Intentional Forgetting: The Role of Retrieval in Encoding

by

Pelin Tan

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Randall Jamieson Professor, Department of Psychology University of Manitoba
Supervisors	Dr. Colin M. MacLeod Distinguished Professor Emeritus Department of Psychology
	Dr. Myra A. Fernandes Professor, Department of Psychology
Internal Members	Dr. Evan F. Risko Associate Professor, Department of Psychology
	Dr. Jonathan Fugelsang Professor, Department of Psychology
Internal-external Member	Dr. W. Richard Staines Professor Department of Kinesiology & Health Sciences

Author’s Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This research was conducted at the University of Waterloo by Pelin Tan under the supervision of Dr. Colin M. MacLeod and Dr. Myra A. Fernandes. Experiments 1-4 form an article that is under review for publication in the journal *Memory & Cognition* (Tanberg et al., under review, see references). This article is co-authored by myself, Myra A. Fernandes, Colin M. MacLeod, and William E. Hockley.

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) Alexander Graham Bell Canada Graduate Scholarship and by a NSERC Michael Smith Foreign Study Supplement, both awarded to me, and by NSERC Discovery Grants A7459 to Colin M. MacLeod, 2020-03917 to Myra A. Fernandes, and RGPIN-2016-46758 to William E. Hockley.

Abstract

Intentional remembering and intentional forgetting are adaptive processes that permit us to exert control over the contents of our memories. These abilities ensure that memory preserves the most goal-relevant information, and that goal-irrelevant information is discarded. Often studied in the literature using the *item-method directed forgetting paradigm,* research continues to debate the cognitive mechanisms that individuals use to intentionally remember and forget information. In this paradigm, during the study phase each presented item receives its own instructional cue—either to-be-forgotten (F) or to-be-remembered (R). The typical finding during the test phase is that memory is poorer for F items than for R items—the *directed forgetting effect*.

In this dissertation, in Chapter 2, I tested the assumptions of a prominent and longstanding account of the directed forgetting effect: the selective rehearsal account. To do so, I manipulated the time available for rehearsal and time-based decay. Four experiments investigated the influence of instructional cue durations of 1, 5, and 10 seconds. Experiments 1a and 1b, with the order of cue durations randomized, showed no effect of cue duration on recognition of either R or F single words. Experiment 2, using unrelated word pairs, again showed no effect of cue duration, here on associative recognition. Experiments 3 and 4 blocked cue duration and showed enhanced recognition of both R and F single words and word pairs with increasing cue duration. To explain this set of findings, I suggested that better memory for R items than for F items across cue duration depends on (1) a rapid retrieval check engaged for R items only and (2) a rapid removal process implemented for F items only. Additionally, any post-cue rehearsal is carried out only when cue duration is predictable and is equally likely for F items and R items.

In Chapter 3, I set out to test this rapid retrieval check mechanism by inducing an act of retrieval for F items. I predicted that if a rapid retrieval check of R items drives the directed forgetting effect, then inducing such a retrieval check mechanism for F items should reduce the magnitude of the directed forgetting effect. Experiment 5 demonstrated that

simply repeating an F item in the encoding list did not force the retrieval of that item. However, in Experiment 6, incorporating a button press for the participant to indicate noticing the repetition of an item revealed that, for items where this repetition was noticed, the directed forgetting effect was eliminated. Experiment 7 induced an act of retrieval using an immediate recognition task following the R/F cue presentation. Results indicated that a retrieval check was successfully induced for F items where the target on the recognition task matched the preceding item, eliminating the directed forgetting effect and confirming the critical role of retrieval in item-method directed forgetting. Experiment 8 replicated these findings while also demonstrating that the removal of F items from working memory did not depend on an active inhibitory mechanism, framing the retrieval-based explanation as a non-inhibitory account of intentional forgetting at encoding.

Ultimately, this dissertation provides a novel account of item-method directed forgetting in the form of a *selective retrieval account*. The account emphasizes the pivotal contribution of a rapid retrieval check mechanism applied only to R items during encoding being primarily responsible for the directed forgetting effect. This new theoretical perspective opens avenues for future research on intentional forgetting to explore the role of retrieval during encoding.

Acknowledgements

In June 2017, Dr. William Hockley, my undergraduate mentor, supervisor, and now good friend, entrusted me, a new research assistant, with the opportunity to present a poster on directed forgetting at my first conference—a subject I had limited knowledge about. At the conference, I had the privilege of meeting Dr. Tyler Ensor, who generously provided a concise overview just 30 minutes before my presentation. While presenting my poster, I had a memorable encounter with Dr. Colin MacLeod, where he asked me a simple question that daunted me for years: "How does the selective rehearsal account explain these results?"

I did not know the answer.

Fast-forwarding seven years, I can confidently say that, not only answer his question but provide new contributions to the directed forgetting literature. Throughout my graduate school journey, little did I anticipate that my journey would culminate in training alongside two eminent and nurturing researchers: Dr. Colin MacLeod and Dr. Myra Fernandes. I am indebted to both of you for your unwavering support, from countless Teams calls to patiently listening to my life's random musings, it has been the utmost honour, and blast, to be co-supervised by you. Whether offering encouragement during the toughest times, providing valuable guidance and feedback, or challenging me because you know I can do better, you have significantly contributed to my growth as a researcher. Your love and passion for what you do as scholars has been infectious and taught me to be a better scholar, while having fun. I can recall numerous occasions where I felt disheartened about my research, but inevitably, a meeting with you was sufficient to reinvigorate my excitement. Because of the both of you, I can confidently say that I had a blast during my PhD, and that I will continue to pay this positive, joyful, kind, yet productive experience forward. Wherever life takes me, I aspire to be half the supervisor you both have been to me. I feel incredibly privileged to have both of you as supervisors.

