Category labels can influence the effects of selective retrieval on nonretrieved items

Michael Wirth & Karl-Heinz T. Bäuml

Springer

Springer

Category labels can influence the effects of selective retrieval on nonretrieved items

Michael Wirth[1] · **Karl-Heinz T. Bäuml**[1]

Abstract

Using lists of unrelated items as study material, recent studies have shown that selective retrieval of some studied items can impair or improve recall of the nonretrieved items, depending on whether the lag between study and selective retrieval is short or long. This study examined whether the results generalize when the items are studied together with their category labels (e.g., BIRD-*magpie*) and the category labels are reexposed as retrieval cues at test (e.g., BIRD-*m___*), a procedure often used in research on the effects of selective retrieval. Two lag conditions were employed in this study: a short 1-min lag between study and selective retrieval, and a longer 15-min lag that included mental context change tasks to enhance the lag-induced contextual drift. Experiment 1 employed lists of unrelated items in the absence of any category labels and replicated both the detrimental effect (after short lag) and the beneficial effect (after long lag) of selective retrieval. Experiment 2 was identical to Experiment 1 but provided the items' category labels during both study and retrieval, and Experiment 3 was identical to Experiment 2 but employed a categorized list. In both experiments, selective retrieval impaired recall in both lag conditions, indicating a critical role of category labels for the effects of selective retrieval. The results of the three experiments are consistent with a two-factor explanation of selective retrieval and the proposal that reexposure of category labels during retrieval can reinstate study context after longer lag.

Keywords Context effects · Interference · Inhibition · Recall

Research over the past decades has demonstrated that selective retrieval of some studied items can impair recall of nonretrieved items, a finding referred to as retrieval-induced forgetting (RIF). RIF has typically been observed with two experimental tasks: the older output-interference task, here more neutrally termed the output-order task, and the more recent retrieval-practice task. In the output-order task, it was examined whether a studied item's serial position in the testing sequence influences its recall chances. Recall chances depended on output position and declined monotonically with an item's serial position at test (Smith, 1971; Tulving & Arbuckle, 1963), indicating that the selective retrieval of the early recalled items impaired recall of the later items. In the retrieval-practice task, subjects studied several items, then repeatedly retrieved a subset of the items, and later at test recalled all previously studied items. The typical finding was that, relative to a control condition without retrieval-practice, selective retrieval improved recall of the retrieved items but impaired recall of the nonretrieved items, which reflects the RIF effect (Anderson et al., 1994; Anderson & Spellman, 1995). For both types of tasks, RIF emerged over a wide range of materials and experimental settings (for recent reviews, see Bäuml and Kliegl (2017) and Storm et al. (2015)).

RIF has often been attributed to inhibition and blocking. According to the inhibition account (Anderson, 2003), the nonretrieved items interfere during selective retrieval and are inhibited to reduce the interference. This inhibition is supposed to impair the memory representation of the nonretrieved items and thus reduce retrieval of the items over a wide range of memory tasks. According to the blocking account (Raaijmakers & Jakab, 2012; Verde, 2013), retrieval practice strengthens the retrieved items and this strengthening leads to blocking of the (weaker) nonretrieved items at test. Such blocking may arise mainly in test formats in which no item-specific cues are provided, like free recall, and less when item-specific cues are

✉ Karl-Heinz T. Bäuml
karl-heinz.baeuml@ur.de

1   Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany

provided, like in item recognition. Both inhibition and blocking can explain a wide range of RIF findings, but none of them seems to be able to capture the whole range of experimental results (e.g., Bäuml and Kliegl (2017) and Storm and Levy (2012)). For instance, inhibition - but not blocking - can explain why RIF is present in both recall and item recognition (e.g., Hicks and Starns (2004) and Spitzer and Bäuml (2007)), blocking - but not inhibition - can explain why RIF is not retrieval specific with certain restudy formats, and the forgetting arises both in response to selective retrieval and selective restudy (e.g., Raaijmakers and Jakab (2012) and Verde (2013)). An account, which assumes a role of both inhibition and blocking in RIF, however, seems to be able to explain most RIF findings (Rupprecht and Bäuml 2016, 2017; see also Anderson and Levy (2007), Aslan and Bäuml (2010), and Schilling et al. (2014); for another, context-based account of RIF, see Jonker et al. (2013)).

Detrimental and beneficial effects of selective retrieval

When we encode new information, we also store the temporal context in which the material is encountered (Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981). Temporal context drifts slowly over time (Bower, 1972; Estes, 1955), so that, after prolonged delay between study and retrieval, the context at retrieval often differs from the context at study. The contextual drift thus reduces the contextual overlap between study and retrieval and, following Tulving and Thomson's (1973) encoding specificity principle, impairs recall of target information. Typical RIF studies employed no lag at all between study and selective retrieval, or employed a short lag (between 30 sec and 5 min) filled with simple counting or calculation tasks to minimize the possible contribution of short-term memory during selective retrieval (e.g., Anderson et al. (1994), Bäuml (2002), Jonker et al. (2013), and Hicks and Starns (2004)). RIF was therefore examined when the contextual overlap between study and selective retrieval was relatively high.

Several recent studies examined the effects of selective retrieval when the lag between study and selective retrieval was increased and the contextual overlap between the two phases was thus reduced. These studies reported the typical RIF effect when lag between study and selective retrieval was relatively short (between 60 sec and 4 min) but reported a beneficial effect of selective retrieval when lag was prolonged (between 30 min and 48 hours; Abel and Bäuml (2015), Aslan et al. (2015), Bäuml and Dobler (2015), and Bäuml and Schlichting (2014)). For instance, Bäuml and Schlichting (2014, Experiment 1) had participants study a

list of unrelated words and after a lag of either 4 minutes or 48 hours asked participants to recall critical target items from the list first or after prior selective retrieval of the list's remaining (nontarget) items. Following typical RIF studies, target items were recalled using their unique initial letters as retrieval cues, whereas nontarget items were recalled using their unique word stems. Whereas, after short lag, selective retrieval induced the typical RIF effect and decreased target recall, selective retrieval improved target recall after long lag. Lag therefore influenced the effects of selective retrieval, creating a pattern of detrimental and beneficial effects of selective retrieval.

The beneficial effects of selective retrieval after longer lag have been attributed to context retrieval (Bäuml & Dobler 2015; Bäuml & Schlichting 2014; see also Bäuml (2019)). According to this view, selective retrieval does not only trigger inhibition or blocking processes, but can also induce context reactivation. Indeed, when the contextual overlap between study and retrieval is reduced, a retrieved item - more or less automatically - reactivates its study context, which then serves as a retrieval cue for the recall of other studied items, thus improving recall performance. The concept of context retrieval has proven successful in explaining a number of recall findings in the literature, including the contiguity effect, i.e., the tendency to successively recall neighboring list items, and the spacing effect, i.e., the beneficial mnemonic effect of spaced over massed learning (e.g., Greene (1989), Howard and Kahana (2002), and Kahana (1996)). The concept is also incorporated in several computational models (e.g., Polyn et al. (2009)) and provides an interpretation of Tulving's (2002) proposal of mental time travel (see Polyn and Kahana (2008)).

