American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2020, Vol. 46, No. 7, 1372–1386
http://dx.doi.org/10.1037/xlm0000782

Selective Retrieval in Categorized Lists: Detrimental, Neutral, and Beneficial Effects on Nonretrieved Items

Karl-Heinz T. Bäuml and Lisa Wallner
Regensburg University

Numerous studies of retrieval-induced forgetting have shown that the selective retrieval of some studied items can impair recall of other nonretrieved items. Varying the lag between study and selective retrieval and using lists of unrelated items as study material, recent work replicated this detrimental effect when the lag between study and selective retrieval was short but reported a beneficial effect of selective retrieval when the lag was long. Here we report the results of 4 experiments in which we examined the influence of lag (3 min vs. 24 hr) for the effects of selective retrieval in categorized lists. When the selectively retrieved and the nonretrieved items shared the same categories (Experiments 1 and 2), we found detrimental effects of selective retrieval regardless of lag between study and selective retrieval. In contrast, when the selectively retrieved and the nonretrieved items belonged to different categories (Experiments 3 and 4), the effects of selective retrieval varied with lag, showing a neutral effect after the short lag and a beneficial effect after the longer lag. A 2-factor explanation is provided that assumes critical roles in selective retrieval of (a) inhibition and blocking and (b) context retrieval. This account captures the present findings as well as the recent results on the effects of selective retrieval with lists of unrelated items.

Keywords: episodic memory, selective retrieval, retrieval-induced forgetting, context, context reinstatement

Since Anderson, Bjork, and Bjork's (1994) seminal work on retrieval-induced forgetting, numerous studies in the literature have shown that selective retrieval of some studied items can impair recall of nonretrieved items. This recall impairment is a very general effect and has been demonstrated over a wide range of materials and experimental settings (for recent reviews, see Bäuml & Kliegl, 2017; Storm et al., 2015). Regarding the possible roles of delay between study and selective retrieval on the one hand and between selective retrieval and test on the other hand, this literature includes a number of studies in which the influence of retention interval—the delay between selective retrieval and test—on the effects of selective retrieval has been examined (Abel & Bäuml, 2014; Chan, 2009; MacLeod & Macrae, 2001; Storm, Bjork, & Bjork, 2012). But the possible role of lag—the delay between study and selective retrieval—on the effects of selective retrieval has mostly been ignored. Indeed, in nearly all studies in the literature, the lag between study and selective retrieval was relatively short or absent altogether. Prolonged lag between study and selective retrieval, however, can change context and thus reduce the contextual overlap between study and selective retrieval (Bower, 1972; Estes, 1955). Different cognitive mechanisms might therefore operate after short versus longer lag, inducing different effects of selective retrieval on nonretrieved items.

Indeed, a few recent studies have examined the effects of selective retrieval when lag between study and selective retrieval was prolonged, using lags of up to 48 hr (for an overview, see Bäuml, 2019). These studies found selective retrieval to improve, rather than to impair, recall of nonretrieved items, suggesting that lag can have a drastic influence on the effects of selective retrieval. To date, these beneficial effects of selective retrieval have mostly been reported for lists of unrelated items, with hardly any investigation of whether the effects may arise for other types of materials as well. This is a critical shortcoming because most work on the effects of selective retrieval in the literature has been conducted with categorized lists of items and, as will be explained below, it is unclear whether the results with lists of unrelated items will generalize to categorized lists. This holds even more because, with categorized lists, effects of selective retrieval can be examined in two very different ways: when the selectively retrieved and the nonretrieved items share the same category (*within-category retrieval practice*), and when the selectively retrieved and the nonretrieved items belong to different categories (*between-categories retrieval practice*).

This article was published Online First November 7, 2019.

Karl-Heinz T. Bäuml and Lisa Wallner, Department of Experimental Psychology, Regensburg University.

All item materials that were applied in the present experiments as well as all data are available on the Open Science Framework (https://osf.io/s8hfd/?view_only=f6a6e6ac04d5436889b3e11892cae0d7). All experiments reported in this article were implemented using the software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, Pennsylvania). The software was run on standard desktop computers with the operating system Windows 7 (Microsoft, Redmond, Washington). All data were analyzed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, New York).

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

It is the first goal of the present study to fill this empirical gap and to examine the role of lag for the effects of selective retrieval with categorized material both when selective retrieval is conducted within category and when it is conducted between categories. Regarding theoretical accounts of the effects of selective retrieval, explanations of the detrimental effect of selective retrieval often have highlighted the role of inhibition and blocking (Anderson, 2003; Raaijmakers & Jakab, 2012) and explanations of the beneficial effect have highlighted the role of context retrieval (Bäuml, 2019). It is the second goal of the present study to improve our understanding of the circumstances under which the two types of cognitive mechanisms contribute to the effects of selective retrieval, thus providing an explanation for why under certain circumstances selective retrieval seems to be detrimental and under other circumstances it seems to be beneficial for nonretrieved items.

Detrimental and Beneficial Effects of Selective Retrieval

The evidence that selective retrieval of some studied items can impair recall of nonretrieved items has mostly arisen from studies that used the retrieval-practice task (for a review of the evidence arising from the output-interference task, see Roediger & Neely, 1982). In this task, subjects study a list of items, then repeatedly retrieve a subset of the items, and later at test are asked to recall all studied items. The typical finding in this task has been that, relative to an appropriate control condition, selective retrieval improved recall of the retrieved items but impaired recall of the nonretrieved items; this recall impairment reflects the effect of retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995; for a recent meta-analysis, see Murayama, Miyatsu, Buchli, & Storm, 2014). Retrieval-induced forgetting is usually attributed to inhibition or blocking processes. According to the inhibition account, the nonretrieved items interfere during selective retrieval and are inhibited to reduce the interference (Anderson, 2003). According to the blocking account, selective retrieval strengthens the memory representation of the retrieved items, which then at test block recall of the weaker nonretrieved items (Raaijmakers & Jakab, 2012). Inhibition and blocking have also been suggested to operate in concert to create this detrimental effect of selective retrieval (see Anderson & Levy, 2007; Bäuml, 2008; Rupprecht & Bäuml, 2016, 2017; Schilling, Storm, & Anderson, 2014; for another, context-based account of retrieval-induced forgetting, see Jonker, Seli, & MacLeod, 2013).

Retrieval-induced forgetting has been reported over a wide range of materials, including visual and verbal stimuli, lists of unrelated items and of categorized items, and also more complex study materials. Similarly, it has been reported across a wide range of experimental settings (see Bäuml & Kliegl, 2017; Storm et al., 2015). In nearly all of these studies, however, the lag between study and selective retrieval was fairly short. Indeed, either selective retrieval immediately followed the study phase of the experiment (e.g., Anderson et al., 1994; Jonker et al., 2013; Román, Soriano, Gómez-Ariza, & Bajo, 2009), or the lag between study and selective retrieval was on the order of only few minutes, mostly filled with simple counting or calculation tasks to minimize the possible contribution of short-term memory (STM) during selective retrieval (e.g., Bäuml, 2002; Cuc, Koppel, & Hirst, 2007; Hicks & Starns, 2004). More recent work now indicates that lag between study and selective retrieval can influence the effects of selective retrieval. Using lists of unrelated items as study material, Bäuml and colleagues (Abel & Bäuml, 2015; Aslan, Schlichting, John, & Bäuml, 2015; Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014) demonstrated that selective retrieval can be both detrimental and beneficial for the recall of the nonretrieved items, depending on lag. The results showed that when lag between study and selective retrieval was relatively short (between 60 s and 4 min), selective retrieval impaired recall of the nonretrieved items, whereas when lag was longer (between 30 min and 48 hr), selective retrieval improved recall performance.

Lag between study and retrieval influences the contextual overlap between the two experimental phases. Indeed, when we encode material, we also store information about the temporal context in which the material is encountered (Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981). But temporal context—the current pattern of activity in an individual's mind that, among others, can be influenced by environmental as well as internal factors—changes gradually over time (Bower, 1972; Estes, 1955), so that the context at retrieval often differs more from the context at study when lag is long, which can impair recall of target information.

There are a number of ways in which contextual mismatches between study and retrieval can be overcome or can at least be reduced. For instance, the mismatch can be reduced by reexposing critical context features that were present during study, like video-recorded scenes of real environments (Smith, 1985; Smith & Manzano, 2010), or by deliberate mental context reinstatement, when subjects are asked to mentally reinstate the original study environment (Jonker et al., 2013; Sahakyan & Kelley, 2002). But retrieval can also help itself when the contextual overlap between study and retrieval is reduced. Indeed, a retrieved item can reactivate its study context, which then serves as a retrieval cue for the recall of other studied items, thus improving recall performance (Polyn & Kahana, 2008). Evidence for such context retrieval has arisen from a wide range of recall findings, including the recency effect (i.e., the enhanced recall of end-of-list items in immediate free recall) and the contiguity effect (i.e., the tendency to successively recall neighboring list items; e.g., Howard & Kahana, 1999; Kahana, 1996; see also Polyn & Kahana, 2008).