Dr. William Hockley: You have been an integral part of my graduate journey, and I cannot express enough gratitude to you for seeing something in me with academic research. You gave me many of my first opportunities to learn, thrive, and grow as a researcher and I aspire to take that with me and pay it forward to another researcher-in-training one day. Although you did not have the obligation to continue to work with me, yet alone provide more opportunities, you continued and never gave up on me. I am grateful for our lunches and to call you a friend. Cheers.

I would also like to thank Dr. Klaus Oberauer, my University of Zurich supervisor, who also significantly contributed to my growth as a scholar and welcomed and trained me in Zurich like one of his own graduate students. I am grateful for your thoughtful comments, challenging questions, and for our continuous collaborations. I deeply cherish my time with you in Zurich and look forward to our future work together.

I would like to thank my internal committee members, Dr. Evan Risko and Dr. Jonathan Fugelsang for engaging so much with my graduate progress, sharing your expertise, and thoughtful comments. You played an essential part in making me feel supported during my time in the department. Thank you for seeing the value in my work and in me. In addition, I would like to give a special shout-out thank you to Dr. Britt Anderson for always coming to my Cognition seminars, even though he knew I was going to present at the CNS seminar anyways. Thank you for sharing your profound knowledge and your perspectives on my work and for challenging me to keep moving it forward. Another special shout-out to Dr. Tyler Ensor for always picking up the phone and spending hours talking to me about research, life, and the future.

I am also grateful to have been surrounded by such amazing, dedicated, and supportive fellow past and present graduate students: Dr. Ryan Yeung, Dr. Brady Roberts, Sophia Tran, Yadu Sivashankar, Adam Cox, Khalil Husein, and Brian Kim. Thank you for all of your heart-to-heart conversations, keeping up with my shenanigans, and giving me encouragement. Beyond my lab, thank you to the following list of people who continue to make me feel supported and cherish me with their friendship: Alex, April, Julianna, Kate, Laura, Megan, Torin, Michelle, and Sophie. We all better stay friends. Lastly, thank you to all devoted research assistants who also made this work possible: Patrick Tsapoitis, Evan Skorski, Jerika Chuong, Kristen Sullivan, and Calahndra Blake. You are all going to great places, and I sincerely thank you for your time and effort.

I had the absolute privilege of spending time during my PhD at the University of Zurich, where the entire lab went, and continue to, above and beyond, in supporting me throughout my research. Thank you to: Gidon, Lea, Ven, Isabel, Ruhi, Chenyu, Joscha, and Philipp for making me feel at home during my time, providing endless support, and including me within your journey in Zurich and at other conferences. Finally, thank you, Hannah Dames, for coming to my online poster and fostering not only a collaboration, but a wonderful friendship, with me.

A huge shout-out some friends who spent years listening to me talk about intentional forgetting and act excited while providing unwavering support in many ways: Jessica, Elbethel, Aamna, Dr. Geoff, Liza, Tegan, Anna, Victoria, Emily, Edwyn, and Carol. Thank you.

Finally, to my husband who deserves to wear that t-shirt “I survived my wife’s PhD”. Moritz, I find it difficult to express my appreciation because it is so boundless and unconditional. You are my most enthusiastic cheerleader; you are my best friend; and an amazingly supportive husband. You are my rock, my shoulder to lean on, and my soulmate. Without your support, I would have been lost, and this dissertation would have taken even longer to complete. You have been there for all the ups and downs, challenging me, cheering me on, taking care of everything else so I can just focus on my PhD, and even more. You have shared this entire journey with me, so it only seems right that I dedicate part of this dissertation to you.

I close off this long acknowledgement section with eternal thank you and love to my parents, Selda and Ufuk. Since my early days, it has been the three of us. Without them, I have no idea who I am. Without them, I would not be here, writing this dissertation in Canada. I have learned so much from you and I appreciate all of the sacrifices that you have made for me. Every time I was sure I could not do something, you both convinced I could. I am grateful that the four of us are now the clingiest family ever. Thank you for everything.

Dedication

From Istanbul to Toronto,

Here's to another accomplishment,

That all began with six red Samsonite luggage and a dream,

Never losing sight of the sacrifices, hard work, and love.

Thank you for everything, Anne & Baba,

For bringing me to Canada,

This is all dedicated to you,

I hope you are proud.

Sizi çok seviyorum.

I love you very much.

♥

Table of Contents

List of Figures

List of Tables

List of Abbreviations

RT = Response Time

ISI = Inter-stimulus interval

Chapter 1:
General Introduction

Forgetting is ordinarily considered to be inconvenient and undesirable. No one wants to forget an important speech when standing before a crowd, or the birthday of a friend or family member. In a conversation with a friend, despite actively listening to them, should you nevertheless forget what they said, your friend might well perceive this forgetfulness as indicating that you were not actually attending. These are, however, instances of incidental 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 (e.g., forgetting irrelevant information that your friend just mentioned). For example, if you are baking with your partner, and you hear him telling you that the recipe requires ½ cup of brown sugar, it would be very important to remember the amount of the brown sugar. In addition, if he tells you that the recipe also requires 1 teaspoon of cinnamon, you need to remember to add it. However, when it is time to add the dry ingredients, he tells you that he was mistaken—that the recipe actually does not require any cinnamon. Thus, it is important for you to intentionally forget that the recipe indicated 1 teaspoon of cinnamon and to intentionally remember that the recipe needs ½ cup of brown sugar.