Consistent with this view, the pattern of detrimental and beneficial effects of selective retrieval has been explained by means of a two-factor account of selective retrieval (Bäuml & Samenieh 2012; see also Bäuml (2019)). This account assumes that selective retrieval generally triggers two types of processes: (i) inhibition and blocking, as they have been suggested to underlie RIF, and (ii) context retrieval, as it has been suggested to underlie beneficial effects of selective retrieval. Critically, the relative contribution of the two types of processes to recall performance is supposed 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 inhibition and blocking may operate, while there is not much need for context retrieval. When the contextual overlap is low - as may occur after longer lag when temporal context has drifted - context retrieval operates, while inhibition and blocking may be reduced due to attenuated interitem interference; indeed, increased lag between study and retrieval has recently

been found to reduce the size of people's mental search set during retrieval (Kliegl et al., 2020), a finding which points to reduced interitem interference with increasing lag (e.g., Rohrer (1996)). The resulting differences in relative contributions of the two types of processes - a higher relative contribution of inhibition and blocking when the contextual overlap is high and a higher relative contribution of context retrieval when the overlap is low - may then create the pattern of detrimental and beneficial effects of selective memory retrieval, for instance, a detrimental effect after short lag and a beneficial effect after prolonged lag.

The role of preceding context reinstatement for the effects of selective retrieval

The two-factor account of selective retrieval suggests that a beneficial effect of selective retrieval arises if the contextual overlap between study and selective retrieval is low, which is typically the case when lag between study and selective retrieval is relatively long (Bower, 1972; Estes, 1955). However, this beneficial effect should be reduced or even reversed when, after longer lag, study context was reinstated immediately before selective retrieval starts. Such reinstatement can be achieved when critical context features that were present during study are reexposed at test as a retrieval cue (Smith, 1985; Smith & Manzano, 2010) or when subjects are asked to deliberately try to mentally reinstate study context (Jonker et al., 2013; Sahakyan & Kelley, 2002). Indeed, in both cases, the contextual overlap between study and retrieval should be enhanced and the situation after longer lag be more close to the one after short lag. The results after longer lag should thus parallel those after short lag.

Wallner and Bäuml (2017, Experiment 1) addressed the issue, examining whether mental context reinstatement conducted prior to selective retrieval can eliminate the beneficial effect of selective retrieval as it has been reported after longer lag. In their experiment, participants studied a list of unrelated words and after a lag of 10 min, which included a mental context change task to enhance the lag-induced contextual drift, were asked to recall predefined target items from the list first or after preceding selective retrieval of the list's remaining (nontarget) items. The effects of selective retrieval were compared between two conditions that differed in whether or not 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.

The results showed that, in the absence of the context reinstatement, selective retrieval induced the expected beneficial effect on target recall, whereas in the presence of the reinstatement, selective retrieval led to a detrimental effect, i.e., RIF. Results in the context reinstatement condition thus simulated those typically reported after short lag, indicating that the relative contribution of inhibition and blocking to recall performance was larger than of context retrieval in this situation.

Context reinstatement before selective retrieval starts may not only arise by effortful mental context reinstatement but may also happen unintentionally. For instance, Smith and Manzano (2010) used video-recorded scenes of real environments as context features during study and reexposure of these features at test induced context reinstatement. Similarly, category labels provided together with single items during study (e.g., BIRD-*magpie*, ARTISAN-*butcher*) and reexposed as retrieval cues at test (e.g., BIRD-*m___*, ARTISAN-*b___*) may induce context reinstatement. Indeed, if the category labels are associated to one experimental context - the study context - only, reexposure of the labels at retrieval may more or less routinely reactivate the study context (Jonker et al. 2013, p. 855). Following this proposal, the study of items together with their category labels plus reexposure of the same category labels as retrieval cues at test may lead to a high contextual overlap between study and retrieval, not only after short lag but also after prolonged lag between study and retrieval. This high contextual overlap should increase interitem interference and induce inhibition and blocking in response to selective retrieval, thus inducing detrimental effects on target recall in both lag conditions. If so, then the effect of lagged selective retrieval with lists of unrelated items should depend on whether category labels surround the items during study and retrieval: in the absence of the category labels, selective retrieval should induce a beneficial effect on target recall, in the presence of the labels it should induce a detrimental effect. The issue is of high relevance for studies on selective retrieval, because most RIF studies in the literature used categorized lists as study material with the items' category labels being present both during study and as retrieval cues at test.

The present study

The present study reports the results of three experiments designed to examine the role of items' category labels for the effects of selective retrieval. For this purpose, subjects studied and retrieved items either in the absence of any category labels or in the presence of such labels. In each of the three experiments, subjects were presented a list of 18 items for study, followed by a short 1-min lag or

a longer 15-min lag that included mental context change tasks. This choice of longer lag followed the work by Wallner and Bäuml (2017), who recently demonstrated that lags of 10 or 30 min between study and selective retrieval can mimic the effects of lags of several hours if mental context change tasks are included during lag to enhance lag-induced context drift. After the lag, participants were asked to recall six predefined target items from the list, either before or after the retrieval of six or all twelve of the list's remaining (nontarget) items. Following typical RIF studies, the items' unique initial letters were provided as retrieval cues during retrieval of the target items, and the items' unique word stems served as retrieval cues during retrieval of the nontarget items (see Fig. 1). In Experiment 1, list items were unrelated and no category labels were presented, neither at study (e.g., *magpie*, *butcher*) nor at test (e.g., *m___*, *b___*). Experiment 2 was identical to Experiment 1 with the only difference that the items' category labels were provided, both during study (e.g., BIRD-*magpie*, ARTISAN-*butcher*) and at test (e.g., BIRD-*m___*, ARTISAN-*b___*). Finally, Experiment 3 was identical to Experiment 2 with the only difference that the studied list was categorized, consisting of six items from each of three different categories. In this case, two items from each single category were defined as the target items and the categories' remaining four items served as nontarget items. Experiment 3 was included in the study because, as mentioned above, many RIF studies in the past used categorized lists to study the effects of selective retrieval.

