Memory & Cognition (2025) 53:896–910
https://doi.org/10.3758/s13421-024-01617-5

How varying cue duration influences item-method directed forgetting: A novel selective retrieval interpretation

Pelin Tanberg[1] · Myra A. Fernandes[1] · Colin M. MacLeod[1] · William E. Hockley[2]

Accepted: 23 July 2024 / Published online: 27 August 2024

Abstract

A series of four experiments tested the assumptions of the most prominent and longstanding account of item-method directed forgetting: the selective rehearsal account. In the item-method directed forgetting paradigm, each presented item is followed by its own instructional cue during the study phase – either to-be-forgotten (F) or to-be-remembered (R). On a subsequent test, memory is poorer for F items than for R items. To clarify the mechanism underlying memory performance, we manipulated the time available for rehearsal, examining instructional cue durations of 1 s, 5 s, and 10 s. Experiments 1a and 1b, where the order of cue durations was randomized, showed no effect of cue duration on item recognition of unrelated single words, for either R or F items. Experiment 2, using unrelated word pairs, again showed no effect of randomized cue duration, this time on associative recognition. Experiments 3 and 4 blocked cue duration and showed equivalent increases in recognition of both R and F single words and word pairs with increasing cue duration. We suggest that any post-cue rehearsal is carried out only when cue duration is predictable, and that such limited rehearsal is equally likely for F items and R items. The consistently better memory for R items than for F items across cue duration depends on selective retrieval involving (1) a rapid retrieval check engaged for R items only and (2) a rapid removal process implemented for F items only.

Keywords Item-method directed forgetting · Selective rehearsal · Retrieval · Attentional withdrawal · Removal

Introduction

Forgetting is ordinarily considered to be inconvenient and undesirable. For example, forgetting anniversaries or birthdays can lead to negative consequences. These are, however, instances of unintended forgetting. In contrast, when performed intentionally, forgetting is an essential function used both to update memory with the most current information (e.g., when changing a password) and to prevent memory from being inundated with irrelevant information (see Bjork, 1989; Fawcett & Hulbert, 2020; Gravitz, 2019).

For decades, researchers have been interested in understanding intentional forgetting as a means to control memory (e.g., Bjork et al., 1968; see Golding & MacLeod, 1998). In the laboratory, over a half century ago, a *directed forgetting* paradigm was developed to investigate how individuals accomplish the forgetting of irrelevant information. Relative to instructions to remember specific information, instructions to forget specific information actually do result in poorer memory for the specified information. This successful forgetting on cue demonstrates the ability to voluntarily control memory and to reduce access to unwanted information on demand, a finding that emphasizes the adaptability of human memory.

Intentional forgetting during initial encoding has been examined in the laboratory primarily by using the *item-method directed forgetting* paradigm. Here, participants study a series of items in anticipation of a later memory test. Immediately after each item, a cue is presented that designates that item as either to-be-remembered (R) or to-be-forgotten (F). Additionally, participants receive a critical instruction that differentiates directed forgetting from other memory procedures: They are told that they will be tested only on the R items and not on the F items. In fact, however, participants ultimately are tested on all of the items.

Numerous studies (see MacLeod, 1998, for a review) have confirmed better memory for R-cued items than for F-cued

✉ Pelin Tanberg
pelin.tan@uwaterloo.ca

[1] Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

[2] Department of Psychology, Wilfrid Laurier University, Waterloo, ON, Canada

items, a performance difference called the *directed forgetting effect*. It is not surprising that we have better memory for information that we want to remember. What might be surprising is that we can reduce memory for information that we do not wish to remember. This reduction highlights the ability to intentionally control the contents of memory and to limit the encoding of F items, minimizing their likelihood of being stored, or to reduce their subsequent accessibility from long-term memory, or both. Thus, much of the research on intentional forgetting has emphasized how we forget on demand and has explored the mechanisms that underlie voluntary memory control.

The literature examining item-method directed forgetting has considered several possible mechanism(s) that could be responsible for intentional forgetting in this paradigm. In the current investigation, we sought to test the assumptions of the most prominent and longstanding account of item-method directed forgetting: the *selective rehearsal* account (Bjork, 1970; see MacLeod, 1998, for review). To do so, we manipulated the time available for rehearsal processes to operate and potentially to contribute to the directed forgetting effect. Because the time course of rehearsal holds direct relevance for theoretical accounts of item-method directed forgetting, understanding the influence of time available to rehearse will help to clarify the mechanisms underlying memory performance under an intent to remember versus forget.

The selective rehearsal account

Since the beginning of research on item-method directed forgetting in the late 1960s, the dominant explanation has been the selective rehearsal account (e.g., Tan et al., 2020; see MacLeod, 1998, for review). Selective rehearsal is seen as serving a dual function: promoting rehearsal of R items to increase the likelihood of their storage in long-term memory and limiting rehearsal of F items to reduce that likelihood. Although there have been multiple explanations of *when* it operates (Marevic & Rummel, 2020; Taylor et al., 2018a, 2018b; Sheard & MacLeod, 2005), the primary mechanism of elaborative rehearsal remains well articulated. Upon receipt of an R cue, the participant actively rehearses the item during the available cue duration. Upon receipt of an F cue, the participant ceases rehearsal of the item, devoting any further rehearsal while the F cue is present exclusively to previous R items. In this way, only R items benefit from rehearsal.

A key assumption of the selective rehearsal account is that to be able to rehearse selectively, participants must alternate between different rehearsal strategies. During item presentation, participants must hold each item in working memory – via maintenance rehearsal, which simply refreshes the item in working memory (cf. Craik & Lockhart, 1972) – in anticipation of the memory cue (Woodward et al., 1973). Maintenance rehearsal ensures the availability of the item in working memory until the instruction to remember or to forget that item is presented. Upon presentation of an R instruction, participants then switch to elaborative rehearsal (e.g., by forming associations) to encode that item into long-term memory. Upon presentation of an F instruction, the switch to elaboration is not made for that item, which is then allowed to decay.

Another critical – although debated – assumption (Dames & Oberauer, 2022; Fawcett & Taylor, 2008) is that as the study progresses, participants engage in cumulative rehearsal in working memory of a "rehearsal set" that contains only the R items, and this ongoing rehearsal improves memory over time. Critically, this single-process selective rehearsal account asserts that the directed forgetting effect arises from the selective (i.e., elaborative) rehearsal of the R items to the exclusion of the F items (Basden et al., 1993). Under this assumption of the selective rehearsal account, a key advantage of an item-based instruction to forget is that participants devote all elaborative rehearsal activity exclusively to R items.

The current investigation

Although the selective rehearsal account is a widely accepted explanation of the elaborative rehearsal of R items, its assumption that F items passively decay from working memory due to the termination of maintenance rehearsal has been challenged, notably in the attentional inhibition account (Anderson & Hulbert, 2021; Fawcett & Taylor, 2012). In fact, there is existing research that demonstrates evidence consistent with both inhibition and rehearsal effects, with inhibition effects observed within the first second of the cue period and rehearsal effects happening about a second later (Fellner et al., 2020). A passive view of forgetting as decay has been a critical component of the selective rehearsal account, yet substantial evidence has been marshalled against decay in working memory (Neath & Brown, 2012; Oberauer & Lewandowsky, 2013, 2014) and particularly in intentional forgetting (Dames & Oberauer, 2022; Fawcett & Taylor, 2008; Fawcett et al., 2016; Lewis-Peacock et al., 2018; Oberauer & Greve, 2022).

Given that a critical assumption of the selective rehearsal account has been heavily criticized, a major goal in the field should be to test the validity of this aspect of the account. Notably, although the elaborative rehearsal of only the R items is generally accepted as being at the core of the selective rehearsal account, other assumptions, in particular the cumulative rehearsal of R items, have not been systematically examined. If the selective rehearsal account is to be accepted as providing a thoroughly sufficient explanation of item-method directed forgetting, these assumptions must be verified. Therefore, a central goal of this investigation is to test these key assumptions of the selective rehearsal account.