For decades, researchers have been interested in understanding intentional forgetting as a means to control memory. In the laboratory, over a half century ago, *directed forgetting* paradigms were developed (e.g., Bjork, LaBerge, & Legrand, 1968) to investigate how individuals accomplish the forgetting of irrelevant or outdated information. Relative to instructions to remember specific information, instructions to forget specific information actually does reduce later memory for that specific 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.

1.1 Paradigms of Intentional Forgetting

In a typical directed forgetting paradigm, participants study a series of items for a later memory test. Each item is followed by a cue designating 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 items, a performance difference called the *directed forgetting effect*.

The finding that memory is better for R items than for F items is not surprising, given that participants believe that only the R items will be tested. The finding that memory is poorer for F items is more surprising. This reduction highlights the ability to intentionally control the contents of memory and either to limit the encoding of F items, and hence minimize their likelihood of them being stored, or to reduce their accessibility from long-term memory, or both. Thus, much of the research on intentional forgetting has emphasized how we forget on demand and the mechanisms underlying voluntary memory control.

Over the years, two now widely used directed forgetting procedures have been developed. In the *list-method* paradigm (see Sahakyan et al., 2013, for a review), participants study items in two lists and are divided into two groups. In the F group, participants are told after the first list that they should forget it. This F instruction can be delivered by telling participants that List 1 was presented in error, simulating a computer-crash that requires restarting the experiment, or by an explicit instruction to forget the List 1 items. Following the F instruction, participants study a second list (i.e., List 2) and are told to remember this list for a later test. In the R group, participants are told to continue remembering List 1 before studying List 2. During a later memory test, in which participants are asked to recall all of the studied words, those in the F group recall fewer List 1 items (i.e., List 1 cost) and more List 2 items (i.e., List 2 benefit) than do participants in the R group. The cost and benefit effects from list-method directed forgetting have been the basis for investigating the mechanisms of intentional forgetting well-encoded information (see MacLeod, 1998; Sahakyan et al., 2013; Sahakyan & Foster, 2016 for reviews).

To date, two principal mechanisms have been proposed to explain these results. Under the context change account (Sahakyan & Kelley, 2002), context continually drifts during study of the two lists, such that the context at the time of test more closely resembles—and hence favours the recall of—List 2 items. Under the inhibition account (Zacks, Radvansky, & Hasher, 1996), the F instruction following List 1 leads the participant to suppress those items.

Although the list-method directed forgetting paradigm has been a useful tool to investigate how we forget information that has been encoded in long-term memory, the focus of this dissertation is to examine how we select to-be-remembered information for encoding into long-term memory and how we remove to-be-forgotten information to prevent its encoding. For example, if you are driving to the gym with your partner, and they tell you to turn right at the next stop sign, you would intentionally remember this information to correctly follow directions. However, if your partner suddenly changes their instruction, and instead tells you to go straight past the stop sign, you would want to intentionally forget turning right . As a consequence, the irrelevant directional instruction does not interfere with your goal of intentionally remembering the relevant directional instruction. This type of control process is essential for an adaptive memory system that monitors the relevance of each event and accordingly selects relevant information for further encoding and irrelevant information for removal. Referred to as *intentional forgetting at encoding*, the focus of this dissertation is to understand this type of selection to control the contents of memory.

Intentional forgetting at encoding is studied in the laboratory using the *item-method directed forgetting* paradigm. In this paradigm, each individual item receives its own R or F cue immediately after its presentation. Critically, participants are told that they will only be tested on the R items, however, they are tested on both R and F items. Decades of research demonstrate robust directed forgetting effects, where performance is better for R items than for F items (for reviews, see MacLeod, 1998; Sahakyan & Foster, 2016). This effect is prominent for tests of recall and for tests of recognition, whether yes/no recognition (see Basden & Basden, 1996, for a review) or 2-alternative forced choice recognition (Ahmad, Tan, & Hockley, 2019; Montagliani & Hockley, 2019; Whitlock et al., 2022). The directed forgetting effect in the item-method paradigm has also been replicated using a variety of stimuli including pictures (Hauswald & Kissler, 2008; Quinlan et al., 2010), scenes and objects (Ahmad et al., 2019; Whitlock et al., 2022), abstract images (Fawcett, Lawrence, & Taylor, 2016; Hourihan, Ozubko, & MacLeod, 2009), and event segments (Fawcett, Taylor, & Nadel, 2013).

In this dissertation, I use the item-method paradigm, given the goal of examining how information is selected to be promoted from working memory to long-term memory. Item-method directed forgetting has enriched our understanding of how information in working memory (i.e., system responsible for holding a limited amount of information available for processing: Baddeley, 1986; Bartsch, Singmann, & Oberauer, 2018) is selected or not selected to be incorporated into long-term memory (Dames & Oberauer, 2022). It is important for irrelevant information to be expunged from working memory so that it does not deplete limited capacity resources and so that it is not further processed and retained (Lewis-Peacock et al., 2018; Popov & Reder, 2020). Thus, as a memory control process, item-method directed forgetting provides a way to examine how capacity is freed to be devoted to processing of relevant information and simultaneously to reducing memory for irrelevant information.

For over a half century, much of the research concerning item-method directed forgetting has sprung from the debate regarding the mechanisms invoked when intentionally remembering and intentionally forgetting information. In the following sections, I provide a review of the current theoretical accounts of the cognitive mechanisms of intentional forgetting at encoding.