We expected to replicate the recent results of Bäuml and colleagues (e.g., Bäuml and Dobler (2015) and Bäuml and Schlichting (2014)) in Experiment 1 and find a detrimental effect of selective retrieval when lag between study and selective retrieval was short but a beneficial effect when lag was prolonged. In fact, following the two-factor account, the relative contribution of inhibition and blocking should be higher than of context retrieval after short lag but the relative contribution of context retrieval be higher than of inhibition and blocking after long lag, resulting in the two opposing effects of selective retrieval in the two lag conditions. Regarding Experiment 2, we expected to replicate the results of Experiment 1 after short lag but not after long lag. Indeed, after long lag, the category labels should induce context reinstatement and thus enhance the contextual overlap between study and retrieval, making contextual overlap after long lag similar to the short lag condition. If so, the relative contribution of inhibition and blocking should be larger than of context retrieval in both lag conditions and induce detrimental effects of selective retrieval. Regarding Experiment 3, we expected to replicate the results of Experiment 2. Because category labels should reinstate context not only when unrelated items are studied together with their category labels but also when a categorized list is provided with the items' category labels, again detrimental effects of selective retrieval should arise in both lag conditions. The results of the three experiments will provide important information on the role of category labels for the effects of selective retrieval.

Fig. 1 Procedure and conditions employed in Experiments 1, 2, and 3. Participants studied a list of unrelated items in Experiment 1 (**a**), the same items in the presence of the items' category labels in Experiment 2 (**b**), or a categorized list of items, again in the presence of the items' category labels, in Experiment 3 (**c**). After a lag of 1 min or 15 min, participants in all three experiments were asked to recall predefined target items from the list. Target items were recalled first (0 previously retrieved nontarget items) or after retrieval of six or twelve of the list's remaining nontarget items (6 or 12 previously retrieved nontarget items). In Experiments 2 and 3, target and nontarget items were recalled in the presence of their category labels. Predefined target items are depicted in bold letters
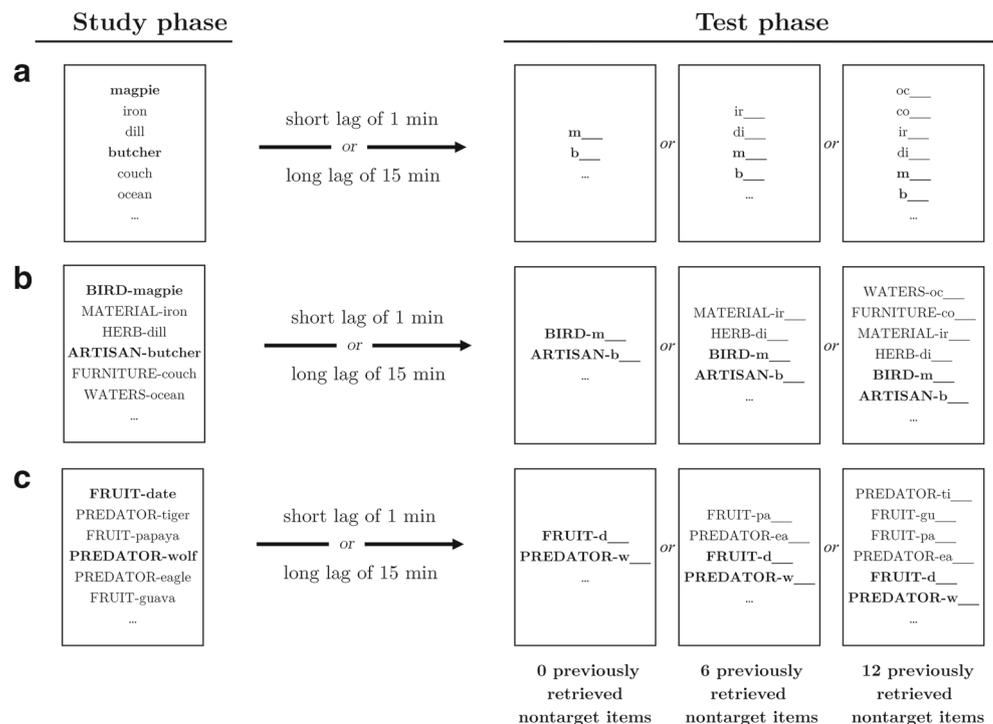
Experiment 1

Method

Participants. 72 students of Regensburg University participated in the experiment ($M = 21.90$ years, range = 18-30 years, 80.6% female). They were randomly assigned and equally distributed across the three between-subjects conditions, resulting in $n = 24$ participants in each condition. We determined the desired sample size based on reported effect sizes in Bäuml and Schlichting (2014, $\eta^2 = 0.12$), 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. All subjects spoke German as native language and received monetary reward or course credit for their participation.

Materials. Two study lists of 18 items each were constructed, each list containing one item from each of 18 different semantic categories. The material was taken from a German word norm (Scheithe & Bäuml, 1995). The items of both lists were unrelated, as indicated by the word norms of Nelson et al. (2004). For both List 1 and List 2, six items were defined as target items and the remaining 12 items as nontarget items. Within each list, target items had unique initial letters, whereas nontarget items had unique word stems. The distinction between target and nontarget items was unknown to the participants.

Design. The experiment used a 2×3 mixed factorial design. LAG (short or long) was manipulated within participants, separating the test phase either by a short 1 min or a longer 15 min lag from the study phase. SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items) was varied between participants, who were either asked to recall target items first (0 previously retrieved nontarget items) or after prior retrieval of six of the twelve nontarget items (6 previously retrieved nontarget items) or after prior retrieval of all twelve nontarget items (12 previously retrieved nontarget items).

Procedure. Each participant completed two experimental blocks, in which the items of one of the two lists were presented successively and in random order on a computer screen for 5 s each, alternated with the presentation of a fixation point in the center of the screen for a duration of 1.5 s. Following the study phase, in the short lag condition, participants counted backwards in steps of three from a random three digit number for 60 s; in the long lag condition, the same participants solved arithmetical problems and engaged in three mental context change tasks (e.g., "imagine [your] parents house, mentally walk through it and describe it [...]", see Sahakyan and Kelley (2002) and Wallner and Bäuml

(2017)) for a total of 15 min. In the context change tasks, participants were encouraged to close their eyes while imagining the scenarios for 45 s, after which they were asked to write down what they had imagined for 120 s. Inclusion of the mental context change tasks followed the experiments by Wallner and Bäuml (2017) and was conducted to enhance the lag-induced contextual drift and thus to further reduce the contextual overlap between study and retrieval. The series of distractors in the long lag condition started with a context change task (3 min), followed by an arithmetical problems task (3 min), another context change task (3 min), more arithmetical problems (3 min), and a final, third context change task (3 min). Depending on the SELECTIVE RETRIEVAL condition, participants at test were either asked to recall the target items first or after prior selective retrieval of six or all twelve of the nontarget items. Participants recalled the cued words verbally. For the target items, the items' initial letters served as retrieval cues, whereas for the nontarget items, the items' first two letters were used as retrieval cues. Cues were presented for 6 s each followed by a 0.5 s blank. Order of the two lists as well as whether the short or the long lag condition was employed first were counterbalanced across participants. Study lists and order of item presentation within the target and nontarget subsets at test were randomized. Between blocks participants took a 4 min break and started the second block with the same parameters as described above.