Suggesting context retrieval as the cognitive mechanism underlying the beneficial effect of selective retrieval, Bäuml and colleagues explained the pattern of detrimental and beneficial effects of selective retrieval by means of a two-factor account of selective retrieval (Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014). This account assumes that selective retrieval generally triggers two types of processes: (a) inhibition and blocking and (b) context retrieval. The relative contributions of the two types of processes are hypothesized to depend on the contextual overlap between study and selective retrieval. When the contextual overlap is high—as may occur after a short lag between study and selective retrieval—interitem interference is often high and mainly inhibition and blocking operate; there is not much need for context retrieval. When the contextual overlap is low—as may occur after a longer lag—mainly context retrieval operates, whereas inhibition and blocking may be reduced due to attenuated interitem interference; indeed, increased lag between study and retrieval has been found to reduce the size of people's mental search set during retrieval (Kliegl, Carls, & Bäuml, in press), a finding which points to reduced interitem interference with increasing lag (e.g., Rohrer, 1996). The account then assumes that the differences in relative contributions of the two types of processes create the pattern of detrimental and beneficial effects of selective memory retrieval—a detrimental effect after short lag and a beneficial effect after prolonged lag.

Selective Retrieval in Categorized Lists

To date the pattern of detrimental and beneficial effects of selective retrieval has mainly been shown for lists of unrelated items, whereas most work on the effects of selective retrieval has been conducted with categorized lists and has emphasized detrimental effects. In the experiments with categorized lists, subjects studied category-exemplar pairs (e.g., spice–*vanilla*, spice–*basil*, instrument–*cello*, instrument–*tuba*) and were then asked to selectively retrieve a subset of the studied items from a subset of the studied categories when provided with the exemplars' category labels and unique word stems as retrieval cues (e.g., spice–*ba*___). At test, the category labels were reexposed and subjects were then asked to recall the studied items when provided with the items' unique initial letters as additional retrieval cues (e.g., spice–*v*____, instrument–*t*____). The focus in nearly all of these experiments was on the effects of selective retrieval on the recall of the practiced categories' nonretrieved items, which is referred to as *within-category retrieval practice* in the following. The critical finding has been that selective retrieval impaired recall of the practiced categories' nonretrieved items (spice–*vanilla*), relative to recall of the (control) items from the unpracticed categories (instrument–*tuba*), thus demonstrating retrieval-induced forgetting.

There is only one study in the literature which also examined the effects of selective retrieval on the recall of the nonretrieved items from the unpracticed categories, which is referred to as *between-categories retrieval practice* in the following. In this study, Shaw, Bjork, and Handal (1995) investigated the potential impact of repeated questioning of a witness, showing subjects slides depicting the aftermath of a theft. Subjects were then asked to retrieve selected details of what they saw. Results showed that information that bore a categorical similarity to the selectively retrieved items—which in a simple categorized list could be *vanilla* if *basil* was selectively retrieved—became less recallable at test, which replicates the presence of retrieval-induced forgetting in a within-category retrieval practice situation; in contrast, recall of information that did not bear such a similarity—which in a simple categorized list could be *tuba* if *basil* was selectively retrieved—was not affected, relative to a baseline condition in which subjects did not engage in any retrieval practice at all. The absence of the forgetting effect when the categorical similarity between selectively retrieved and nonretrieved items is low is consistent with the assumption of a generally low level of interference between items from different categories (e.g., Rundus, 1973). Because of this low interference level, the effects of inhibition and blocking should be small and may not be sufficient to induce forgetting in between-categories retrieval practice situations.

Nearly all of the findings on the effects of selective retrieval with categorized lists—a detrimental effect when retrieval practice was conducted within category and a neutral effect when it was conducted between categories—emerged under conditions in which the lag between study and selective retrieval was relatively short or absent altogether (see above). The question therefore arises whether a longer lag will influence the effects of selective retrieval and, if so, whether it will induce the same beneficial effects after long lag as have recently been reported for lists of unrelated items.

Possible Expectations Regarding the Role of Lag in Categorized Lists

Following the results on the effects of lag between study and selective retrieval with unrelated lists and given the two-factor account of these effects, a first expectation regarding the effects of lag in categorized lists might be that the effects observed with unrelated lists directly generalize to categorized lists. Critically, also with categorized lists, prolonged lag between study and selective retrieval should reduce the contextual overlap between the two experimental phases and context retrieval might then operate during selective retrieval, reinstating the study context. Regarding the within-category retrieval practice effect, this would mean that, when the lag between study and selective retrieval is short, the relative contributions of inhibition and blocking should be larger than that of context retrieval, inducing a detrimental effect of within-category retrieval practice. In contrast, when selective retrieval is lagged, the relative contribution of context retrieval should be larger than that of inhibition and blocking, inducing a beneficial effect of within-category retrieval practice. Regarding the between-categories retrieval practice effect, the roles of inhibition and blocking should generally be small given the low level of interference between items from different categories, inducing a neutral effect of between-categories retrieval practice when the lag between study and selective retrieval is short. When the lag is long, however, between-categories retrieval practice may also trigger context retrieval and thus induce a beneficial effect of selective retrieval (see Table 1). Thus, with regard to both within-category and between-categories retrieval practice, selective retrieval might improve recall performance when selective retrieval is lagged, which would parallel the results with lists of unrelated items.

A second, different expectation regarding possible effects of lagged selective retrieval in categorized lists takes into account that reexposure of category labels during selective retrieval may already reinstate study context. Because in a typical experiment on selective retrieval, the category labels that are presented as retrieval cues during selective retrieval are associated to one experimental context—the study context—only, reexposure of the category labels at retrieval may more or less routinely reactivate study context (Jonker et al., 2013, p. 855). If so, even when selective retrieval is lagged, the contextual overlap between study and selective retrieval may be relatively high with categorized lists, which would reduce the need for (further) context retrieval.

Inclusion of the assumption that category labels reinstate study context into the two-factor explanation of the effects of selective retrieval changes expectations for lagged retrieval. Regarding within-category retrieval practice, the context reinstatement induced by the category labels should lead to a revival of interitem interference and a minor need for context retrieval as induced by selective retrieval. The relative contributions of inhibition and blocking on the one hand and of context retrieval on the other should thus be relatively close to those for unlagged selective

Table 1
Overview of Expectations Regarding the Role of Lag for the Effects of Selective Retrieval in Categorized Lists

Type of retrieval practice	Short lag	Long lag when category labels do not reinstate context	Long lag when category labels reinstate context
WC-RP			
Mechanisms	High contributions of inhibition and blocking, lower contribution of context retrieval	High contribution of context retrieval, lower contributions of inhibition and blocking	High contributions of inhibition and blocking, lower contribution of context retrieval
Effect	Detrimental effect	Beneficial effect	Detrimental effect
BC-RP			
Mechanisms	Low contributions of inhibition, blocking, and context retrieval	High contribution of context retrieval, low contributions of inhibition and blocking	Moderate contribution of context retrieval, low contributions of inhibition and blocking
Effect	Neutral effect	Beneficial effect	(Small) beneficial effect

Note. WC-RP = within-category retrieval practice; BC-RP = between-categories retrieval practice.

retrieval and lead to a detrimental—rather than a beneficial—effect of selective retrieval. Regarding between-categories retrieval practice, the context reinstatement induced by the category labels should also limit context retrieval. However, because techniques to reinstate study context when the contextual overlap has been reduced seldom lead to perfect context reinstatement (see Sahakyan & Kelley, 2002; Smith & Manzano, 2010), the context reinstatement induced by the category labels may also be incomplete and may leave room for context retrieval. Given that the contributions of inhibition and blocking to recall should be low in this type of situation (see above), this context retrieval may be sufficient to induce a (small) beneficial effect of selective retrieval (see Table 1). Following this rationale, within-category retrieval practice may thus induce a detrimental effect and between-categories retrieval practice a (small) beneficial effect on nonretrieved items when selective retrieval is lagged.

The Present Study

In a series of four experiments, this study examined the influence of lag between study and selective retrieval on the effects of selective retrieval in categorized lists, both when retrieval practice was conducted within categories (i.e., when the selectively retrieved and the nonretrieved items belonged to the same categories) and when it was conducted between categories (i.e., when the nonretrieved items belonged to other categories than the practiced items). Experiments 1 and 2 examined the effects of selective retrieval for within-category retrieval practice. In each of the two experiments, subjects studied category-exemplar pairs (e.g., spice–*basil*, spice–*vanilla*, instrument–*cello*, instrument–*tuba*) and then, after a short lag of 3 min or a longer lag of 24 hr, either repeatedly retrieved half of the items of each single category (e.g., spice–*ba*___, instrument–*tu*___; *retrieval practice present*), or retrieved no items at all (*retrieval practice absent*). At test, subjects in both conditions were asked to recall all studied items, beginning with those items that were not subject to retrieval practice in the intermediate phase (e.g., spice–*v*___, instrument–*c*___); these items are referred to as the target items in the following (see Figure 1A).[1]

On the basis of prior work on the effects of selective retrieval, we expected the typical detrimental effect of selective retrieval for the target items when the lag between study and selective retrieval was short. Expectations when the lag was long depended on the suggested role of the category labels for context reinstatement. If the category labels did not induce much context reinstatement, then, according to the two-factor account, selective retrieval should later trigger context retrieval and induce a beneficial effect on the target items. In contrast, if the category labels reinstated study context, then selective retrieval should not show such beneficial effect and, due to inhibition and blocking, might even impair recall of the target items, thus replicating the detrimental effect that is expected when the lag between study and selective retrieval is short.

