Items that are Subject to Retrieval-induced Forgetting show Slowed Forgetting over Time

Sarah R. Meixensperger & Karl-Heinz T. Bäuml

Department of Experimental Psychology

Regensburg University, Germany

To appear in:

Memory

This paper is not the copy of record and may not exactly replicate the authoritative document published in the journal.

Correspondence address:

Sarah R. Meixensperger, Department of Experimental Psychology,

Regensburg University, 93040 Regensburg, Germany.

E-mail: sarah.meixensperger@ur.de

Phone: +49-941-943-3780

FAX: +49-941-943-3872

Items that are Subject to Retrieval-induced Forgetting show Slowed Forgetting over Time

Abstract

Retrieval-induced forgetting (RIF) refers to the finding that selective retrieval practice of a subset of studied items can impair recall of related unpracticed items, relative to recall of unrelated control items. Using categorized study material, we examined in two experiments how the retention interval between retrieval practice and test (1 min, 4 h, 24 h, or 7 d) influences the size of this RIF effect. Across both experiments, the RIF effect was robust at shorter retention intervals (1 min, 4 h) but disappeared at longer intervals (24 h, 7 d). Unlike prior work, we also fitted power functions of time to the recall rates of unpracticed and control items to examine forgetting rates of the two types of items. Analysis of the function's parameters showed that while unpracticed items were initially impaired, they were forgotten more slowly over time than control items, making RIF disappear at longer retention intervals. Results are discussed with respect to prominent accounts of RIF.

Keywords

retrieval - retrieval practice - forgetting - retrieval-induced forgetting - time-dependent forgetting

Items that are Subject to Retrieval-induced Forgetting show Slowed Forgetting over Time

Retrieval practice shortly after encoding enhances memory performance for the studied material relative to conditions in which the studied material is reexposed for further study or is not repeated at all. This beneficial effect of retrieval practice is often referred to as the testing effect and has been demonstrated over a wide range of materials and experimental settings (for reviews, see Karpicke, 2017; Roediger & Butler, 2011). Quite often, the testing effect is not yet observable when the retention interval between practice and test is short and in the order of minutes but emerges with prolonged retention interval (e.g., Roediger & Karpicke, 2006; Toppino & Cohen, 2009). Consistent with this finding, analysis of time-dependent forgetting of retrieval practiced, restudied, and not repeated studied items has shown that, when measuring recall performance for the three item types at several retention intervals and fitting a power function of time to these data, forgetting rates for retrieval practiced items turn out to be reduced compared to the forgetting rates of restudied or not repeated items (Carpenter et al., 2008; Nickl & Bäuml, 2023). Retrieval practice thus seems to attenuate time-dependent forgetting of studied material.

Retrieval practice in daily life, be it in educational settings or eyewitness testimony situations, is often selective and only a subset of the encoded material is retrieved. Such selective retrieval, when conducted shortly upon encoding, has been shown to not only influence memory performance for the practiced material but to also influence memory performance for the unpracticed material. Indeed, while selective retrieval practice improves memory performance for the practiced material, which parallels the testing effect, it often impairs memory performance for the unpracticed material. In the typical experimental setup demonstrating this finding, participants study several items from different semantic categories (e.g., FRUIT-orange, FURNITURE-chair, FRUIT-banana, FURNITURE-table) and shortly after study practice retrieval of half of the items from half of the categories (e.g., FRUIT-or___). On a cued recall test few minutes after retrieval practice, in which participants are asked to recall all studied items (e.g., FRUIT-b___, FRUIT-o___, FURNITURE-t___, FURNITURE-c___), this retrieval practice typically leads to enhanced recall for the practiced items (orange) and reduced recall for the unpracticed items from the practiced categories (banana), referred to as unpracticed items in the following. Both effects arise relative to recall for the unpracticed items from the unpracticed categories (chair, table), which serve as control items in this task (Anderson et al., 1994; Anderson & Spellman, 1995). This detrimental effect, known as retrieval-induced forgetting (RIF), has been replicated across a wide range of materials and experimental settings (Bäuml & Kliegl, 2017; Storm et al., 2015).

While research on the testing effect suggests that the size of the testing effect increases with length of retention interval between practice and test, to date no conclusive evidence exists on how retention interval influences the size of the RIF effect. The few studies that examined the issue in prior work mostly compared recall of unpracticed and control items after a longer retention interval of typically 24 h or 7 d with recall of the items after a short retention interval of 5 min or less (for a short review, see Bäuml & Kliegl, 2017). Among these studies, five studies employed a 24-h retention interval as the long retention interval. They found RIF to be present after the short but to be absent after the long retention interval (Abel & Bäuml, 2014; Carroll et al., 2007; Chan, 2009; MacLeod & Macrae, 2001; Saunders & MacLeod, 2002). In contrast, two studies employed a 7-d retention interval as the long retention interval. They found RIF to be present after both the short and the long retention interval and the size of the RIF effect to not diminish statistically with retention interval (Garcia-Bajos et al., 2009; Storm et al., 2012). Only one study (Tandoh & Naka, 2007) used more than two retention intervals (immediate recall, 10 min, 1 h, 7 d) to examine how retention interval influences RIF. This study found RIF effects of similar size across retention intervals, which, however, may have been influenced by the rather small RIF effect in the immediate recall condition (~6 percentage points). To date, no study in the literature analyzed time-dependent forgetting of the single item types by fitting power functions of time to the recall rates and comparing forgetting rates of unpracticed and control items.

Forgetting over time typically shows a curvilinear nature, with memory performance declining rapidly soon after encoding followed by a long, much slower decline in performance (Ebbinghaus, 1885). Such forgetting often has been captured by a power function of time, $r(t) = a(1 + t)^{-b}$, where $r(t)$ represents proportion of remembered items at time t, parameter a represents performance level directly after encoding, and parameter b represents the relative forgetting rate as time passes. While the power function is able to describe time-dependent forgetting over a wide range of experimental situations (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991, 1997), time-dependent forgetting of unpracticed items after selective retrieval has not been studied yet. It is thus unclear whether unpracticed and control items show similar or different forgetting rates as time after retrieval practice passes. Critically, the power function has a focus on relative forgetting that contrasts with the focus on absolute forgetting in the ANOVA-based approach (see Wixted, 2022). Indeed, equality of absolute forgetting can differ from equality of forgetting in relative terms. For instance, if retention drops from 50 to 30 percentage points for one item type (-40%) and from 30 to 10 percentage points for another item type (-67%), then both show the same absolute drop (-20 percentage points) but very different relative forgetting. This illustrates why power functions provide insights beyond standard ANOVA comparisons (see Carpenter et al., 2008; Nickl & Bäuml, 2023; Wixted, 2022).

The two most prominent explanations of RIF are the inhibition and blocking accounts of RIF. The inhibition account proposes that RIF arises as a consequence of the necessity to overcome retrieval competition during selective retrieval practice (Anderson, 2003). This account assumes that when a subset of the studied items is practiced (e.g., orange), the not-to-be-practiced items (e.g., banana) interfere and compete for conscious recall. To reduce the interference and facilitate selection of the to-be-practiced items, the memory representation of the not-to-be-practiced items becomes suppressed, leading to weakened representations of these items and impaired recall on a later memory test. In contrast, the blocking account assumes that retrieval practice during selective retrieval strengthens the associations between the practiced items (e.g., orange) and their category cue (e.g., FRUIT), and such strengthening leads to blocking of the category's (not strengthened) unpracticed items (e.g., banana) at test, creating RIF (Raaijmakers & Jakab, 2013). Thus, according to inhibition, RIF arises as a result of inhibitory action during selective retrieval, whereas according to blocking, RIF arises as a result of strength-based competition at test. The two accounts are not mutually exclusive and there is indeed evidence that both can contribute to RIF under certain experimental conditions (see Rupprecht & Bäuml, 2015).

