrieval of some (nontarget) items strengthens the memory representations of those items and thus increases their interference potential for later recall of the yet-to-be-recalled (target) items (e.g., Raaijmakers & Jakab, 2012; Roediger & Neely, 1982); the inhibition account of retrieval-induced forgetting and output interference assumes that during selective retrieval, other not-to-be-retrieved items interfere and are inhibited to reduce the interference (e.g., Anderson et al., 1994; Bäuml, 1998). If the distinct context features view on retrieval practice effects was correct and nonselective retrieval practice created an enriched set of (unique) context features for the retrieved items, then nonselective retrieval practice should improve the distinctiveness of the items and benefit retrieval by allowing subjects to consider information that is specific to a particular item. Nonselective retrieval practice should therefore reduce response competition between items, possible blocking effects at recall, and the need for any inhibitory action during selective memory retrieval (e.g., Hunt & McDaniel, 1993; Hunt & Smith, 1996; Smith & Hunt, 2000), and thus attenuate the list-length effect, output interference, and retrieval-induced forgetting, which is consistent with the present results.

The present finding that nonselective retrieval practice reduces intralist interference generalizes findings from prior work, which examined the effects of nonselective retrieval practice on interlist interference. This work, for instance, demonstrated that retrieval practice after study of a target list can reduce retroactive interference (Abel & Bäuml, 2014; Halamish & Bjork, 2011) and that retrieval practice after study of single nontarget lists can reduce proactive interference (Pastötter et al., 2011; Szpunar et al., 2008). These findings were explained through enhanced list segregation processes, arguing that retrieval practice after the first studied list(s) can help distinguishing the retrieved information from the other, nonretrieved information. Although this segregation account does not speak directly to the single-list learning situations used in the present study (see previous paragraphs), the present distinctive context features view on retrieval practice effects may be able to explain the effects of retrieval practice on both intralist and interlist interference. Indeed, if nonselective retrieval practice on a list created distinct context features for the single list items by updating the items' original study context with elements of the current retrieval context, such an enriched set of retrieval cues should both make the single list items more distinctive within their list, thus reducing the list-length effect, output interference, and retrieval-

induced forgetting, and, in multiple list learning, enhance list segregation by adding distinct context features to the items of the retrieved list(s) but not to the items of other, nonretrieved list(s), thus reducing proactive and retroactive interference of target list items. Future work may like to examine this proposal in more depth, for instance, by using computational modeling techniques.<sup>5</sup>

### Alternative Views on Retrieval Practice Effects

While the distinctive context features view can explain the present results, the question arises of whether the results may be equally well described by alternative views on retrieval practice effects, like the elaborative retrieval hypothesis or the bifurcation model. According to the elaborative retrieval hypothesis, retrieval practice improves recall by inducing more elaborative processing than restudy does (Carpenter, 2009; Pyc & Rawson, 2010). When attempting to retrieve a target item from memory, semantically related items may be activated while searching for the target information and become linked to the target item. Such extra information may be activated mainly during more difficult retrieval tasks, when the target information is less readily retrievable and more extensive memory search is required, and may not be activated at all during restudy opportunities, when the target information is reexposed intact. Support for the account, for instance, comes from studies showing that the testing effect is more beneficial when the initial retrieval practice is made more difficult (Carpenter, 2009; Pyc & Rawson, 2010) and when the delay between study and retrieval practice is increased (Rawson, Vaughn, & Carpenter, 2015), and from studies showing that semantic mediators are more likely to be activated during retrieval than during restudy cycles (Carpenter, 2011).

However, the account cannot easily explain the present results, because there is reason to expect that elaboration should enhance rather than reduce the list-length effect, output interference, and retrieval-induced forgetting. Such reasoning is suggested by the results of a recent study by Lehman, Smith, and Karpicke (2014), in which subjects were presented several nontarget lists and a final target list in succession and, after study of each of the nontarget lists, were asked to either retrieve the list items, perform an unrelated

<sup>5</sup> At first glance, the results of the present Experiment 1 may look similar to results from the list-before-last paradigm (e.g., Jang & Huber, 2008; Shiffrin, 1970). However, a more careful look at the two lines of work reveals that the earlier work addressed a very different aspect of retrieval practice effects than does this study. Indeed, in a typical list-before-last experiment, subjects basically study three item lists in succession: a preceding nontarget list (List 1), a target list (List 2), and an intervening nontarget list (List 3). Subsequently, there is recall of all target list (List 2) items. The core finding in this paradigm is that, if the preceding nontarget list (List 1) is retrieved between study of the target list (List 2) and the intervening nontarget list (List 3), later recall of the target list is not affected by the length of the intervening list (List 3), whereas it is affected if there is no such retrieval practice. Two differences between these previous multiple-list studies and the present single-list study stand out: (1) In our study, the target list is subject to retrieval practice, in the list-before-last paradigm the preceding nontarget list is practiced and there is no retrieval practice of the target list at all. (2) In the list-before-last paradigm, retrieval practice (on the preceding nontarget list) eliminates the effect of list length of the intervening list, whereas in our study retrieval practice (on the target list) eliminates the effect of list length of the target list. Thus, there is no simple relationship between the present Experiment 1 and the prior work using the list-before-last paradigm.