Experiments 3 and 4 examined the effects of selective retrieval for between-categories retrieval practice. In each of the two experiments, subjects studied category-exemplar pairs (e.g., spice–*basil*, spice–*vanilla*, instrument–*cello*, instrument–*tuba*) and then, after a short lag of 3 min or a longer lag of 24 hr, repeatedly retrieved all items from half of the categories (e.g., spice–*ba*___, spice–*van*___; *retrieval practice present*), or retrieved no items at all (*retrieval practice absent*). At test, subjects in both conditions were asked to recall all studied items, again beginning with those (target) items that were not subject to retrieval practice in the intermediate phase (e.g., instrument–*t*___; instrument–*c*___; see Figure 1B).

Following Shaw et al. (1995), we expected to find a neutral effect of selective retrieval when the lag between study and selective retrieval was short. In contrast, we expected a beneficial effect of selective retrieval when the lag was long. Expectations on the possible size of the beneficial effect depended on whether category labels already reinstated context. If the category labels induced no such reinstatement, the beneficial effect should be roughly comparable to the beneficial effect observed with lists of unrelated items and be of medium or even large size (e.g., Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). In contrast, if the category labels did induce context reinstatement, there should be a reduced need for further context retrieval and the beneficial effect of selective retrieval should be of small size only. As a whole, the

[1] In prior work on within-category retrieval practice, usually some of the items of some of the categories were selectively retrieved, and the items from the remaining unpracticed categories served as a control (e. g., Anderson et al., 1994). In the present experiments, a separate no-retrieval-practice condition was used as a control. With this choice, it can be excluded that recall of the control items can be influenced by the prior retrieval of items from the practiced categories (see also below).

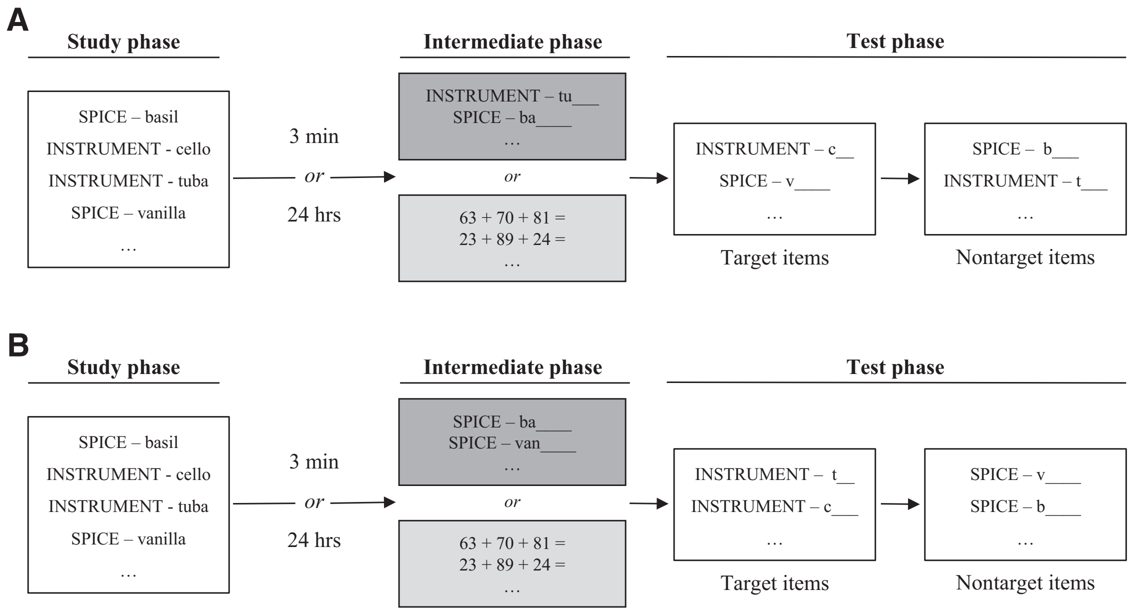

Figure 1. A: Procedure and conditions employed in Experiments 1 and 2 (within-category retrieval practice). Participants studied a list of category-exemplar pairs. After a lag of 3 min or 24 hr, half of the participants engaged in two retrieval practice cycles of half of the exemplars from each single category. The other half of the participants solved arithmetic problems as a control. At test, participants were asked to recall first the unpracticed (target) items and then the practiced (nontarget) items. B: Procedure and conditions employed in Experiments 3 and 4 (between-categories retrieval practice). Participants studied a list of category-exemplar pairs. After a lag of 3 min or 24 hr, half of the participants engaged in two retrieval practice cycles of all exemplars from half of the studied categories. The other half of the participants solved arithmetic problems as a control. At test, participants were asked to recall first the unpracticed (target) items and then the practiced (nontarget) items.

results of the four experiments will provide important information on the role of lag between study and selective retrieval for the effects of selective retrieval in categorized lists, with regard to both within-category and between-categories retrieval practice. This information will improve our understanding of the circumstances under which selective retrieval is detrimental and under which selective retrieval is beneficial for other memories.

Experiment 1

Experiment 1 examined the effects of lag for within-category retrieval practice. Participants studied a list of 20 category-exemplar pairs, with 10 items from each of two semantic categories. After a short lag of 3 min or a prolonged lag of 24 hr, half of the participants selectively retrieved half of the studied exemplars of each of the two categories (*retrieval practice present*)—these practiced items are referred to as the nontarget items in the following—whereas the other half of the participants solved unrelated distractor tasks for the same duration of time (*retrieval practice absent*). At test, subjects were cued to recall all studied items, with the (unpracticed) target items being recalled first and the (practiced) nontarget items being recalled second (see again Figure 1A).

To our knowledge, there is only one experiment in the literature in which the effects of lag between study and selective retrieval for within-category retrieval practice effects were examined to date. Using an impression formation task, MacLeod and Macrae (2001; Experiment 2) showed participants personality characteristics of two men before participants were asked to retrieve half of the personality traits of one of the two characters, either immediately after study or after a lag of 24 hr. After retrieval practice, a final cued-recall test took place, in which participants tried to recall all the personality traits that they had previously seen. The results showed the expected forgetting effect when lag between study and selective retrieval was short; that is, recall of the unpracticed traits from the practiced person was impaired relative to recall of the unpracticed traits from the unpracticed person. When lag was long, the forgetting was numerically reduced—6% versus 14% with unlagged practice—but remained significant, indicating that the sheer presence of the forgetting may not depend on lag.

MacLeod and Macrae (2001) followed prior studies on retrieval-induced forgetting and used recall of the personality traits from the unpracticed character as a control to measure the effects of within-category retrieval practice. This procedure appears appropriate in the condition when lag between study and selective retrieval was short (see Shaw et al., 1995) but may no longer be appropriate when lag was long. In such case, recall of the control traits might have benefitted from context retrieval induced by the prior selective retrieval of the traits of the practiced character. If so, recall levels of the control traits might have been overestimated and thus have indicated a detrimental effect that would no longer be present when a no-retrieval-practice control condition was employed. To avoid this potential problem, the experiments in the present study used such a no-retrieval-practice control condition.

Method

Participants. Eighty students of Regensburg University participated in the experiment (M = 20.74 years, range = 18–30 years, 83.8% female). They were equally distributed across the two between-subjects conditions, resulting in n = 40 participants in each of the two conditions. We determined the desired sample size based on counterbalancing purposes and the results of an analysis of test power conducted with the G*Power program (Version 3, Faul, Erdfelder, Lang, & Buchner, 2007). For this analysis, we set $\alpha = .05$ and $\beta = .20$ and assumed a medium effect size of retrieval practice (d = .40; Murayama et al., 2014). All subjects spoke German as native language and received monetary reward or course credit for their participation.

Materials. Four study lists were constructed (A, B, C, D). Half of the participants studied lists A and B, the other half studied lists C and D. Each of the four lists consisted of 20 exemplars with 10 items from each of two semantic categories (list A: spices, instruments; list B: vegetables, professions; list C: quadrupeds, metals; list D: car components, trees). Among all items within a category, each item had a unique initial letter. The items were drawn from German word norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995) or were translated from English word norms (Van Overschelde, Rawson, & Dunlosky, 2004). Regarding the words' frequency in the norms, the four items with the highest frequencies in each category were excluded to reduce guessing.

Design. The experiment had a 2 × 2 design with the within-subject factor of lag (3 min, 24 hr) and the between-subjects factor of retrieval practice (present, absent). For each participant, the experiment consisted of two conditions that were identical apart from materials and lag between study and retrieval practice. Order of lags and lists were counterbalanced across participants. After study and the subsequent lag, half of the participants were asked to retrieve half of the items of each category (retrieval practice present), whereas the other half of the participants performed an unrelated distractor task for the same period of time (retrieval practice absent). Assignment of conditions was counterbalanced as well as which items of the categories were subject to retrieval practice.

