Routledge
Taylor & Francis Group

Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Retrieval practice reduces relative forgetting over time

Anna T. Nickl & Karl-Heinz T. Bäuml

To cite this article: Anna T. Nickl & Karl-Heinz T. Bäuml (2023) Retrieval practice reduces relative forgetting over time, Memory, 31:10, 1412-1424, DOI: 10.1080/09658211.2023.2270735

To link to this article: https://doi.org/10.1080/09658211.2023.2270735

MEMORY
2023, VOL. 31, NO. 10, 1412–1424
https://doi.org/10.1080/09658211.2023.2270735

Routledge
Taylor & Francis Group

Retrieval practice reduces relative forgetting over time

Anna T. Nickl and Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University, Regensburg, Germany

ABSTRACT

There is overwhelming evidence in the literature that retrieval practice of studied material can lead to better final recall than restudy of the same material. Far less clear is whether this recall benefit is accompanied by reduced subsequent forgetting over time. This study revisited the issue in two experiments by comparing the effects of retrieval practice – with and without feedback –, restudy, and a no-practice condition on recall across different delay intervals ranging between three minutes and several days. We fitted power functions of time to the recall rates of each practice condition and compared relative forgetting rates between conditions. The comparisons showed that relative forgetting was reduced after retrieval practice relative to restudy, the relative forgetting rate after retrieval practice was unaffected by the presence of feedback, and forgetting after restudy did not differ from the no-practice condition. Together with other findings in the literature, the results provide evidence that retrieval practice reduces relative forgetting over time.

ARTICLE HISTORY
Received 4 April 2023
Accepted 5 October 2023

KEYWORDS
episodic memory; retrieval practice; time-dependent forgetting; forgetting

As already noted by Ebbinghaus (1885) in his pioneering work, forgetting over time follows a distinct course, with a steep drop in memory performance shortly after study that levels off as time between study and test increases. Ever since then, memory research has been concerned with the question of how to reduce such forgetting over time (e.g., Abbott, 1909; Spitzer, 1939), with a large number of the studies focusing on the effects of retrieval practice. In these studies, retrieval practice, where previously studied information is retrieval practiced before a final test, is often compared to restudy, where previously studied information is reexposed for further study. However, despite a long research history on the issue (see Roediger & Karpicke, 2006a), a definitive answer on the question of whether retrieval practice reduces forgetting over time is still missing. The goal of the present study was to contribute to this research question and help gain a more clearcut answer on the issue.

Studies on retrieval practice effects typically include an initial study phase, during which, for instance, words, paired associates, or some prose material are exposed to participants, immediately followed by a practice phase, during which the studied material is either retrieved or restudied; in the retrieval-practice condition, feedback may be provided. After a delay interval of few minutes up to several days, often a final recall test is conducted, in which participants are asked to recall the initially studied material. In most studies, recall in the retrieval practice condition – with and without feedback – has been found to be higher than recall in the restudy condition. Retrieval practice effects have been reported for a wide range of materials, participant groups, and practice protocols (for reviews, see Karpicke, 2017; Roediger & Butler, 2011).

Retrieval practice effects across longer delays

Most of the studies on retrieval practice effects compared final recall of retrieval practiced and restudied material at two delay intervals between practice and test: a short delay of few minutes and a longer delay often between 30 min and several days. It was then of interest whether the decrease in recall performance from the short to the longer delay was similar for retrieval practiced and restudied material, or whether the decrease differed between practice conditions. Typically, the question was addressed by using analysis of variance (ANOVA) and examining the presence of a significant interaction between delay (short, long) and type of practice (retrieval practice, restudy). A significant interaction was taken as evidence that the two practice formats differed in amount of forgetting over time.

The studies comparing the effects of retrieval practice without feedback and restudy across two delay intervals reported mixed results. In some studies, significant interactions between the two factors were found, with a lower amount of forgetting in the retrieval practice than the restudy condition (e.g., Abel et al., 2019, Experiment 1a; Mulligan & Picklesimer, 2016, Experiment 2; Thompson et al., 1978, Experiment 2; Toppino & Cohen, 2009; Wheeler

CONTACT Anna T. Nickl anna.nickl@ur.de Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany

et al., 2003), whereas in other studies no such interactions arose (e.g., Abel et al., 2019, Experiment 2b; Agarwal et al., 2017; Thompson et al., 1978, Experiment 3). However, overall more significant than non-significant interactions have been reported. A similar picture arises for the studies employing retrieval practice with feedback. Some studies reported significant interactions between delay and type of practice with reduced forgetting in the retrieval practice condition (e.g., Abel & Bäuml, 2020; Abel & Roediger, 2018; Mulligan & Peterson, 2015), whereas other studies did not find any significant interactions (e.g., Abel et al., 2019, Experiment 1a; Carrier & Pashler, 1992). Again, the majority of the studies has reported interactions.

One of the few studies that examined retrieval practice effects across more than two delay intervals is the study by Roediger and Karpicke (2006b). In this study, participants were exposed to prose materials that they restudied or retrieval practiced without feedback shortly after study. Five minutes, 2 days, or 1 week after practice, participants were asked to recall as much content as possible from the originally studied material. ANOVA showed a significant interaction between delay and type of practice, with a steeper decrease in recall performance over time for the restudied material.[1] Thus, when using ANOVA to analyze recall performance, at least in tendency, retrieval practice, both with and without feedback, seems to attenuate forgetting relative to restudy.

Using power function analysis to capture time-dependent forgetting

Apart from using ANOVA to examine forgetting over time, alternative methods have been employed to describe such time-dependent forgetting and compare the forgetting across different practice conditions. Time-dependent forgetting usually follows a characteristic trajectory, with a steep initial drop in memory performance followed by a long, more shallow decline in performance, and numerous researchers have fit mathematical functions to corresponding data in order to capture the curvilinear nature of forgetting (e.g., Ebbinghaus, 1885; Wickelgren, 1972, 1974). On the basis of this work, a power function of time, $r(t) = a(1 + ct)^{-b}$, has since been widely accepted to account for forgetting over time, where a is recall at time $t = 0$, c is a scaling unit, and b is the relative rate of forgetting over time.[2] The power function with its focus on relative forgetting contrasts with the focus on absolute forgetting in the ANOVA-based approach, which indicates that the two approaches will not always lead to the same conclusions when comparing time-dependent forgetting across different experimental conditions (see Wixted, 2022). Indeed, equality of absolute forgetting (which would result in a non-significant interaction between delay and type of practice in an ANOVA) can differ dramatically from equality of forgetting in relative terms. For instance, if retention drops from 60 percentage points to 40 percentage points in one condition, and from 30 percentage points to 10 percentage points in another condition, the two conditions show the same amount of absolute forgetting (−20 percentage points), but differ vastly in relative forgetting (−33% versus −67%). Although the power function can describe a wide range of forgetting data well (e.g., Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991, 1997), so far it has rarely been used to compare forgetting after retrieval practice with forgetting after restudy.