On the basis of the inhibition and blocking accounts of RIF, no unequivocal expectations arise on whether relative forgetting rates differ between unpracticed and control items. According to the inhibition account, unpracticed items show a reduced memory representation compared to control items. Because items of different memory strength typically show similar absolute forgetting (Slamecka & McElree, 1983; Wixted, 2022), the (stronger) control items and the (weaker) unpracticed items might also show similar absolute forgetting and thus a comparable size of the RIF effect across retention interval. As outlined above, the (weaker) unpracticed might then show enhanced relative forgetting compared to the (stronger) control items. However, it has repeatedly been argued that recovery from inhibition may be an intrinsic property of inhibition (see Carroll et al., 2007; MacLeod & Macrae, 2001; Storm et al., 2012). If so, unpracticed items should show reduced absolute forgetting compared to control items. Whether unpracticed items would also show reduced relative forgetting depends on detail. For instance, if retention dropped from 50 to 30 percentage points for control items, and from 30 to 18 percentage points for unpracticed items, then the two item types would differ in absolute forgetting (-20 versus -12 percentage points) but show the same relative forgetting (-40%); if unpracticed items dropped only from 30 to 24 percentage points, then unpracticed items would show both reduced absolute and reduced relative forgetting compared to control items (-40% versus -20%); if unpracticed items dropped from 30 to even 12 percentage points, then unpracticed items would show reduced absolute but enhanced relative forgetting compared to control items (-40% versus -60%).

On the basis of the blocking account, expectations on how retention interval influences RIF are also not unequivocal and depend on whether interference between list items is assumed to persist or diminish with retention interval. If interference effects persist with retention interval, blocking effects should remain present with retention interval and might induce RIF effects of similar size across retention interval. Unpracticed items might then show an enhanced relative forgetting rate compared to control items (see above). However, results from several experimental tasks suggest that interference effects between list items can diminish with retention interval (e.g., Kriechbaum & Bäuml, 2023; Lehmer & Bäuml, 2018). In such case, blocking effects should also decrease with retention interval, which would lead to a reduction in absolute forgetting for unpracticed items. Analogous to the inhibition account when recovery from inhibition is assumed to be present, unpracticed items might then show a relative forgetting rate that is reduced, similar, or even enhanced compared to control items, dependent on how the two item types differ in absolute forgetting over time.

The present study had two goals. The first goal was to use an experimental setup that is typical for RIF studies in the literature and measure recall rates of unpracticed and control items over more than two retention intervals. Beyond a short 1-min and a longer 4-h retention interval, both a 24-h and a 7-d retention interval were employed, as prior work reported a RIF effect after the 7-d retention interval but not the 24-h retention interval. We used ANOVA to compare unpracticed and control items' absolute forgetting over time and *t*-tests to analyze whether the RIF effect was still present in the longer retention interval conditions. The second goal of the study was to compare unpracticed and control items' relative forgetting rates across retention intervals. For this purpose, power functions of time were fitted to the recall rates of the two item types and the estimated relative forgetting rate parameter of the function (b) was compared across item types.

Two experiments were conducted, each employing categorized word lists as study material. In Experiment 1, item type was varied within participants, as has been done in most prior RIF work, and retention interval was varied between participants. All participants studied a list of 36 items, consisting of six items from each of six different categories. Each participant practiced half of the items from half of the categories shortly after study. At test, each participant was asked to recall all studied items 1 min, 4 h, 24 h, or 7 d after retrieval practice. Both at practice and at test, cued recall formats were employed and each item's category name and unique word stem (at practice) or each item's category name and unique initial letter (at test) were provided as retrieval cues.

Experiment 2 was similar to Experiment 1 but employed other categorized material and varied item type (unpracticed vs. control items and practiced vs. control items) between participants and retention interval within participants, thus following prior work on time-dependent forgetting that also varied retention interval within participants (Carpenter et al., 2008; Siler & Benjamin, 2020). All participants studied a list of 48 items, consisting of six items from each of eight categories. Half of the participants practiced half of the items from each single category, whereas the other half of the participants did not engage in any retrieval practice. At test, each participant recalled all six items of two categories 1 min after retrieval practice, all six items of two other categories 4 h after retrieval practice, all six items of two further categories 24 h after retrieval practice, and, finally, all six items of the last two categories 7 d after retrieval practice. The same cued recall formats were employed at practice and at test as were used in Experiment 1.

The results of the two experiments may help resolve the inconsistencies arising from prior RIF work and provide more clearcut evidence on whether unpracticed and control items differ in absolute forgetting over time. Above all, this study is the first to examine unpracticed and control items' relative forgetting rates across multiple retention intervals using power function analysis. The results will thus provide first evidence on whether unpracticed and control items show similar or different relative forgetting rates. The findings will impose important restrictions on current accounts of RIF.

Experiment 1

Methods

Ethical considerations

Both experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants

A total of 192 participants took part in the experiment (M = 23.68 years, SD = 3.97, range = 18-35 years, 149 female, 1 divers). They were recruited from Regensburg University, as well as by placing online advertisements in students' groups of different German universities. Participants were distributed equally across the four between-participants conditions, yielding n = 48 participants per condition. We determined the desired sample size based on reported effect sizes in prior work on the effects of selective retrieval practice and retention interval on unpracticed and control items (e.g., Abel & Bäuml, 2014; Chan, 2009; Storm et al., 2012), counterbalancing purposes, and the results of analyses of test power conducted with the G*Power program (version 3.1; Faul et al., 2007). For these analyses, we set α at .05, power at .80, and (i) d at 0.50 to detect at least a medium-sized RIF effect after short retention interval and (ii) f at 0.15 to detect at least a small- to medium-sized effect for the interaction between the factors of item type (unpracticed vs. control items) and retention interval. All participants spoke German as their native language and reported no neurological or psychiatric disease. In both experiments, all participants gave their spoken informed consent and took part in the experiment in return for either course credit or a compensatory amount of money. In both experiments, there were no participants failing to show up when testing was delayed and there were no participants that were excluded from analysis.

Materials

The material consisted of 36 concrete German nouns which belonged to six different semantic categories (translated into English: type of fabric, kitchen utensil, four-footed animal, weather phenomenon, part of a building, tree), with each category comprising six exemplars (Van Overschelde et al., 2004). Within categories, all items had unique initial letters. For instance, translated into English, the exemplars in the category type of fabric were cashmere, flannel, wool, polyester, leather, and sheepskin, and the exemplars in the category tree were spruce, beech, larch, maple, palm, and chestnut (for more details, see Meixensperger & Bäuml, 2025). For each participant, half of the categories served as practiced categories, whereas the other half served as control categories. Each category's six words were randomly divided into two sets of three words each. For the practiced categories, the one set of words was selectively retrieved during retrieval practice (practiced items, P+ items), whereas the other set was left unpracticed (unpracticed items, P- items). At test, a category's P- items were always recalled before the category's P+ items. For the control categories, no items were selectively retrieved during practice but the one set (C- items) was recalled first at test and thus matched the output position of the P- items, whereas the other set (C+ items) was recalled after the first set and thus matched the output position of the P+ items. It was counterbalanced across participants which categories served as practiced and control categories and, within categories, which items served as P+ and P- items respectively C+ and C- items.

Design

The experiment followed a 4 (RETENTION INTERVAL: 1 min vs. 4 h vs. 24 h vs. 7 d) x 4 (ITEM TYPE: P- vs. P+ vs. C- vs. C+) mixed factorial design. Retention interval was manipulated between participants, item type was manipulated within participants.