With this theoretical background, our goal was to investigate how manipulating the time allocated for remembering or forgetting information influences memory performance. More time ought to permit more selective rehearsal, which should increase the divergence in memory for the R and F items. Our manipulation of cue duration differs from that of Bancroft et al. (2013), who examined only quite short cue durations (Experiments 1 and 2: 300, 600, or 900 ms; Experiment 3: 1, 2, or 3 s; Experiment 4 with associative memory: 2, 4, or 6 s), and reported a robust directed forgetting effect. Their longest cue duration for item memory was 3 s. Here, we introduce a larger range of cue durations to examine their influence on the directed forgetting effect across short (e.g., 1 s), medium (e.g., 5 s), and long (e.g., 10 s) durations of the instruction cue. Given that a cumulative rehearsal process suggests an ongoing and continuous enhancement of the R-only rehearsal set, increasing the duration of the cue should increasingly benefit memory for R items.

Experiment 1a

In the first experiment, we examined the influence of this broader range of cue durations in the typical single-word list-learning procedure. Our goal was to explore whether more extended cue durations would progressively enlarge the advantage of R items over F items, consistent with increasing the available time for differential rehearsal. Would providing increasingly greater time to remember an item enhance its representation in long-term memory, and might providing more time to forget an item reduce its representation in long-term memory?

To date, there have been conflicting findings regarding the influence of cue duration in the item method. Several studies have reported enhanced item recognition of both R and F items with increasing cue duration (Bancroft et al., 2013; Lee et al., 2007; Wetzel & Hunt, 1977) but several other studies have found no effect of cue duration (1 s vs. 3 s) on R and F items (Allen & Vokey, 1998) or on F items (Dames & Oberauer, 2022). These investigations have featured several methodological differences. First, in a study by Allen and Vokey (1998), cue duration was manipulated between subjects: Participants either saw all cues presented at 1 s or all cues presented at 3 s. However, in the Bancroft et al. (2013) and Lee et al. (2007) studies, each study phase was blocked by cue-duration condition with the order of the conditions counterbalanced across participants.

An investigation by Wetzel and Hunt (1977) remains one of the most in-depth examinations given that they systematically varied the cue duration between 1 s and 12 s. They observed results similar to those of Bancroft et al. (2013): Recall of both R and F items increased with increasing cue duration. Their recognition and recall results were consistent, but their recognition findings are difficult to interpret because the recognition test followed two recall test phases. The conflicting findings of cue duration on recognition memory are yet to be resolved with respect to the assumptions of the selective rehearsal account. Overall, it is quite possible that how cue duration is manipulated influences intentional remembering and forgetting. One of the principal purposes of the current investigation was to help resolve these conflicting findings in the literature.

If participants strategically switch to elaborative rehearsal when given an R cue, then as more time becomes available for selective rehearsal, memory for R items should increasingly diverge from memory for F items. This should be reflected in an interaction between instruction and cue duration. With respect to F items, there are three possible outcomes. First, Bancroft et al. (2013) found an equivalent increase in F-item and R-item memory with increasing cue duration. If their findings extend to our longer cue durations, then we should replicate their finding. Second, if F items are simply left unattended in working memory, then a longer F-cue duration could provide more time for decay to operate (i.e., the time-based decay view of selective rehearsal: Basden et al., 1993; Tan et al., 2020). Possibly, then, memory for F items should decrease with increasing cue duration. It is also possible that variation in cue duration would be irrelevant for F items: Interference from ongoing, cumulative rehearsal of R items would be continuous, such that individual cue durations would not influence memory for F items, if F items are rapidly removed from working memory to prevent their further encoding (i.e., the removal explanation: Dames & Oberauer, 2022).

Method

All experiments were preregistered. The complete Open Science Framework (OSF) page for this is posted at https://osf.io/ph3m5/

Participants

We conducted a sample size calculation using G*Power software (Faul et al., 2007, 2009). We calculated sample size using a medium effect size from Bancroft et al. (2013; Exp. 1), $\eta^2_p = .032$, which yielded an effect size f = .18 using the G*Power calculation, to power for a within-factors interaction with a desired power of ~.80 ($\alpha = .05$, two-tailed). This yielded a required sample size of $n = 63$.

We recruited 69 participants from Prolific (https://prolific.co/). Their mean age was 28.45 (*SD* = 1.41) years, and 86% were female. Six participants were excluded because they reported in the post-experimental questionnaire that they had not followed the instructions to the best of their ability. The final sample for analysis therefore consisted of 63 participants.

The following criteria were used on Prolific: (1) native English speaker, (2) approval rating of at least 90% on prior Prolific participation, and (3) between 18 and 35 years old. We also included further exclusion criteria to remove inattentive participants: (1) responding "yes" to a post-experiment question "Were you doing anything else while completing this task (e.g., Netflix, music, other)?", (2) responding "no" to a post-experiment question "Did you follow directions to the best of your ability?", and (3) responding "yes" to a post-experiment question "Did you use any aid (e.g., writing down the words) to remember the words?" Participants were asked to be honest when answering these questions and were told that their compensation would not be affected based on their answers.

All participants had either normal or corrected-to-normal vision (by self-report) and were compensated £2.50 (~ US $3.55) for their participation. The study was approved by the Wilfrid Laurier University Research Ethics Board (REB #6647).

Materials

We created a master word list of 200 common, high-frequency English words from the SenticNet 4-word corpus (Cambria et al., 2016). From this master list, 144 single study words were randomly selected without replacement to generate unique stimulus combinations across conditions and participants. These randomizations were carried out according to instructions provided by Taylor et al. (2018a, 2018b). Stimuli for all tasks were presented using jsPsych (de Leeuw, 2015). Our master word list is available via the OSF at: https://osf.io/gez4u/

Procedure

After signing up on Prolific, participants were directed to a secure website hosting the online experiment. They were told to use a computer for running the experiment and were prevented from running any phase of the experiment using a mobile device.

In the study phase, each trial began with a 1-s fixation in the center of the screen. Immediately following fixation offset, an individual word appeared for 3 s. Each word was presented in Open Sans font size 14 pt. The word was immediately followed by an R or F cue in the center of the screen for 1 s, 5 s, or 10 s. Cues were presented in Open Sans font size 18 pt. The order of the cue duration trials was random. The encoding phase included 72 single words, with half followed by an R cue and half followed by an F cue. Of the 36 R-cued words, 12 had a 1-s cue duration, 12 had a 5-s cue duration, and 12 had a 10-s cue duration. The F-cued words were divided in the same way. In addition, three words (all R-cued) were inserted at the beginning and another three at the end of the study list to serve as primacy and recency buffers; these items were not tested. Participants were instructed to remember all words followed by R and to forget all words followed by F. Critically, participants were informed that there would be a memory test after the study phase, but that they would be tested only on the R items.

In the test phase, the recognition test list consisted of 144 words, 72 old and 72 new. Participants were told that they would see one word at a time and that they should respond "old" to any word that had been presented during the study phase – regardless of the instruction that the word had received. Participants initiated the test list when they were ready and were told that they should answer the question: "Did you see the word during the first part?" For old words, participants were to press the Z key; for new words, they were to press the M key. The test phase was self-paced, and participants were encouraged to be as accurate as possible. A post-experiment questionnaire was included to assess self-reports of whether participants followed the instructions. The experimental instructions and the post-experiment questionnaire are available via the OSF at: https://osf.io/gez4u/

Data analysis

For all experiments, reported analyses were pre-registered. Mixed-effects logistic regression analyses were conducted on recognition performance on a trial-by-trial basis using R (R Core Team, 2019) and using the *lme4* package (Bates et al., 2015).[1] Lure recognition trials were removed because the false alarm rate was the same for all conditions. The structure of the random effects models followed the recommendations of Brown (2021). For the random effects structure, we included a model with by-participant and by-item random intercepts. We additionally included by-participant and by-item random slopes for the effects of Cue Type and Cue Duration only if including these improved model fit. If the model was singular, we continued to reduce the model complexity to report the best-fitting maximal model, and we state the random effects structure used for that model. After fitting the models, we performed hypothesis tests using the *Anova* function of the *car* package (Fox & Weisberg, 2011). Follow-up analyses to the models were performed using the *emmeans* package (Lenth et al., 2018) using Tukey's method to control for multiple comparisons. All logistic mixed-effects models were run using the *bobyqa* optimizer. Effect sizes are reported in terms of generalized η^2 (η_G^2)

[1] We thank Gidon Frischkorn and Ven Popov for helping with the mixed-effects logistic regression data preparation and R code.