Procedure. Each participant completed both lag conditions, with each condition consisting of three main phases: an initial study phase, an intermediate phase, and a final test phase. The two conditions were conducted on two successive days. Half of the participants started with the short lag condition. These participants conducted all three phases of the short lag condition on Day 1, and, after a break, also took place in the study phase of the long lag condition. The intermediate and test phases of this condition were completed on Day 2. The other half of the participants started with the long lag condition. These participants took part in the study phase of this condition on Day 1 and completed the intermediate and test phases of this condition on Day 2. After a break, the participants conducted all three phases of the short lag condition the same day.

In the study phase, participants studied category-exemplar pairs at a 5-s rate (interstimulus interval [ISI] = 500 ms) displayed on a computer screen. The order of the word pairs in the list was blocked randomized, so that no more than two word pairs of the same category were presented in succession. Doing so, we compiled 10 item blocks with each item block comprising one exemplar from each of the two categories of the list. Order of item blocks as well as order of word pairs within the item blocks were random. The study phase was followed by a lag of 3 min, in which participants counted backward in steps of three from a three-digit number in the one lag condition, or participants were disbanded and asked to come back after 24 hr, that is, same time next day, in the other lag condition.

In the subsequent intermediate phase, half of the participants solved simple math tasks (addition of three two-digit numbers) for 2 min, whereas the other half of the participants retrieved half of the exemplars (five items) from each of the two categories: the category label and the word stem of an exemplar were presented as retrieval cues for 5 s (ISI = 500 ms) and participants were instructed to recall the matching exemplar from the study list orally while the experimenter logged the data. Again, presentation of the retrieval cues was blocked randomized. The 10 exemplars were practiced twice in consecutive cycles. No feedback was provided. Before the final test, participants in both conditions counted backward in steps of three from a three-digit number for 3 min.

At test, all participants engaged in a cued-recall test, in which they were asked to recall all previously studied items. Consistent with prior work on the effects of selective retrieval with categorized lists, recall was blocked by category. The studied items were cued with the category label and the initial letter of the exemplar and participants were asked to orally respond with the corresponding item within 5 s (ISI = 500 ms). The experimenter recorded the answers. In each condition, the unpracticed (target) items were tested first and the practiced (nontarget) items second. Order of categories and output order of target and nontarget items within a category were random.

Results

Success rates during retrieval practice were high, with higher recall after the short lag (92.9%) than the long lag (87.5%), $t(39) = 2.22$, $p = .033$, $d = 0.53$.

Figure 2A shows mean recall rates for the target items in each single condition. A 2 × 2 analysis of variance (ANOVA) with the between-subjects factor of retrieval practice (present, absent) and the within-subjects factor of lag (3 min, 24 hr) showed significant main effects of retrieval practice (59.4% vs. 68.3%), $F(1, 78) = 10.89$, $MSE = 289.41$, $p = .001$, $\eta^2 = 0.12$, and lag (68.3% vs. 59.5%), $F(1, 78) = 9.50$, $MSE = 313.37$, $p = .003$, $\eta^2 = 0.11$, indicating reduced recall in the presence of retrieval practice and reduced recall after longer lag. There was no interaction between the two factors, $F(1, 78) < 1$. Consistently, planned comparisons showed that there was a detrimental effect after both short lag (73.0% vs. 63.3%), $t(78) = 2.50$, $p = .014$, $d = 0.68$, and long lag (63.5% vs. 55.5%), $t(78) = 2.07$, $p = .042$, $d = 0.45$. A comparison of target recall rates in the condition without retrieval practice between the short and the long lag conditions showed significant time-dependent forgetting (73.0% vs. 63.5%), $t(39) = 2.59$, $p = .013$, $d = 0.60$.

Table 2 shows mean recall rates for the nontarget items. A 2 × 2 ANOVA with the between-subjects factor of retrieval practice

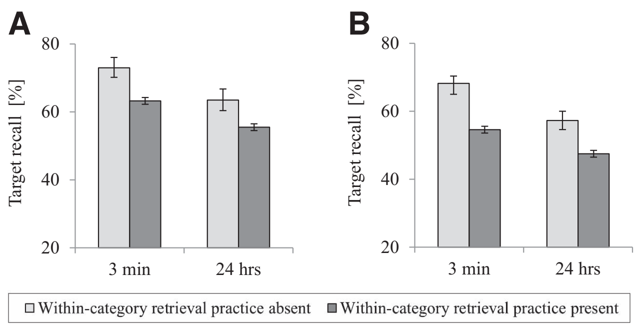

Figure 2. Results of Experiment 1(A) and Experiment 2(B), which examined the effects of within-category retrieval practice. Percentage of recalled target items is shown as a function of lag between study and selective retrieval (3 min, 24 hr) and retrieval practice (within-category retrieval practice absent, within-category retrieval practice present). Error bars represent standard errors.

(present, absent) and the within-subject factor of lag (3 min, 24 hr) revealed a significant main effect of retrieval practice (82.2% vs. 58.5%), $F(1, 78) = 78.24$, $MSE = 279.34$, $p < .001$, $\eta^2 = 0.50$, affirming successful retrieval practice, and a marginally significant main effect of lag (72.8% vs. 67.6%), $F(1, 78) = 3.57$, $MSE = 294.22$, $p = .063$, $\eta^2 = 0.04$, indicating a trend toward time-dependent forgetting. The interaction between the two factors was not significant, $F(1, 78) < 1$.

Half of the participants in this experiment completed the short lag condition first, whereas the other half completed the long lag condition first. Lag order influenced recall performance, as is reflected by the fact that target recall was higher when subjects completed the short lag condition first (67.0% vs. 60.6%), $F(1, 76) = 5.94$, $MSE = 273.78$, $p = .017$, $\eta^2 = 0.07$. Critically, however, there was no interaction of lag order with any of the other factors, all $Fs(1,76) < 1$. In addition, lists were varied and counterbalanced across participants in this experiment. Yet, material did not influence recall performance: There was no main effect of material, and no interaction of material with any of the other factors, all $Fs(1,76) < 1$.

Discussion

The results replicate prior work on retrieval-induced forgetting by showing that, when the lag between study and selective retrieval is short, selective retrieval of some category exemplars can impair recall of the same categories' nonretrieved items (Anderson et al., 1994). More important, the results show that the same effect also arises when lag is prolonged, which indicates that lag may not influence the effect of selective retrieval with categorized lists. This finding is inconsistent with an interpretation of the two-factor account of selective retrieval that associates a short lag between study and selective retrieval with a high contextual overlap between the two experimental phases, and a long lag with a low contextual overlap. Indeed, with such interpretation, within-category retrieval practice should induce a detrimental effect after short lag but a beneficial effect after long lag, which is not what the results show. Rather, the results are consistent with the account if the assumption is included that reexposure of the category labels at retrieval reinstates study context (Jonker et al., 2013). In such case, the long-lag situation should be roughly comparable to the short-lag situation and a relatively high contextual overlap between study and selective retrieval should be present in both lag conditions. Following the two-factor account, the relative contributions of inhibition and blocking should thus be high in both lag conditions and induce a detrimental effect, which is what the present results show.

The results of the experiment parallel those reported in MacLeod and Macrae (2001; Experiment 2). Using an impression formation task, these researchers found that retrieval practice on some previously presented personality traits of one of two characters can impair later recall of that character's other traits, both when the lag between presentation of the traits and selective retrieval was short and when it was long. MacLeod and Macrae used recall of the personality traits from the second, unpracticed character as a control to measure the effects of within-category retrieval practice, whereas in the present experiment a no-retrieval-practice control condition was used. The no-retrieval-practice control condition is principally the better choice to examine retrieval practice effects after longer lag because, with this control, recall of the control items cannot be influenced by the prior retrieval of the items from the practiced category and thus cannot bias estimates of the effects of selective retrieval (see above). The similarity of results between the two studies, however, indicates that, at least with the settings used, findings do not vary much with choice of control condition.

Experiment 2

There were two goals with Experiment 2. One goal was to replicate the findings of Experiment 1 with a list structure that is more typical for prior work on the effects of selective retrieval. Indeed, in the prior work, often study lists were used that consisted of a larger number of categories but a smaller number of items within categories. Accordingly, in Experiment 2, we used lists consisting of six exemplars from each of four different categories. The second goal was to examine the effects of selective retrieval when the delay between the intermediate and the test phases was eliminated. In Experiment 1, we had used a delay interval of 3 min between the two experimental phases. Prior work, however, suggests that context reactivation processes may contribute more

Table 2
Mean Nontarget Recall (Plus Standard Deviations) in Experiments 1 and 2 as a Function of Lag Between Study and Selective Retrieval (3 min, 24 hr) and Retrieval Practice (Within-Category Retrieval Practice Absent, Within-Category Retrieval Practice Present)

	3-min lag		24-hr lag	
Experiment	*M*	*SD*	*M*	*SD*
Experiment 1				
WC-RP absent	60.5	16.3	56.5	20.8
WC-RP present	85.5	14.5	78.8	15.4
Experiment 2				
WC-RP absent	63.7	19.0	50.2	18.5
WC-RP present	86.0	13.9	83.3	13.3

Note. WC-RP = within-category retrieval practice.

easily to recall performance when the test immediately follows selective retrieval (Dobler & Bäuml, 2012; Polyn, Norman, & Kahana, 2009), which raises the possibility that the results for longer lag found in Experiment 1 may not generalize to undelayed testing.