To the best of our knowledge, only two studies to date have used the power function to compare forgetting after retrieval practice with forgetting after restudy. In the one study, Carpenter et al. (2008) asked subjects to study obscure facts (Experiments 1 and 2) or vocabulary pairs (Experiment 3). After study, subjects either retrieval practiced the materials with feedback or restudied the materials, once (Experiment 1) or three times in succession (Experiments 2 and 3). After each of six different delay intervals between practice and test (5 min or 1, 2, 7, 14, or 42 days), subjects were then asked to recall one of six different subsets of the studied material. For each subject and practice type, individual power functions were fit to the final recall data. The researchers reported a significantly reduced forgetting rate parameter b after retrieval practice compared to restudy in Experiments 2 and 3, whereas Experiment 1 showed a numerical trend in this direction only. At the same time, ANOVA yielded a significant interaction between delay and type of practice only in Experiment 1, illustrating that ANOVA and power function analysis do not always coincide in their implications regarding possible differences in forgetting over time.

In the other study, Siler and Benjamin (2020, Experiment 2) asked their participants to study pictures of different birds and their taxonomic families. Like Carpenter et al. (2008), they employed a within-subject manipulation of delay interval, restudy and retrieval practice with feedback as types of practice, and delay intervals of few minutes, 1, 7, and 25 days. During practice, some taxonomic bird families were restudied and some other families retrieval practiced followed by feedback. Siler and Benjamin were interested in a number of issues in this experiment. However, of particular interest for the present study was that they conducted an old/new-recognition test for the studied items and fitted the three-parametric power function to the recognition data. Results showed a numerical trend towards a reduced forgetting rate parameter b after retrieval practice with feedback compared to restudy, but the difference in parameter estimates did not reach significance. Also, no significant interaction between delay and type of practice arose when using ANOVA. The two studies by Carpenter et al. (2008) and Siler and Benjamin (2020) thus yielded somewhat mixed results, with a significant difference in forgetting rates between restudy and retrieval practice with feedback

in two of the four experiments, and a non-significant numerical difference between conditions in the remaining experiments.

The present study

In light of the mixed results reported by Carpenter et al. (2008) and Siler and Benjamin (2020), this study aimed to revisit the issue of whether, when using power function analysis, restudy and retrieval practice with feedback lead to different forgetting rate parameters. Going beyond the two previous studies, it was also examined whether a reduced forgetting rate parameter would arise in response to retrieval practice without feedback, which also allows a comparison of forgetting rates between retrieval practice with and without feedback. Somewhat surprisingly, no study has yet examined forgetting rates when retrieval practice is without feedback. Doing so thus will fill a critical empirical gap in the retrieval practice literature. Finally, we aimed to contrast the restudy condition with a no-practice condition in order to examine whether restudy in its own can also slow forgetting.

The results of two experiments are reported designed to address these research questions. Experiment 1 followed up on the studies by Carpenter et al. (2008) and Siler and Benjamin (2020) and included both a restudy and a retrieval-practice-with-feedback condition. In addition, the experiment contained a retrieval-practice-without-feedback condition to allow both a comparison in forgetting rates between restudy and retrieval practice without feedback and a comparison in forgetting rates between the two retrieval practice formats. Because Experiment 1 is the first experiment in the literature that examined relative forgetting when retrieval practice is without feedback, Experiment 2 again included such a condition and compared forgetting rate in a retrieval-practice-without-feedback condition with forgetting rate in a restudy condition. However, instead of the retrieval-practice-with-feedback condition, a no-practice condition was included to examine whether restudy in its own also reduces time-dependent forgetting. In both experiments, a list of paired associates served as study material. Four delay intervals were employed in each experiment: delays of 3 min, 1, 3, and 7 days in Experiment 1 and delays of 3 min, 1, 2, and 3 days in Experiment 2. Delay was manipulated between participants in order to prevent that forgetting rates after longer delays were influenced by the prior recall at earlier delay intervals (see also General Discussion). The results of the two experiments will provide insights into whether (i) retrieval practice – with and without feedback – reduces forgetting rates compared to restudy, (ii) the two retrieval practice formats differ in forgetting rates, and (iii) restudy reduces forgetting compared to a no-practice condition.

Experiment 1

Method

Ethical considerations

All reported studies were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants

144 participants took part in the experiment ($M = 22.32$ years, $SD = 2.59$, range = 18-30 years, 102 female). They were recruited mainly from Regensburg University, as well as by placing online advertisements in students' groups in Germany. Participants were distributed equally across the four between-subjects conditions, yielding $n = 36$ participants per delay condition. Sample size was based on a power analysis conducted in G*Power (version 3.1.9.7; Faul et al., 2007) – with $\alpha = .05$, $\beta = .20$, and $\eta_p^2 = 0.06$, as previous studies often had reported small to medium effect sizes (η_p^2s of 0.05–0.11) for the ANOVA interaction between delay and type of practice (e.g., Abel & Bäuml, 2020; Mulligan & Peterson, 2015; Toppino & Cohen, 2009) – as well as counterbalancing purposes. Doing so, our sample size was also similar to studies from other research areas that also compared relative forgetting across different experimental conditions (e.g., Bäuml & Trißl, 2022). All subjects spoke German as their native language and reported no neurological or psychiatric disease. In both experiments, all subjects gave their spoken informed consent and took part in the experiment in return for either course credit or a compensatory amount of money. There were no participants failing to show up when testing was delayed.

Materials

Materials consisted of 24 unrelated word pairs of concrete nouns, which were drawn from Van Overschelde et al. (2004) and already used in Bäuml et al. (2014). Individual nouns were chosen from different semantic categories, each with 1–2 syllables. Within each pair, one word was always used as the cue word while the other word served as the target word. The list of 24 pairs was divided into three sets of 8 pairs each (A, B, and C). Assignment of sets to practice conditions and order of sets during the practice phase were counterbalanced across participants within each delay condition, resulting in 36 different combinations of type of practice and set sequence. All materials as well as the data from both experiments are available on the Open Science Framework (https://osf.io/e8cg5/?view_only=6865e077babf42fea39ba1cf713107f8).

Design

The experiment followed a 4 (DELAY: 3 min vs. 1 day vs. 3 days vs. 7 days) × 3 (TYPE OF PRACTICE: restudy vs. retrieval practice with feedback vs. retrieval practice without

feedback) mixed design. Delay was manipulated between-subjects, type of practice was manipulated within-subject.

Procedure

Data collection took place via individual zoom meetings. Both participants and experimenters were required to keep their cameras and microphones on during the meetings. For participants in the long delay conditions (1 day, 3 days, or 7 days), the experiment consisted of two sessions that were scheduled for the same time of day (± 2 hrs). Individual participants were tested at various times of day during working hours, but time of day did not vary significantly across delay conditions, $\chi^2(12) = 10.06$, $p = .611$. The software OpenSesame (version 3.3; Mathôt et al., 2012) was used for stimulus presentation and balancing. During sessions, experimenters shared their screen for stimulus presentation and instructed participants orally.