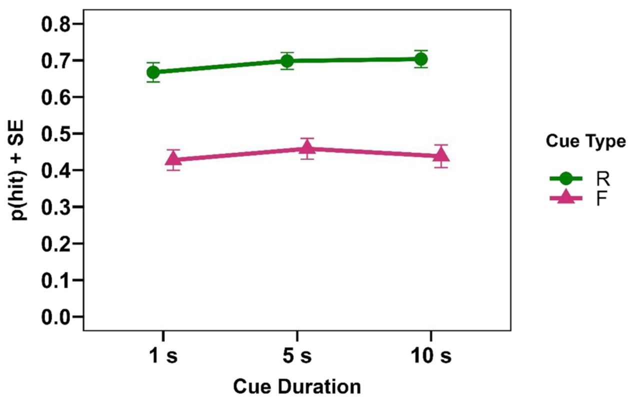

Fig. 1 Experiment 1a: Mean recognition performance represented as proportion hit on the yes/no recognition test. Descriptive proportion hit is shown for Cue Type (R vs. F) as a function of Cue Duration (1 s, 5 s, vs. 10 s). Error bars denote descriptive standard errors

and Cohen's *d*. In addition, odds ratio statistics for the interpretation of effect sizes (OR; Szumilas, 2010), along with model-based 95% confidence intervals (CIs), are reported for all mixed-effects logistic regression models on recognition performance.

Within the text, means and standard errors for all experiments are presented on the response scale, back-transformed from the log odds ratio scale used in the mixed-effects logistic regression analyses. Plots included visualizations of the results for all experiments using descriptive parameters of proportion hit, calculated as Hit Rate (i.e., number of items correctly selected as old/total number of old items), and standard error of the mean for each condition. Fixed effects included Cue Type (R vs. F) and Cue Duration (1 s vs. 5 s vs. 10 s) and were coded in all models using sum-to-zero contrasts.

The principal analysis was a trend analysis to determine how increasing cue duration influenced memory of R items and of F items. Cue Type was coded in all models using sum-to-zero contrasts and Cue Duration was coded in all models using orthogonal polynomials to permit observing any linear or quadratic trends across conditions. All trend analysis models converged with random effects including by-participant and by-item intercepts. Follow-up analyses to the models were performed using the *emtrends* function from the *emmeans* package. Data and analysis code for all experiments are available via the OSF (data: https://osf.io/n2d5j/; analysis: https://osf.io/h7msw/).

Results

Figure 1 presents *proportion hit* as a function of Cue Type and Cue Duration. The mean collective False Alarm Rate for all conditions was .20 (SE = .02).

There was a significant effect of Cue Type, χ^2 (1) = 230.72, $p < .001$, where participants overall recognized more R items (M = .78, SE = .01) than F items (M = .55, SE = .03), representing the familiar directed forgetting effect. There was, however, no main effect of Cue Duration, χ^2 (2) = .34, p = .843, indicating that recognition performance averaged over the levels of Cue Type did not significantly differ across the levels of Cue Duration at 1 s (M = .67, SE = .03), at 5 s (M = .68, SE = .03), or at 10 s (M = .67, SE = .03). The Cue Type × Cue Duration interaction was also non-significant, χ^2 (2) = 4.15, p = .126.

Follow-up comparisons to further investigate the main effect of Cue Type revealed significant directed forgetting effects at the 1-s duration (R: M = .76, SE = .02; F: M = .56, SE = .03), OR = .40, CI = .31 – .50, $p < .001$, the 5-s duration (R: M = .78, SE = .02; F: M = .56, SE = .03), OR = .40, CI = .29 – .47, $p < .001$, and the 10-s duration (R: M = .79, SE = .02; F: M = .53, SE = .03), OR = .29, CI = .23 – .36, $p < .001$. Collapsed across Cue Type, memory performance did not differ significantly across the three levels of Cue Duration, ORs > .84, CIs = .62 – 1.52, ps > .330.

The trend analysis examined the overall pattern across Cue Duration as a function of Cue Type. There was a marginally significant trend across both Cue Types (R: M_{trend} = .09, SE_{trend} = .06; F: M_{trend} = .07, SE_{trend} = .06), b = .16, p = .055.

Discussion

As expected, there was a robust directed forgetting effect at each of the three levels of cue duration. However, in direct conflict with the selective rehearsal account, increasing cue duration did not influence recognition of R items – nor indeed of F items. These results with longer cue durations do not replicate the pattern observed by Bancroft et al. (2013) for shorter durations. Instead, they support two conclusions. First, although under the selective rehearsal account memory for R items should increase with increasing cue duration, this did not occur. It is possible that the cue durations were sufficiently long and unpredictable that elaborative rehearsal – to the extent that it was invoked – ceased quite quickly. Second, there also was no effect of increasing cue duration on F item recognition. The decay view of the selective rehearsal account might have expected a decline in memory for F items as cue duration increased, as more time was provided for decay – or for interference from the anticipated cumulative rehearsal of R items – but this also did not happen. Given that the central finding – no effect of increasing cue duration on R items – is inconsistent the most fundamental prediction of the selective rehearsal account, we decided that a replication was warranted.

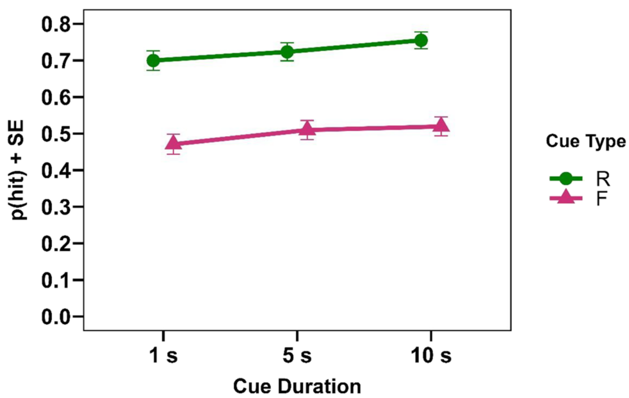

Fig. 2 Experiment 1b: Mean recognition performance represented as proportion hit on the yes/no recognition test. Descriptive proportion hit is shown for Cue Type (R vs. F) as a function of Cue Duration (1 s, 5 s, vs. 10 s). Error bars denote descriptive standard errors

Experiment 1b: Replication

Method

Participants

The sample size calculation was identical to Experiment 1a, although we fell a little short. Fifty-nine participants from Prolific participated. Their mean age was 31.23 (*SD* = 2.43) years and 75% were female. Data from six participants were removed from analyses because they reported writing down the words. Data from one participant were excluded because they reported being distracted during the study (i.e., switching between multiple applications). The final sample for analyses included 52 participants.

Results

Figure 2 presents proportion hit as a function of Cue Type and Cue Duration. The mean collective False Alarm Rate for all conditions was .17 (*SE* = .02).

The results of Experiment 1a replicated in their entirety. There was a significant effect of Cue Type, χ^2 (1) = 198.00, *p* < .001, where participants overall recognized more R items (*M* = .77, *SE* = .02) than F items (*M* = .54, *SE* = .03). There was no main effect of Cue Duration, χ^2 (2) = .09, *p* = .956, indicating that memory did not differ between the 1-s duration (*M* = .66, *SE* = .03), the 5-s duration (*M* = .66, *SE* = .03), and the 10-s duration (*M* = .67, *SE* = .03). There was no Cue Type × Cue Duration interaction, χ^2 (2) = .02, *p* = .991.