Results

Target recall rates are depicted in Fig. 2a. They were analyzed using a 2×3 repeated measures analysis of variance (ANOVA) with the within-participants factor of LAG (short or long) and the between-participants factor of SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items). The ANOVA yielded a significant main effect of LAG, $F(1, 69) = 7.12$, $MSE = .02$, $p = .010$, $\eta^2 = 0.93$, with higher recall in the short than the long lag condition (44.9% vs. 38.0%), and a significant interaction between LAG and SELECTIVE RETRIEVAL, $F(2, 69) = 19.42$, $MSE = .02$, $p < .001$, $\eta^2 = 0.36$, indicating that prior nontarget retrieval affected recall differently in the two lag conditions. There was no main effect of SELECTIVE RETRIEVAL (43.8% vs. 41.3% vs. 39.2%), $F(2, 69) = 0.57$, $MSE = .04$, $p = .569$, $\eta^2 = 0.02$. Follow-up comparisons employing separate one-way-ANOVAs for the short and long lag conditions revealed a significant effect of SELECTIVE RETRIEVAL in both the short lag condition (57.6% vs. 43.7% vs. 33.3%), $F(2, 69) = 8.90$, $MSE = .04$, $p < .001$, $\eta^2 = 0.20$, and the long lag condition (29.9% vs. 38.9% vs. 45.1%), $F(2, 69) = 5.15$,
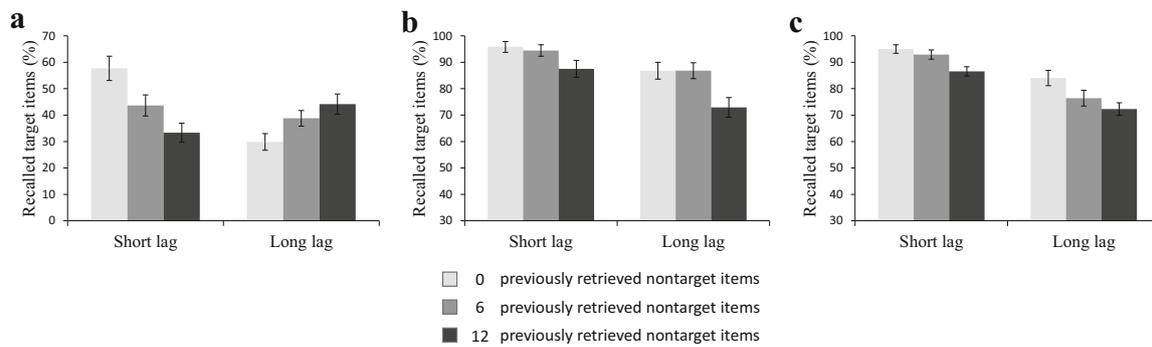
Fig. 2 Results of Experiment 1 (**a**), Experiment 2 (**b**), and Experiment 3 (**c**). Percentage of recalled target items is shown as a function of lag between study and selective retrieval (short or long) and prior selective retrieval of nontarget items (0 or 6 or 12 nontarget items). Error bars represent standard errors

$MSE = .03$, $p = .008$, $\eta^2 = 0.13$. The significant effect in the short lag condition was detrimental in nature and showed impaired target recall when nontarget items were selectively retrieved. In contrast, the significant effect in the long lag condition was beneficial in nature and showed improved target recall when nontarget items were selectively retrieved.

Table 1 shows success rates for the nontarget items when these items were selectively retrieved before recall of the target items. A 2 × 2 ANOVA with the factors of LAG (short or long) and SELECTIVE RETRIEVAL (prior retrieval of 6 or 12 nontarget items) yielded a significant main effect of LAG, $F(1, 46) = 14.80$, $MSE = .04$, $p < .001$, $\eta^2 = 0.24$, with higher recall in the short than the long lag condition (64.4% vs. 48.8%), but no main effect of SELECTIVE RETRIEVAL (57.3% vs. 55.9%), $F(1, 46) = 0.10$, $MSE = .05$, $p = .752$, $\eta^2 < 0.01$, and no significant interaction between the two factors, $F(1, 46) = 0.59$, $MSE = .04$, $p = .446$, $\eta^2 = 0.01$. These findings indicate that the selective retrieval of the nontarget items was reasonably successful and, as expected, was higher after short than long lag between study and selective retrieval. When the nontarget items were recalled after the target items, a similar picture arose, with a significant main effect of LAG only (64.9% vs. 50.0%), $F(1, 46) = 14.30$, $MSE = .04$, $p = .001$, $\eta^2 = 0.24$.

Half of the participants in this experiment started with the short lag and the other half with the long lag condition. Follow-up analyzes showed no main effect of testing order both on target and nontarget recall, both $Fs < 0.15$, $MSEs < .05$, $ps > .696$, $\eta^2s < 0.01$, and no interaction of testing order with any of the other variables, all $Fs < 0.77$, $MSEs < .05$, $ps > .385$, $\eta^2s < 0.01$.

Discussion

The results replicate those from previous studies with lists of unrelated items, both in pattern and in size. They show a detrimental effect of selective retrieval after short lag

and thus indicate the typical RIF effect. In contrast, they show a beneficial effect of selective retrieval after long lag. The results therefore add to the list of studies pointing to a critical role of lag between study and retrieval for the effects of selective retrieval (Abel & Bäuml, 2015; Aslan et al., 2015; Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). The findings of the experiment are consistent with the two-factor account of selective retrieval. On the basis of this account, we expected a higher relative contribution of inhibition and blocking than of context retrieval after short lag but a higher relative contribution of context retrieval than of inhibition and blocking after long lag, thus leading to the observed pattern of detrimental and beneficial effects of selective retrieval.

Experiment 2

It was the goal of Experiment 2 to examine whether the pattern of results observed in Experiment 1 would generalize if the same material was studied and tested as in Experiment 1 but the items' category labels were provided both during study (e.g., BIRD-*magpie*) and as a retrieval cue at test (e.g., BIRD-*m___*). On the basis of the two-factor account and the proposal that reexposure of category labels during retrieval can reinstate study context, we expected that selective retrieval would impair recall of the nonretrieved items in Experiment 2 regardless of lag. If so, the results of Experiment 1 would generalize to Experiment 2 after short lag but not after long lag between study and retrieval.