Procedure

Data collection took place via Zoom meetings (Zoom Video Communications), in which participants and experimenters were connected by live webcam and microphone feeds. For participants in the 1-min retention interval condition, the experiment took place during a single session, while for all other participants the experiment consisted of two sessions with 4-h, 24-h, and 7-d retention intervals between them. During each session, the experimenter shared their screen and instructed participants orally. The software PowerPoint 2019 (Microsoft Corporation) was used for stimulus presentation.

The experiment consisted of three main phases: study phase, retrieval practice phase, and test phase. For all participants, the experiment started with the study phase, during which all 36 items were presented together with their category labels centrally on the computer screen for 5 s each. They were displayed individually and in a pseudorandomized order, with no two items of the same category following each other. Afterwards, participants were asked to count backwards in steps of three from a randomly selected three-digit number for 30 s before retrieval practice started. During retrieval practice, all participants were asked to recall half of the items from three of the six categories (P+ items) in three successive practice cycles. The words' category labels and unique word stems were provided as retrieval cues. The participants had 6 s to recall each single item. Answers were given orally. Retrieval practice was followed by a 1-min distractor task, in which participants sorted number triples by their value.

For the participants in the 1-min retention interval condition, the test phase followed immediately after the number triples sorting, while the participants in the three conditions with long retention interval (4 h, 24 h, and 7 d) were dismissed and returned for the second session after the corresponding retention interval. The second session started with a 1-min distractor task, in which participants rated the pleasantness of places which were presented one after the other on the computer screen. The test phase was the same for all participants. The words' category labels and unique first letters were provided as retrieval cues and were presented successively in a category blocked manner positioned centrally on the computer screen. The sequence of categories was pseudorandomized, with no more than two practiced or control categories following each other. The items of a category were tested successively in a randomized order with the restriction that for the practiced categories, the P- items were always tested first followed by the P+ items, and for the control categories, the C- items were always tested first followed by the remaining C+ items. The participants had 6 s to recall each single item and gave their response orally before the next retrieval cue appeared on the screen.

Fitting the power function to the recall rates

The procedure followed prior work from our lab (Bäuml et al., 2025, Bäuml & Trissl, 2022; Nickl & Bäuml, 2023). We fitted a power function of time, $r(t) = a(1 + t)^{-b}$, to practiced, unpracticed, and control items' recall rates using maximum likelihood methods (Riefer & Batchelder, 1988; Wickens, 1982). For each item type, the goodness-of-fit of the power function was compared to the goodness-of-fit of a statistical baseline model, which describes recall rates in the four retention interval conditions (1 min, 4 h, 24 h, and 7 d after practice) as the product of four independent binomial distributions. The comparison of the two models – power function versus statistical baseline model – is based on the calculation of a likelihood ratio and leads to an approximative χ^2-test with two degrees of freedom. The parameters of the power function were estimated by maximizing the likelihood function of the power function. For this maximization, we employed hours since practice as units of measurement of time.

If the power function was able to describe recall rates for each item type, we examined whether the parameters *a* and *b* of the power function varied significantly between pairs of item types. For this analysis, we combined two sets of recall rates (e.g., the recall rates of unpracticed and control items) and compared the goodness-of- fit of the general power function model – with two free *a*-parameters and two free *b*-parameters – with a power function model that imposes the restriction that parameters *a* or *b* are constant across conditions. Again, the comparison between the two models is based on the calculation of a likelihood ratio and leads to a χ^2-test with one degree of freedom. The parameters of the power functions were estimated by maximizing the likelihood function of the power functions.

Results

Success rates during retrieval practice

Participants recalled 88.89% (*SD* = 11.37) of the items in the first practice cycle, 91.26% (*SD* = 9.59) in the second practice cycle, and 91.96% (*SD* = 9.18) in the third practice cycle. A 3 × 4 mixed-factors ANOVA with the within-participants factor of practice cycle (first, second, third) and the between-participants factor of retention interval (1 min, 4 h, 24 h, 7 d) showed a main effect of practice cycle, $F(2, 376) = 30.21$, $MSE = 16.44$, $p < .001$, $\eta_p^2 = 0.14$, but no main effect of retention interval condition, $F(3, 188) = 0.03$, $MSE = 276.56$, $p = .994$, $\eta_p^2 = 0.00$, and no interaction between the two factors, $F(6, 376) = 1.53$, $MSE = 16.44$, $p = .169$, $\eta_p^2 = 0.02$. Success rates during practice thus increased over the three practice cycles (first vs. second cycle: $t(191) = 5.27$, $p < .001$, $d = 0.38$; second vs. third cycle: $t(191) = 2.73$, $p = .007$, $d = 0.20$) and did not vary with retention interval condition.

Recall of unpracticed and control items at test

Figure 1A shows percentage of recalled P- (unpracticed) and C- (control) items at all four retention intervals together with the best fitting power functions for the two item types. A 2 x 4 mixed-factors ANOVA with the within-participants factor of item type (P- items, C- items) and the between-participants factor of retention interval (1 min, 4 h, 24 h, 7 d) showed a main effect of item type, $F(1, 188) = 16.07$, $MSE = 184.35$, $p < .001$, $\eta_p^2 = 0.08$, indicating lower recall for P- than C- items and thus a general RIF effect, a main effect of retention interval condition, $F(3, 188) = 17.66$, $MSE = 292.35$, $p < .001$, $\eta_p^2 = 0.22$, indicating typical time-dependent forgetting, and an interaction between the two factors, $F(3, 188) = 3.63$, $MSE = 184.35$, $p = .014$, $\eta_p^2 = 0.05$, suggesting a decline of the size of the RIF effect with retention interval. Planned comparisons between recall of P- and C- items at all four retention intervals revealed a significant RIF effect at the 1-min, $t(47) = 5.42$, $p < .001$, $d = 0.78$, $BF_{01} = 0.00$, and 4-h retention intervals , $t(47) = 2.11$, $p = .041$, $d = 0.30$, $BF_{01} = 1.10$, but no significant RIF effects at the 24-h, $t(47) = 1.23$, $p = .227$, $d = 0.18$, $BF_{01} = 4.28$, and 7-d retention intervals, $t(47) = 0.09$, $p = .927$, $d = 0.01$, $BF_{01} = 8.82$.

The power function described the time-dependent forgetting of the P- and C- items well, as is reflected by the $\chi^2(2)$-values of 1.56 for the P- items and 3.40 for the C- items. A direct comparison of the function's parameters *a* and *b* between item types revealed that parameter *a* differed significantly between item types and was higher for the C- than P- items, $\chi^2(1) = 16.69$. Importantly, forgetting rate parameter *b* also differed significantly between the two item types, and indicated a lower relative forgetting rate for the P- items, $\chi^2(1) = 5.08$. P- items thus showed lower recall shortly after practice accompanied by a reduced relative forgetting rate.

Recall of practiced and control items at test

Figure 2A shows percentage of recalled P+ (practiced) and C+ (control) items at all four retention intervals. A 2 x 4 mixed-factors ANOVA with the within-participants factor of item type (P+ items, C+ items) and the between-participants factor of retention interval (1 min, 4 h, 24 h, 7 d) showed a main effect of item type, $F(1, 188) = 350.14$, $MSE = 215.99$, $p < .001$, $\eta_p^2 = 0.65$, indicating higher recall for P+ than C+ items, and a main effect of retention interval condition, $F(3, 188) = 26.57$, $MSE = 270.33$, $p < .001$, $\eta_p^2 = 0.30$, indicating time-dependent forgetting. There was no interaction between the two factors, $F(3, 188) = 1.15$, $MSE = 215.99$, $p = .331$, $\eta_p^2 = 0.02$.