distractor task, or generate semantically related words for each list item. After study of the target list, all subjects were asked to recall the list. Whereas, compared with the distractor condition, interpolated retrieval practice increased recall totals and reduced recall latencies, which replicated prior work (Bäuml & Kliegl, 2013), the interpolated elaboration task reduced recall totals and increased response latencies, indicating that elaboration enhances interference. Because the present results suggest that retrieval practice reduces interference, extra processing of semantic information during retrieval cycles should not have mediated the effects of retrieval practice in the present experiments.

Another recent account of the testing effect is the bifurcation model (Kornell et al., 2011), which assumes that, on a scale of memory strength, retrieval practice creates a bifurcated distribution of items, with the successfully retrieved items being strengthened to a very high degree and the nonretrieved items remaining at their original strength level. In contrast, when subjects are provided with the opportunity to restudy the previously encoded items, all restudied items are supposed to be strengthened about equally, though to a lower degree than successful retrieval does for the retrieval-practiced items. The bifurcation model, for instance, can explain the role of delay between retrieval practice and final test for the testing effect (Kornell et al., 2011), it is in line with results showing that the testing effect can be present with difficult testing formats but be absent with easy testing formats (Halamish & Bjork, 2011), and it provides an explanation for the finding that sleep immediately after study can eliminate the testing effect (Bäuml et al., 2014).

Like the elaborative retrieval hypothesis, however, the bifurcation model cannot easily explain the present results. Indeed, if nonselective retrieval practice created higher strengthening levels for the successfully practiced items than restudy does for restudied items, as is suggested by the bifurcation model, then the list-length effect, output interference, and retrieval-induced forgetting should not be eliminated after retrieval practice. Corresponding evidence, for instance, arises from prior work on the list-length effect, which found that the size of the effect is largely independent from the strength level of the list items (Roberts, 1972; see also Raaijmakers & Shiffrin, 1981). Similarly, studies on retrieval-induced forgetting and output interference reported evidence that items with a higher strength level suffer more from the preceding selective retrieval of other list items than items with a lower strength level (e.g., Anderson et al., 1994; Bäuml, 1998; see also Murayama, Miyatsu, Buchli, & Storm, 2014). Because the present results suggest that retrieval practice reduces, and can even eliminate, the list-length effect, output interference, and retrieval-induced forgetting, the suggested high level of strengthening of retrieved items should not have mediated the effects of retrieval practice in the present experiments. Thus, though both the elaborative retrieval hypothesis and the bifurcation model may capture important aspects of the beneficial effects of nonselective retrieval practice (see previous text), they should not underlie the finding that nonselective retrieval practice can reduce and even eliminate interference between items.

## Two Final Remarks on the Beneficial and Detrimental Effects of Retrieval Practice

Although most previous studies on the testing effect reported evidence that nonselective retrieval practice improves later recall of practiced items more than restudy does (e.g., Roediger & Butler, 2011), a number of studies also showed that, under certain condi-

tions, the testing effect may not arise. Accordingly, no testing effect or even improved recall in the restudy relative to the retrieval practice condition were observed if the retention interval between practice and final test was short or difficulty of the final recall test was low (e.g., Bäuml et al., 2014; Halamish & Bjork, 2011; Kornell et al., 2011; Roediger & Karpicke, 2006). In the present series of experiments, both of these conditions were met. In all three experiments, the retention interval between practice and test was on the order of few minutes only and unique initial letter cues were provided as item-specific cues at test to control the output order of the studied items. The present finding that recall levels of studied targets in the retrieval condition were below those of the restudy condition in the short-list condition of Experiment 1, in the targets-first condition of Experiment 2, and for the control items of Experiment 3 thus is in line with the results of the previous studies and supports the claim that retention interval and recall format can modulate and even reverse the testing effect.

Although the results of all three present experiments include conditions that show a reversed testing effect, each of them clearly indicates beneficial effects of nonselective retrieval practice on intralist interference, thus adding to the list of benefits that can result from nonselective retrieval practice. Doing so, the present results also reveal a new boundary condition on retrieval-induced forgetting. To date, research on the effects of nonselective retrieval practice and research on the effects of selective retrieval practice have followed very different research lines, with the one line focusing on the benefits of (nonselective) retrieval practice, and the other line focusing on possible detrimental effects of (selective) retrieval practice. The present research makes a step toward bridging this gap by demonstrating that nonselective retrieval practice can directly influence the detrimental effect of selective retrieval practice. Consistently, the results of Experiment 3 on retrieval-induced forgetting showed that preceding nonselective retrieval practice can eliminate the detrimental effect of selective memory retrieval, and the results of Experiment 2 on output interference showed that the preceding retrieval of the nontargets of a list does no longer impair subsequent target recall if nonselective retrieval practice occurs after study. When studied in isolation, both the beneficial and the detrimental effects of retrieval practice have been demonstrated over a wide range of materials and experimental settings (e.g., Roediger & Butler, 2011; Storm et al., 2015). Future work may therefore like to generalize the present findings to some of the circumstances used in the prior work on selective and nonselective retrieval practice effects.