Method

Participants. A total of 72 participants took part in the experiment ($M = 22.40$ years, range = 18-33 years, 76.4% female). They were randomly assigned and equally distributed across the three between-subjects conditions, resulting in $n = 24$ participants in each

condition. Sample size followed Experiment 1. None of the participants had been tested in Experiment 1. Again, all of the participants spoke German as native language, were tested individually, and received monetary reward or course credit for participation.

Materials. Materials were identical to List 1 and List 2 in Experiment 1. Each list consisted of the same six target and the same twelve nontarget items as used in Experiment 1.

Design. Experiment 2 employed the same 2×3 mixed factorial design as Experiment 1. LAG (short or long) was varied within participants, whereas SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items) was manipulated between participants. In the short lag condition, testing occurred 1 min after study, in the long lag condition, it occurred after a lag of 15 min.

Procedure. The procedure was identical to Experiment 1 with the following exceptions: (i) in the study phase, the unrelated words were presented along with their category labels as taken from the Scheithe and Bäuml (1995) norms; (ii) in the test phase, the same category labels were provided together with the items' unique initial letters (target items) or together with the items' unique word stems (nontarget items) to serve as additional retrieval cues (see Fig. 1b).

Results

Target recall rates are depicted in Fig. 2b. A 2×3 repeated measures ANOVA with the within-participants factor of LAG (short or long) and the between-participants factor of SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items) yielded a significant main effect of LAG, $F(1, 69) = 28.31$, $MSE = .01$, $p < .001$, $\eta^2 = 0.29$, with higher recall in the short lag than the long lag condition (92.6% vs. 82.2%), and a significant main effect of SELECTIVE RETRIEVAL, $F(2, 69) = 6.64$, $MSE = .03$, $p = .002$, $\eta^2 = 0.16$, with lower target recall when there was prior selective retrieval of nontarget items (91.3% vs. 90.6% vs. 80.2%). The interaction between LAG and SELECTIVE RETRIEVAL was nonsignificant, $F(2, 69) = 1.17$, $MSE = .01$, $p = .315$, $\eta^2 = 0.03$. Follow-up analyses employing separate one-way-ANOVAs for the short and long lag conditions showed a marginally significant effect of SELECTIVE RETRIEVAL after short lag (95.8% vs. 94.4% vs. 87.5%), $F(2, 69) = 3.10$, $MSE = .02$, $p = .051$, $\eta^2 = 0.08$, and a significant effect after long lag (86.8% vs. 86.8% vs. 72.9%), $F(2, 69) = 5.85$, $MSE = .03$, $p = .005$, $\eta^2 = 0.15$. Both effects were in the same direction showing reduced target recall when there was prior selective retrieval of nontarget items.

Table 1 shows success rates for the nontarget items when these items were selectively retrieved before recall of the target items. A 2×2 ANOVA with the factors of LAG (short or long) and SELECTIVE RETRIEVAL (prior retrieval of 6 or 12 nontarget items) yielded a significant main effect of LAG, $F(1, 46) = 7.61$, $MSE = .004$, $p < .001$, $\eta^2 = 0.14$, with higher recall in the short than the long lag condition (98.6% vs. 95.1%), but no main effect of SELECTIVE RETRIEVAL (97.5% vs. 96.1%), $F(1, 46) = 1.40$, $MSE = .003$, $p = .242$, $\eta^2 = 0.03$, and no interaction between the two factors, $F(1, 46) = 1.22$, $MSE = .004$, $p = .276$, $\eta^2 = 0.03$. The selective retrieval of the nontarget items was thus successful and was higher after short than long lag between study and retrieval. When the nontarget items were recalled after the target items, a similar picture arose, showing a significant main effect of LAG only (97.5% vs. 93.7%), $F(1, 46) = 6.64$, $MSE = .005$, $p = .013$, $\eta^2 = 0.13$.

Again, half of the participants started the experiment with the short lag condition and the other half with the long lag condition. Follow-up analyzes showed no main effect of testing order on target and nontarget recall, both $Fs < 1.54$,

Table 1 Success rates for the nontarget items when these items were selectively retrieved before recall of the target items

		Short lag		Long lag	
		M	SD	M	SD
Experiment 1	6 nontarget items	66.7	26.5	47.9	22.1
	12 nontarget items	62.2	15.1	49.7	16.9
Experiment 2	6 nontarget items	98.6	4.7	96.5	6.9
	12 nontarget items	98.6	4.0	93.8	7.5
Experiment 3	6 nontarget items	94.4	8.0	91.0	13.0
	12 nontarget items	93.8	7.1	84.8	13.2

Rates (plus standard deviations) in Experiments 1, 2, and 3 are shown as a function of lag (short, long) and selective retrieval (prior retrieval of 6 or 12 nontarget items). M = mean recall rate; SD = standard deviation

$MSEs < .03$, $ps > .219$, $\eta^2 s < 0.02$, and no interaction of testing order with any of the other factors, all $Fs < 2.78$, $MSEs < .01$, $ps > .102$, $\eta^2 s < 0.06$.

Discussion

The results show a detrimental effect of selective retrieval both after short and after long lag between study and selective retrieval. The finding after short lag mimics the one reported in Experiment 1, indicating that the detrimental effect arises independently of whether category labels are provided during study and are reexposed as a retrieval cue at test. In contrast, the finding after long lag differs from the one observed in Experiment 1, indicating that the presence of items' category labels can turn the beneficial effect, as it was observed in the absence of the labels, into a detrimental effect. This finding is consistent with the two-factor account and the proposal that the presentation of category labels as retrieval cues at test can reinstate study context when the items have previously been studied with the labels. Such context reinstatement should increase the contextual overlap between study and retrieval and thus make context after long lag comparable to context after short lag. As a result, detrimental effects of selective retrieval should arise in both lag conditions, which is what the present results show.

The detrimental effects of selective retrieval observed in Experiment 1 versus Experiment 2 differ considerably in size, with the effects in Experiment 2 being only about half the size than the effect found in Experiment 1. The reduction of detrimental effects in Experiment 2 may reflect a reduced interference level between the list items. In fact, presentation of the category labels during study should have made the single list items more distinct and thus have reduced interference between the single items, which should have attenuated the contribution of inhibition and blocking to recall performance (see Anderson and McCulloch (1999) and Smith and Hunt (2000)). Reduced interitem interference may also have contributed to the increase in recall levels from Experiment 1 to Experiment 2, although the sheer presence of the category labels as additional retrieval cues in Experiment 2 will have contributed to this effect as well (see General Discussion).

Experiment 3

It was the goal of Experiment 3 to replicate the results of Experiment 2, this time using a categorized list consisting of several exemplars from several semantic categories. Some items from each of the single categories were selectively retrieved and it was examined how this would influence recall of the categories' remaining (target) items, a procedure following typical research on RIF. In addition,

Experiment 3 differed from Experiment 2 by varying lag between study and retrieval between subjects rather than within subjects. Although lag order did not show a statistical influence on results in Experiments 1 and 2 (see above), Experiment 3 was aimed to exclude any possible effects of lag order on the results.