The power function described the time-dependent forgetting of the C+ items well, as is reflected by the $\chi^2(2)$-value of 1.52. By contrast, the power function could not describe the time-dependent forgetting of the P+ items ($\chi^2(2) = 7.17$). As a result, it did not make sense to statistically compare parameters *a* and *b* between the two item types.

Discussion

The results replicate prior RIF work by showing that selective retrieval practice shortly upon encoding can impair recall of related unpracticed items when the retention interval between practice and test is on the order of few minutes. Besides, the results also show a RIF effect when the retention interval was prolonged up to 4 h, but fail to find a RIF effect for the longer retention intervals of 24 h and 7 d. These findings suggest reduced absolute forgetting over time for unpracticed compared to control items. The results also show that unpracticed items' forgetting over time can be well described by a power function of time. Analysis of the forgetting rate parameter of the function indicated that unpracticed items also show reduced relative forgetting compared to control items, pointing to slowed forgetting over time for unpracticed items. The goal of Experiment 2 was to replicate these findings with other study material and when retention interval was varied within and item type (P- vs. C- items and P+ vs. C+ items) between participants.

Experiment 2

Methods

Participants

Another 128 participants took part in the experiment ($M = 23.17$ years, $SD = 2.91$, range = 18-30 years, 96 female). They were recruited from Regensburg University, as well as by placing online advertisements in students' groups of different German universities. Participants were distributed equally across the two between-participants conditions, yielding $n = 64$ participants per condition. Sample size was again determined on the basis of power analyses conducted in G*Power (version 3.1, Faul et al., 2007) as well as counterbalancing purposes. For the power analyses, we again set α at .05, power at .80, and (i) *d* at 0.50 to detect at least a medium-sized RIF effect after short retention interval and (ii) *f* at 0.15 to detect at least a small- to medium-sized effect for the interaction between the factors of item type (unpracticed versus control items) and retention interval. All participants spoke German as their native language and reported no neurological or psychiatric disease.

Materials

Two new word lists each consisting of 48 concrete German nouns were used as study material (Scheithe & Bäuml, 1995; Van Overschelde et al., 2004). Each list consisted of eight different semantic categories, with each category comprising six exemplars with unique initial letters (for details, see Meixensperger & Bäuml, 2025). Like in Experiment 1, a category's six words were randomly divided into two sets of three words each. For the group of participants who engaged in retrieval practice, a category's one set was used as P+ items and the other set as P- items, with P- items being recalled first and P+ items second at test. Analogously, for the group of participants who did not engage in any retrieval practice, a category's one set was used as C+ items and the other set as C- items, with C- items being recalled first and C+ items second at test. It was counterbalanced within categories, which items served as P+ and P- items respectively C+ and C- items. For half of the participants, the one item list served as study material, whereas for the other half of the participants, the other item list was used for study.

Design

The experiment followed a 4 (RETENTION INTERVAL: 1 min vs. 4 h vs. 24 h vs. 7 d) x 2 (PRACTICE: retrieval practice vs. no retrieval practice) mixed factorial design. Practice was manipulated between participants. Participants who engaged in retrieval practice produced P- and P+ items, participants who did not engage in any retrieval practice produced C- and C+ items. Retention interval was manipulated within participants, so that every participant ran all four tests 1 min, 4 h, 24 h, and 7 d after retrieval practice.

Procedure

Like in Experiment 1, data collection took place via Zoom meetings (Zoom Video Communications), in which participants and experimenters were connected by live webcam and microphone feeds. All participants took part in four sessions. After the first test 1 min after retrieval practice, the participants were dismissed and returned for the following sessions 4 h, 24 h, and 7 d after retrieval practice. During all sessions of the experiment, the experimenter shared their screen and instructed participants orally. The software PowerPoint 2019 (Microsoft Corporation) was used for stimulus presentation.

The experiment consisted of a study phase, a retrieval practice phase, and an extended test phase that took place 1 min, 4 h, 24 h, and 7 d after retrieval practice. For all participants, the experiment started with the study phase, during which all 48 items of one of the two study lists were presented together with their category labels centrally on the computer screen for 5 s each. They were displayed individually and in a pseudorandomized order, with no two items of the same category following each other. Afterwards, participants were asked to count backwards in steps of three from a randomly selected three-digit number for 30 s before retrieval practice started. During retrieval practice, half of the participants were asked to recall half of the items from all eight categories in three successive retrieval cycles. The words' category labels and unique word stems were provided as retrieval cues. The participants had 6 s to recall each single item. Answers were given orally. The other half of the participants completed a neutral distractor task for the same period of time, in which they engaged in decision tasks. Retrieval practice respectively the neutral distractor task were followed by a 1-min distractor task (sorting number triples) for all participants.

The test phase consisted of four successive tests 1 min, 4 h, 24 h, and 7 d after retrieval practice. During each test, participants recalled the items of two of the eight studied categories. It was counterbalanced which categories were recalled in which of the four tests. The words' category labels and unique first letters were provided as retrieval cues and were presented successively in a category blocked manner positioned centrally on the computer screen. The sequence of the categories was randomized. Within each category, the P- items were always tested first followed by the P+ items for participants with retrieval practice, and the C- items first followed by the C+ items for participants without retrieval practice. Participants had 6 s to recall each item and gave their response orally before the next retrieval cue appeared on the screen. The first session ended after the first test and a 1-min distractor task, in which participants calculated cross sums before they were dismissed. The three subsequent sessions 4 h, 24 h, and 7 d after retrieval practice followed the same basic procedure. Each session started with a 1-min distractor task, in which the participants rated the pleasantness of pictures (places, food, faces). Afterwards, participants were asked to recall the words of two more of the studied categories. The procedure of the test was the same as for the first test 1 min after retrieval practice. Like at the end of the first test, at the end of the tests conducted 4 h and 24 h after retrieval practice, participants calculated cross sums before they were dismissed.

Fitting the power function to the recall rates

The fitting of power functions to the recall rates of the single item types followed the procedure employed in Experiment 1.

Results

Success rates during retrieval practice

Participants recalled 87.70% (*SD* = 8.08) of the items in the first practice cycle, 90.17% (*SD* = 6.23) in the second practice cycle and 91.34 (*SD* = 6.05) in the third practice cycle. A one-way ANOVA with the within-participants factor of practice cycle (first, second, third) showed a main effect of practice cycle, $F(2, 126) = 26.62$, $MSE = 8.33$, $p < .001$, $\eta_p^2 = 0.30$. Like in Experiment 1, success rates during practice increased over the three practice cycles (first vs. second cycle: $t(63) = 4.66$, $p < .001$, $d = 0.58$; second vs. third cycle: $t(63) = 3.44$, $p = .001$, $d = 0.43$).

Recall of unpracticed and control items at test

Figure 1B shows percentage of recalled P- (unpracticed) and C- (control) items at all four retention intervals together with the best fitting power functions for the two item types. A 2 x 4 mixed-factors ANOVA with the between-participants factor of item type (P- items, C- items) and the within-participants factor of retention interval (1 min, 4 h, 24 h, 7 d) showed a main effect of item type, $F(1, 126) = 9.28$, $MSE = 733.70$, $p = .003$, $\eta_p^2 = 0.07$, indicating a general RIF effect, a main effect of retention interval condition, $F(3, 378) = 11.86$, $MSE = 367.41$, $p < .001$, $\eta_p^2 = 0.09$, indicating time-dependent forgetting, and an interaction between the two factors, $F(3, 378) = 2.63$, $MSE = 367.41$, $p = .050$, $\eta_p^2 = 0.02$, suggesting a decline of the size of the RIF effect with retention interval. Again, planned comparisons between recall rates of P- and C- items at all four retention intervals revealed a significant RIF effect at the 1-min, $t(126) = 3.78$, $p < .001$, $d = 0.67$, $BF_{01} = 0.01$, and 4-h retention intervals, $t(126) = 2.33$, $p = .022$, $d = 0.41$, $BF_{01} = 0.58$, but no significant RIF effect at the 24-h, $t(126) = 1.24$, $p = .216$, $d = 0.22$, $BF_{01} = 3.51$, and 7-d retention intervals, $t(126) = 0.40$, $p = .689$, $d = 0.07$, $BF_{01} = 6.76$.