The experiment started with a study phase consisting of two study cycles. During each cycle, all 24 word pairs were presented individually for 5 s. Subjects were asked to remember the pairs for an upcoming test at the end of the experiment. Between the two study cycles and after the second study cycle, subjects sorted triples of two-digit numbers in ascending or descending order for 1 min. After the study phase, the practice phase started. For the restudy condition, subjects were informed that they could now study some of the previously presented word pairs again. Accordingly, 8 pairs (either set A, B, or C) were reexposed for 7 s each. For the two retrieval practice conditions, participants were told that they should try to remember some of the studied word pairs. In the retrieval-practice-without-feedback condition, subjects were shown the cue word and the first two letters of 8 further pairs for 7 s and were asked to type the target word into an empty document that was made accessible on the screen. We chose to use initial-letter cues for the target words in order to increase recall rates. The retrieval-practice-with-feedback condition mirrored the retrieval-practice-without-feedback condition, with the only exception that the retrieval cues were present for 5 s only, during which subjects were asked to type in the pair's target item followed by 2 s during which the complete pair was provided as feedback. Order of practice conditions was counterbalanced across participants.

After practice, participants in the long delay conditions were instructed to count backwards in steps of 7 from a three-digit number for 2 min as a recency control. They were then dismissed and asked to return to their second scheduled zoom meeting 1, 3, or 7 days later. The second session began with a 3 min distractor task during which subjects solved Raven's Standard Progressive Matrices (Raven et al., 2000). Subjects in the 3-min delay condition proceeded to this task immediately after the practice phase. In all delay conditions, subjects performed the final test for all originally studied word pairs. Participants were presented with one cue word at a time for 7 s and were asked to type in the associated target word. The initial letter of each target word was provided as a retrieval cue.

Fitting the power function to the recall rates

We used maximum likelihood methods to fit the three-parametric power function of time, $r(t) = a(1 + ct)^{-b}$, to the recall rates of the three practice conditions using group average data. Time was measured in days since the end of the practice phase. For each practice condition, we generated a separate power function model and compared it to a statistical baseline model that described the recall rates of the four delay conditions as the product of four independent binomial distributions. Parameters were estimated by maximising the likelihood of the power function model. This likelihood was then compared to the likelihood of the baseline model using likelihood ratio, which led to an approximate χ^2-test to examine whether the power function described the recall rates well (Bäuml & Trißl, 2022; Trißl & Bäuml, 2022; see also Riefer & Batchelder, 1988; Wickens, 1982).

Before running such analyses, however, we followed the previous studies by Carpenter et al. (2008) and Siler and Benjamin (2020) and estimated a common scaling parameter c for the three practice conditions. For this, we compared a restricted power function model, in which parameters a and b were allowed to vary freely between conditions but parameter c was restricted to be the same for all three conditions, to a more general model, in which all three parameters in all three conditions were allowed to vary freely. We tested the fit of this restricted model using maximum likelihood methods and a χ^2-test with two degrees of freedom. As it turned out, the restricted model described the data equally well as the more general model, $\chi^2(2) = 0.34$, $p = .844$, with $c = 0.45$ as the best fitting scaling parameter. This parameter estimate was then used for all further analyses. As a result, a χ^2-test with two degrees of freedom was used to evaluate the fit of the power function (with its remaining two free parameters a and b) to the (four) recall rates of a practice condition.

In the next step, we examined whether parameters a and b of the power function varied significantly between two practice conditions, for instance, the retrieval-practice-with-feedback and the restudy condition. For this, we combined the data sets of the two practice conditions and compared the fit of a more general power function model that allowed parameters a and b to vary freely between conditions to that of a more restricted power function model in which either parameter a or parameter b were fixed to be the same for the two conditions. Again, the comparison was based on maximum likelihood methods, resulting in a χ^2-test with one degree of freedom (see also Bäuml & Trißl, 2022; Trißl & Bäuml, 2022). All fitting procedures were written in R (R Core Team, 2021) and implemented in R Studio (RStudio Team, 2020), using optim() from the R package *stats*

(version 4.1.1) with a Nelder-Mead method for maximisation.

Results

Success rates during retrieval practice

Mean recall rates during retrieval practice were .75 (*SD* = .23) in the condition with feedback and .76 (*SD* = .23) in the condition without feedback. A 4 × 2 mixed-factors ANOVA with the between-subjects factor of DELAY (3 min vs. 1 day vs. 3 days vs. 7 days) and the within-subject factor of TYPE OF PRACTICE (retrieval practice with feedback vs. retrieval practice without feedback) produced no main effect of DELAY, $F(3, 140) = 0.48$, $MSE = .08$, $p = .696$, $\eta_p^2 = 0.01$, no main effect of TYPE OF PRACTICE, $F(1, 140) = 0.46$, $MSE = .02$, $p = .498$, $\eta_p^2 < 0.01$, and no interaction between the two factors, $F(3, 140) = 2.04$, $MSE = .02$, $p = .111$, $\eta_p^2 = 0.04$, indicating that success rates did not vary across practice condition and delay.

Final recall: analysis of variance

Figure 1 shows the percentage of recalled target items for all three practice conditions across the four delay intervals. Using ANOVA, we compared recall rates in the restudy condition with those in the retrieval-practice-with-feedback and the retrieval-practice-without-feedback conditions, and finally compared recall rates between the two retrieval practice conditions. Regarding the comparison between restudy and retrieval practice with feedback, a 4 × 2 mixed-factors ANOVA with the between-subjects factor of DELAY (3 min vs. 1 day vs. 3 days vs. 7 days) and the within-subject factor of TYPE OF PRACTICE (restudy vs. retrieval practice with feedback) revealed a main effect of DELAY, $F(3, 140) = 49.93$, $MSE = .10$, $p < .001$, $\eta_p^2 = 0.52$, with lower recall after longer than shorter delay, a main effect of TYPE OF PRACTICE, $F(1, 140) = 14.34$, $MSE = .02$, $p < .001$, $\eta_p^2 = 0.09$, with lower recall in the restudy condition, as well as an interaction between the two factors, $F(3, 140) = 3.06$, $MSE = .02$, $p = .030$, $\eta_p^2 = 0.06$, reflecting the fact that the detrimental effect of delay on recall was larger in the restudy than the retrieval-practice-with-feedback condition.

The comparison between restudy and retrieval practice without feedback showed a main effect of DELAY, $F(3, 140) = 46.88$, $MSE = .10$, $p < .001$, $\eta_p^2 = 0.50$, no main effect of TYPE OF PRACTICE, $F(1, 140) = 0.39$, $MSE = .02$, $p = .535$, $\eta_p^2 < 0.01$, and a marginally significant interaction between the two factors, $F(3, 140) = 2.39$, $MSE = .02$, $p = .071$, $\eta_p^2 = 0.05$, indicating a trend towards a larger detrimental effect of delay in the restudy condition. Finally, the comparison between the two retrieval practice conditions revealed a main effect of DELAY, $F(3, 140) = 40.83$, $MSE = .10$, $p < .001$, $\eta_p^2 = 0.47$, a main effect of TYPE OF PRACTICE, $F(1, 140) = 8.21$, $MSE = .02$, $p = .005$, $\eta_p^2 = 0.06$, with overall higher recall after retrieval practice with feedback, but no interaction between the two factors, $F(3, 140) = 0.38$, $MSE = .02$, $p = .765$, $\eta_p^2 < 0.01$, pointing to a similarly sized delay effect for the two retrieval practice formats.