Method

Participants. Another 80 students of Regensburg University participated in the experiment (M = 22.91 years, range = 18–29 years, 72.5% female). Again, all subjects spoke German as native language and received monetary reward or course credit for their participation.

Materials. Two study lists were constructed (A, B). Both lists consisted of 24 exemplars with six exemplars from each of four different semantic categories (list A: body parts, sports, professions, birds; list B: fruits, toys, flowers, furniture). Again, the items were drawn from several different word norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995; Van Overschelde et al., 2004); each item within a category had a unique initial letter; the four items with the highest frequencies in each category were excluded to reduce guessing.

Design and procedure. The design was identical to Experiment 1 and the procedure followed the one used in Experiment 1, differing only in the delay between the intermediate and test phases. In the intermediate phase, half of the participants retrieved half of the exemplars (three items) from each of the 4 categories (a total of 12 items), whereas the other half of the participants performed an unrelated distractor task. At test, in each condition, the 12 unpracticed target items were tested first and the 12 practiced nontarget items were tested second. For both types of items, recall was blocked by category. There was no delay between the intermediate and test phases.

Results

Success rates during retrieval practice were high, but they were higher after short than long lag (92.3% vs. 87.3%), $t(39) = 2.28$, $p = .028$, $d = 0.46$.

Figure 2B shows mean recall rates for the target items in each condition. A 2 × 2 ANOVA with the between-subjects factor of retrieval practice (present, absent) and the within-subject factor of lag (3 min, 24 hr) showed significant main effects of retrieval practice (51.1% vs. 62.8%), $F(1, 78) = 15.66$, $MSE = 350.40$, $p < .001$, $\eta^2 = 0.17$, and lag (61.4% vs. 52.4%), $F(1, 78) = 13.48$, $MSE = 240.60$, $p = .001$, $\eta^2 = 0.15$, suggesting reductions in recall in the presence of retrieval practice as well as after longer lag. There was no interaction between the two factors, $F(1, 78) < 1$. Planned comparisons confirm these results by showing that there was retrieval-induced forgetting after both the short lag (54.6% vs. 68.2%), $t(78) = 3.52$, $p = .001$, $d = 0.79$, and the long lag (47.5% vs. 57.3%), $t(78) = 2.57$, $p = .012$, $d = 0.57$. In the condition without retrieval practice, a comparison of target recall rates between the short and the long lag conditions showed significant time-dependent forgetting (68.2% vs. 57.3%), $t(39) = 3.22$, $p = .003$, $d = 0.70$.

Table 2 shows mean recall rates for the nontarget items. A 2 × 2 ANOVA with the between-subjects factor of retrieval practice (present, absent) and the within-subject factor of lag (3 min, 24 hr) revealed a significant main effect of retrieval practice (84.7% vs. 57.0%), $F(1, 78) = 94.56$, $MSE = 324.79$, $p < .001$, $\eta^2 = 0.55$, affirming the beneficial effect of retrieval practice, a significant main effect of lag (74.9% vs. 66.8%), $F(1, 78) = 12.36$, $MSE = 213.73$, $p = .001$, $\eta^2 = 0.14$, reflecting time-dependent forgetting, and a significant interaction between the two factors, $F(1, 78) = 9.17$, $MSE = 213.73$, $p = .022$, $\eta^2 = 0.07$. Consistently, planned comparisons showed a larger beneficial effect after the long lag, $t(78) = 5.98$, $p < .001$, $d = 2.05$, than the short lag, $t(78) = 5.49$, $p < .001$, $d = 1.40$.

Each participant in this experiment completed both lag conditions in counterbalanced order. Lag order, however, did not influence recall performance: There was no main effect of lag order and no interaction of lag order with any of the other factors, all $Fs(1,76) < 1.23$, $MSEs > 350.749$, $ps > .270$, $\eta^2s < 0.02$. In addition, the assignment of lists to experimental blocks was counterbalanced across conditions. Yet material did not influence target recall: There was no main effect of material and no interaction of material with any of the other factors, all $Fs(1,76) < 3.26$, $MSEs < 338.63$, $ps > .075$, $\eta^2s < 0.04$.

Discussion

The results of Experiment 2 replicated those of Experiment 1, both in pattern and in size. Again, there was a detrimental effect of selective retrieval on the practiced categories' nonretrieved items, and again the presence of this effect did not depend on lag between study and retrieval practice. Together with the results of Experiment 1, these findings indicate that the effects with both short and long lag do not vary much with list structure and do not depend on delay between selective retrieval and test, at least for delay intervals of up to few minutes. Like the results of Experiment 1, the results of Experiment 2 are consistent with the two-factor account of selective retrieval if the assumption is included that the presence of the category labels during retrieval reinstated study context. In such case, the results for long lag should be similar to those for short lag and show a detrimental effect of selective retrieval in both lag conditions, which is what the present results show.

The detrimental effect of selective retrieval found in this experiment did not vary statistically with lag, although there was a slight numerical reduction with lag (13.6% after short lag, 9.8% after long lag). Exactly the same pattern arose in Experiment 1 and was also present in MacLeod and Macrae's (2001) study (see above). The slight numerical reduction of the forgetting effect with increasing lag might reflect the slightly different success rates during selective retrieval in the two lag conditions, for instance, with a higher success rate of 92.6% after short lag and a lower success rate of 87.4% after long lag in Experiment 2. One may argue therefore that if success rates had been equated perfectly between lag conditions, the size of the detrimental effect after long lag would have been numerically identical to that after short lag, or would even have exceeded it in size. However, prior work on retrieval-induced forgetting failed to find any dependence of the size of the detrimental effect on success rate during retrieval practice (see Murayama et al., 2014). Moreover, Storm, Bjork, Bjork, and Nestojko (2006) reported evidence that the attempt to retrieve, even if unsuccessful, produces the forgetting effect. These findings suggest that the size of the detrimental effects found in the present study would not have been much different if success rates had been equated, for instance, by providing stronger item-specific cues after long lag than after short lag. Rather, the numerical difference may indeed reflect the influence of lag.[2]

Experiment 3

Whereas Experiments 1 and 2 examined the effects of lag for within-category retrieval practice, that is, when the nonretrieved items shared the same categories as the practiced items, Experiments 3 and 4 examined the effects of lag for between-categories retrieval practice, that is, when the nonretrieved items belonged to other categories than the practiced items. Like in Experiment 1, participants in Experiment 3 studied a list of 20 category-exemplar pairs, with 10 items from each of two categories. Unlike in Experiment 1, half of the participants selectively retrieved all studied exemplars of one of the two categories (*retrieval practice present*)—these practiced items are again referred to as the nontarget items in the following—whereas the other half of the participants solved unrelated distractor tasks for the same duration of time (*retrieval practice absent*). Again, there was either a short lag of 3 min or a prolonged lag of 24 hr between study and selective retrieval. At test, subjects recalled the (unpracticed) target items, that is, the items from the unpracticed category, first and the (practiced) nontarget items second (see again Figure 1B). Like in Experiment 1, there was a delay of 3 min between selective retrieval and test.

Following Shaw et al. (1995), we expected to find a neutral effect of selective retrieval when the lag between study and selective retrieval was short. Because interference between items from different categories is typically low, the contributions of inhibition and blocking to recall should also be low and therefore should not induce a detrimental effect on target recall. In contrast, we expected a beneficial effect of selective retrieval when the lag between study and selective retrieval was long, assuming that, with longer lag, context retrieval will contribute to recall performance. This beneficial effect, however, was expected to be of small size only. Indeed, because context reinstatement has often been found to be incomplete (Sahakyan & Kelley, 2002; Smith & Manzano, 2010), the context reinstatement induced by category labels was also expected to be incomplete. Context retrieval should therefore contribute to recall performance, but the contribution should be limited and create only a small beneficial effect. As a whole, the effects of between-categories retrieval practice should thus differ from those of within-category retrieval practice, after both short and long lag between study and selective retrieval.

Method

Participants. One hundred twenty-eight students of Regensburg University participated in the experiment (M = 22.0 years, range = 18–30 years, 81.3% female). They were equally distributed across the two between-subjects conditions, resulting in n = 64 participants in each of the two conditions. Like in Experiment 1, we determined the desired sample size based on counterbalancing purposes and the results of an analysis of test power conducted with the G*Power program (Version 3; Faul et al., 2007). For this analysis, we set $\alpha = .05$ and $\beta = .20$ and assumed a relatively small effect size of selective retrieval (d = 0.30). Again, all participants spoke German as native language and received course credit or monetary reward for participation.

Materials and design. Materials and design were identical to Experiment 1.