The power function described the time-dependent forgetting of the P- and C- items well, as is reflected by the $\chi^2(2)$-values of .14 for the P- items and .43 for the C- items. A direct comparison of the function's parameters *a* and *b* between item types revealed that both parameter *a* and parameter *b* were higher for the C- than P- items (parameter *a*: $\chi^2(1) = 21.18$, parameter *b*: $\chi^2(1) = 5.24$). Thus, like in Experiment 1, P- items showed lower recall shortly after practice accompanied by a reduced relative forgetting rate.

Recall of practiced and control items at test

Figure 2B shows percentage of recalled P+ (practiced) and C+ (control) items at all four retention intervals. A 2 x 4 mixed-factors ANOVA with the between-participants factor of item type (P+ items, C+ items) and the within-participants factor of retention interval (1 min, 4 h, 24 h, 7 d) showed a main effect of item type, $F(1, 126) = 109.94$, $MSE = 614.93$, $p < .001$, $\eta_p^2 = 0.47$, indicating higher recall rates for P+ than C+ items, and a main effect of retention interval condition, $F(3, 378) = 31.98$, $MSE = 319.45$, $p < .001$, $\eta_p^2 = 0.20$, indicating time-dependent forgetting. There was no interaction between the two factors, $F(3, 378) = 0.35$, $MSE = 319.45$, $p = .792$, $\eta_p^2 = 0.00$.

The power function described the time-dependent forgetting of the P+ items well, as is reflected by the $\chi^2(2)$-value of 4.51. However, the function could not describe the time-dependent forgetting of the C+ items, which is reflected by the $\chi^2(2)$-value of 9.02. Therefore, we did not statistically compare parameters *a* and *b* between item types.

Discussion

The results replicate the findings of Experiment 1. Again, RIF arose after the 1-min and 4-h retention intervals but not after the longer 24-h and 7-d retention intervals, suggesting reduced absolute forgetting over time for unpracticed compared to control items. Also, the recall rates of both unpracticed and control items were again well described by a power function of time, with the forgetting rate parameter of the function being reduced for unpracticed compared to control items. Like the findings of Experiment 1, these results suggest slowed forgetting over time of unpracticed items.

General Discussion

This is the first study in the literature that examined the effect of selective retrieval practice on unpracticed items' recall rates over more than two retention intervals, analyzing both unpracticed and control items' absolute forgetting over time and unpracticed and control items' relative forgetting rates. In both experiments, ANOVA of recall rates of unpracticed and control items showed a gradual attenuation of absolute forgetting over time for unpracticed compared to control items, with the RIF effect being statistically present after the short and 4-h retention intervals but statistically absent after the 24-h and 7-d retention intervals. Recall rates of both unpracticed and control items were well described by a power function of time, with the relative forgetting rate parameter of the function being reduced for unpracticed compared to control items, which suggests slower time-dependent forgetting for unpracticed items. Unpracticed items thus revealed both reduced absolute and reduced relative forgetting, a pattern that arose regardless of whether retention interval was varied within participants and item type between participants, or vice versa.

The present ANOVA findings are consistent with the results from prior work in which recall rates of unpracticed and control items were compared between a short and a 24-h retention interval. Like the present experiments, this prior work found a reduction of the size of the RIF effect with retention interval, with a significant RIF effect after the short retention interval but no such effect after the prolonged retention interval (e.g., Abel & Bäuml, 2014; Chan, 2009; MacLeod & Macrae, 2001). By contrast, the present findings disagree with the results from two prior studies, in which significant RIF effects arose after both a short and a 7-d retention interval, and the size of the RIF effect was statistically unaffected by retention interval (Garcia-Bajos et al., 2009; Storm et al., 2012). This holds while, in both of these studies, at least a numerical reduction of the size of the RIF effect with retention interval arose (see also below). Critically, none of the prior studies fit power functions of time to the recall rates from multiple retention intervals to estimate and compare relative forgetting rates across item types. The results from the present study fill this gap. They demonstrate that unpracticed items are forgotten more slowly than control items, making RIF disappear at longer retention intervals.

The present results are consistent with the inhibition account of RIF if recovery from inhibition is regarded an intrinsic property of inhibition. In such case, the size of the RIF effect should attenuate with retention interval and unpracticed items show reduced absolute forgetting compared to control items. Such pattern was indeed observed in the present results. If intralist interference is assumed to attenuate with retention interval, the present results are also consistent with the blocking account. An attenuation of intralist interference should reduce blocking effects and thus reduce the size of the RIF effect with retention interval as well as unpracticed items' absolute forgetting over time, which is what the present results show. Critically, the present results disagree with inhibition if recovery from inhibition is assumed to be absent and they disagree with blocking if intralist interference is assumed to persist with retention interval. The present findings thus impose restrictions on both accounts of RIF.

Besides inhibition and blocking, the RIF effect has also been attributed to context change (Jonker et al., 2013). This account assumes that retrieval practice introduces a shift in participant's internal context, creating distinct study and practice contexts. With categorized lists, control categories are therefore encountered in the study phase only, whereas retrieval-practice categories are encountered in both the study and retrieval-practice contexts. At test, participants are then assumed to inappropriately access the more recent practice context when searching for unpracticed items, but access the study context when searching for control items, which is supposed to result in RIF. Because such context change effects are known to attenuate rather quickly with retention interval (Abel & Bäuml, 2017; Divis & Benjamin, 2014), according to the context-based account the size of the RIF effect and unpracticed items' absolute forgetting should reduce with retention interval. The present findings thus are basically in line with the context-based account.

This study employed an experimental setup that is typical for RIF studies in the literature, with respect to both study material and experimental procedure. The question therefore arises of whether results would generalize to other experimental setups. The findings by Storm et al. (2012) suggest that, to some extent, results may indeed vary with procedural detail. In their experiment, these researchers employed five retrieval practice cycles compared to the more typical three retrieval practice cycles, which were used in many prior RIF studies as well as in the present experiments. Doing so, they found a RIF effect that was larger than after three or one retrieval practice cycle and only a numerical, but not a statistical reduction of RIF with retention interval. Number of retrieval practice cycles or the size of the RIF effect shortly after practice may thus influence the reduction of RIF with retention interval (see also Storm et al., 2012), as may many other factors, including study material (e.g., Garcia-Bajos et al., 2009). Discovering whether experiments that employ setups less typical for RIF studies in the literature lead to similar or different results than reported here might become an important issue in future research on RIF.