## Conclusion

Across 3 experiments we showed that nonselective retrieval practice after study can reduce items' intralist interference and thus attenuate the list-length effect, output interference, and retrieval-induced forgetting. Indeed, whereas all three forms of forgetting were present after restudy cycles, they were all absent after nonselective retrieval practice. Theoretically, the results are consistent with a combination of contextual variability theory and a variant of study-phase retrieval theory that assumes that retrieval can create more distinct context features for retrieved items than restudy does for restudied items, thus reducing items' susceptibility to interference. More generally, the findings add to the view that retrieval

practice can stabilize and consolidate memories (Roediger & Butler, 2011).

## References

- Abel, M., & Bäuml, K.-H. T. (2014). The roles of delay and retroactive interference in retrieval-induced forgetting. *Memory & Cognition*, *42*, 141–150.
- Abel, M., & Bäuml, K.-H. T. (in press). Retrieval practice can eliminate list-method directed forgetting. *Memory & Cognition*.
- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445.
- 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.
- Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68–100.
- Arnold, K. M., & McDermott, K. B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review*, *20*, 507–513.
- Bäuml, K.-H. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review*, *5*, 459–463.
- Bäuml, K.-H. T., & Dobler, I. M. (2015). The two faces of selective memory retrieval: Recall specificity of the detrimental but not the beneficial effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 246–253.
- Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect - it enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1568–1581.
- Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, *68*, 39–53.
- Bäuml, K.-H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science*, *21*, 793–795.
- Bäuml, K.-H. T., & Samenieh, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 488–494.
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*, 228–247.
- Bower, G. H. (1972). A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 93–137). New York, NY: Academic Press.
- Butler, A. C. (2010). Repeated testing produces improved transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563–1569.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1547–1552.
- Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1403–1414.
- Delaney, P. F., Verkoeijen, P. P., & Spigler, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, *53*, 63–147.
- 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.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145–154.
- Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371–377.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 801–812.
- 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.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445.
- Hunt, R. R., & Smith, R. E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition*, *24*, 217–225.
- 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.
- Jonker, T. R., Seli, P., & MacLeod, C. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review*, *4*, 852–872.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103–109.
- Karpicke, J. D., Lehman, M., & Aue, R. W. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation*: Vol. 61 (pp. 237–284). San Diego, CA: Elsevier Academic Press.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*, 85–97.
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*, 155–189.
- Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1787–1794.
- Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.
- Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie*, *1*, 1–300.
- Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*, 1383–1409.
- Murdock, B. B. (1962). The serial position curve of free recall. *Journal of Experimental Psychology*, *64*, 482–488.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287–297.
- 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.
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, *17*, 132–138.

- Potts, R., & Shanks, D. R. (2012). Can testing immunize against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1780–1785.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335–335.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.
- Raaijmakers, J. G. W., & Jakab, E. (2012). Retrieval-induced forgetting without competition: Testing the retrieval specificity assumption of the inhibition theory. *Memory & Cognition*, *40*, 19–27.
- Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition*, *43*, 619–633.
- Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, *92*, 365–372.
- Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, *12*, 644–657.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford, England: Elsevier.
- Roediger, H. L., III, & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, *36*, 213–242.
- Rundus, D. (1973). Negative effects of using list items as retrieval cues. *Journal of Verbal Learning and Verbal Behavior*, *12*, 43–50.
- Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601–1603.
- Siegel, L. L., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 755–764.
- Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 400–408.
- Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, *28*, 503–508.
- Storm, B. C., Angello, G., Buchli, D. R., Koppel, R. H., Little, J. L., & Nestojko, J. F. (2015). A review of retrieval-induced forgetting in the contexts of learning, eye-witness memory, social cognition, autobiographical memory, and creative cognition. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 141–194). Academic Press: Elsevier Inc.
- 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.
- Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*, 529–536.
- Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 89–104.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*, 49–60.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335.
- Verde, M. F. (2009). The list-strength effect in recall: Relative-strength competition and retrieval inhibition may both contribute to forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 205–220.
- Watkins, O. C., & Watkins, M. J. (1975). Build up of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology Human Learning and Memory*, *1*, 442–452.
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, *18*, 518–523.
- Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*, 995–1008.

Received February 10, 2015  
 Revision received June 10, 2015  
 Accepted June 15, 2015 ■