Method

Participants.   144 students participated in the experiment ($M = 21.83$ years, range = 18-34 years, 75.0% female). They were equally distributed across the six between-subjects conditions, resulting in $n = 24$ participants in each condition. Like in Experiment 1, we determined the desired sample size based on reported effect sizes in Bäuml and Schlichting (2014), 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. None of the participants had taken part in Experiment 1 or Experiment 2. They spoke German as their native language, were tested individually, and received monetary reward or course credit for participation.

Materials.   Items were taken from three semantic categories (PREDATOR, EXOTIC FRUIT, U.S. STATE) of the Scheithe and Bäuml (1995) norms and consisted of 18 concrete German nouns, six exemplars per category. Two of the six items of each category were defined as the target items, and the remaining four items as the nontarget items, resulting in six target items and twelve nontarget items. Again, target items had unique initial letters and nontarget items unique word stems.

Design.   Experiment 3 employed the same 2×3 design as Experiments 1 and 2. In contrast to Experiment 2, however, both LAG (short or long) and SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items) were manipulated between participants.

Procedure.   The procedure of Experiment 3 was identical to the one used in Experiment 2. In the study phase, the words were again presented along with their category labels and, in the test phase, the same category labels were used for each single item as an additional retrieval cue (see Fig. 1c). During study, presentation order of the single items was random. At test, presentation order was random within the target and nontarget item sets.

Results

Target recall rates are shown in Fig. 2c. A 2×3 ANOVA with the between-participants factors of LAG (short or long) and SELECTIVE RETRIEVAL (prior retrieval of 0 or 6 or 12 nontarget items) showed a significant main effect of LAG,

$F(1, 138) = 56.82$, $MSE = .01$, $p < .001$, $\eta^2 = 0.29$, with higher recall in the short lag than the long lag condition (91.7% vs. 77.5%), as well as a significant main effect of SELECTIVE RETRIEVAL, $F(2, 138) = 9.64$, $MSE = .01$, $p < .001$, $\eta^2 = 0.12$, with lower recall when there was prior selective retrieval of nontarget items (89.6% vs. 84.7% vs. 79.5%). There was no significant interaction between the two factors, $F(2, 138) = 0.75$, $MSE = .01$, $p = .474$, $\eta^2 = 0.01$. Follow-up separate one-way ANOVAs indicated a significant effect of SELECTIVE RETRIEVAL for both the short lag condition, $F(2, 69) = 6.69$, $MSE = .01$, $p = .002$, $\eta^2 = 0.16$, and the long lag condition, $F(2, 69) = 4.64$, $MSE = .02$, $p = .013$, $\eta^2 = 0.12$. The two effects were in the same direction and reflected a detrimental effect of prior selective retrieval (short lag: 95.1% vs. 93.1% vs. 86.8%; long lag: 84.0% vs. 76.4% vs. 72.2%).

Table 1 again shows success rates for the nontarget items when these items were selectively retrieved before recall of the target items. A 2 × 2 ANOVA with the factors of LAG (short or long) and SELECTIVE RETRIEVAL (prior retrieval of 6 or 12 nontarget items) yielded a significant main effect of LAG, $F(1, 92) = 8.22$, $MSE = .01$, $p = .005$, $\eta^2 = 0.08$, with higher recall in the short than the long lag condition (94.1% vs. 87.9%), but no main effect of SELECTIVE RETRIEVAL (92.7% vs. 89.2%), $F(1, 92) = 2.54$, $MSE = .01$, $p = .115$, $\eta^2 = 0.03$, and no significant interaction between the two factors, $F(1, 92) = 1.63$, $MSE = .01$, $p = .206$, $\eta^2 = 0.02$. Thus, like in Experiments 1 and 2, the selective retrieval of the nontarget items was successful, with higher recall after short than long lag between study and retrieval. When nontarget items were recalled after the target items, significant main effects of LAG (92.9% vs. 85.6%), $F(1, 92) = 9.63$, $MSE = .01$, $p = .003$, $\eta^2 = 0.10$, and SELECTIVE RETRIEVAL (92.7% vs. 85.8%), $F(1, 92) = 8.71$, $MSE = .01$, $p = .004$, $\eta^2 = 0.09$, arose.

Discussion

Using a categorized list rather than a list of unrelated items to examine the effects of selective retrieval of some items from each single category on the categories' nonretrieved items, the results of Experiment 3 replicate those of Experiment 2, both in pattern and in size. The results show a detrimental effect of selective retrieval after short lag as well as after long lag. Size of effects was well comparable between the two experiments in both lag conditions. The findings indicate that, as long as the items are studied together with their category labels and the labels are reexposed as a retrieval cue at test, the effects of selective retrieval do not much depend on whether several exemplars or just a single exemplar from a category has been studied. This holds although in the one case (lists of unrelated items)

interitem interference can occur between all list items, whereas in the other case (categorized lists) it occurs mostly between the items of a category (e. g., Rundus (1973) and Shaw et al. (1995)). The results provide another case for the proposal that, in the presence of category labels, context is reinstated after longer lag and the contextual overlap between study and retrieval thus becomes comparable to the short-lag condition. On the basis of the two-factor account, it is this increase in contextual overlap between study and retrieval that mediated the detrimental effect in the long-lag condition.

General discussion

The findings of the present experiments replicate and extend the results from prior work. They replicate the prior work by showing that, with lists of unrelated items and in the absence of the items' category labels during both study and retrieval, selective retrieval decreases recall when the lag between study and selective retrieval is short but increases recall when lag is long (Abel & Bäuml, 2015; Aslan et al., 2015; Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). The findings extend the prior work by showing that, with both lists of unrelated items and categorized item lists, results change when the items' category labels are provided during study and the same labels are reexposed as retrieval cues at test. In such case, selective retrieval decreases recall regardless of whether lag between study and selective retrieval is short or long. Thus, while the effect of selective retrieval after short lag does not vary with the presence versus absence of the items' category labels, providing such labels during study and as retrieval cues at test influences the effect of selective retrieval after longer lag.

A two-factor explanation of the results

Bäuml and colleagues suggested a two-factor account to explain the different effects of selective retrieval in different experimental conditions (Bäuml & Samenieh 2012; Bäuml & Schlichting 2014; see also Bäuml (2019)). According to this account, selective retrieval triggers two types of processes: (i) inhibition and blocking and (ii) context retrieval. Critically, the relative contribution of the two types of processes to recall performance is supposed to depend on the contextual overlap between study and retrieval. When the contextual overlap is high, like typically after a short lag between study and selective retrieval, the relative contribution of inhibition and blocking is higher than of context retrieval, inducing a detrimental effect of selective retrieval. In contrast, when the contextual overlap is low, like typically after longer lag, the relative contribution of

context retrieval is higher than of inhibition and blocking, resulting in a beneficial effect of selective retrieval.