The present study employed retention intervals between 1 min and 7 d to study RIF and thus followed previous studies on the role of retention interval for RIF. Using this range of intervals, the present experiments suggest reduced forgetting over time for unpracticed compared to control items, with recall of the two item types being nearly identical after the longer 7-d retention interval. The question is how forgetting over time of the two item types would proceed if even longer retention intervals, like 10 or 14 d, were employed. Indeed, if the estimated parameters of the power function were still valid for such longer intervals, the functions should intersect around the interval of 7 d and recall gradually become higher for unpracticed than control items. By contrast, if the results for unpracticed items during the first 7 d were influenced by recovery from inhibition, the gradual elimination of intralist interference, or diminishing effects of context change, then recall after the 7-d interval should become more or less identical for unpracticed and control items and recall of unpracticed items follow the trajectory of the control items. Future work may address the issue by measuring recall rates of unpracticed and control items over longer retention intervals than the intervals employed in the present study. Such measurements will provide critical tests for current accounts of RIF.

The analysis of time-dependent forgetting conducted in the present study follows typical prior work on time-dependent forgetting and does not take possible distributional effects into account. Such effects would arise if the forgetting function of a distribution of items differed from the forgetting function of individual items in that distribution. The reduction in the rate of forgetting for unpracticed relative to control items might thus be due to differences in forgetting of a relatively small number of items in the distribution, as opposed to general effects for all items. For instance, if RIF was mediated by inhibition and inhibition weeded out the more difficult to recall (weak) items from the distribution, this might lead to a distribution of items more robust to forgetting over time, given that weaker compared to stronger items show higher forgetting rates (see Wixted, 2022). Although exactly this scenario appears unlikely, because RIF effects shortly after retrieval practice have been found to be larger for stronger than weaker items (see Anderson et al., 1994; Bäuml, 1998), distributional effects may nonetheless contribute to the effects of retention interval on RIF and may be of high priority for future work on the issue.

On the basis of the testing effect literature, the expectation may arise that selective retrieval practice enhances recall of the practiced items relative to the control items and this beneficial effect increases with retention interval (e.g., Roediger & Karpicke, 2006). Consistent with some previous RIF studies (e.g., Garcia-Bajos et al., 2009; MacLeod & Macrae, 2001; but see Abel & Bäuml, 2014; Chan, 2009), however, the present ANOVA findings show that the size of the beneficial effect of retrieval practice on practiced items did not increase with retention interval. A possible reason for this discrepancy with the testing effect literature may be that, while in typical testing effect studies quite often a free recall format is employed during practice (see Karpicke, 2017), here items' word stems were provided as retrieval cues during practice. The presentation of such item-specific cues is typical for RIF studies (e.g., Anderson et al., 1994), but may make retrieval practice less demanding than it is in typical testing effect studies. Difficulty of practice, however, can be a critical factor for the size of the testing effect (e.g., Carpenter, 2009) and may also influence the forgetting rate of practiced items. Results for practiced items that are more similar to those found in many testing effect studies may arise if items' initial letters rather than word stems were provided as retrieval cues during selective retrieval, which would make the practice task more demanding than it was in the present study. Future research may address the issue.

In both Experiment 1 and Experiment 2 the power function described recall rates of unpracticed and corresponding control (C-) items well, which replicates results from prior work (Bäuml et al, 2025; Bäuml & Trißl, 2022). However, only in Experiment 2 the power function was able to describe recall rates of practiced items, and only in Experiment 1 it was able to describe recall rates of corresponding control (C+) items, which contrasts with prior work, in which the power function fit recall rates of both item types well (Bäuml & Trißl, 2022). The likely reason for the two failures in the present study is an unexpectedly high level of noise in the recall rates. Indeed, recall rates of both practiced items in Experiment 1 and corresponding control (C+) items in Experiment 2 turned out to be higher in the 24-h than 4-h retention interval condition, which deviates from typical time-dependent forgetting and raises problems for all candidate functions to account for forgetting over time. Although a priori chosen sample size appeared sufficient for the present experiments (see above), higher numbers of participants would have been necessary to reduce the noise in the data and create more typical time-dependent forgetting for all item types.

Selective retrieval practice often induces RIF for unpracticed material, but there are exceptions to this ”rule”. One line of exceptions are studies showing that, under certain conditions, selective retrieval practice shortly upon encoding does not affect unpracticed items' recall performance. Negative mood, stress, or divided attention during practice provide such examples (Bäuml & Kuhbandner, 2007; Koessler et al., 2009; Roman et al., 2009), but the best-studied example arises from research showing that, when participants interrelate study items during learning, RIF is absent (Anderson et al., 2000; Bäuml & Hartinger, 2002; Goodmon & Anderson, 2011). Intriguingly, this null effect can even turn into a beneficial effect of retrieval practice on unpracticed items if recall is measured 24 h after practice (Chan, 2009; Chan et al., 2006). The finding of a neutral effect of selective retrieval on unpracticed items after short retention interval and a beneficial effect after the long retention interval suggests reduced forgetting over time of unpracticed compared to control items also in the absence of an initial RIF effect.

Another line of exceptions to the ”rule” of detrimental effects of selective retrieval on unpracticed items are findings showing that selective retrieval can enhance recall of unpracticed items if practice is time-lagged and does not occur shortly upon encoding (Bäuml & Schlichting, 2014; Kriechbaum & Bäuml, 2023; see also Bäuml, 2019). Interestingly, in such case the forgetting over time of unpracticed items is enhanced relative to a control condition in which retrieval practice is absent (Bäuml et al., 2025; Bäuml & Trissl, 2022). These findings suggest that selective retrieval does not always reduce time-dependent forgetting of unpracticed items, and the reduction finding may rather be tied to conditions in which practice occurs shortly upon encoding.

The present study varied retention interval both between participants (Experiment 1) and within participants (Experiment 2). The experiments thus mirror prior work on time-dependent forgetting, in which retention interval was varied between participants in some studies (e.g., Bäuml & Trißl, 2022; Nickl & Bäuml, 2023) but within participants in others (e.g., Carpenter et al., 2008; Siler & Benjamin, 2020). Critically, each of the two designs comes with a mixture of advantages and disadvantages. On the one hand, the within-participants design has the advantage that no averaging of recall rates across participants is required before fitting the function to recall data, and thus no averaging artifacts can arise (see Estes, 1956). On the other hand, the problem may arise that recall of some items at early retention intervals can enhance recall of other items at later retention intervals, which can lead to an underestimation of forgetting rates. By contrast, in the between-participants design, potential influences of recall at early retention intervals on recall at later retention intervals cannot arise, but averaging artifacts could emerge, which might distort the results (see Nickl & Bäuml, 2023; Wixted & Ebbesen, 1997).The present study does not rule out the possible problems that may come with each of the two designs. However, by using both designs and finding the same pattern of results, it indicates that results may not depend much on design and the results arising with the one design may well generalize to the other design.

To conclude, numerous studies have shown that selective retrieval shortly upon encoding often enhances recall of the practiced material and impairs recall of the unpracticed material, relative to control conditions. The present results show that these beneficial and detrimental effects of retrieval practice differ in persistence with retention interval. While the beneficial effect on practiced items remains present with retention interval – and, under some circumstances, may even increase with retention interval –, the detrimental effect on unpracticed items reduces with retention interval and can already be absent 24 h after practice. This pattern of results is mediated by slowed forgetting over time of unpracticed compared to control items and, as reported in the prior testing effect literature, slowed forgetting over time of practiced compared to control items. For longer retention intervals, selective retrieval practice primarily produces benefits and imposes only minimal, short-lived costs on recall of studied information.

Disclosure of interest

The authors report there are no competing interests to declare.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data and materials are available on the Open Science Framework (https://osf.io/629ft/?view_only=42c5d7898b7245cb96de036c13f69802). All experiments reported in this study were implemented using the software PowerPoint 2019 (Microsoft Corporation) and the software Zoom (Zoom Video Communications). The software was run on standard desktop computers with the operating system Windows 10 (Microsoft, Redmond, WA). Data were analyzed using IBM SPSS Statistics for Windows, Version 29.0.1 (IBMCorp., Armonk, NY), G*Power 3.1 (Faul et al., 2007) as well as C program code that was used to fit power functions to target recall rates.