1.2 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, Ensor, Hockley, Harrison, & Wilson, 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 *where* it operates (Marevic & Rummel, 2020; ; Sheard & MacLeod, 2005; Taylor, Cutmore, & Pries, 2018), the primary mechanism remains well-articulated: R items are rehearsed more than F items.

Upon item presentation, in anticipation of the memory instruction, participants retain each item in working memory via maintenance rehearsal (Bartsch, Singmann, & Oberauer, 2018; Greene, 1987). Following an instruction to remember the just-presented item, participants engage in elaborative rehearsal of that item to further encode it into long-term memory. This elaborative rehearsal is critical in strengthening the representation of R items in long-term memory. In contrast, following an instruction to forget, the early view of the selective rehearsal account was that participants terminated maintenance rehearsal of the item allowing it to deteriorate, or decay, in working memory (Basden, Basden, & Gargano, 1993). Given the considerable evidence against decay in working memory, however, more recent explanations of the selective rehearsal account instead state that upon receipt of an instruction to forget, participants cease maintenance rehearsal of the F item, devoting any further rehearsal while the F cue is present exclusively to the elaborative of previous R items, a process referred to as cumulative rehearsal of R items. In this way, only R items benefit from elaborative rehearsal and forgetting occurs as a by-product of not elaboratively rehearsing F items. Critically, the selective rehearsal account asserts that the directed forgetting effect in the item-method paradigm arises from the selective (i.e., elaborative and cumulative) rehearsal of the R items, to the exclusion of the F items (Basden, Basden, & Gargano, 1993; Tan et al., 2020).

Research comparing the memorial representations of R and F items has demonstrated that participants recognize more specific or perceptual details of R items than of F items regardless of whether the stimuli were words (Lee, Lee, & Fawcett, 2013; Montagliani & Hockley, 2019), complex pictures of scenes and objects (Ahmad, Tan, & Hockley, 2019), abstract images (Fawcett, Lawrence, & Taylor, 2016), or event segments (Fawcett, Taylor, & Nadel, 2013). Thus, R items receive elaborative encoding with more specific details whereas F items are only remembered at a gist-like level. In addition, Golding, Long, and MacLeod (1994) also supported the selective rehearsal account by examining the influence of semantically related compound word pairs in the item-method paradigm. On some trials, words from earlier in the list were semantically related to the words from later in the list (e.g., crab – leg). When the first word of the pair was followed by an R cue but the second was followed by an F cue, memory for both words was similar to when both words were followed by an R cue. Thus, when a subsequent F word was semantically related to a prior R word, participants were predisposed to elaboratively rehearse the second word from the pair because of this association, even when instructed to forget the later word.

Pandey et al. (2023) also provided support for the cumulative rehearsal of R items. They found that replacing an F item with a new to-be-remembered item did not improve intentional forgetting relative to presenting only an instruction to forget. In addition, memory for R items and new to-be-remembered items did not differ, indicating that participants engaged in similar rehearsal strategies for R items and for the new items presented together with an F cue. If previously presented R items are cumulatively rehearsed during the F cue presentation, then integrating the new to-be-remembered item provides just another item to add to the rehearsal set. Further, Hourihan (2021) found that receiving F items from a mostly remember source (i.e., picture of an individual with a higher probability of presenting R items) improved memory for those F items relative to receiving F items from a mostly forget source. Contrarily, receiving R items from a mostly forget source reduced memory for those R items compared to receiving the items from a mostly remember source.[1]

Finally, in my work, my colleagues and I tested the selective rehearsal account using a novel variant of the item-method directed forgetting paradigm (Tan et al., 2020). Participants studied two unrelated items at a time. For some pairs, both items were followed by the same cue, and participants were instructed to remember or to forget both items (pure condition). On other trials, participants were to remember one word but to forget the other word (mixed condition). In the mixed condition, the task is either remembering another item or forgetting another item. Performance for one of the words in a pair, whether remembering or forgetting, should decline as the cognitive demand associated with the other word increases (from F to R). The principal assumptions of the selective rehearsal account state that intentional remembering via differential rehearsal is effortful but that intentional forgetting is passive and does not demand cognitive resources. We reasoned that, if true, memory for one of the words in a pair should be poorer when the task for the other word is remembering rather than forgetting because remembering imposes a heavier cognitive load than forgetting. This would support the key assumptions of the selective rehearsal account.

[1] Although Hourihan (2021) found this decreased memory for R items from a mostly forget source compared to from a mostly remember source, memory for these R items did not drop below memory for F items from a mostly forget source: Hourihan still demonstrated robust directed forgetting effects.

We found that, for R items, participants recognized more R items in the mixed condition than in the pure condition. Similarly, for F items, participants recognized more F items in the mixed condition than in the pure condition. Additionally, in another experiment that incorporated a baseline instruction (i.e., neutral/no-cue condition), memory for F items in both the pure and mixed conditions did not differ from neutral item memory. Consistent with the selective rehearsal account, this confirmed that the intentional remembering of R items—accomplished via elaborative rehearsal—requires more cognitive resources than does the intentional forgetting of F items.

1.3 Rationale for the Current Investigations

Although the selective rehearsal account is generally accepted as explaining 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 (Fawcett & Taylor, 2010; Taylor & Fawcett, 2010; Fawcett & Taylor, 2012; Anderson & Hulbert, 2021). The passive view of forgetting has been a critical component of the selective rehearsal account, yet substantial evidence has been proposed against decay in working memory (Barrouillet et al., 2004; Neath & Brown, 2012; Oberauer & Lewandowsky, 2013, 2014; Ricker et al., 2016) and intentional forgetting (Dames & Oberauer, 2022; Fawcett & Taylor, 2008; Fawcett, Lawrence, & Taylor, 2016; Lewis-Peacock et al., 2018; Oberauer & Werner, 2022). Given that a critical assumption of this account has been heavily criticized, a major goal in the field should be to test the validity of this aspect of the account. Lastly, although the elaborative rehearsal of R items is generally accepted, other assumptions of the selective rehearsal account—notably the cumulative rehearsal of R items—are yet to be systematically examined. If the selective rehearsal account is to be accepted as the dominant explanation of item-method directed forgetting, these assumptions must be verified. Therefore, one of the goals of this dissertation is to test these key assumptions of the selective rehearsal account.