Separated by Cue Duration, there were significant and equivalent directed forgetting effects at the 1-s duration (R: *M* = .77, *SE* = .03; F: *M* = .53, *SE* = .04), *OR* = .34, *CI* = .26 – .44, *p* < .001, the 5-s duration (R: *M* = .77, *SE* = .03; F: *M* = .54, *SE* = .04), *OR* = .34, *CI* = .26 – .44, *p* < .001, and the 10-s duration (R: *M* = .77, *SE* = .02; F: *M* = .54, *SE* = .04), *OR* = .35, *CI* = .27 – .45, *p* < .001. Collapsed across Cue Type, memory performance did not differ significantly across the three levels of Cue Duration, *ORs* > .97, *CIs* = .71 – 1.39, *ps* > .956.

The trend analysis examined the overall pattern across Cue Duration as a function of Cue Type. There was no significant trend across both Cue Type conditions (R: M_{trend} = .01, SE_{trend} = .07; F: M_{trend} = .02, SE_{trend} = .06), *b* = .01, *p* = .921, indicating that recognition did not increase across Cue Duration.

Discussion

We replicated the pattern seen in Experiment 1a, again observing no influence of cue duration on memory for either R items or F items. This confirmed that increasing cue duration did not influence the directed forgetting effect, with respect to either the R items or the F items. We again found that providing participants more time did not enhance their recognition of R items nor did it bring about more forgetting of F items. Critically, instructional cue did not interact with cue duration as would be expected if, unlike F items, R items were being cumulatively rehearsed, gaining progressively as cue duration lengthened.

The prediction that increasing cue duration should increase memory for R items derived from the assumption that participants selectively rehearse the R items and may even cumulatively rehearse earlier R items in the rehearsal set when they are given more time to do so during cue presentation. Following an R cue, the current item should be entered into the cumulative rehearsal set; following an F cue, the ongoing rehearsal set of prior R items should be reactivated. Thus, R items should benefit increasingly from increasing cue durations, but that is not what we have observed in these two experiments.

The finding that memory for F items was unaffected by increasing cue duration is in line with the prediction (Popov et al., 2019; Tan et al., 2020) that variation in cue duration would be irrelevant for F items (because they do not receive elaborative rehearsal). Resources should instead be devoted to ongoing cumulative rehearsal of R items in the rehearsal set. But the finding that memory for the R items did not increase with increasing cue duration contradicts this cumulative rehearsal process. Our findings for F items are, however, consistent with models of working memory in which irrelevant information is removed from working memory before it can enter long-term memory. Muter (1980), Ecker et al. (2014), and Oberauer (2018) presented evidence that forgetting from working memory can be very rapid. Previously, Bancroft et al. (2013) found that recognition of F items increased with increasing cue durations at 300 ms, 600 ms, and 900 ms. Whatever is happening – deterioration or removal from working memory – is a rapid process, likely occurring in less than 1 s. This would explain why we still observe above-chance performance for F items: Removal of irrelevant information only operates to stop further processing of an irrelevant item.

A novel retrieval-based explanation

To explain this collection of findings, we propose a new retrieval-based account which can be seen as analogous to the testing effect in long-term memory (see Rowland, 2014, for a review). In calling this new account "selective retrieval," we are mindful of the potential confusion with selective rehearsal, but we believe that this is still the best label for our new explanation. The core idea is that when an R cue is presented, the just-presented (but no longer visible) item is accessed in working memory; this access does not occur for an item when an F cue is presented. We see this as a quick retrieval check for each R item, which can also be seen as redirecting attention to that item in working memory. This check is effectively a kind of test that the just-presented item is accessible. It is consequently a retrieval that is selective in that it is performed only on R items. In contrast, because an F item need not be remembered, it is quickly removed from working memory (Dames & Oberauer, 2022; Oberauer, 2018). The result is a quick boost for R items only, which therefore is visible as a consistent memory benefit for R items over F items, independent of cue duration and without depending on intentional or ongoing rehearsal. Under this account, there is neither selective rehearsal nor cumulative rehearsal of R items, only immediate selective retrieval.

Experiment 2

In Experiments 1a and 1b, we examined memory for individual items. Investigating associative memory provides another way to examine the mechanism underlying better memory for R items and the role played by available rehearsal time as a function of cue duration. In Experiment 2, our goal was to investigate how associative recognition is influenced by increasing cue duration, and to provide converging evidence for the pattern of results found in Experiments 1a and 1b. After studying a series of unrelated word pairs under instructions to associate the members of each pair, participants were asked at test to discriminate intact pairs – pairs of words that had been presented together at study – from rearranged pairs – words that had not been presented together at study, having been presented instead in separate pairs. This is the classic associative recognition test (see, e.g., Hockley et al., 2016; Hockley, 1992). Here, to respond accurately on the test, participants must rely on memory for the associations that they formed between pairs of items during study: Memory for the individual words cannot aid in discriminating between intact and rearranged word pairs because all test items were studied.

Bancroft et al. (2013; Exp.4) varied the cue duration between 2 s, 4 s, and 6 s in their examination of the effect of R cues versus F cues on associative recognition. They found that performance increased with cue duration for both R and F pairs. Given that we did not reproduce the Bancroft et al. findings for longer cue durations in Experiments 1a and 1b, our principal goal in Experiment 2 was to investigate whether associative recognition would be influenced by our longer cue durations. We foresaw three possible outcomes. First, we could replicate the findings of Bancroft et al., with memory for both R and F items increasing over our cue duration manipulation (1 s, 5 s, and 10 s). Second, if forming associations between R items is an important strategy in the item-method paradigm, then providing more time to continue strengthening these associations (by increasing cue duration) should selectively improve memory for R items over that for F items. However, according to the selective rehearsal account, if participants terminate rehearsal following an F cue, then providing more time (by increasing cue duration) for F items should not affect their associative recognition. Of course, it is also possible that, similar to our Experiments 1a and 1b, we would find no influence of increasing cue duration on associative recognition in a mixed-list design.

Method

Participants

We calculated sample size using a medium effect size from Bancroft et al. (2013; Exp. 4), $\eta^2_p = .021$, which yielded an effect size f = .15 using the G*Power calculation, to power for a within-factors interaction with a desired power of ~.80 ($\alpha = .05$, two-tailed). This yielded a required sample size of $n = 96$.

We recruited 106 participants from Prolific. Their mean age was 32.41 ($SD = 2.65$) years, and 71% were female. Five participants were excluded because they reported in the post-experiment questionnaire that they had not followed the instructions to the best of their ability. Three participants were excluded due to a programming error. Two participants were excluded because they switched tabs during the study more than three times. The final sample for analysis included 96 participants. All other Prolific and participant criteria were identical to the previous experiments.

Materials

Our master word list was identical to that in the previous experiments.

Procedure

The study phase was very similar to that of Experiments 1a and 1b, with the exception that instead of presenting individual words for 3 s, we presented two unrelated words together for 4 s. There was always a fixation point in the center of the screen, with one word above the fixation point and one word below the fixation point on each study trial. The encoding phase included 72 word pairs, with half followed by an R cue and half followed by an F cue. The cue duration conditions were identical to those in Experiments 1a and 1b – 1 s, 5 s, and 10 s. Critically, the order of the cue duration trials was again random. Participants were instructed to form associations between the two words upon the presentation of each word pair. Similar to Bancroft et al. (2013), participants were urged to strengthen this association when the word pair was followed by an R cue because this would help them remember that the two words were shown together. Participants were informed that there would be a memory test after the study phase, and that they would be tested only on the R pairs.

In the test phase, the test list consisted of 72 word pairs, 36 intact (old) pairs and 36 rearranged (new) pairs, the latter pairs created by combining two words that had been presented in different pairs at study. Assignment of rearranged pairs always remained within the same R or F Cue Type and Cue Duration conditions. The top-bottom study order of the words was also preserved in both intact and rearranged test pairs. Pair order on the test was random. Participants were told that they should answer the question: "Did you study these words together?" For intact pairs, participants were to press the Z key; for rearranged pairs, they were to press the M key. The test phase was self-paced with participants encouraged to be as accurate as possible. A post-experiment questionnaire followed the test phase to assess self-reports of whether participants followed the instructions. The post-experiment questionnaire and experimental instructions are available via the OSF at: https://osf.io/gez4u/

Data analysis

The data analysis tools, model structures, fixed effects, and packages used were identical to those in Experiments 1a and 1b.