Final recall: power function analysis

When fitting the power function to the recall rates of each single practice condition, results showed that the function described the recall rates of all three practice conditions well (Figure 1), as is reflected in $\chi^2(2)$ values of ≤ 4.82, $p \geq .090$, for the single practice conditions (Table 1). Indeed, the power function explained most of the variance in the data, as is represented in r^2 values of $\geq .983$ in the three conditions. We compared parameters *a* and *b* of the function in the restudy condition with those in the two retrieval practice conditions, and finally compared parameters between the two retrieval practice formats. Regarding the comparison between restudy and retrieval practice with feedback, estimates for parameter *a* were found to not differ significantly between conditions, $\chi^2(1) < 0.01$, $p = .920$, suggesting similar initial recall levels in the two conditions. In contrast, the two conditions differed in parameter *b*, $\chi^2(1) = 10.15$, $p = .001$, reflecting a higher forgetting rate for restudy than retrieval practice with feedback.

Regarding the comparison between restudy and retrieval practice without feedback, parameter *a* was found to be significantly smaller for retrieval practice than restudy, $\chi^2(1) = 4.58$, $p = .032$, indicating lower initial recall in the retrieval practice condition. Parameter *b* was also smaller in the retrieval practice condition, $\chi^2(1) = 6.92$, $p = .008$, reflecting a higher forgetting rate for restudy than retrieval practice without feedback. Finally, when comparing the two retrieval practice conditions, parameter *a* was found to be smaller in the retrieval-practice-without-feedback condition, $\chi^2(1) = 4.40$, $p = .036$, indicating lower initial recall in this condition, whereas forgetting rate parameter *b* did not differ between the two retrieval practice formats, $\chi^2(1) = 0.17$, $p = .680$.

Discussion

Time-dependent forgetting in all three practice conditions was well described by the three-parametric power function of time, which is consistent with the previous studies by Carpenter et al. (2008) and Siler and Benjamin (2020). Regarding the comparison between restudy and retrieval practice with feedback, our findings replicate the findings from two of the three experiments reported in Carpenter et al. (2008), in that a lower forgetting rate parameter (*b*) was found in the retrieval-practice-with-feedback than in the restudy condition. Going beyond this previous study, the results also show that the reduction in forgetting rate parameter generalizes from retrieval practice with feedback to retrieval practice without feedback. Moreover, forgetting rates were comparable in size between the two retrieval practice

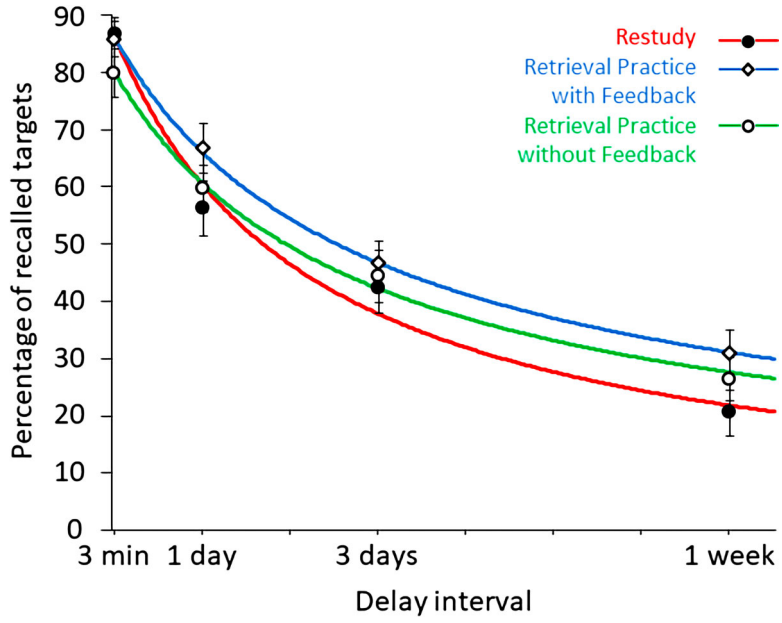

Figure 1. *Results of Experiment 1.* Recall rates for the three practice conditions are displayed along with the best-fitting power functions. The power function described the recall rates well, showing a larger forgetting rate parameter in the restudy condition than in the two retrieval practice conditions, which did not differ in forgetting rates (see Table 1). Error bars represent ± 1 standard error.

formats. Retrieval practice – with and without feedback – thus seems to slow time-dependent forgetting.

As pointed out by both Carpenter et al. (2008) and Wixted (2022), due to the difference in focus on absolute versus relative forgetting, the conclusions drawn from ANOVAs and power function analyses do not necessarily coincide when it comes to comparing time-dependent forgetting across experimental conditions. Nevertheless, we found the results from the ANOVAs to largely mimic those from the power function analysis. Consistently, ANOVA revealed a larger amount of (absolute) time-dependent forgetting for restudy compared to retrieval practice with feedback, at least a trend towards a larger amount of forgetting for restudy compared to retrieval practice without feedback, and no difference in amount of forgetting between the two retrieval practice formats.

The finding of a reduced (relative) forgetting rate after retrieval practice without feedback relative to restudy is the first of this kind in the literature on retrieval practice effects, so we aimed to replicate the result in Experiment 2. Because free recall, rather than initial-letter cued recall – as it has been employed in Experiment 1 –, is the typical procedure for final recall of paired associates in studies on retrieval practice effects (e.g., Roediger & Karpicke, 2006a; see also Carpenter et al., 2008), in Experiment 2 the target items' initial letters were no longer provided as retrieval cues at test and recall was rather conducted in the absence of any item-specific cues. We also replaced the retrieval-practice-with-feedback condition by a no-practice condition to examine whether restudy in itself also has a retarding effect on time-dependent forgetting. Finally, instead of the 7-days delay condition we included a 2-days delay condition into the experimental setup. This was done because forgetting curves tended to stabilize by Day 3 and thus more information about the underlying forgetting rate may be gained if a shorter delay interval was included.

Experiment 2

Method

Participants

144 new participants were recruited for the experiment (M = 22.43 years, SD = 2.91, range = 18–30 years, 86 female), mainly from Regensburg University but also by placing online advertisements in students' groups in Germany. Participants were distributed equally across the four between-subjects conditions, yielding n = 36 participants per delay condition. Sample size followed Experiment 1. Like in Experiment 1, there were no participants failing to show up when testing was delayed.

Materials

24 new unrelated word pairs of concrete nouns from different semantic categories (1–2 syllables) were used, drawn from Van Overschelde et al.'s (2004) norms. Pairs were partially sampled from Bäuml et al. (2014). Like in Experiment 1, the pairs were divided into three sets of 8 pairs each (A, B, and C). Assignment of sets to practice conditions and order of sets during the practice phase were again counterbalanced across participants within each delay condition. Experiment 2 had only 12 different combinations of type of practice and set sequence, as only two (instead of all three) sets were reexposed during practice.

Table 1. Best power function fits and explained variance for Experiment 1.

Condition	a	b	$\chi^2(2)$	*p*	r^2
Restudy	0.863	0.965	4.82	.090	.983
Retrieval Practice with Feedback	0.860	0.714	0.08	.961	.999
Retrieval Practice without Feedback	0.799	0.746	0.88	.644	.995

Design

Again, we used a 4 (DELAY: 3 min vs. 1 day vs. 2 days vs. 3 days) × 3 (TYPE OF PRACTICE: restudy vs. retrieval practice without feedback vs. no practice) mixed design. Delay was manipulated between-subjects, type of practice was manipulated within-subject.