In the second chapter of this dissertation, I present a systematic examination of the two key assumptions of the selective rehearsal account by exploring how processes to intentionally remember and forget are influenced over time. Little to no previous research has explored the time course of processing in the item-method directed forgetting paradigm. In this chapter, I investigate how manipulating the time allocated to remember or to forget information influences recognition performance. I introduce a novel, larger range of cue duration 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. Here, I reveal that the time course of remembering and forgetting during cue presentation is important: Participants employ different strategies to intentionally remember and forget depending on their awareness of the duration of the cue. Finally, I discuss my findings with implications for the accounts of item-method directed forgetting, particularly questioning the critical assumptions of the selective rehearsal account.

In the third chapter, I propose a novel account of item-method directed forgetting to address the shortcomings of the selective rehearsal account revealed the second chapter. Here, I explore a fundamental process in human memory (i.e., retrieval) yet to be characterized within the item-method paradigm. Specifically, I examine the possibility that for R item representation to be differentially strengthened relative to F items, R items must receive a rapid boost in the form of a retrieval check. This check occurs only following an R instruction, bringing back the just-presented item to the focus of attention to ensure its registration. The act of quickly checking the retrieval of an R item improves its elaborative encoding, ultimately providing the benefit to the R items in the directed forgetting effect. Consequently, there is active retrieval of the R items that is not accorded to the F items.

Across four experiments, I address the question: Is such a retrieval check for R items what underlies the directed forgetting effect? This possibility leads to a testable implication: If this minimal act of retrieval ordinarily provides R items with a memory boost, would inducing this same retrieval check for the F items reduce—or perhaps even eliminate—the directed forgetting effect? This work introduces a new mechanism to a literature that has predominantly focused on a single theoretical debate: whether intentional forgetting is active or passive. In this work, I accentuate the importance of how strengthening R items via a single selective retrieval contributes to the directed forgetting effect.

Chapter 2:
On the Time Course of Intentional Remembering and Forgetting

For decades, experimental research on item-method directed forgetting was almost exclusively conducted with the same limited set of cue durations (i.e., ~2-3 s; Basden, Basden, & Gargano, 1993; MacLeod, 1999; Hourihan & Taylor, 2006; Fawcett & Taylor, 2008; Quinlan et al., 2010; Burgess et al., 2017; Ahmad et al., 2019; Hourihan, 2021; Tan et al., 2020; Tanberg et al., 2022). Consequently, theoretical accounts were usually evaluated in terms of how well they explained findings within this time course. Yet, in the list-method directed forgetting literature, decreasing the length of the retention interval (i.e., duration between study phase and test phase) challenged the selective rehearsal account. List 1 forgetting after a short retention interval was observed not only for intentionally studied items but also for incidentally encoded items (Sahakyan & Delaney, 2010). Because participants should not have the incentive to selectively rehearse the incidentally encoded items, researchers proposed that the selective rehearsal account could not be the sole mechanism to explain list-method directed forgetting (Geiselman et al., 1983). In contrast, increasing the retention interval in an item-method directed forgetting paradigm revealed support for differential rehearsal processes: Relative time-dependent forgetting rate was higher for F items than for R items even up to one week retention intervals.

Although, as discussed earlier, directed forgetting effects within the two types of paradigms have been attributed to different mechanisms (i.e., context-change account of list-method, Sahakyan & Kelley, 2002; selective rehearsal account of item-method, Basden, Basden, & Gargano, 1993; Pandey et al., 2023; Tan et al., 2020; attentional inhibition account of item-method, Fawcett & Taylor, 2008; 2012), collectively, these findings suggest that the time course of intentional remembering and intentional forgetting is critical for understanding directed forgetting. In this chapter, my goal was to manipulate the time course of these processes by varying the duration of the R/F cue presentation.

To date, there have been conflicting findings regarding the influence of cue duration in the item-method paradigm. Several studies have reported enhanced item recognition of both R and F items with increasing cue duration (Bancroft et al., 2013; Lee, Lee, & Tsai, 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). For example, for item recognition, Bancroft et al. (2013) used yes/no recognition and R/F cue source judgment tests to investigate whether increasing cue duration contributes to item-method directed forgetting. Despite consistently observing robust directed forgetting effects at each cue duration, they found that increasing cue duration benefited performance for both R and F item performance. Lee et al. (2007) concurred, finding that increasing cue duration improved recall and recognition for both R and F items. Contrarily, Allen and Vokey (1998) found no effects of cue duration for the recognition of R and F items at durations of 1 s and 3 s.

These three studies featured several methodological differences. First, in the Allen and Vokey (1998) study, cue duration was a between-subjects condition: 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. Wetzel and Hunt (1977) also examined how increasing cue duration influences item-method directed forgetting. Currently, their investigation remains one of the most in depth examinations of length of time in this paradigm 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), where recall memory for both R and F items increased with increasing cue duration. However, although their recognition memory results were consistent with their recall performance results, their results for recognition are problematic to interpret because their recognition tests followed two recall test phases. Therefore, the conflicting findings on recognition memory and cue duration are yet to be resolved with respect to the assumptions of the selective rehearsal account. Overall, it is possible that *how* cue duration is manipulated contributes to the mechanisms of intentional remembering and intentional forgetting. One of the principal purposes of the current experiments was to help resolve these conflicting findings in the literature.