Procedure. The procedure followed Experiment 1, with only one major deviation: Whereas in Experiment 1, half of the participants were asked to practice retrieval on half of the items of each of the two categories (a total of 10 items), in Experiment 3, all items of one of the two categories and no items from the other category were practiced (also a total of 10 items). Order of items within a category was random. At test, all subjects were tested on the unpracticed target items first and the practiced nontarget items second. Again, order of items within a category was random. In all other aspects, Experiment 3 was identical to Experiment 1.

Results

Like in Experiments 1 and 2, success rates during retrieval practice were high and, at least in tendency, were higher after short than long lag (90.0% vs. 85.7%), $t(63) = 1.78$, $p = .079$, $d = 0.30$.

Figure 3A depicts mean recall rates for the target items in each condition. A 2 × 2 ANOVA with the between-subjects factor of retrieval practice (present, absent) and the within-subject factor of lag (3 min, 24 hr) showed a significant main effect of lag (65.0% vs. 56.7%), $F(1, 126) = 13.36$, $MSE = 328.58$, $p < .001$, $\eta^2 = 0.10$, suggesting reduced recall after longer lag, but no main effect of retrieval practice (62.2% vs. 59.5%), $F(1, 126) < 1$. There was, however, a significant interaction between the two factors, $F(1, 126) = 4.00$, $MSE = 328.58$, $p = .048$, $\eta^2 = 0.03$, indicating that retrieval practice influenced target recall differently in the two lag conditions. Planned comparisons in fact showed that although there was no effect of retrieval practice after the short lag (65.1% vs. 65.9%), $t(126) < 1$, there was a significant beneficial effect after the long lag (60.3% vs. 53.1%), $t(126) = 2.03$, $p = .045$, $d = 0.36$. A comparison of target recall rates in the condition without retrieval practice between the short and the long lag conditions showed expected time-dependent forgetting (65.9% vs. 53.1%), $t(63) = 4.19$, $p < .001$, $d = 0.69$.

Mean recall rates for the nontarget items are depicted in Table 3. ANOVA revealed a significant main effect of retrieval practice (81.7% vs. 58.7%), $F(1, 126) = 87.33$, $MSE = 389.264$, $p < .001$, $\eta^2 = 0.41$, affirming the beneficial effect of practice, as well as a marginally significant main effect of lag (72.1% vs. 68.3%), $F(1, 126) = 2.75$, $MSE = 340.50$, $p = .099$, $\eta^2 = 0.02$, suggesting a trend toward time-dependent forgetting. The interaction between the two factors was not significant, $F(1, 126) < 1$.

Each participant ran through both lag conditions, but lag order did not affect target recall. Indeed, there was no main effect of lag order and no interaction of lag order with any of the other factors, all $Fs(1,124) < 1$. Similarly, there was no influence of material on target recall, all $Fs(1,122) < 1.22$, $MSEs < 459.72$, $ps > .298$, $\eta^2s < 0.02$.

[2] Both in the present study and in the prior work by MacLeod and Macrae (2001), the slight numerical difference in the size of the detrimental effect between lag conditions arose when recall levels of the control items were not taken into account. Alternatively, one may measure size of the forgetting effect in the two lag conditions using proportion decrease—that is, proportion of items forgotten relative to the control items' recall levels. As it turned out, however, the numerical differences between lag conditions remained more or less unchanged by that measure.

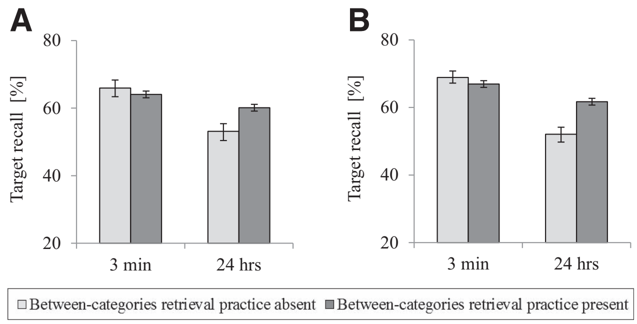

Figure 3. Results of Experiment 3(A) and Experiment 4(B), which examined the effects of between-categories retrieval practice. Percentage of recalled target items is shown as a function of lag between study and selective retrieval (3 min, 24 hr) and retrieval practice (between-categories retrieval practice absent, between-categories retrieval practice present). Error bars represent standard errors.

Discussion

The results replicated prior work on the effects of selective retrieval in between-categories retrieval practice situations by showing that, when the lag between study and selective retrieval was short, selective retrieval of the items of one category left recall of the items from the other category unaffected (Shaw et al., 1995). More important, the results showed that the same neutral effect was not evident when lag was increased. Rather, selective retrieval from one category improved recall of the items from the other category when lag was long. These findings are consistent with the two-factor account of selective retrieval. On the basis of this account, the neutral effect after short lag was expected because interference between items from different categories is generally low and inhibition and blocking should therefore contribute very little to recall performance. The beneficial effect after long lag was expected because, although the category labels per se should already induce some context reinstatement, this reinstatement should be incomplete (Sahakyan & Kelley, 2002; Smith & Manzano, 2010), thus leaving room for additional context retrieval as induced by selective retrieval. The results of the experiment agree with these expectations.

Experiment 4

Just like the goal of Experiment 2 was to replicate the results of Experiment 1 with a different list structure than was employed in Experiment 1 and an elimination of the delay between selective retrieval and test, it was the goal of Experiment 4 to replicate the results of Experiment 3 using the same types of modifications. Participants in Experiment 4 therefore studied a list of 24 items with six items from each of four different categories. In the intermediate phase, half of the participants then retrieved all studied exemplars from two of the four categories, whereas the other half solved unrelated distractor tasks for the same duration of time. Immediately after this phase, all study items were tested with the (target) items from the two unpracticed categories being tested first and the (nontarget) items from the two practiced categories being tested second. On the basis of prior work on the effects of selective retrieval, we expected that list structure would not affect the results, and selective retrieval would therefore induce a neutral effect after short lag but a (small) beneficial effect after long lag. Because, in this experiment, the test followed selective retrieval immediately—so that context retrieval might contribute more easily to recall performance (Dobler & Bäuml, 2012; Polyn et al., 2009)—we further expected that the beneficial effect would be at least as large as in Experiment 3.

Method

Participants. Another 128 students of Regensburg University participated in the experiment ($M = 21.57$ years, range = 18–33 years, 78.9% female). All subjects spoke German as native language. In exchange for participation, course credit or monetary reward was provided.

Materials and design. Materials and design were identical to Experiment 2.

Procedure. The procedure largely followed the one used in Experiment 3, differing only in the list structure of the material and the delay between retrieval practice and test. In the study phase, participants studied all 24 category-exemplar pairs, which, after a lag of 3 min or a lag of 24 hr, was followed by the intermediate phase. In this phase, half of the participants retrieved all exemplars (6 items) from 2 of the 4 categories (a total of 12 items), whereas the other half solved unrelated distractor tasks. Immediately after this phase, all participants were asked to recall the 12 unpracticed (target) items first and the 12 practiced (nontarget) items second, with both target and nontarget recall being blocked by category. In all other aspects, Experiment 4 was identical to Experiment 3.

Results

Success rates during retrieval practice were high, with higher recall after the short lag (91.8%) than the long lag (82.2%), $t(64) = 5.10$, $p < .001$, $d = 0.83$.

Figure 3B shows mean recall rates for the target items in each condition. A 2 × 2 ANOVA with the between-subjects factor of retrieval practice (present, absent) and the within-subject factor of lag (3 min, 24 hr) showed a significant main effect of lag (67.9% vs. 56.9%), $F(1, 126) = 34.16$, $MSE = 227.02$, $p < .001$, $\eta^2 = 0.21$, suggesting time-dependent forgetting, a marginally significant main effect of retrieval practice (64.3% vs. 60.5%), $F(1,$

Table 3
Mean Nontarget Recall (Plus Standard Deviations) in Experiments 3 and 4 as a Function of Lag Between Study and Selective Retrieval (3 min, 24 hr) and Retrieval Practice (Between-Categories Retrieval Practice Absent, Between-Categories Retrieval Practice Present)

	3-min lag		24-hr lag	
Experiment	*M*	*SD*	*M*	*SD*
Experiment 3				
BC-RP absent	60.8	21.8	56.6	18.7
BC-RP present	83.4	14.5	80.0	20.5
Experiment 4				
BC-RP absent	59.6	17.5	51.6	19.6
BC-RP present	83.3	12.6	76.6	16.7

Note. BC-RP = between-categories retrieval practice.

126) = 3.20, *MSE* = 295.33, p = .076, η^2 = 0.02, suggesting a trend toward higher recall in the presence of retrieval practice, and a significant interaction between the two factors, $F(1, 126) = 9.45$, *MSE* = 227.02, p = .003, η^2 = 0.07, indicating that retrieval practice affected target recall differently in the two lag conditions. Planned comparisons showed that there was no effect of retrieval practice after the short lag (66.9% vs. 68.9%), $t(126) < 1$, but that retrieval practice enhanced target recall after the long lag (61.7% vs. 52.1%), $t(126) = 3.08$, $p = .003$, $d = 0.54$. A comparison of target recall rates in the condition without retrieval practice between the short and the long lag conditions showed significant time-dependent forgetting (68.9% vs. 52.1%), $t(63) = 6.68$, $p <$ .001, $d = 1.04$.