Procedure

Experiment 2 closely followed Experiment 1 with the following exceptions: (a) between the two study cycles and after the second cycle, we used a different 1 min distractor task, during which subjects first added up the digits for each of two two-digit numbers and then either added or subtracted the two interim results; (b) instead of the-retrieval-practice-with-feedback condition, we introduced a no-practice condition; the set of items assigned to this condition (A, B, or C) was not presented again after the two initial study cycles; (c) during the final test, subjects were asked to recall the targets in the absence of any item-specific retrieval cues.

Fitting the power function to the recall rates

The fitting procedure was identical to the one employed in Experiment 1. Again, before running more detailed analyses, we estimated a common scaling parameter *c* for the three practice conditions using the same procedure as in Experiment 1. Like in Experiment 1, the restricted power function model with a common scaling parameter *c* for the three practice conditions described the recall rates equally well as the more general power function model in which the parameter varied freely across conditions, $\chi^2(2) = 1.70$, $p = .427$, with $c = 27.78$ as the best fitting parameter. Like in Experiment 1, this parameter estimate was used for all further analyses.

Results

Success rates during retrieval practice

Mean recall rate during retrieval practice was .83 (*SD* = .20). Recall performance did not vary across delay conditions, $F(3, 140) = 0.37$, $MSE = .04$, $p = .778$, $\eta_p^2 < 0.01$.

Final recall: analysis of variance

Figure 2 shows the percentage of recalled target items for all three practice conditions across the four delay intervals. Using ANOVA, we compared recall rates in the restudy condition with those in the retrieval-practice-without-feedback condition as well as with those in the no-practice condition. For the comparison between restudy and retrieval practice without feedback, a 4 × 2 mixed-factors ANOVA with the between-subjects factor DELAY (3 min vs. 1 day vs. 2 days vs. 3 days) and the within-subject factor TYPE OF PRACTICE (restudy vs. retrieval practice without feedback) revealed a main effect of DELAY, $F(3, 140) = 10.61$, $MSE = .13$, $p < .001$, $\eta_p^2 = 0.19$, with lower recall after longer delay, no main effect of TYPE OF PRACTICE, $F(1, 140) = 2.37$, $MSE = .02$, $p = .126$, $\eta_p^2 = 0.02$, and a significant interaction between the two factors, $F(3, 140) = 3.06$, $MSE = .02$, $p = .030$, $\eta_p^2 = 0.06$, indicating that the detrimental effect of delay on recall rates was larger in the restudy condition. Comparing restudy and no-practice conditions, we found a main effect of DELAY, $F(3, 140) = 14.61$, $MSE = .11$, $p < .001$, $\eta_p^2 = 0.24$, a main effect of TYPE OF PRACTICE, $F(1, 140) = 101.31$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.42$, with higher recall in the restudy condition, but no significant interaction between the two factors, $F(3, 140) = .94$, $MSE = .03$, $p = .422$, $\eta_p^2 = 0.02$.

Final recall: power function analysis

Again, the power function described the recall rates in the three practice conditions well (Figure 2), as is reflected in $\chi^2(2)$ values of ≤ 2.52, $p \geq .284$, for the single conditions (Table 2). Consistent with this finding, in all three conditions the power function also explained most of the variance in the data, as is reflected in r^2 values $\geq .975$ in the three conditions. Comparing parameters *a* and *b* between restudy and the retrieval-practice-without-feedback condition, we found parameter *a* to not differ between the two conditions, $\chi^2(1) = 2.70$, $p = .100$, suggesting similar initial recall, whereas forgetting rate parameter *b* was found to be higher in the restudy than the retrieval-practice-without-feedback condition, $\chi^2(1) = 7.96$, $p = .005$. Comparing parameters *a* and *b* between restudy and the no-practice condition, we found parameter *a* to be smaller in the no-practice condition, $\chi^2(1) = 35.15$, $p < .001$, suggesting lower initial recall in this condition, whereas no difference was found in parameter *b*, $\chi^2(1) = 2.83$, $p = .093$, pointing to similar forgetting rates in the two conditions.[3]

Discussion

Like in Experiment 1, the power function of time described time-dependent forgetting in all three practice conditions well. In particular, we replicated the finding of Experiment 1 of a smaller forgetting rate parameter after retrieval practice without feedback relative to restudy. The finding supports the view that not only retrieval practice with feedback but also retrieval practice without feedback may reduce time-dependent forgetting, which indicates that successful retrieval during practice potentiates later recall at test (Glover, 1989; Halamish & Bjork, 2011; Pashler et al., 2005). Additionally, we found that restudy on its own did not significantly reduce the forgetting

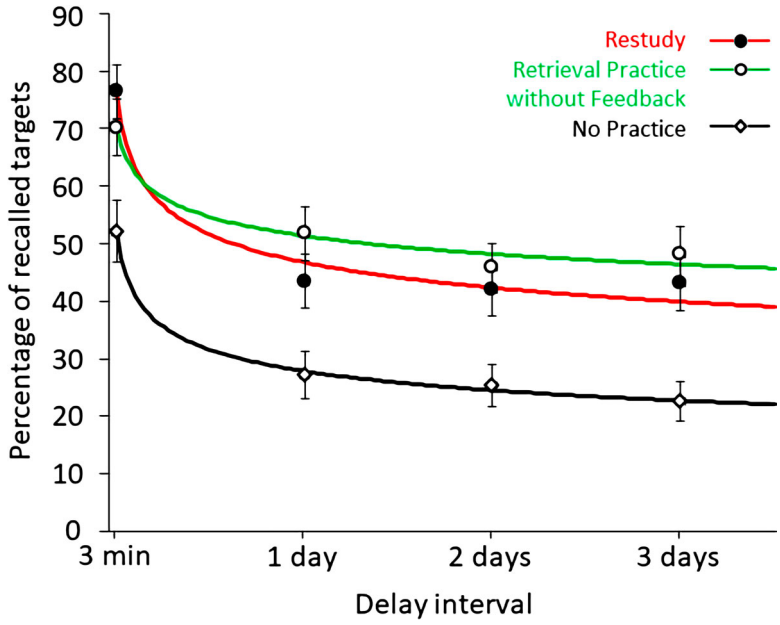

Figure 2. *Results of Experiment 2.* Recall rates for the three practice conditions are displayed along with the best-fitting power functions. The power function described the recall rates well, showing a larger forgetting rate parameter in the restudy than the retrieval practice without feedback condition. Forgetting rates did not differ between the restudy and the no-practice conditions (see Table 2). Error bars represent ± 1 standard error.

rate parameter compared to a no-practice condition, even though the parameter value was at least numerically reduced after restudy. The ANOVA results turned out to be analogous to those of the power function analyses. Consistently, a significant interaction between delay and type of practice emerged for the comparison between restudy and retrieval practice without feedback, indicating less (absolute) forgetting after retrieval practice than restudy. Regarding the comparison between restudy and the no-practice condition, no significant interaction was found, suggesting similar amounts of (absolute) forgetting in the two conditions.