The findings of Experiment 1 are consistent with the two-factor account when assuming that temporal context is the primary retrieval cue with lists of unrelated items and there is no prior context reinstatement - deliberate or unintentional - before selective memory retrieval starts. In such case, selective retrieval should induce a higher relative contribution of inhibition and blocking than context retrieval when the lag between study and selective retrieval is short and the contextual overlap between study and selective retrieval is high, and induce a higher relative contribution of context retrieval than inhibition and blocking when the lag is long and the contextual overlap is low, thus creating the pattern of detrimental and beneficial effects in the two lag conditions.

The findings of Experiments 2 and 3 are inconsistent with an interpretation of the two-factor account that directly 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, selective retrieval in Experiments 2 and 3 should have impaired recall after short lag but improved recall after long lag, which is not what the results show. The results fit with the account, however, if the assumption is included that presentation of the items' category labels during study and reexposure of the same labels as retrieval cues at test reinstated study context (Jonker et al., 2013). In such case, the long-lag situation should be similar to the short-lag situation and the contextual overlap between study and retrieval be high in both lag conditions. Following the two-factor account, the relative contribution of inhibition and blocking to recall should thus be high after both short and long lag and induce detrimental effects, which is what the results of the two experiments show.

The results of Experiment 1 differ from those of Experiments 2 and 3 not only in retrieval dynamics after long lag but also in time-dependent forgetting between the two lag conditions. In fact, target items' time-dependent forgetting decreased from about 30% in Experiment 1 to less than 15% in Experiments 2 and 3. This reduction is in line with the view that the presentation of the category labels in Experiments 2 and 3 induced context reinstatement. Because such context reinstatement should increase the contextual overlap between study and retrieval, the recall impairment after long lag should be reduced and amount of time-dependent forgetting be attenuated in Experiments 2 and 3 relative to Experiment 1, which is what the present results show. The results on time-dependent forgetting thus support the view entertained above that the category labels induced context reinstatement in the present experiments.

Following typical RIF studies, Experiments 2 and 3 in this study provided items' category labels both during study and at retrieval. Doing so, the beneficial effect of selective retrieval, as it was observed after longer lag when category labels were absent (see Experiment 1), turned into a detrimental effect, which we interpreted as evidence for a reinstatement of study context. One may alternatively argue that the presence of the category labels at test just changed the nature of the memory test, for instance, by making retrieval less sensitive to interference effects, without reinstating study context. While this proposal can explain why recall levels after long lag increased drastically from Experiment 1 to Experiments 2 and 3, it can not easily explain why this increase was much smaller when nontargets were previously retrieved (about 30%) than when no nontargets were previously retrieved (about 60%). To explain the difference, the proposal, for instance, may assume that the beneficial effect of selective retrieval, as it was observed in the absence of the category labels, was accompanied by a strong reduction in interitem interference, so that the category labels might have led to a smaller increase in recall levels when nontargets were previously retrieved. While this assumption can explain the present results, it is not covered by prior work on the effects of selective retrieval (e. g., Bäuml and Kliegl (2017) and Storm and Levy (2012)). Future work may address the issue more directly and examine the effects of selective retrieval when the items' category labels are absent during study but are provided as retrieval cues at test.

Relation to prior work on selective retrieval and context reinstatement

The present indication that context reinstatement prior to selective retrieval can turn the beneficial effect of selective retrieval into a detrimental one fits with one of Wallner and Bauml's (2017) recent results. In their Experiment 1, these researchers used lists of unrelated items as study material and, similar to the present study, employed a 10-min lag between study and selective retrieval that included a mental context change task to enhance the lag-induced contextual drift. The effects of selective retrieval were compared between two conditions, which differed in whether there was mental context reinstatement prior to selective retrieval. 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, whereas in the no-context-reinstatement condition, subjects solved arithmetic problems for the same duration of time. The results showed the expected beneficial effect of selective retrieval in the absence of the mental context reinstatement but a reversal of the effect into a detrimental one in its

presence. These findings suggest that deliberate mental context reinstatement prior to selective retrieval can induce detrimental effects of selective retrieval also after longer lag between study and selective retrieval, which parallels the findings of present Experiment 2, in which lists of unrelated items were studied as well but the items' category labels were provided during study and retrieval.

Besides deliberate mental context reinstatement and the reexposure of items' category labels, there is a number of further methods that have been employed in the literature to (unintentionally) enhance the contextual overlap between study and retrieval when the contextual overlap has become poor (see Baddeley et al. (2015)). For instance, in some studies a certain visual stimulus, some odor, or music surrounded the encoding of study material and these context features were then reexposed at test to improve the contextual overlap between study and test. Relative to a condition in which no such reexposure occurred, the reexposure of the context features enhanced recall levels after longer lag (e. g., Cady et al. (2008), Herz (2006), Smith (1985), and Smith and Manzano (2010)). Critically, the results of the present study together with those reported in Wallner and Bäuml (2017) suggest that reexposure of such context features may not only lead to higher recall levels after longer lag but may also change the effects of selective retrieval: Whereas selective retrieval may induce a higher relative contribution of context retrieval and improve recall of the nonretrieved contents when no such context features are provided and temporal context is the primary cue to retrieve the studied items, selective retrieval may induce a higher relative contribution of inhibition and blocking and impair recall of the nonretrieved contents when such context features are reexposed. Future work may address the issue and examine whether the many context reinstatement techniques reported in the literature influence overall recall levels only or influence the effects of selective retrieval as well.

The results of the present study arrive at a time when, to the best of our knowledge, there is only one experiment in the literature yet, in which the influence of lag on the effects of selective retrieval in categorized lists has been examined. In this experiment, MacLeod and Macrae (2001, Experiment 2) employed an impression formation task in which participants were instructed that their task was to form impressions of two virtual men, named JOHN and BILL. For impression formation, participants were shown personality traits of the two men on index cards. After a short lag of 5 min or a long lag of 24 hrs, the participants were asked to selectively retrieve half of the traits of one of the two men (e. g., JOHN), providing the traits' unique word stems as retrieval cues. The test was conducted 5 min after selective retrieval and participants were asked to recall all previously exposed traits of the two men. Relative to recall of the nonretrieved traits of the unpracticed man (e. g., BILL), which served as a control in this experiment, results showed reduced recall of the practiced man's (e. g., JOHN'S) nonretrieved traits, i. e., RIF, in both lag conditions. Although this experiment differs in a number of ways from the experiments reported in the present study, the results are broadly consistent with those of present Experiment 3. This consistency suggests that reexposure of one of the two men's names during selective retrieval in MacLeod and Macrae's experiment reinstated study context very similar to how reexposure of the category labels supposedly did in the present experiment.