Mean nontarget recall is shown in Table 3. ANOVA showed a significant main effect of retrieval practice (80.0% vs. 55.6%), $F(1, 126) = 100.06$, *MSE* = 379.23, $p < .001$, $\eta^2 = 0.44$, reflecting a beneficial effect of retrieval practice, and a significant main effect of lag (71.5% vs. 64.1%), $F(1, 126) = 18.93$, *MSE* = 186.20, $p < .001$, $\eta^2 = 0.13$, indicating time-dependent forgetting. There was no interaction between the two factors, $F(1, 126) < 1$.

Finally, potential confounding factors like lag order and material were also considered in this experiment. There was no main effect of lag order on target recall and no interaction of lag order with any of the other factors, all Fs(1,124) < 2.88, *MSE*s > 225.44, ps > .092, η^2s < 0.02. Similarly, material also did not influence target recall: There was no main effect of material and no interaction of material with any of the other factors, all Fs(1,124) < 1.86, *MSE*s < 227.04, ps > .175, η^2s < 0.02.

Discussion

The results of Experiment 4 replicated those of Experiment 3, both in pattern and in size. Again, we found no effect of between-categories retrieval practice after the short lag but a beneficial effect after the long lag, which suggests an influence of lag on the effects of between-categories retrieval practice. Together with the results of Experiment 3, these findings suggest that there are no major roles of list structure and delay between selective retrieval and test for the effects of between-categories retrieval practice, after both short and long lags between study and selective retrieval. Like the results of Experiment 3, the results of Experiment 4 are in line with the two-factor account of selective retrieval, on the basis of which we had expected the observed neutral effect of selective retrieval after short lag and the observed beneficial effect after long lag.

General Discussion

This study examined the influence of lag between study and selective retrieval on the effects of selective retrieval in categorized lists, both when the selectively retrieved and the nonretrieved items shared the same category (within-category retrieval practice) and when the two types of items belonged to different categories (between-categories retrieval practice). The results for within-category retrieval practice showed a detrimental effect of selective retrieval that was present after both short and long lag between study and selective retrieval, indicating that the negative effect of selective retrieval within category does not depend much on lag. The results for between-categories retrieval practice showed a neutral effect of selective retrieval after short lag and a beneficial effect after long lag, thus pointing to an influence of lag on the effect of selective retrieval between categories. The results therefore reveal differences between within-category and between-categories retrieval practice, both after short lag and after long lag. These differences can be captured by a two-factor explanation of the effects of selective retrieval, as will be described next.

A Two-Factor Explanation of the Effects of Selective Retrieval

Bäuml and colleagues suggested a two-factor account to explain the pattern of detrimental and beneficial effects of selective retrieval as it has recently been reported with lists of unrelated items as a function of lag (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). According to this account, selective retrieval triggers two types of processes-inhibition/blocking and context retrieval-but the relative contributions of the two types of processes varies with the contextual overlap between study and selective retrieval. When the contextual overlap is high, the relative contributions of inhibition and blocking are higher than of context retrieval and induce a detrimental effect of selective retrieval; when the contextual overlap is low, the relative contribution of context retrieval is higher than of inhibition and blocking and induces a beneficial effect. Obviously, the present results with categorized lists are inconsistent with an interpretation of this account that associates a short lag between study and selective retrieval with a high contextual overlap and a long lag with a low contextual overlap between the two experimental phases. Indeed, in this case, the prediction for within-category retrieval practice would be that the effect of selective retrieval should be beneficial after long lag, which is not what the results show.

The findings of the present study, however, are consistent with the account if the assumption is included that reexposure of the category labels at retrieval reinstates study context but the reexposure does not lead to a complete reinstatement (Jonker et al., 2013; Sahakyan & Kelley, 2002; Smith & Manzano, 2010). Such context reinstatement would make the long-lag situation roughly comparable to the short-lag situation, creating a relatively high contextual overlap between study and selective retrieval in both lag conditions. With regard to within-category retrieval practice, the assumption thus suggests that, in both lag conditions, the relative contributions of inhibition and blocking should be larger than of context retrieval (see Figure 4A), creating retrieval-induced forgetting after both short and long lag, which is what the present results show. With regard to between-categories retrieval practice, the assumption that category labels induce incomplete context reinstatement suggests that context retrieval should contribute to recall performance. Because interference between items from different categories is generally low (Rundus, 1973; Shaw et al., 1995), however, inhibition and blocking should influence recall very little and thus should not mask any beneficial effects of context retrieval (see Figure 4B). As a result, there should be a neutral effect of selective retrieval after short lag and a (small) beneficial effect after long lag, which is what the results show.

The results of the present experiments provide another demonstration that selective retrieval does not only induce detrimental effects on nonretrieved items but can also induce beneficial effects. Accounts of the effects of selective retrieval that focus on

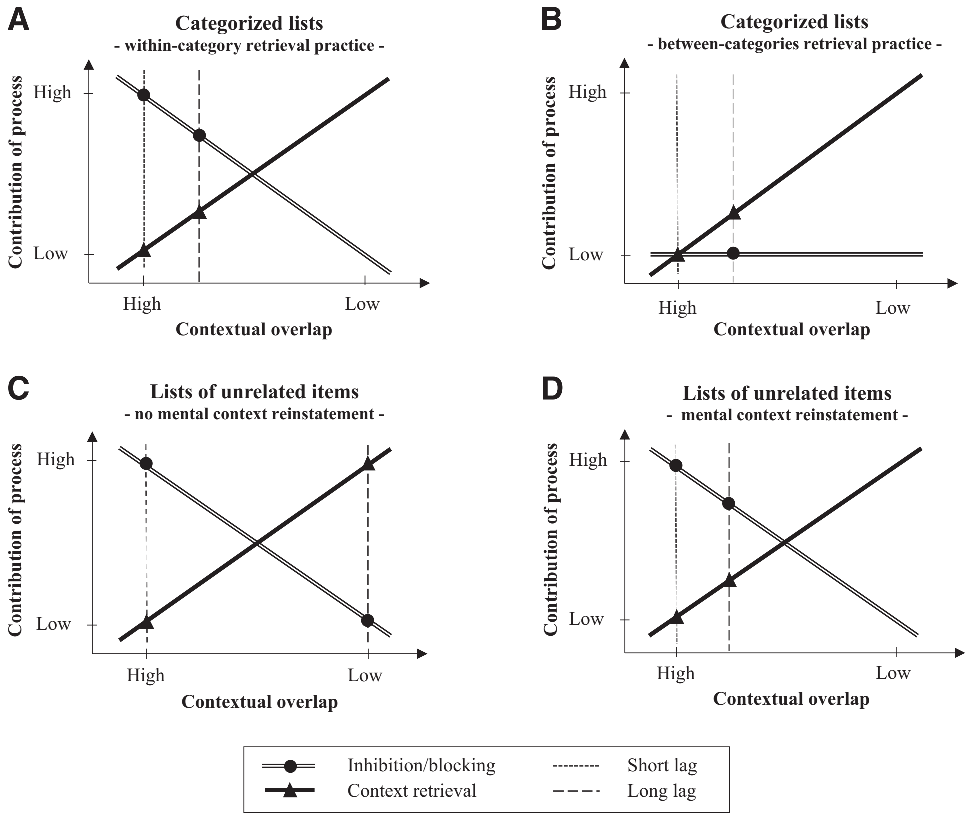

Figure 4. A two-factor explanation of the effects of selective retrieval after both short and long lag between study and selective retrieval: Simplified depiction of the possible contributions of (a) inhibition and blocking and (b) context retrieval (high, low) to recall as a function of the contextual overlap between study and selective retrieval (high, low). The suggested locations of short-lag and long-lag conditions are depicted for four different experimental conditions. A: Categorized lists when the selectively retrieved and the nonretrieved items share the same categories (within-category retrieval practice) and category labels during retrieval induce (incomplete) context reinstatement. B: Categorized lists when the selectively retrieved and the nonretrieved items belong to different categories (between-categories retrieval practice) and category labels during retrieval induce (incomplete) context reinstatement. C: Lists of unrelated items when there is no mental context reinstatement before retrieval starts. D: Lists of unrelated items when there is prior mental context reinstatement. When the relative contributions of inhibition and blocking are larger than that of context retrieval, selective retrieval induces a detrimental effect; when the relative contribution of context retrieval is larger than those of inhibition and blocking, selective retrieval induces a beneficial effect.

retrieval-induced forgetting and the underlying operations of inhibition and blocking are therefore incomplete. Indeed, the results of Experiments 3 and 4 show that selectively retrieving (nontarget) items of some studied categories during retrieval practice can create a contextual cue for recall of the (target) items from the unpracticed categories at test. This holds although the category labels that were present at test also provided context information for the target items. This context information was incomplete, however, and the selective retrieval of the nontarget items provided further context information for the target items, thus increasing the items' recall levels.