Additional analyses

Arithmetical versus Geometrical Averaging. Using arithmetical averaging to analyze group data can produce averaging artifacts, for example, a group function with mathematical properties that differ from those of the individual subject data (Estes, 1956; Sidman, 1952). To evaluate a possible contribution of averaging artifacts to the pattern of results presented above, we re-ran all analyses using geometrically averaged data (see Anderson & Tweney, 1997; Wixted & Ebbesen, 1997). For both experiments, the pattern of results remained largely the same when using geometrical rather than arithmetical averaging of individual recall rates.[4]

Like with the arithmetically averaged data, the power function described the recall rates for all three conditions of each experiment well (see Table 3). In Experiment 1, estimates for parameter *a* did not differ for restudy and retrieval practice with feedback, $\chi^2(1) = 0.02$, $p = .887$, whereas estimates for forgetting rate parameter *b* did, $\chi^2(1) = 11.20$, $p < .001$. Comparing restudy and retrieval practice without feedback, both parameter *a* and parameter *b* were found to be higher after restudy than retrieval practice (*a*: $\chi^2(1) = 5.58$, $p = .018$, *b*: $\chi^2(1) = 8.24$, $p = .004$). Finally, regarding the two retrieval practice conditions, parameter *a* was found to be higher after feedback, $\chi^2(1) = 5.19$, $p = .023$, whereas forgetting rates did not differ between conditions, $\chi^2(1) = 0.10$, $p = .752$. In Experiment 2, parameter *a* did not differ between restudy and retrieval practice without feedback, $\chi^2(1) = 2.88$, $p = .090$, but forgetting rate parameter *b* was larger after restudy, $\chi^2(1) = 8.66$, $p = .003$. Comparing restudy and the no-practice condition, parameter *a* was found to be higher in the

Table 2. Best power function fits and explained variance for Experiment 2.

Condition	a	b	$\chi^2(2)$	*p*	r^2
Restudy	0.766	0.147	2.52	.284	.975
Retrieval Practice without Feedback	0.705	0.094	1.06	.589	.975
No Practice	0.526	0.189	0.20	.905	.998

Table 3. Best power function fits and explained variance for geometrically averaged data.

Condition	a	b	$\chi^2(2)$	*p*	r^2
Experiment 1					
Restudy	0.854	0.993	5.62	.060	.982
Retrieval Practice with Feedback	0.851	0.727	0.07	.966	.999
Retrieval Practice without Feedback	0.782	0.751	0.77	.681	.996
Experiment 2					
Restudy	0.743	0.155	2.49	.288	.976
Retrieval Practice without Feedback	0.677	0.097	0.70	.705	.983
No Practice	0.492	0.191	0.26	.878	.997

restudy condition, $\chi^2(1) = 36.84$, $p < .001$, whereas forgetting rates did not differ, $\chi^2(1) = 1.87$, $p = .171$. In light of these results, it seems unlikely that the pattern reported above has been caused by averaging artifacts (Anderson & Tweney, 1997; Wixted & Ebbesen, 1997).

Forgetting Rates for Successfully and Unsuccessfully Practiced Items. Studies on retrieval practice effects in the absence of feedback have shown that mostly the successfully practiced items are recalled at test (Glover, 1989; Halamish & Bjork, 2011; Pashler et al., 2005). A similar pattern can arise when feedback is provided, although previously unsuccessfully practiced items may benefit more from feedback than successfully practiced items (Butler et al., 2008; Butler & Roediger, 2008; Pashler et al., 2005). On the basis of these findings, one may expect differences in forgetting rates between successfully and unsuccessfully practiced items, with successfully practiced items showing reduced forgetting relative to unsuccessfully practiced items.

Table 4 reports the percentage of items recalled on the final test across the four delay intervals separately for successfully and unsuccessfully practiced items. As expected, successfully practiced items showed higher recall levels than unsuccessfully practiced items, and both item types showed typical forgetting over time. When fitting the power function to the recall rates of the two item types, results showed a numerically reduced forgetting parameter *b* for the successfully practiced items relative to the unsuccessfully practiced items, both when retrieval practice was with and when it was without feedback (see again Table 4), which again is consistent with a priori expectations.

In the present experiments, success rates during retrieval practice were rather high – .75 (with feedback) and .76 (without feedback) in Experiment 1, and .83 in Experiment 2 (without feedback) –, so that only relatively few items were not successfully practiced. From these unsuccessfully practiced items, even in the short 3-min delay condition, less than a quarter of the items were recalled at test when feedback was absent, and only about half of these items were recalled when feedback was provided, meaning that, on average, less than one unsuccessfully practiced item was recalled at test per participant. Because numbers even decreased with increasing delay, we did not run any statistical analysis on these recall rates and thus did also waive comparing forgetting rates statistically between successfully and unsuccessfully practiced items. Future work may revisit the issue using experimental setups that result in lower success rates during practice (see also below).

General discussion

Consistent with the results of many previous studies in the literature (e.g., Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991, 1997), the results from the present experiments show that time-dependent forgetting can be well described by a power function of time. Besides, the comparison of the function's forgetting rate parameter across the two experiments' different practice conditions revealed a couple of interesting findings. First, the forgetting rate was reduced after retrieval practice with feedback relative to the restudy condition (Experiment 1). Second, the forgetting rate was also reduced after retrieval practice without feedback relative to restudy (Experiments 1 and 2). Third, forgetting rates after retrieval practice were unaffected by whether feedback was provided during practice or not (Experiment 1). And fourth, the forgetting rate after restudy did not differ significantly from the forgetting rate when practice was absent (Experiment 2).

Carpenter et al. (2008) and Siler and Benjamin (2020) had already employed power function analysis to examine whether retrieval practice with feedback reduces time-dependent forgetting relative to restudy. Whereas Carpenter et al. found a corresponding reduction in forgetting rate in two of the three experiments they reported (Experiments 2 and 3), Siler and Benjamin did not find a difference in their experiment. The results of the present experiments thus are consistent with those from Carpenter et al.'s Experiments 2 and 3, indicating that retrieval practice with feedback may indeed reduce forgetting rates relative to restudy. At least numerically, the present results are also consistent with the results from the two experiments that did not find a significant reduction in forgetting rate in response to retrieval practice with feedback – Experiment 1 in Carpenter et al.'s study and Experiment 2 reported in Siler and Benjamin – , both of which showed a numerical difference in the same direction.

The present experiments are the first ones in the literature reporting power function analysis of time-dependent forgetting in response to retrieval practice without feedback. The results mirrored those for retrieval practice with feedback and indicate that also in the absence of feedback retrieval practice can reduce forgetting rates. Retrieval practice with feedback typically induces higher recall rates on a later memory test than retrieval practice without feedback, and this pattern also emerged in the present experiments. This difference in recall level, however, was not accompanied by a difference in forgetting rate. Rather, retrieval practice induced quite similar forgetting rates regardless of whether feedback was provided during practice or not, indicating that successful retrieval during practice potentiated successful recall on the final test.

Recall rates in the present study were not only analyzed using power function analysis but also ANOVA. Because the two types of analysis are based on very different views of forgetting – relative forgetting in the case of power function analysis and absolute forgetting in the case of ANOVA (see above and Wixted, 2022) – results on comparisons of forgetting in different experimental conditions can well differ between methods (Carpenter et al., 2008; see also Wixted, 2022). Still, in the present study, the conclusions arising from ANOVA largely coincided with those from power function analysis. Thus,

Table 4. Mean final recall and best fitting parameter estimates for successfully and unsuccessfully practiced items.