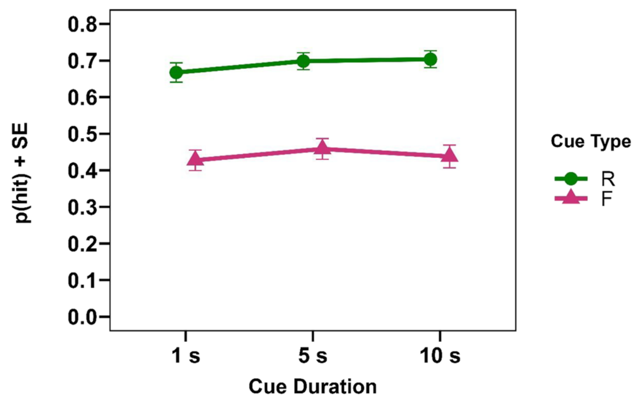

Fig. 3 Experiment 2: Mean recognition performance represented as proportion hit on the associative recognition test. Descriptive proportion hit is shown for Cue Type (R vs. F) as a function of Cue Duration (1 s, 5 s, vs. 10 s). Error bars denote descriptive standard errors

Results

False alarm rate model

Because positive responses to rearranged trials constitute false alarms in an associative recognition task, we examined whether this false alarm rate differed as a function of Cue Type and Cue Duration. There was no significant main effect of Cue Type, χ^2 (1) = 3.29, *p* = .071, no main effect of Cue Duration, χ^2 (2) = 2.40, *p* = .302, and no Cue Type × Cue Duration interaction, χ^2 (2) = .20, *p* = .909, on false alarm rates. Because false alarm rates from rearranged trials did not significantly differ across conditions, we removed rearranged trials from the rest of the analyses of recognition performance.

Hit rate model

Figure 3 presents proportion hit as a function of Cue Type and Cue Duration. There was a significant effect of Cue Type, χ^2 (1) = 218.29, *p* < .001, where participants overall recognized more R pairs (*M* = .72, *SE* = .02) than F pairs (*M* = .46, *SE* = .03), the usual directed forgetting effect. Again, there was no main effect of Cue Duration, χ^2 (2) = .50, *p* = .780, indicating that memory did not differ between the 1-s duration (*M* = .60, *SE* = .03), the 5-s duration (*M* = .59, *SE* = .03), and the 10-s duration (*M* = .60, *SE* = .03). There also was no Cue Type × Cue Duration interaction, χ^2 (2) = 2.71, *p* = .258.

Separated by Cue Duration, there were significant directed forgetting effects at the 1-s duration (R: *M* = .70, *SE* = .03; F: *M* = .48, *SE* = .04), *OR* = .40, *CI* = .31 – .51, *p* < .001, the 5-s duration (R: *M* = .72, *SE* = .03; F: *M* = .44, *SE* = .03), *OR* = .30, *CI* = .24 – .39, *p* < .001, and the 10-s duration (R: M = .73, SE = .03; F: M = .46, SE = .03), OR = .35, CI = .24 – .40, p < .001. Collapsed across Cue Type, memory performance did not differ significantly across the three levels of Cue Duration, ORs > .88, CIs = .65 – 1.60, ps > .305.

The trend analysis examined the overall pattern across Cue Duration as a function of Cue Type. There was no significant trend across either Cue Type condition (R: M_{trend} = .06, SE_{trend} = .06; F: M_{trend} = .06, SE_{trend} = .06), b = . 21, p = .178, indicating that performance did not increase with increasing Cue Duration.

Discussion

There was a robust directed forgetting effect across all levels of cue duration when participants were explicitly instructed to encode associative information. In typical associative memory and directed forgetting studies, participants are told to form an association between the two words after the presentation of the cue if it is a cue to remember (e.g., Bancroft et al., 2013; Exp. 4). In our experiment, we instructed participants to form an association between the members of each word pair upon presentation (i.e., before the instructional cue). This was done to provide a purer test of forgetting and associative memory, where all pairs were intended to be associated before they were cued.

Of critical interest was whether the cue duration manipulation would influence associative recognition. Similar to Experiments 1a and 1b with item recognition, we found no effect of increasing cue duration on associative recognition. This finding did not replicate the finding of Bancroft et al. (2013; Exp. 4). Moreover, we again found no support for continuous, cumulative rehearsal: Providing more time to continue strengthening associative information did not selectively improve memory for R items relative to F items, which would have been visible, had it occurred, as an interaction with R pairs showing a sharper rise in memorability than F pairs with increasing cue duration.

Experiment 3

Upon further consideration of the longstanding selective rehearsal account, it occurred to us that randomizing the cue durations as was done in Experiments 1a and 1b and Experiment 2 could have worked against any kind of rehearsal because participants did not know for any given item how long they might have available for rehearsal. As a consequence, they might not have rehearsed much at all given the unpredictability of the cue duration that would be available. Previous investigations of cue duration and item-method directed forgetting had either blocked the study list by cue duration (Bancroft et al., 2013; Lee et al., 2007) or manipulated cue duration between subjects (Allen & Vokey, 1998; Wetzel & Hunt, 1977). Experiments 1a, 1b, and 2 in this investigation were the first examinations of manipulating cue duration intermixed within the study list in the item-method paradigm. The goal of this was to provide a more direct test of the time course of processing because in an intermixed study list, participants must either rapidly engage in intentional remembering or forgetting – given the uncertainty of time allocated – or deliberately switch their strategy for each cue duration condition.

Regardless, how cue duration is manipulated in the item-method directed forgetting paradigm is potentially important for examining the time course of these processes. To provide a more straightforward comparison of the current intermixed investigations to the existing literature, in Experiment 3, we switched the study phase to feature three blocks, each with a unique cue duration – again 1 s, 5s, or 10 s. Blocking cue duration should make clearer to the participant the time that they will have available to rehearse, possibly allowing them to take greater advantage of the cue duration.

Method

Participants

We increased our sample size relative to Experiments 1a and 1b because our manipulation of cue duration had yielded a smaller effect size than that reported by Bancroft et al. (2013). We recruited 106 participants from Prolific, none of whom had participated in Experiments 1a, 1b, or 2. Their mean age was 28.65 (SD = 3.47) years, and 65% were female. Five participants were excluded because they reported not following the instructions to the best of their ability. Three participants were removed due to a coding error that led to those participants studying only the 5-s cue duration. Four participants were excluded because they reported writing all of the words down (and hence scoring 100% on the recognition test). The final sample for data analysis included 94 participants.

Procedure

The procedure was identical to that of Experiments 1a and 1b except that there now were three separate blocks of cue duration in the study phase. Six counterbalanced versions of the cue duration blocking order were used with approximately 16 participants assigned to each. Participants were given the same instructions regarding the cue presentation, but they now were told in addition that the study phase was divided into three parts. At the beginning of each block,

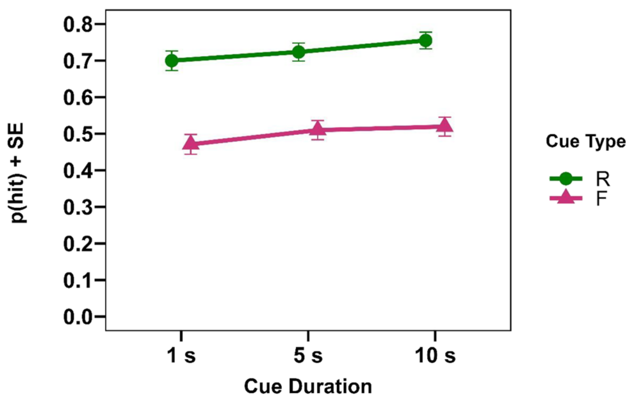

Fig. 4 Experiment 3: Mean recognition performance represented as proportion hit on the yes/no recognition test. Descriptive proportion hit is shown for Cue Type (R vs. F) as a function of Cue Duration (1 s, 5 s, vs. 10 s). Error bars denote descriptive standard errors

they were told the cue duration for that block (i.e., 1 s, 5 s, or 10 s).

Results

Data analysis was identical to that of Experiments 1a and 1b. Figure 4 presents proportion of hits as a function of Cue Type and Cue Duration. The mean collective False Alarm Rate for all conditions was .17 (*SE* = .02).