Output-order "versus" retrieval-practice task

In the literature, effects of selective retrieval have often been examined with the retrieval-practice task. In this task, subjects typically study a categorized list, then in a retrieval practice phase, selectively retrieve a subset of the items from a subset of the categories, before in a final test phase they try to recall all previously studied items. The task leads to three different types of items: practiced items (from the practiced categories), unpracticed items from the practiced categories, and control items from the unpracticed categories. The difference in recall levels between unpracticed items and control items then reflects the effect of selective retrieval, for instance, RIF (e.g., Anderson et al. (1994)).[1] In contrast, the output-order task consists of a study and a test phase only, for instance, manipulating whether predefined target items from the study list are recalled first at test, or are recalled after previous selectice retrieval of other (nontarget) list items. This procedure leads to two item types only: target items that were recalled after previous selective retrieval of nontarget items - mimicking the unpracticed items in the retrieval-practice task; and target items that were recalled without previous selective retrieval of the nontarget items - mimicking the control items in the retrieval-practice task. The difference in recall levels between the two types of items then reflects the effect of selective retrieval, for instance, RIF (e. g., Bäuml and Samenieh (2010)).[2]

---

[1] The three item types may also result from two different intermediate conditions: a retrieval-practice condition, in which subjects selectively retrieve a subset of the items from each single studied category; and a no-retrieval-practice condition, in which subjects engage in an unrelated distractor task and there is no selective retrieval at all (see Shaw et al. (1995)).

[2] The two basic item types may also result by analyzing whether a studied item's serial position in the testing sequence influences its recall chances. Recall levels of items tested early in the testing sequence may then be compared to the recall levels of items tested late in the testing sequence (see Smith (1971) and Tulving and Arbuckle (1963)).

Most RIF studies in the literature employed the retrieval-practice task (see Murayama et al. (2014)), whereas beneficial effects of selective retrieval were mostly demonstrated using the output-order task (for a review, see Bäuml et al. (2017); for demonstrations of the beneficial effect employing the retrieval-practice task, see Bäuml and Dobler (2015) and Chan et al. (2006)). While findings typically converge between the two tasks with regard to RIF and can also do so with regard to the beneficial effect of selective retrieval (see Bäuml and Dobler (2015)), the two tasks may not always lead to the same results. For instance, in the output-order task, target retrieval follows selective retrieval immediately, whereas, in the retrieval-practice task, there is typically a delay between selective retrieval and the final recall test (see Anderson et al. (1994)). Such delay can potentially induce context drift processes, which may reduce the benefits of study context reactivation during selective retrieval and thus, on the basis of the two-factor account, reduce possible beneficial effects of selective retrieval (but see Chan et al. (2006)). Beneficial effects of selective retrieval might therefore be easier to find in the output-order than the retrieval-practice task, an issue we examine in ongoing work (e.g., Bäuml and Wallner (2020)).

Conclusions

The presentation of items' category labels during study and the reexposure of the labels as retrieval cues at test increases studied items' recall levels, reduces amount of time-dependent forgetting, and reduces the size of possible detrimental effects of selective retrieval. Moreover, it can change retrieval dynamics. When lag between study and retrieval is prolonged, selective retrieval induces a beneficial effect on nonretrieved items when category labels are absent but induces a detrimental effect when the labels are present. These findings are consistent with a two-factor explanation, which assigns roles for both inhibition/blocking and context retrieval in selective memory retrieval, and the proposal that reexposure of category labels at test can reinstate study context.

Category-exemplar pairs as they were employed in the present study represent a special case of paired associates in which higher-order semantic information is used as the cue part for a pair's target item. The present findings therefore leave it open how the effects of selective retrieval on nonretrieved items may look like with other types of paired associates, like pairs of unrelated words or paired associates that consist of face-name pairs or foreign language vocabulary words. Whether the present results for longer lag generalize to such paired associates will, among other factors, depend on the degree to which reexposure of the cue items of such pairs during retrieval induces

reinstatement of the study context. Future studies may investigate this research question.

References

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 Gorfein, D. S., & MacLeod, C. M. (Eds.) *Inhibition in cognition*, (pp. 81–102). Washington: American Psychological Association.

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., & Bäuml, K.-H. T. (2010). Retrieval-induced forgetting in young children. *Psychonomic Bulletin and Review*, *17*, 704–709.

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.

Baddeley, A. D., Eysenck, M., & Anderson, M. C. (2015). *Memory*, (2nd ed.). Hove: Psychology Press.

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

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., Aslan, A., & Abel, M. (2017). The two faces of selective memory retrieval - cognitive, developmental, and social processes. In Ross, B. (Ed.) *The psychology of learning and motivation*, (Vol. 66, pp. 167–209). Elsevier Inc: Academic Press.

Bäuml, K.-H. T., & Dobler, I. M. (2015). The two faces of memory 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 Wixted, J. T. (Ed.) *Cognitive psychology of memory, Vol. 2 of Learning and Memory: A comprehensive reference, 2nd edition, Byrne, J. H. (Ed.)*, (pp. 27–51). Oxford: Academic Press.

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.

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

Bäuml, K.-H. T., & Wallner, L. (2020). Selective retrieval in categorized lists: detrimental, neutral, and beneficial effects on nonretrieved Items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

Cady, E. T., Harris, R. J., & Knappenberger, J. B. (2008). Using music to cue autobiographical memories of different lifetime periods. *Psychology of Music*, *36*, 157–177.

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

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). GPOwer 3: A fexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. Journal of Experimental Psychology: Learning. *Memory, & Cognition*, *15*, 371–377.

Herz, R. S. (2006). *I know what i like: understanding odor preferences. In J. Drobnick (Ed.), The smell culture reader (pp 190-203)*. New York: Berg.

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

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*, *120*, 852–872.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, *24*, 103–109.

Kliegl, O., Carls, T., & Bäuml, K.-H. T. (2020). 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.

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.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407.

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 and Cognition*, *40*, 19–27.

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

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

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. *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 & Review*, *2*, 249–253.

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.

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.

Spitzer, B., & Bäuml, K.-H. T. (2007). Retrieval-induced forgetting in item recognition: Evidence for a reduction in general memory strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 863–875.

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, eyewitness memory, social cognition, autobiographical memory, and creative cognition. In Ross, B. (Ed.) *Psychology of learning and motivation*, (pp. 141–194). Elsevier Inc: Academic Press.

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory and Cognition*, *40*, 827–843.

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

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*, 321–334.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1433–1448.

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.