Experiment 1	3 min	1 day	3 days	7 days	a	b
Retrieval Practice with Feedback						
Successfully Practiced	.94	.77	.56	.38	0.942	0.614
Unsuccessfully Practiced	.59	.36	.18	.10	0.585	1.308
Retrieval Practice without Feedback						
Successfully Practiced	.94	.73	.58	.35	0.939	0.646
Unsuccessfully Practiced	.22	.07	.09	.01	0.203	1.505
Experiment 2	**3 min**	**1 day**	**2 days**	**3 days**	**a**	**b**
Retrieval Practice without Feedback						
Successfully Practiced	.84	.62	.53	.56	0.846	0.101
Unsuccessfully Practiced	.14	.02	.09	.04	0.139	0.256

also when employing ANOVA, retrieval practice reduced the forgetting relative to restudy and did so similarly in the presence and absence of feedback. The present ANOVA findings thus parallel findings from many previous studies, which also found reduced (absolute) forgetting in response to retrieval practice with feedback (e.g., Abel & Roediger, 2018; Mulligan & Peterson, 2015) and reduced (absolute) forgetting in response to retrieval practice without feedback (e.g., Roediger & Karpicke, 2006b; Wheeler et al., 2003). In consequence, they disagree with the other studies that did not report significant interactions between delay and type of practice (e.g., Agarwal et al., 2017; Carrier & Pashler, 1992).

The observation that retrieval practice can slow forgetting over time ties in with related findings from the literature suggesting that retrieval practice can reduce retroactive interference (e.g., Halamish & Bjork, 2011), interitem interference (e.g., Kliegl & Bäuml, 2016), and even (list-method) directed forgetting (Abel & Bäuml, 2016). Several accounts of retrieval practice effects have been proposed to explain some of these findings. For instance, the elaborative retrieval account, according to which retrieval practice can activate elaborative information related to the target items (e.g., Carpenter, 2009), has been suggested to explain the reduction in forgetting over time. The episodic context account, according to which retrieval practice can create an enriched set of contextual cues (Karpicke et al., 2014), has been suggested to explain the reductions in time-dependent forgetting, intralist interference, and directed forgetting. Based on a very different set of assumptions, the bifurcation model of retrieval practice has been argued to explain both the reduction in forgetting over time and the reduction in retroactive interference (Halamish & Bjork, 2011; Kornell et al., 2011). Last not least, retrieval practice might also be assumed to induce memory consolidation in itself (e.g., Antony et al., 2017) or be interpreted as an indicator of subjective relevance (following Wixted, 2022), with both views suggesting an attenuation of time-dependent forgetting after retrieval practice. The present study was not designed to improve our understanding of which mechanism(s) underlie(s) retrieval practice effects. The results, however, strengthen the point that any candidate mechanism should explain why time-dependent forgetting is reduced in response to retrieval practice and, at least under certain circumstances, the reduction in forgetting can be similar in the presence versus absence of feedback during practice.

Previous studies have shown that feedback predominantly impacts items that were not successfully retrieved during practice (Butler et al., 2008; Pashler et al., 2005; see also Soderstrom et al., 2016) and that the effects vary for example depending on type (Pashler et al., 2005) and timing of feedback (Butler et al., 2007; Butler & Roediger, 2008). Feedback during retrieval practice has also been found to reverse initial retrieval practice benefits over restudy into restudy benefits, at least when success rates during retrieval practice are relatively low (Pastötter & Bäuml, 2016; Storm et al., 2014; Racsmány et al., 2020), indicating that success rate during retrieval practice together with feedback can have moderating effects on long-term retention. In the present experiments, success rates during retrieval practice were relatively high, leaving not much room for positive effects of feedback on overall recall performance. Consistently, only few unsuccessfully practiced items were recalled on the final test, suggesting that the positive effects of retrieval practice mostly pertained to successfully practiced items (see Additional Analysis above). Future studies may examine the influence of feedback on forgetting rates in more detail, for instance, when lower success rates during retrieval practice are implemented and there is thus more room for feedback to influence recall performance. In such case, the forgetting rate might increase relative to when feedback during retrieval practice is absent, at least if the effect of feedback on unsuccessfully practiced items mimicked effects of restudy on studied items.

The relatively high success rates during retrieval practice in the present study reflect the fact that a relatively easy test format during retrieval practice was employed, in which participants were provided the cue item of each paired associate together with the target item's first two initial letters serving as retrieval cues. The literature on retrieval practice effects suggests that the effects are generally more pronounced if retrieval practice is more demanding (Carpenter & Delosh, 2006; Glover, 1989; see also Bjork, 1994). More demanding retrieval practice tasks, however, may not only influence the size of the

retrieval practice effect but may also influence forgetting rates, which may be further reduced relative to restudy if impoverished retrieval cues were provided during practice. Future work may address this issue and directly compare forgetting rates after more difficult versus more easy retrieval practice tasks.

Experiment 2 of the present study compared forgetting rates between a restudy condition and a condition in which practice was absent. Results showed no significant difference in forgetting rates but a numerically reduced forgetting rate after restudy. Wixted (2022) recently reanalyzed data from several previous studies, all of which had examined the role of item strength – operationalised through differences in items' study time or differences in number of item presentations during study – for the amount of time-dependent forgetting. Using power function analysis, he found in all cases that stronger items showed a numerically reduced forgetting rate relative to weaker items.[5] On the basis of these findings, the numerical difference in forgetting rates between restudy and the no-practice condition observed in the present study might indicate that, like other forms of increasing items' memory strength, restudy – conducted in a separate experimental phase after study – also reduces forgetting rates. Further studies, preferably with increased statistical power, should revisit the issue and provide a more definitive answer on whether restudy in its own can reduce forgetting over time.

Both Carpenter et al. (2008) and Siler and Benjamin (2020) employed a within-subject design, in which each subject was tested on a different subset of the studied items across the different delay conditions. In contrast, in the present study, a between-subjects design was used, in which each subject was tested on all studied items in only one of the delay conditions. Both designs come with a mixture of advantages and disadvantages. The within-subject design employed by Carpenter et al. and Siler and Benjamin has the advantage of getting individual estimates of forgetting rates, so that no averaging of recall rates across subjects is required before fitting the function to recall data, and thus no averaging artifacts can arise. This advantage, however, comes with the potential problem that individual function fits can be poor (see Carpenter et al., 2008) and thus bias the results. Also, recall of some items at early delay intervals can improve recall of other items at later delay intervals (see Bäuml & Trißl, 2022), thus possibly leading to an underestimation of forgetting rates. In the between-subjects design employed in the present study, the problems of poor fits of individual recall rates and potential influences of recall at early delays on recall at later delays cannot arise. However, averaging artifacts could emerge, which can distort the results (e.g., Estes, 1956). We addressed this problem in the present study through reanalysis of recall rates using geometric averaging, which, if results are similar to those with arithmetical averaging, effectively rules out the problem. Ideally, results would just not vary much with employed design, and the findings reported by Carpenter et al. (2008) using the within-subject design together with those reported in the present study using the between-subjects design may roughly point into such direction. Still, future work should address the issue more directly by examining time-dependent forgetting after retrieval practice when using both a within-subject and a between-subjects design.