There was a significant effect of Cue Type, χ^2 (1) = 461.19, $p < .001$: As usual, participants overall recognized more R items (*M* = .78, *SE* = .02) than F items (*M* = .50, *SE* = .03). Now, however, there was also a significant main effect of Cue Duration, χ^2 (2) = 15.51, $p < .001$, where recognition increased from the 1-s duration (*M* = .62, *SE* = .03), to the 5-s duration (*M* = .65, *SE* = .03), to the 10 s duration (*M* = .68, *SE* = .03). As before, however, there was no Cue Type × Cue Duration interaction, χ^2 (2) = .940, *p* = .625.

Follow-up comparisons aggregated across Cue Duration condition revealed significant directed forgetting effects at the 1-s duration (R: *M* = .75, *SE* = .02; F: *M* = .47, *SE* = .03), *OR* = .30, *CI* = .25 – .36, $p < .001$, the 5-s duration (R: *M* = .77, *SE* = .02; F: *M* = .51, *SE* = .03), *OR* = .31, *CI* = .25 – .37, $p < .001$, and the 10-s duration (R: *M* = .80, *SE* = .02; F: *M* = .52, *SE* = .03), *OR* = .27, *CI* = .22 – .33, $p < .001$.

Trend analysis confirmed that recognition performance increased with increasing cue duration, and that it did so for both the R condition (M_{trend} = .16, SE_{trend} = .05) and the F condition (M_{trend} = .11, SE_{trend} = .05), *bs* > .19, *ps* < .001. A comparison of slopes demonstrated that the increases for the two conditions did not significantly differ, *b* = .05, *p* = .477.

Discussion

We again observed robust directed forgetting effects across all cue duration levels. Blocking by cue duration did, however, highlight the available time for potential rehearsal, and recognition now improved, albeit not dramatically, as cue duration lengthened. This was supported by the trend analysis: There was a significant linear trend, where increasing cue duration increased memory performance. These findings replicate Lee et al. (2007) and Bancroft et al. (2013), and extend their findings to cue durations beyond 3 s for item memory. There was, however, a remaining puzzle: Blocking by cue duration benefitted memory equivalently for R items and F items, confirming the findings of Bancroft et al. This equivalence runs contrary to the selective rehearsal account: If participants rehearse only (or at least primarily) R items, and if they rehearse R items more as cue duration lengthens, then the memory improvement should be increasingly greater for R items relative to F items. We reason that blocking by cue duration permitted "rehearsal slippage" where both R items and F items occasionally benefitted from rehearsal. That the functions for R items and F items are essentially parallel across cue duration does imply, however, that there is no differential cumulative rehearsal favoring R items.

Experiment 4

In Experiment 4, we again used a blocked procedure like that in Experiment 3 but this time with item pairs, like the materials in Experiment 2. The goal was to determine whether associative recognition, like item recognition, would now show improvement with increasing cue duration. Randomizing cue durations in Experiment 2 may have discouraged participants from elaboratively forming associations between the word pairs because they would not know how long they had to do so. If so, then blocking cue duration should provide a better test of how increasing cue duration affects intentional remembering and forgetting for associative memory.

Method

Participants

We calculated sample size using our effect size from Experiment 2, given that it was larger than that of Bancroft et al. (2013), η^2_p = .039, which yielded an effect size f = .20 using the G*Power calculation, to power for a within-factors interaction with a desired power of ~.80 (α = .05, two-tailed). This yielded n = 52 participants.

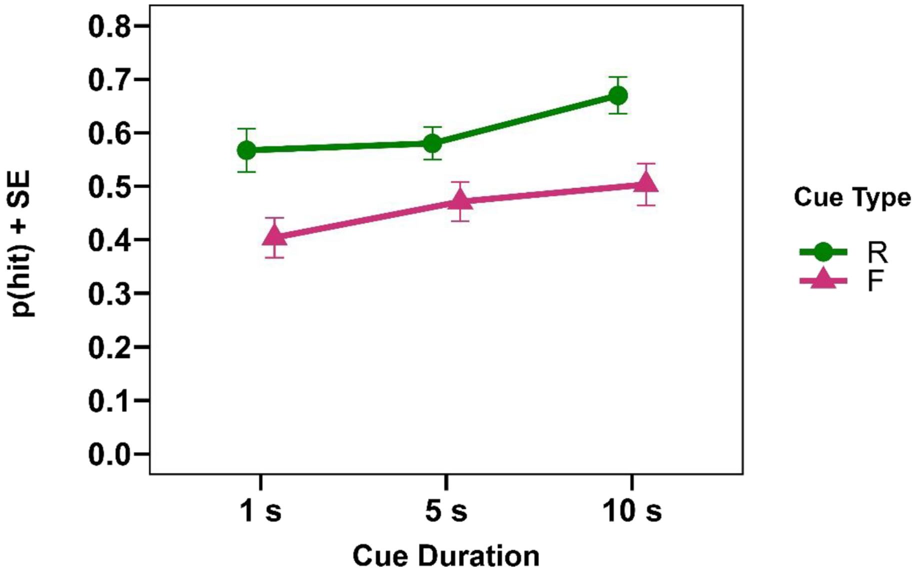

Fig. 5 Experiment 4: Mean recognition performance represented as proportion hit on the associative recognition test. Descriptive proportion hit is shown for Cue Type (R vs. F) as a function of Cue Duration (1 s, 5 s, vs. 10 s). Error bars denote descriptive standard errors

We recruited 60 participants from Prolific. Their mean age was 30.32 (*SD* = 2.34) years, and 68% were female. Four participants were excluded because they switched tabs during the study more than three times. Three participants were removed for having 100% recognition performance and reporting having written down all of the words. The final sample for analysis included 53 participants. All other Prolific and participant criteria were identical to the previous experiments.

Procedure

The procedure was identical to that of Experiment 2 except that there now were three separate blocks of cue duration in the study phase, analogous to Experiment 3. Six counterbalanced versions of the cue duration blocking order were used with approximately eight participants assigned to each. Participants were given the same instructions regarding the cue presentation, but they now were told that the study phase was divided into three parts. At the beginning of each block, they were explicitly told the cue duration for that block (i.e., 1 s, 5 s, or 10 s).

Results

Data analysis was identical to that of Experiment 2.

False alarm rate model

There was a significant main effect of Cue Type, χ^2 (1) = 8.45, *p* = .004, where participants false alarmed to F pairs (*M* = .77, *SE* = .03) more than to R pairs (*M* = .71, *SE* = .04). There was no main effect of Cue Duration, χ^2 (2) = 1.29, *p* = .525, and no Cue Type × Cue Duration interaction, χ^2 (2) = 1.88, *p* = .391, on false alarm rates.

Hit rate model

Figure 5 presents proportion of hits as a function of Cue Type and Cue Duration. As usual, there was a significant effect of Cue Type, χ^2 (1) = 42.71, *p* < .001, where participants overall recognized more R pairs (*M* = .62, *SE* = .03) than F pairs (*M* = .46, *SE* = .03). This time, however, there was also a significant main effect of Cue Duration, χ^2 (2) = 14.26, *p* < .001, where recognition differed between the 1-s duration (*M* = .49, *SE* = .03), the 5-s duration (*M* = .53, *SE* = .03), and the 10-s duration (*M* = .60, *SE* = .03). There was no Cue Type × Cue Duration interaction, χ^2 (2) = 1.65, *p* = .439.

Separated by Cue Duration, there were significant directed forgetting effects at the 1-s duration (R: *M* = .70, *SE* = .03; F: *M* = .48, *SE* = .04), *OR* = .48, *CI* = .34 – .67, *p* < .001, the 5-s duration (R: *M* = .72, *SE* = .03; F: *M* = .44, *SE* = .03), *OR* = .62, *CI* = .44 – .87, *p* < .001, and the 10-s duration (R: *M* = .73, *SE* = .03; F: *M* = .46, *SE* = .03), *OR* = .47, *CI* = .33 – .66, *p* < .001.