References

Abel, M., & Bäuml, K.-H.T. (2014). The roles of delay and retroactive interference in retrieval-induced forgetting. *Memory & Cognition, 42*, 141–150. https://doi.org/10.3758/s13421-013-0347-0

Abel, M., & Bäuml, K.-H. T. (2017). Testing the context-change account of list-method directed forgetting: The role of retention interval. *Journal of Memory and Language, 92,* 170-182. https://doi.org/10.1016/j.jml.2016.06.009

Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

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*(5), 1063-1087. https://doi.org/10.1037/0278-7393.20.5.1063

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1141-1159. https://doi.org/10.1037/0278-7393.26.5.1141

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*(1), 68-100. https://doi.org/10.1037/0033-295X.102.1.68

Bäuml, K.-H. T. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5*(3), 459-463. https://doi.org/10.3758/BF03208822

Bäuml, K.-H. T. (2019). Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science, 28*(2), 177-182. https://doi.org/10.1177/0963721419827520

Bäuml, K.-H. T., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10*(3), 215-224. https://doi.org/10.1080/09658210143000362

Bäuml, K.-H. T., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In J. T. Wixted, & J. H. Byrne (Eds.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (2nd ed., pp. 27–51). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21048-1

Bäuml, K.-H. T., & Kuhbandner, C. (2007). Remembering can cause forgetting – but not in negative moods. *Psychological Science*, *18*(2), 111-115. https://doi.org/10.1111/j.1467-9280.2007.01857.x

Bäuml, K.-H. T., Meixensperger, S. R., & Hirsch, M. L. (2025). Reinstating memories' temporal context at encoding causes Sisyphus-like memory rejuvenation. *Proceedings of the National Academy of Sciences*, *122*(32), e2505120122. https://doi.org/10.1073/pnas.2505120122

Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition*, *132*(1), 16-21. https://doi.org/10.1016/j.cognition.2014.03.007

Bäuml, K.-H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences, 119*(8), e2114377119. https://doi.org/10.1073/pnas.2114377119

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569. https://doi.org/10.1037/a0017021

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*(2), 438-448. https://doi.org/10.3758/MC.36.2.438

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology*, *19*(4/5), 580-606. https://doi.org/10.1080/09541440701326071

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*(2), 153-170. https://doi.org/10.1016/j.jml.2009.04.004

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553-571. https://doi.org/10.1037/0096-3445.135.4.553

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*(7), 1049-1062. https://doi.org/10.3758/s13421-014-0425-y

Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*(2), 134–140. https://doi.org/10.1037/h0045156

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*(2), 175-191. https://doi.org/10.3758/BF03193146

Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory*, *17*(1), 92-103. https://doi.org/10.1080/09658210802572454

Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 416-436. https://doi.org/10.1037/a0021963

Jonker, T. R., Seli, P., & MacLeod, C.M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review, 120*(4), 852–872. https://doi.org/10.1037/a0034246

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted, & J. H. Byrne (Eds.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (2nd ed., pp. 487-514). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Koessler, S., Engler, H., Riether, C., & Kissler, J. (2009). No retrieval-induced forgetting under stress. *Psychological Science*, *20*(11), 1356-1363. https://doi.org/10.1111/j.1467-9280.2009.02450.x

Kriechbaum, V. M., & Bäuml, K.-H. T. (2023). The critical importance of timing of retrieval practice for the fate of nonretrieved memories. *Scientific Reports*, *13*, 6128. https://doi.org/10.1038/s41598-023-32916-7

Lehmer, E.-M., & Bäuml, K.-H. T. (2018). Part-List cuing can impair, improve, or not influence recall performance: The critical roles of encoding and access to study context at test. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1186-1200. https://doi.org/10.1037/xlm0000517

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*(2), 148-152. https://doi.org/10.1111/1467-9280.00325

Meixensperger, S. R., & Bäuml, K.-H. T. (2025). Items that are Subject to Retrieval-induced Forgetting show Slowed Forgetting over Time. Open Science Framework. https://osf.io/629ft/?view_only=42c5d7898b7245cb96de036c13f69802

Nickl, A. T., & Bäuml, K.-H. T. (2023). Retrieval practice reduces relative forgetting over time. *Memory*, *31*(10), 1412-1424. https://doi.org/10.1080/09658211.2023.2270735

Raaijmakers, J. G., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*(2), 98-122. https://doi.org/10.1016/j.jml.2012.10.002

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318-339. https://doi.org/10.1037/0033-295X.95.3.318

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science*, *20*(9), 1053-1058. https://doi.org/10.1111/j.1467-9280.2009.02415.x

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734-760. https://doi.org/10.1037/0033-295X.103.4.734

Rupprecht, J., & Bäuml, K.-H. T. (2016). Retrieval-induced forgetting in item recognition: Retrieval specificity revisited. *Journal of Memory and Language*, *86*, 97-118. https://doi.org/10.1016/j.jml.2015.09.003

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied*, *8*(2), 127-142. https://doi.org/10.1037/1076-898X.8.2.127

Scheithe, K., & Bäuml, K.-H. T. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien. *Sprache & Kognition*, *14*, 39-43.

Siler, J., & Benjamin, A. S. (2020). Long-term inference and memory following retrieval practice. *Memory & Cognition,* *48*(4), 645-654. https://doi.org/10.3758/s13421-019-00997-3

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 384-397. https://doi.org/10.1037/0278-7393.9.3.384

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. *Psychology of Learning and Motivation*, *62*, 141-194. https://doi.org/10.1016/bs.plm.2014.09.005

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology*, *24*(5), 617-629. https://doi.org/10.1080/20445911.2012.674030

Tandoh, K., & Naka, M. (2007). Durability of retrieval-induced forgetting. *Shinrigaku Kenkyu: The Japanese Journal of Psychology, 78*(3), 310-315. https://doi.org/10.4992/jjpsy.78.310

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252-257. https://doi.org/10.1027/1618-3169.56.4.252

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*(3), 289-335. https://doi.org/10.1016/j.jml.2003.10.003

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

Wixted, J. T. (2022). Absolute versus relative forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1775-1786. https://doi.org/10.1037/xlm0001196

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*(6), 409-415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, *25*(5), 731-739. https://doi.org/10.3758/BF03211316