My manipulation of cue duration differs from that of Bancroft et al. (2013) who examined only quite short cue durations in item memory (Experiments 1 and 2: 300, 600, or 900 ms; Experiment 3: 1, 2, or 3 s), and that of Lee et al. (2007) who extended the range of durations (Experiments 1 and 2: 1 or 5 s) and found a robust directed forgetting effect. The longest cue duration for item memory in these earlier studies was 5 s. Here, I 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 rehearsal set, where the rehearsal set contains only R items, increasing the cue duration should increasingly benefit memory for R items particularly when cue durations are quite long.

Previous research has been limited in that little to no research has provided a systematic examination of how time allocated to remember or to forget information influences recognition performance in the item-method directed forgetting paradigm. In doing so, these experiments will help to further unravel the mechanisms mediating item-method directed forgetting. More important, manipulating the time course of intentional remembering and intentional forgetting provides a direct test of the assumptions of the selective rehearsal account. As cue duration increases, so should memory for R items due to the increasing opportunity to rehearse them elaboratively. This should not be true for F items, however, which are supposedly not being elaboratively rehearsed at all.

2.1 Experiment 1a

In this first experiment, I examined the influence of this broader range of cue durations in the typical single-word list learning procedure. The overarching goal was to explore whether more extended cue presentations would progressively enlarge the advantage of R items over F items, consistent with increasing the available time for differential elaborative 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?

If participants strategically switch to elaborative rehearsal when given an R cue, then memory for R items should increasingly diverge from memory for F items, as more time becomes available for selective rehearsal. This should be reflected in an interaction between cue instruction and cue duration. With respect to F items, there are three possible outcomes. First, Bancroft et al. (2013) found an increase in F item performance with increasing cue duration. If their findings extend to the longer cue durations, then I 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. 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, and that interference from ongoing, cumulative rehearsal of R items would be continuous, such that individual cue durations would not influence memory for F items.

2.1.1 Method

All experiments were preregistered. The complete OSF page for this investigation can be found at https://osf.io/ph3m5/.

Participants.

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

Sixty-nine participants were recruited from Prolific (https://prolific.com/). Their mean age was 28.45 (*SD* = 1.41), 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. I 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 (~ USD $3.55) for their participation. All experiments in Chapter 2 received ethics approval by the Wilfrid Laurier University Research Ethics Board (REB #6647) and University of Waterloo Office of Research Ethics (Protocol #41459).

Materials.

A master word list of 200 common, high-frequency English words was generated from the SenticNet 4-word corpus (Cambria, Poria, Bajpai, & Schuller, 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, Quinlan, and Vullings (2018).

Stimuli for all tasks were presented using jsPsych (de Leeuw, 2015). The master word list is available at OSF: 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. The word was immediately followed by a R or F cue in the center of the screen for either 1 s, 5 s, or 10 s. Cues were presented in Open Sans font size 18. The study phase included 72 single words, with half followed by a R cue and half followed by a 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 were inserted at the beginning and another three at the end of the study list to serve as primacy and recency buffers; these items, all of which were R-cued, were not tested. The order of the cue duration trials was random. Participants were instructed to remember all words followed by a R and to forget all words followed by a 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 following the test phase to assess self-reports of whether participants followed the instructions. The post-experiment questionnaire and experimental instructions are available on the OSF page, https://osf.io/gez4u/.

Data Analysis.

For all experiments in Chapter 2, 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). Lure recognition trials were removed because the false alarm rate was the same for all conditions[2]. The structure of the random effects models followed the recommendations of Brown (2021). For the random effects structure, I included a model with by-participant and by-item random intercepts. I additionally included by-participant and by-item random slopes for the effects of Cue Type and Cue Duration only if including these random slopes improved model fit. If the model was singular, I continued to reduce the model complexity to report the best-fitting maximal model, and I state the random effects structure used for that

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

model. After fitting the models, I 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., 2019) with Tukey 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) and Cohen's *d*. In addition, odds ratio statistic 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 for 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 so I could observe the 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 at OSF (Data: https://osf.io/n2d5j/; Analysis: https://osf.io/h7msw/).