To conclude, employing power function analysis, the findings of the present study indicate that retrieval practice can reduce forgetting over time relative to restudy. This finding arose both when retrieval practice was accompanied by feedback and when feedback was absent. Moreover, the two retrieval practice formats led to very similar forgetting rates, indicating that, at least under certain circumstances, the presence of feedback does not influence rates of forgetting over time. The findings impose important restrictions on accounts of retrieval practice effects and thus may help to uncover the mechanisms mediating this powerful beneficial effect on memory performance.

Notes

1. A few other studies measured retention after retrieval practice without feedback across more than two delays, but did not include a restudy condition. Instead, they compared recall in the retrieval practice condition to recall in a no-practice condition, which introduces a confound with total exposure to material (Chan, 2010; Runquist, 1983; Slamecka & Katsaiti, 1988; Spitzer, 1939).
2. A two-parametric version of the power function of the form, $r(t) = at^{-b}$, has also seen wide-spread use in the forgetting literature (e.g., Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991, 1997). Under most conditions, the two-parametric and three-parametric versions behave very similarly and lead to largely the same conclusions (see Wixted, 2004).
3. The parameter estimates for b were considerably higher in values in Experiment 1 than Experiment 2. This is related to the fact that the two experiments also differed considerably in best fitting scaling parameter c. Indeed, Experiment 1 showed a relatively low parameter c with relatively high parameters b, whereas Experiment 2 showed a relatively high parameter c with relatively low parameters b. When fitting the two-parametric power function model, $r(t) = at^{-b}$, in which no scaling parameter is included, to the recall rates of the two experiments, the values of parameter b were found to be much more similar between experiments. In Experiment 1, estimates of b were 0.113 for restudy, 0.084 for retrieval practice with feedback, and 0.089 for retrieval practice without feedback; in Experiment 2, estimates of b were 0.085 for restudy, 0.054 for retrieval practice without feedback, and 0.110 for the no-practice condition.
4. Like for the arithmetically averaged data, we estimated a common c parameter also for each experiment's geometrically averaged data that we then used for all further analyses. Again, the two identified parameters differed clearly between experiments (0.48 for Experiment 1 and 28.13 for Experiment 2), but were numerically quite similar to the ones estimated for the arithmetically averaged data (0.45 for Experiment 1 and 27.78 for Experiment 2).
5. Wixted (2022) did not report any statistical analysis on whether forgetting parameters differed between stronger and weaker

items. The conclusions therefore rest on the found numerical differences in forgetting rates between stronger and weaker items.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

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

Credit authorship contribution statement

A. T. Nickl: Data curation, Methodology, Visualization, Formal analysis, Writing – original draft, Investigation **K.-H. T. Bäuml:** Conceptualization, Methodology, Supervision, Resources, Formal analysis, Writing – review & editing, Investigation

This research is part of A. T. Nickl's dissertation.

Declaration of interests

The authors have no conflicts of interest with respect to the authorship or the publication of this article.

Open practices statement

Materials and data are available on the Open Science Framework (https://osf.io/e8cg5/?view_only=6865e077babf42fea39ba1cf713107f8). Further requests for the data or materials can be sent via email to the corresponding author at [anna.nickl@ur.de].

References

Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, *11* (1), 159–177. https://doi.org/10.1037/h0093018

Abel, M., & Bäuml, K.-H. T. (2016). Retrieval practice can eliminate list method directed forgetting. *Memory & Cognition*, *44*(1), 15–23. https://doi.org/10.3758/s13421-015-0539-x

Abel, M., & Bäuml, K.-H. T. (2020). Would you like to learn more? Retrieval practice plus feedback can increase motivation to keep on studying. *Cognition*, *201*, Article 104316. https://doi.org/10.1016/j.cognition.2020.104316

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K.-H. T. (2019). Sleep reduces the testing effect—But not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 272–287. https://doi.org/10.1037/xlm0000576

Abel, M., & Roediger, H. L. (2018). The testing effect in a social setting: Does retrieval practice benefit a listener? *Journal of Experimental Psychology: Applied*, *24*(3), 347–359. https://doi.org/10.1037/xap0000148.

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory (Hove, England)*, *25*(6), 764–771. https://doi.org/10.1080/09658211.2016.1220579

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, *25*(5), 724–730. https://doi.org/10.3758/BF03211315

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21* (8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect: It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1568–1581. https://doi.org/10.1037/xlm0000025.

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

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

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., & Delosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268–276. https://doi.org/10.3758/BF03193405

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

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633–642. https://doi.org/10.3758/BF03202713

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory (Hove, England)*, *18*(1), 49–57. https://doi.org/10.1080/09658210903405737

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

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392–399. https://doi.org/10.1037/0022-0663.81.3.392

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801–812. https://doi.org/10.1037/a0023219

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In *Learning and memory: A comprehensive reference* (pp. 487–514). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kliegl, O., & Bäuml, K.-H. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 202–214. https://doi.org/10.1037/xlm0000172

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mulligan, N. W., & Peterson, D. J. (2015). The negative testing and negative generation effects are eliminated by delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1014–1025. https://doi.org/10.1037/xlm0000070

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 938–950. https://doi.org/10.1037/xlm0000227

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Pastötter, B., & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 473–488. https://doi.org/10.3758/s13415-016-0407-6

Racsmány, M., Szőllősi, A., & Marián, M. (2020). Reversing the testing effect by feedback is a matter of performance criterion at practice. *Memory & Cognition*, *48*(7), 1161–1170. https://doi.org/10.3758/s13421-020-01041-5

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Psychology Press.

R Core Team. (2021). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. https://www.Rproject.org/.

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, 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, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced learning. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

RStudio Team. (2020). *RStudio: Integrated development for R [Computer software]*. RStudio, PBC. http://www.rstudio.com/

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

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*(6), 641–650. https://doi.org/10.3758/BF03198289

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, *49*(3), 263–269. https://doi.org/10.1037/h0063643

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., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 716–727. https://doi.org/10.1037/0278-7393.14.4.716

Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval—And spacing—For learning. *Psychological Science*, *27*(2), 223–230. https://doi.org/10.1177/0956797615617778

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*(9), 641–656. https://doi.org/10.1037/h0063404

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 115–124. https://doi.org/10.1037/a0034252

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(3), 210–221. https://doi.org/10.1037/0278-7393.4.3.210

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

Trißl, L., & Bäuml, K.-H. T. (2022). Selective restudy can reset recall of forgotten information. *Psychonomic Bulletin & Review*, *29*(6), 2202–2210. https://doi.org/10.3758/s13423-022-02131-y

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

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory (Hove, England)*, *11*(6), 571–580. https://doi.org/10.1080/09658210244000414

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, *9*(4), 418–455. https://doi.org/10.1016/0022-2496(72)90015-6

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, *2*(4), 775–780. https://doi.org/10.3758/BF03198154

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

Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review*, *111*(4), 864–879. https://doi.org/10.1037/0033-295X.111.4.864

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