The focal linear trend analysis was significant, confirming that memory for both R pairs (M_{trend} = .22, SE_{trend} = .06) and F pairs (M_{trend} = .24, SE_{trend} = .08) increased with increasing cue duration, *b* = .33, *p* < .001. However, a comparison of the trends revealed no difference between the R pairs trend and the F pairs trend, *b* = .01, *p* = .928.

Discussion

There again was a robust and consistent directed forgetting effect across all levels of cue duration. Blocking by cue duration caused recognition to increase from the shortest cue duration to the longest cue duration for both R items and F items, as shown by the significant linear trend. Similar to Experiment 3 with single words, this suggests that, when cue duration is blocked and participants can reliably predict the time available for rehearsal, they further elaborate the associations that they had formed prior to the cue presentation, enhancing memory. Critically, though, they apparently do this for both R pairs and F pairs, inconsistent with the selective rehearsal account. Instead, our findings fit with the explanation that we put forward after Experiments 1a and 1b that this improvement is dependent on participants knowing how much time that they have available to dedicate to further processing.

General discussion

Over the past 50 years, the directed forgetting effect has become one of the most well-established phenomena in the memory literature (for reviews, see MacLeod, 1998; Sahakyan & Foster, 2016). This effect signifies the ability to voluntarily control the contents of memory during the processing of information, choosing what information is promoted to long-term memory and what information is not promoted. Although investigators continue to debate how F items are intentionally removed from memory (Dames & Oberauer, 2022; Fawcett & Taylor, 2008), theory has long supported the role of a differential rehearsal mechanism that favors R items in item-method directed forgetting.

In the present investigation, we examined how varying the time allocated to an intent to remember or to forget information influences directed forgetting. Specifically, we carried out an in-depth investigation of the time course of potential rehearsal mechanisms, given the previous disagreements in the literature concerning how time is used for each process (Allen & Vokey, 1998; Bancroft et al., 2013; Dames & Oberauer, 2022; Lee et al., 2007). According to the longstanding selective rehearsal account (e.g., Bjork, 1970), providing participants with more time should differentially improve memory for R items over memory for F items because participants engage in elaborative rehearsal of items – and potentially in cumulative elaborative rehearsal of an R-only rehearsal set – throughout the study phase. However, when we introduced a larger range of cue durations to examine the hypothetical differential rehearsal, time did not strongly affect specifically the rehearsal of R items; instead, the pattern was identical for R items and for F items, and this was true both for item memory and for associative memory.

Across four experiments, we used a consistent range of cue durations: 1 s, 5 s, and 10 s. In Experiments 1a, 1b, and 3, we examined how increasing cue duration would influence item memory. We observed a memory benefit with increasing cue duration only when participants were aware of how much time they could allocate to each item – that is, when the cue durations were blocked in Experiment 3 – and this benefit was equivalent for R and F items. In Experiments 2 and 4, we examined whether associative memory was influenced by increasing cue duration and observed exactly the same pattern of results.

Why were there different effects of cue duration depending on the design of the study phase – intermixed versus blocked cue durations? In an intermixed design, participants cannot predict how long they will have to remember or forget the just-presented item; in a blocked design, they know how long they have. We suggest that in the intermixed case, participants do not rehearse at all; rather, they are hamstrung by the unpredictability. In contrast, in the blocked case, participants occasionally do rehearse, although not differentially: They are just as likely to rehearse an R item as an F item. Consequently, increasing cue duration affects both types of items in the same way.

Why, then, is there a directed forgetting effect? Under our retrieval-based interpretation, we suggest that R items – and only R items – benefit from a rapid boost due to a quick retrieval check, and that this alone drives the directed forgetting effect. Immediately upon receipt of an R cue, participants routinely attend to the just presented R item, a step that is omitted for F items. Such a selective retrieval check does not require a rehearsal process to boost an R item in memory; rather, quickly retrieving the item boosts its representation (Dames & Oberauer, 2022). When the length of the cue duration is unpredictable, this retrieval check for R items, coupled with the removal of F items, may be all that participants are able to do.

Now consider a blocked design where participants know the length of time available for elaborative rehearsal. Here, we observed that memory increased for R items and for F items with increasing cue duration. Additional rehearsal appears to be more likely when participants are aware that they have a longer duration to rehearse. Intriguingly, though, this rehearsal seems to benefit R items and F items equivalently, suggesting to us that it is a kind of "slippage" – that the likelihood of non-strategic rehearsal becomes greater with the passage of time but that this is independent of the item's cue. At any rate, there is no evidence in our data of differential rehearsal favoring R items: The functions for R and F items were parallel over cue duration in all four of our experiments.

If F items were simply left to decay over time, increasing cue duration could have provided more time for items to decay in memory. That is not what we observed. Instead, our findings are in line with a removal process which has been proposed to operate in working memory (Oberauer, 2018). This removal is proposed to operate rapidly (i.e., under 1 s) so that items do not remain in the focus of attention for too long. If F items are quickly removed from working memory, then memory for F items should be unaffected by increasing cue duration within the 1-s to 10-s range, precisely what we found. Some researchers have gone on to suggest that this removal process requires an active (potentially inhibitory) and resource-demanding mechanism that expunges irrelevant F items from memory (Fawcett & Taylor 2008, 2012; Festini & Reuter-Lorenz, 2013). Contrarily, other researchers have questioned the idea that an intention to forget imposes a heavier cognitive demand than does an intention to remember (Nickl & Bäuml, 2024; Pandey et al., 2023; Popov et al., 2019; Tan et al., 2020). Their findings challenge the resource-demanding assumption of the removal mechanism but do not, of course, rule it out.

Our experiments additionally extend the findings of Bancroft et al. (2013) to a broader range of cue durations. Using blocked cue durations, Bancroft et al. found that increasing cue duration (item recognition: 1–3 s; associative recognition: 2–6 s) led to increased item and associative recognition. This benefit with longer cue durations was expressed to a similar extent for both R items and F items, just as we found for even longer cue durations in our Experiments 3 and 4. It would be interesting in future studies to have participants report the strategies that they use to remember or forget depending on the duration of the cue in item-method directed forgetting: It is possible that these strategies shift as cue duration lengthens.

It is also important to note that our longest cue duration at 10 s may have resulted in additional free time in working memory. Mızrak and Oberauer (2022) demonstrated that increasing available processing time (here via longer cue durations) in working memory improves overall performance by giving an encoding resource more time to recover. After the resource is replenished during a long interval, memory for subsequent items is better compared to when the resource had less time to recover. This encoding-resource account (see also Popov & Reder, 2020) could explain the overall better memory – for both F items and R items – as cue duration increases. In the blocked design, because subsequent cues would also be of 10-s duration, resources have an even longer time to recover throughout the study phase, in contrast to the intermixed condition, where the next item could be of any duration.

In summary, in four experiments, we have documented the role of increasing instructional cue duration on item recognition and on associative recognition beyond the common shorter durations. Although a selective rehearsal account predicts that increasing cue duration should selectively increase memory only for R items and not for F items, we never observed such an interaction. Instead, memory was influenced in the same way for R items and for F items: Both were unaffected by cue duration when durations were randomly mixed, whereas both increased, albeit modestly, when durations were blocked, and this was true whether the material consisted of single words and item recognition or unrelated word pairs and associative recognition.

These findings cannot readily be explained by ongoing cumulative rehearsal of R items versus time-based decay of F items. Instead, when cue durations are random, we suggest that the directed forgetting effect derives from a quick retrieval check selectively done once only for R items but not for F items. This check, analogous to a one-trial testing effect in working memory, boosts the R items but, because it is quick, it is insensitive to cue duration. Our suggestion of a differential retrieval-based process provides an alternative explanation to item-method directed forgetting that better accords with the working memory literature. A similarly quick process of removal may operate on F items (Dames & Oberauer, 2022; Oberauer, 2018). Although our investigation is entirely consistent with this removal process being rapid, future research should continue to examine the nature of such a removal process, in particular focusing on the extent to which it is cognitively demanding and/or inhibitory. In addition, contrary to the solely rehearsal-based explanation in the longstanding selective rehearsal account, our new retrieval-based explanation suggests that the benefit to the R items does not depend on intentional or ongoing continuous rehearsal and that it happens quickly regardless of cue duration. We did see evidence of limited rehearsal modestly improving memory when cue durations were blocked, but that rehearsal appeared to be applied equivalently to R and F items, perhaps reflecting occasional unintended rehearsal of the most recent item, regardless of its cue. We saw no evidence, however, of cumulative rehearsal selectively devoted to a rehearsal set containing only R items.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-024-01617-5.