2.1.2 Results

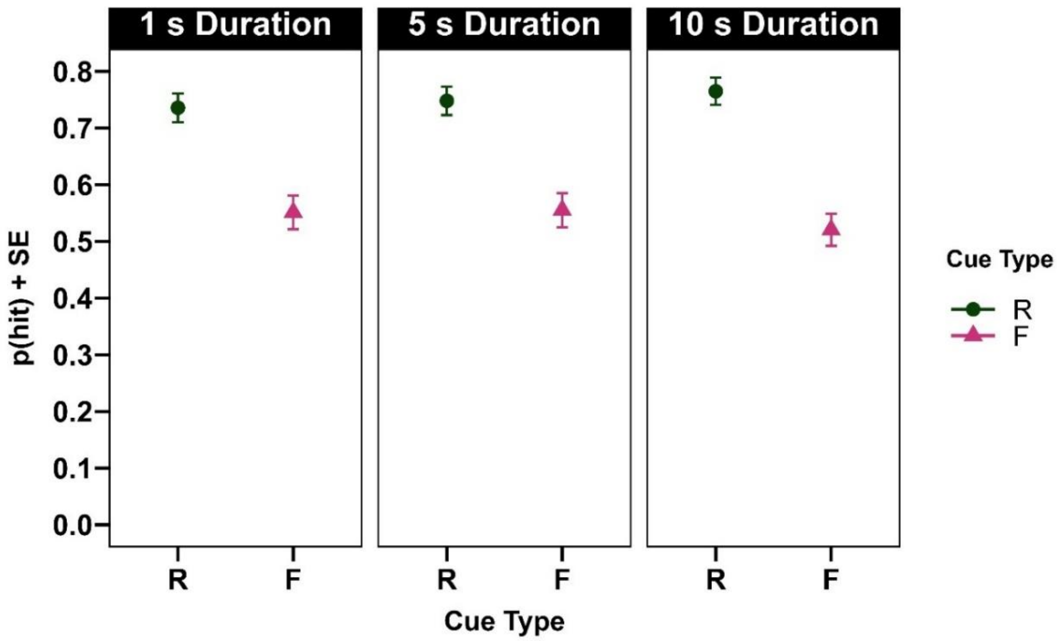
Figure 1 presents the *proportion hit* scores as a function of Cue Type and Cue Duration. The mean collective False Alarm Rate for all conditions was .20 (*SE* = .02).

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). There was no Cue Type x Cue Duration interaction, 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. However, collapsed across Cue Type, memory performance remained not significantly different across levels of the Cue Duration condition, *ORs* > .84, *CIs* = .62 – 1.52, *ps* > .330.

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

Figure 1. Experiment 1a: Mean recognition performance represented as proportion hit on the yes/no recognition test.



Note. 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 error.

2.1.3 Discussion

As expected, there was a robust directed forgetting effect across all 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 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—was unexpected under the selective rehearsal account, I decided that a replication was warranted.

2.2 Experiment 1b: Replication

2.2.1 Method

Participants.

The sample size calculation was identical to Experiment 1a, although the final sample fell a little short. Fifty-nine participants from Prolific participated. Their mean age was 31.23 (*SD* = 2.43) 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.

2.2.2 Results

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

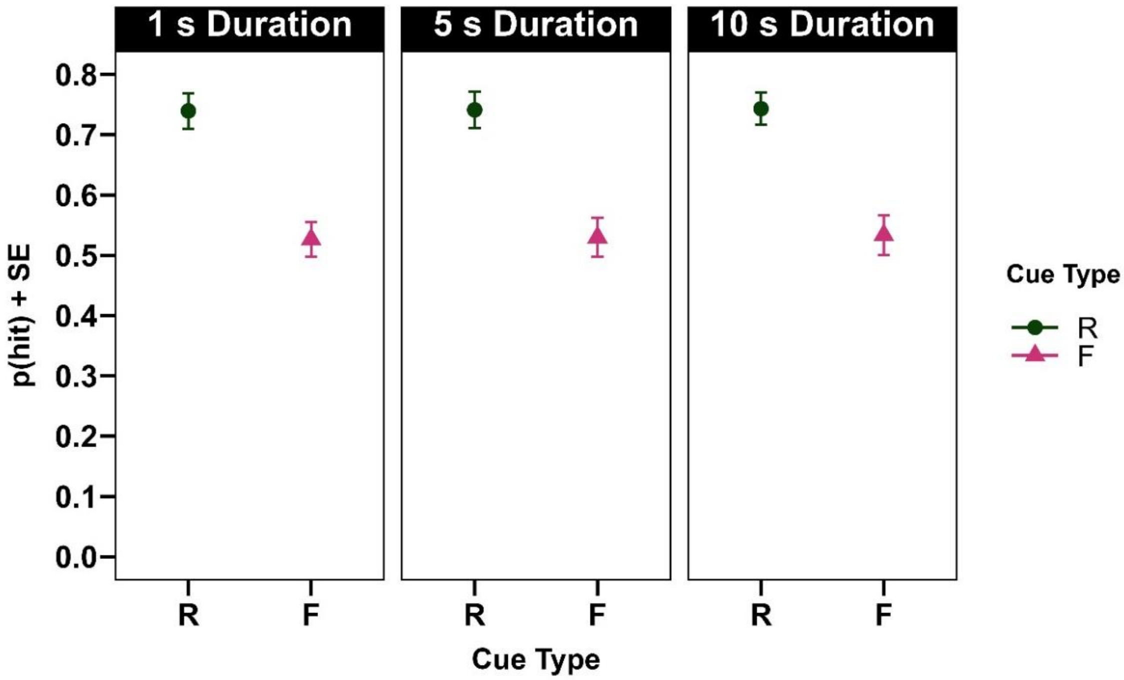
The results of Experiment 1a replicated in their entirety. There was a significant effect of Cue Type, χ^2 (1) = 198.00, $p < .001$, where participants overall recognized more R items (*M* = .77, *SE* = .02) than F items (*M* = .54, *SE* = .03). There was no main effect of Cue

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 x Cue Duration interaction, χ^2 (2) = .02, *p* = .991.

Separated by Cue Duration, there were significant 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. However, collapsed across Cue Type, memory performance remained not significantly different between levels of the Cue Duration condition, *ORs* > .97, *CIs* = .71 – 1.39, *ps* > .956.

The trend analysis examined the overall pattern across the Cue Duration condition as a function of Cue Type condition. 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 conditions.

Figure 2. Experiment 1b: Mean recognition performance represented as proportion hit on the yes/no recognition test.



Note. 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 error.