In the present experiments, the relative amount of contextual support coming from the category labels was larger than that induced by selective retrieval, as is indicated by the fact that the effect of within-category retrieval practice was still detrimental after longer lag (see Figure 4A). This pattern may not generalize to all experiments on selective retrieval, however, and, under other circumstances than were used in the present study, the contextual support induced by selective retrieval may be larger than that coming from the category labels. Although the contextual support of the category labels may thus vary with experimental situation, the results clearly indicate that explanations of the effects of selective retrieval with categorized lists have to take the contribution of the category labels into account to come up with a full explanation of the effects of selective retrieval.

Relation to Prior Work With Lists of Unrelated Items

Prior work with lists of unrelated items reported a detrimental effect of selective retrieval after short lag and a beneficial effect after long lag (Abel & Bäuml, 2015; Aslan et al., 2015; Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). These findings are consistent with the two-factor account when assuming that temporal context is the primary retrieval cue with this type of material and there is no prior context reinstatement before selective memory retrieval starts. In this case, selective retrieval should trigger mainly inhibition and blocking when the lag between study and selective retrieval is short (i.e., when the contextual overlap between study and selective retrieval is high) but should trigger mainly context retrieval when the lag is long (i.e., when the contextual overlap is low; see Figure 4C), thus creating the pattern of detrimental and beneficial effects in the two lag conditions.

However, the contextual overlap between study and selective retrieval at longer lag can also be high with lists of unrelated items. This can occur, for instance, when participants try deliberately to mentally reinstate study context before selective retrieval starts. Wallner and Bäuml (2017, Experiment 1) reported an experiment in which subjects studied a list of unrelated items and, after longer lag, were asked to recall predefined target items from the list either before or after selective retrieval of the list's remaining (nontarget) items. The effect of selective retrieval was compared between two conditions that differed in whether the study context was mentally reinstated before selective recall started. In the context-reinstatement condition, subjects were told to take a minute to recall their thoughts, feelings, and emotions prior to the beginning of the study phase (see also Jonker et al., 2013; Sahakyan & Kelley, 2002), whereas in the no-context-reinstatement condition, subjects solved arithmetic problems for the same duration of time. Mental context reinstatement was expected to enhance the contextual overlap between study and selective retrieval, making results in this long lag condition roughly comparable to those typically seen in a short lag condition (see Figure 4D). Indeed, whereas in the absence of mental context reinstatement selective retrieval induced a beneficial effect on recall performance, in the presence of the reinstatement it induced a detrimental effect. The detrimental effect of selective retrieval with categorized lists as it occurs in within-category retrieval practice situations after long lag can thus be simulated with unrelated lists when there is mental context reinstatement before selective retrieval starts.

Relation to Prior Work With Coherent Prose Material

The fact that both material (lists of unrelated items vs. categorized lists) and type of retrieval practice (within-category vs. between-categories retrieval practice) can influence the effects of selective retrieval raises the question as to how the effects of selective retrieval would look if more complex material was used for study, for instance, coherent prose material. There are some a priori expectations for this type of material. First, because coherent prose material may represent fairly integrated study material, and because integration has been found to reduce the level of interitem interference (Anderson & McCulloch, 1999; Bäuml & Kuhbandner, 2003; Chan, McDermott, & Roediger, 2006), the expectation may emerge that inhibition and blocking would play minor roles for this type of material and selective retrieval therefore would induce a neutral effect on nonretrieved items after short lag. Second, because recall of coherent prose material should also suffer when the contextual overlap between study and retrieval is reduced, context retrieval should occur after longer lag and induce a beneficial effect of selective retrieval.

Bäuml and Schlichting (2014, Experiment 2) addressed the issue, using two text passages, *The Big Bang* and *The Shaolin Temple*, which had already been used in prior work (Chan et al., 2006). Subjects studied one of the two text passages and after a lag of 5 min or a prolonged lag of 48 hr were then tested on some target questions, either before or after the prior selective answering of nontarget questions. For both target and nontarget questions, gapped sentences were provided and subjects were asked to fill in the correct item from the previously studied text (e.g., "The word Shaolin means young _____________" [Answer: forest]; "The Hubble telescope found the heavy element _____________ in extremely ancient stars" [Answer: boron]). Results showed that, when lag was short, correct answers to the target questions were largely unaffected by the prior answering of the nontarget questions, consistent with the view that coherent prose material can reflect fairly integrated study material. In contrast, when lag was long, correct answers to the target questions were more frequent when the nontarget sentences had been answered before, indicating a beneficial effect of selective retrieval (for a replication and extension of the findings, see Wallner & Bäuml, 2017, Experiment 4). Future work should address the issue in more depth and examine using a wider range of more complex study materials how selective retrieval influences recall, both after short and after long lag between study and selective retrieval. The results from such work would be of considerable relevance, from a theoretical but also from a more applied perspective.

Conclusions

In this series of experiments we showed that, with categorized lists, within-category retrieval practice and between-categories retrieval practice induce very different effects of selective retrieval. We found within-category retrieval practice to induce a detrimental effect on nonretrieved items, after both short and long lag between study and selective retrieval. In contrast, we found between-categories retrieval practice to induce a neutral effect on nonretrieved items after short lag and a beneficial effect after long lag. These findings deviate from the results of prior work with lists of unrelated items, which showed detrimental effects of selective retrieval after short lag and beneficial effects after long lag. Still, all of these results are consistent with a two-factor framework of selective retrieval that assumes critical roles in selective retrieval of (a) inhibition and blocking and (b) context retrieval.

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. (2015). Selective memory retrieval in social groups: When silence is golden and when it is not. *Cognition, 140,* 40–48.

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., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 107–132). New York, NY: Psychology Press.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 608–629.

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.

Aslan, A., Schlichting, A., John, T., & Bäuml, K.-H. T. (2015). The two faces of selective memory retrieval: Earlier decline of the beneficial than the detrimental effect with older age. *Psychology and Aging, 30,* 824–834.

Bäuml, K.-H. T. (2002). Semantic generation can cause episodic forgetting. *Psychological Science, 13,* 357–361.

Bäuml, K.-H. T. (2008). Inhibitory processes. In H. L. Roediger, III (Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (pp. 195–220). Oxford, UK: Academic Press.

Bäuml, K.-H. T. (2019). Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science, 28,* 177–182.

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., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In J. T. Wixted (Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (pp. 27–51). Oxford, UK: Academic Press.

Bäuml, K.-H. T., & Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Memory & Cognition, 31,* 1188–1197.

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,* 388–394.

Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition, 132,* 16–21.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. *Coding Processes in Human Memory,* 85–123.

Chan, J. C. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153–170.

Chan, J. C., McDermott, K. B., & Roediger, H. L. III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571.

Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science, 18,* 727–733.

Dobler, I. M., & Bäuml, K.-H. T. (2012). Dissociating the two faces of selective memory retrieval. *Memory, 20,* 478–486.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62,* 145–154.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39,* 175–191.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin and Review, 11,* 125–130.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 923–941.

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

Jonker, T. R., Seli, P., & MacLeod, C. M. (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.

Kliegl, O., Carls, T., & Bäuml, K.-H. T. (in press). How delay influences search processes at test. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12,* 148–152.

Mannhaupt, H. R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien [Production norms for verbal reactions to 40 familiar categories]. *Sprache und Kognition, 2,* 264–278.

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,* 1383–1400.

Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences, 12,* 24–30.

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.

Raaijmakers, J. G. W., & Jakab, E. (2012). Retrieval-induced forgetting without competition: Testing the retrieval specificity assumption of inhibition theory. *Memory & Cognition, 40,* 19–27.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Roediger, H. L. III., & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology, 36,* 213–242.

Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition, 24,* 188–201.

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20,* 1053–1058.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43–50.

Rupprecht, J., & Bäuml, K.-H. T. (2016). Retrieval-induced forgetting in item recognition: Retrieval specificity revisited. *Journal of Memory and Language, 86,* 97–118.

Rupprecht, J., & Bäuml, K.-H. T. (2017). Retrieval-induced versus context-induced forgetting: Can restudy preceded by context change simulate retrieval-induced forgetting? *Journal of Memory and Language, 93,* 259–275.

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.

Scheithe, K., & Bäuml, K.-H. T. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien [German norms for members of 48 categories]. *Sprache und Kognition, 14,* 39–43.

Schilling, C. J., Storm, B. C., & Anderson, M. C. (2014). Examining the costs and benefits of inhibition in memory retrieval. *Cognition, 133,* 358–370.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin and Review, 2,* 249–253.

Smith, S. M. (1985). Environmental context and recognition memory reconsidered. *Bulletin of the Psychonomic Society, 23,* 173–176.

Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods, 42,* 292–301.

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). Oxford, UK: Academic Press.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology, 24,* 617–629.

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review, 13,* 1023–1027.

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.

Wallner, L., & Bäuml, K.-H. T. (2017). Beneficial effects of selective item repetition on the recall of other items. *Journal of Memory and Language, 95,* 159–172.

Received March 19, 2019
Revision received July 17, 2019
Accepted September 6, 2019 ■