Authors' contribution Pelin Tanberg: Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization.

Myra A. Fernandes: Conceptualization, Methodology, Writing – Review and Editing, Funding, Supervision.

Colin M. MacLeod: Conceptualization, Methodology, Writing – Review and Editing, Funding, Supervision.

William E. Hockley: Conceptualization, Methodology, Writing – Review and Editing, Supervision, Funding, Resources.

Funding This research was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada Alexander Graham Bell Canada Graduate Scholarship (PT) and by NSERC Discovery Grants RGPIN-2016-46758 (WEH), 2020-03917 (MAF), and A7459 (CMM).

Availability of data, materials, and code Please see the *Methods* section of each experiment for the Open Science Framework links.

Declarations

Ethics approval The study was approved by the Wilfrid Laurier University Research Ethics Board (REB #6647).

Consent to participate Informed consent was obtained from all individuals included in the study.

Consent for publication Not applicable.

Conflicts of interest/competing interests The authors declare no conflicts of interests or competing interests.

References

Allen, S. W., & Vokey, J. R. (1998). Directed forgetting and rehearsal on direct and indirect memory tests. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 173–195). Lawrence Erlbaum Associates.

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*, *72*, 1–36. https://doi.org/10.1146/annurev-psych-072720-094140

Bancroft, T. D., Hockley, W. E., & Farquhar, R. (2013). The longer we have to forget the more we remember: The ironic effect of postcue duration in item-based directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 691–699. https://doi.org/10.1037/a0029523

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603–616. https://doi.org/10.1037/0278-7393.19.3.603

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III. & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Lawrence Erlbaum Associates.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning & Verbal Behavior, 9*(3), 255–268. https://doi.org/10.1016/S0022-5371(70)80059-7

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10*(2), 55–56. https://doi.org/10.3758/BF03331404

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1), 2515245920960351. https://doi.org/10.1177/2515245920960351

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analyses based on conceptual primitives. *Proceedings of COLING 2016, the 26th conference on computational linguistics: Technical papers* (pp. 2666–2677). The COLING 2016 Organizing Committee.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology: General, 15112*, 2990–3008. https://doi.org/10.1037/xge0001256

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77–90. https://doi.org/10.1016/j.jml.2013.09.003

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fawcett, J. M., & Hulbert, J. C. (2020). The many faces of forgetting: Toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory and Cognition, 9*(1), 1–18. https://doi.org/10.1016/j.jarmac.2019.11.002

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 36*(6), 1168–1181. https://doi.org/10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica, 139*(1), 84–90. https://doi.org/10.1016/j.actpsy.2011.10.001

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General, 145*(1), 56–81. https://doi.org/10.1037/xge0000128

Fellner, M.-C., Waldhauser, G. T., & Axmacher, N. (2020). Tracking selective rehearsal and active inhibition of memory traces in directed forgetting. *Current Biology, 30*(13), 2638–2644.e4. https://doi.org/10.1016/j.cub.2020.04.091

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short- and long-term consequences of directed forgetting in a working memory task. *Memory, 21*(7), 763–777. https://doi.org/10.1080/09658211.2012.754900

Festini, S. B., & Reuter-Lorenz, P. A. (2017). Rehearsal of to-be-remembered items is unnecessary to perform directed forgetting within working memory: Support for an active control mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(1), 94–108. https://doi.org/10.1037/xlm0000308

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Sage Publishing.

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Lawrence Erlbaum Associates.

Gravitz, L. (2019). The forgotten part of memory. *Nature, 571*(7766), S12–S14. https://doi.org/10.1038/d41586-019-02211-5

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1321–1330. https://doi.org/10.1037/0278-7393.18.6.1321

Hockley, W. E., Ahmad, F. N., & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory and Cognition, 44*(2), 220–228. https://doi.org/10.3758/s13421-015-0557-8

Lee, Y.-S., Lee, H.-M., & Tsai, S. (2007). Effects of post-cue interval on intentional forgetting. *British Journal of Psychology, 98*(2), 257–272. https://doi.org/10.1348/000712606X120410

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, AKA least-squares means. R Package Version 1.1. 3; 2018.

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences, 1424*(1), Article 1. https://doi.org/10.1111/nyas.13714

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software, 3*(26), 772. https://doi.org/10.21105/joss.00772

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Lawrence Erlbaum Associates.

Marevic, I., & Rummel, J. (2020). Retrieval-mediated directed forgetting in the item-method paradigm: The effect of semantic cues. *Psychological Research*, *84*(3), 685–705. https://doi.org/10.1007/s00426-018-1085-5

Mızrak, E., & Oberauer, K. (2022). Working memory recruits long-term memory when it is beneficial: Evidence from the Hebb effect. *Journal of Experimental Psychology: General, 151*(4), 763–780. https://doi.org/10.1037/xge0000934

Muter, P. (1980). Very rapid forgetting. *Memory & Cognition, 8*(2), 174–179. https://doi.org/10.3758/BF03213420

Neath, I., & Brown, G. D. A. (2012). Arguments against memory trace decay: A SIMPLE account of Baddeley and Scott. *Frontiers in Psychology, 3*, Article 35. https://doi.org/10.3389/fpsyg.2012.00035

Nickl, A. T., & Bäuml, K.-H. T. (2024). To-be-forgotten information shows more relative forgetting over time than to-be-remembered information. *Psychonomic Bulletin & Review, 31*(1), 156–165. https://doi.org/10.3758/s13423-023-02330-1

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences, 1424*(1), 239–255. https://doi.org/10.1111/nyas.13603

Oberauer, K., & Greve, W. (2022). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology: General, 151*(3), 513–541. https://doi.org/10.1037/xge0001106

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General, 142*(2), 380–411. https://doi.org/10.1037/a0029588

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language, 73*, 15–30. https://doi.org/10.1016/j.jml.2014.02.003

Pandey, A., Michaud, N., Ivanoff, J., & Taylor, T. (2023). Let me give you something to think about: Does needing to remember something new make it easier to forget something old? *Consciousness and Cognition, 115*, 103581. https://doi.org/10.1016/j.concog.2023.103581

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. https://doi.org/10.1037/rev0000161

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting is a feature, not a bug: Intentionally forgetting some things helps us remember others by freeing up working memory resources. *Psychological Science, 30*(9), 1303–1317. https://doi.org/10.1177/0956797619859531

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing https://www.R-project.org/

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Sahakyan, L., & Foster, N. L. (2016). The need for metaforgetting: Insights from directed forgetting. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 341–355). Oxford University Press.

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 219–248). Springer Tokyo.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229.

Tan, P., Ensor, T. M., Hockley, W. E., Harrison, G. W., & Wilson, D. E. (2020). In support of selective rehearsal: Double-item presentation in item-method directed forgetting. *Psychonomic Bulletin & Review, 27*(3), 529–535. https://doi.org/10.3758/s13423-020-01723-w

Taylor, T. L., Cutmore, L., & Pries, L. (2018). Item-method directed forgetting: Effects at retrieval? *Acta Psychologica, 183*, 116–123. https://doi.org/10.1016/j.actpsy.2017.12.004

Taylor, T. L., Quinlan, C. K., & Vullings, K. C. H. (2018). Decomposing item-method directed forgetting of emotional pictures: Equivalent costs and no benefits. *Memory & Cognition, 46*(1), 132–147. https://doi.org/10.3758/s13421-017-0751-y

Wetzel, C. D., & Hunt, R. E. (1977). Cue delay and the role of rehearsal in directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 3*(2), 233–245. https://doi.org/10.1037/0278-7393.3.2.233

Woodward, A. E., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 608–617. https://doi.org/10.1016/S0022-5371(73)80040-4