Figure 1

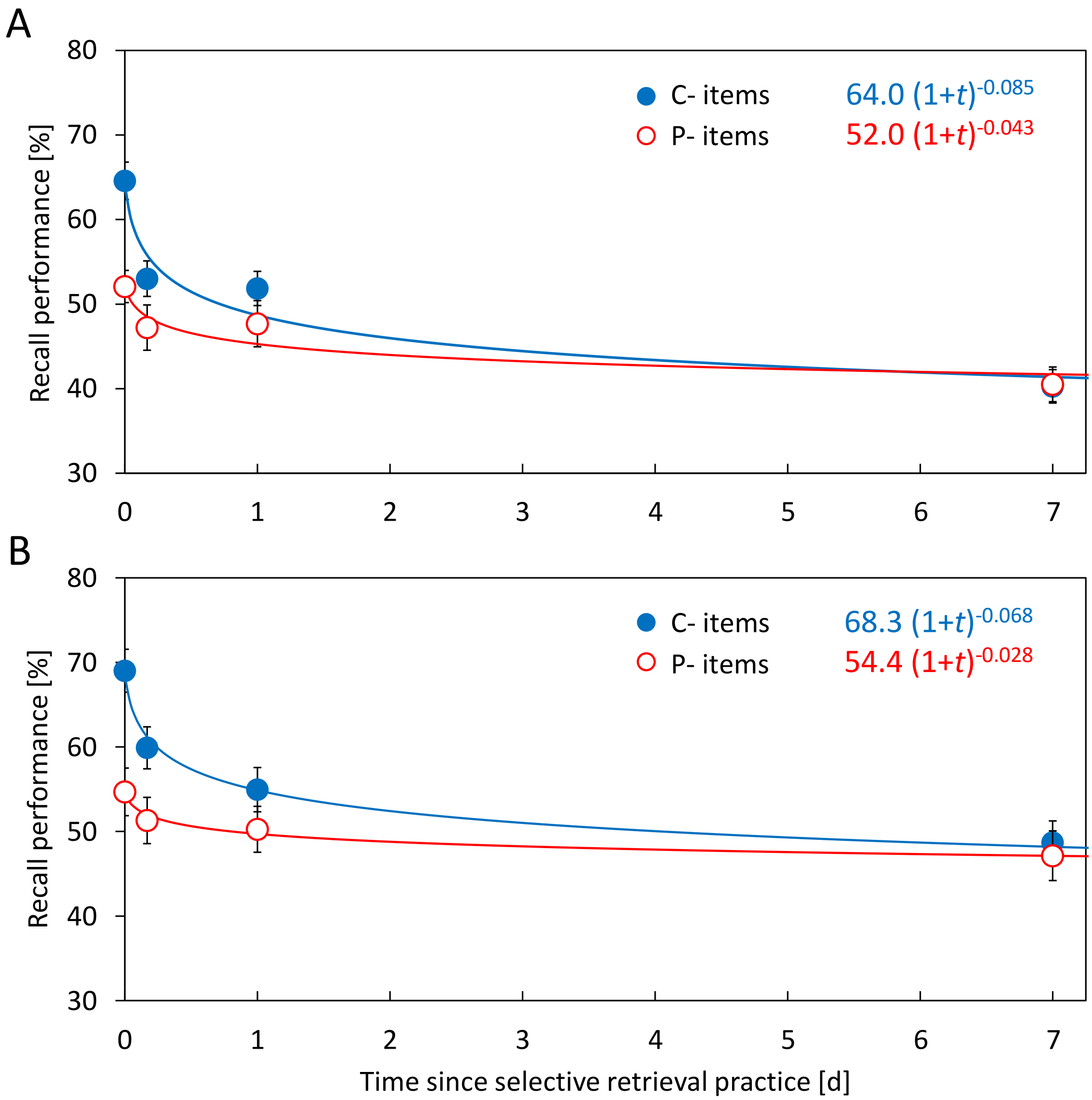
A
B
80
70
60
50
40
30
Recall performance [%]
C- items
P- items
64.0 (1+t)-0.085
52.0 (1+t)-0.043
68.3 (1+t)-0.068
54.4 (1+t)-0.028
0
1
2
3
4
5
6
7
Time since selective retrieval practice [d]

Figure 2

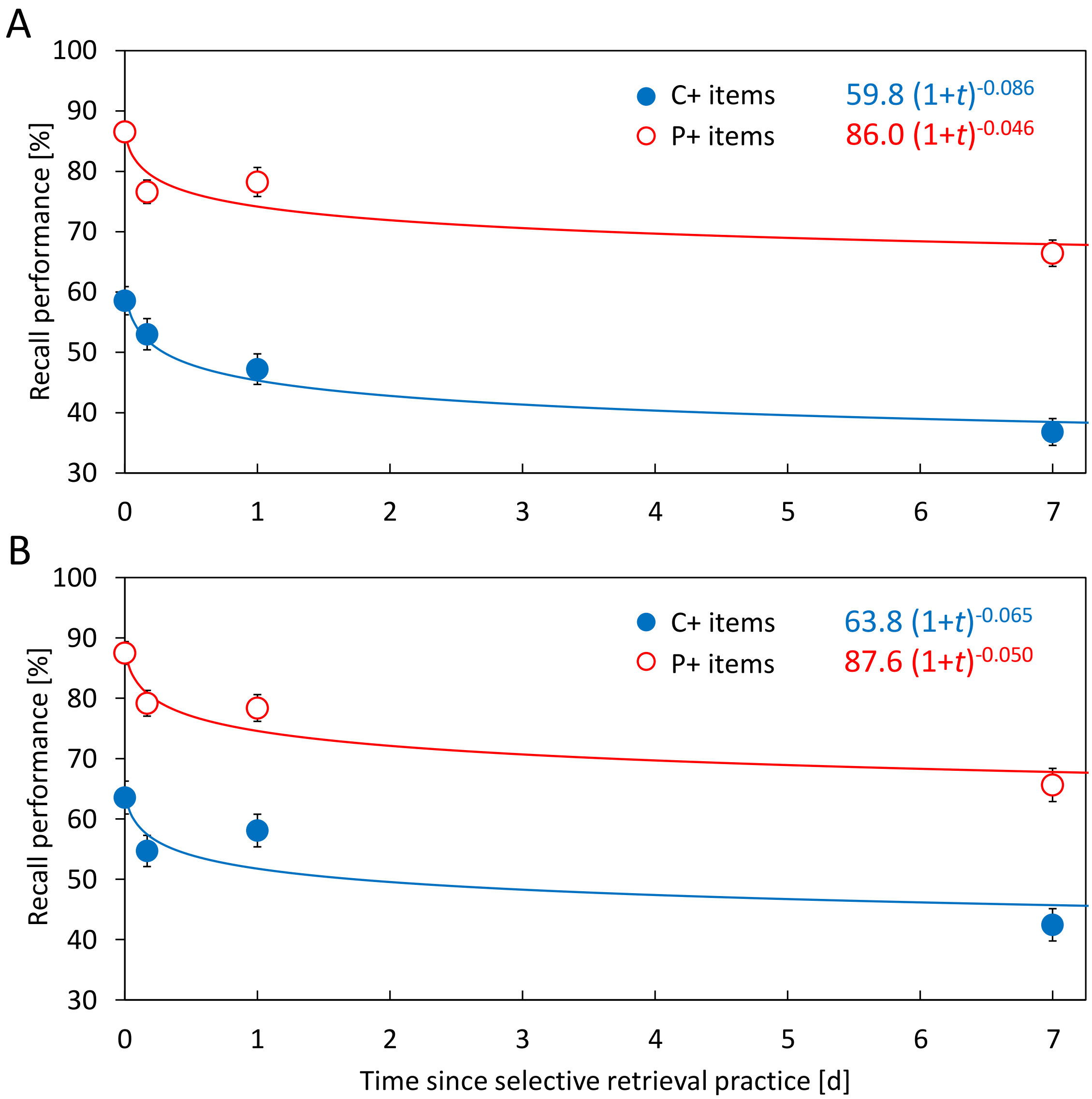
A
100
90
80
70
60
50
40
30
Recall performance [%]
C+ items
P+ items
59.8 (1+t)-0.086
86.0 (1+t)-0.046
0
1
2
3
4
5
6
7
B
C+ items
P+ items
63.8 (1+t)-0.065
87.6 (1+t)-0.050
Time since selective retrieval practice [d]

Figure captions

Figure 1: Percentage of recalled P- and C- items at all four retention intervals are displayed together with the best fitting power functions for the two item types for Experiment 1 (A) and Experiment 2 (B). Error bars represent ± 1 standard error.

Figure 2: Percentage of recalled P+ and C+ items at all four retention intervals are displayed together with the best fitting power functions for the two item types for Experiment 1 (A) and Experiment 2 (B). Error bars represent ± 1 standard error.