2.2.3 Discussion

In Experiment 1b, I again observed no influence of cue duration on memory for either R items or F items, replicating the pattern seen in Experiment 1a. This confirmed that increasing cue duration did not influence the directed forgetting effect, with respect to either the R items or the F items. Again, providing participants more time for R items did not enhance their recognition of those items, and providing more time for F items did not bring about more forgetting of them. 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 elaboratively 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 a R cue, the current item should be entered into the cumulative rehearsal set; following a 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 I 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 are instead expected to 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. The 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, Lewandowsky, and Oberauer (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 there is still above-chance performance for F items: Removal of irrelevant information only operates to stop additional processing of an irrelevant item.

A Novel Retrieval-Based Explanation.

To explain this collection of findings, I 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). When a R cue is presented, to ensure that the just-presented item is accessible in working memory because it now must be remembered, the participant performs a quick retrieval check, basically asking "what is the current item?" This retrieval check, which can also be seen as redirecting attention to the item in working memory, is effectively a kind of recognition test of the just-presented item that is now to-be-remembered. It is selective in that it is performed only on R items. 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 (Dames & Oberauer, 2022), which therefore is visible as a consistent memory benefit for R items over F items, independent of cue duration. Under this account, there is neither selective rehearsal nor cumulative rehearsal of R items, only selective retrieval of them. In Chapter 3, I test the assumptions of this account.

2.3 Experiment 2

In Experiments 1a and 1b, I examined memory for individual items. *Item memory* refers to remembering representations of individual events. However, human memory is exquisitely associative. Linking a name to a newly met face helps to remember this individual at a later date; associating a series of events in the order in which they occurred (e.g., remembering the order of the stores that you visited at the mall) helps to keep track of errands. This type of *associative memory* refers to the ability to create relations between individual events, or to bind two otherwise separate representations.

Forming bindings is an essential component of human cognition often requiring the association of unrelated individual items to each other to create a coherent story of an everyday event. For example, if you go to the mall to shop for a gift for your partner, and you remember them needing new socks, it would be very important to remember which of those stores at the mall had the best and worst options for socks: If you had visited the first store, and did not find the right size, then you need to associate this store with the inadequate socks option—binding the specific details of these socks (e.g., incorrect size) with the first store (e.g., location of the store)—so you can remember not to go back to that store. But, if you had visited a second store and found the exact material, colour, and size of their preferred socks, then you need to bind the location of this store to the adequate socks you purchased as a gift, so you can remember where to go for their favourite socks in the future. This example demonstrates that it is important to form associations between individual events to accumulate relevant information in long-term memory.

Typically, the time course of intentional remembering and intentional forgetting has been investigated using item memory. However, most models of working memory have proposed that forming bindings between representations of individual items (e.g., unrelated words presented together) is essential to promote an efficient working memory system. The *binding hypothesis* suggests that working memory has a limited capacity because of the number of bindings, rather than the number of individual items, that it can maintain at a time (Bartsch & Oberauer, 2023; Cowan, 1995; Dames & Oberauer, 2022; Oberauer, 2005; Oberauer, 2009a,b, 2019; Oberauer & Vockenberg, 2009). Given that, in the item-method directed forgetting paradigm, participants make the selection to intentionally remember versus forget while the item is being maintained in working memory, it is important to understand how an instruction to Remember or Forget influences associative memory for bindings.

Although previous investigations of associative memory and item-method directed forgetting have demonstrated that participants can, indeed, intentionally remember and forget associative information, demonstrating robust directed forgetting effects, it is not clear whether the time course of processing influences intentional remembering and intentional forgetting of associative information. Investigating associative, or binding, 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, my 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 that they were instructed to associate, participants were asked at test to discriminate intact pairs—words that had been presented together at study—from rearranged pairs—words that had not been presented together at study (i.e., they had been presented, but in separate pairs). This is, the classic associative recognition test (see, e.g., Hockley, 1992; Hockley, Ahmad, & Nicholson, 2016). Here, participants must rely on memory for the associations that they formed between pairs of items to answer accurately: Memory for the individual words cannot aid in discriminating between intact and rearranged word pairs because all of the individual test items were studied.

Bancroft et al. (2013; Experiment 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 equivalently with cue duration for both R and F pairs. Given that the Bancroft et al. pattern for longer cue durations did not replicate in Experiments 1a and 1b, my principal goal in Experiment 2 was to investigate whether associative recognition would be influenced by the longer cue durations. I foresaw three possible outcomes. First, I could replicate the findings of Bancroft et al., with memory for both R and F items increasing over my cue duration manipulation (1 s, 5 s, and 10 s). Second, if forming associations between the two members of an R pair 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) should not affect associative recognition of F pairs. Of course, it is also possible that, similar to Experiments 1a and 1b, I would find no influence of increasing cue duration on associative recognition in a mixed-list design. According to my retrieval-based explanation, providing more time should not affect associative recognition because participants perform an immediate retrieval check for the members of an R pair. In this case, rather than elaborating on the association upon receiving an R cue, participants boost the representation of the R pair via retrieval. However, because this retrieval check is quick, associative memory for R pairs should not differ across the cue duration conditions.

2.3.1 Method

Participants.

A sample size calculation was conducted 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 power of ~.80 ($\alpha = .05$, two-tailed). This yielded a required sample size of *n* = 96.

One hundred and six participants were recruited from Prolific. Their mean age was 32.41 (*SD* = 2.65), 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 therefore included 96 participants. All other Prolific and participant criteria were identical to the previous experiments.

Materials.

The master word list was identical to 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, I 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. The study phase included 72 word pairs, with half followed by a R cue and half followed by a F cue. The cue duration conditions were identical to 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 a 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 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 pair 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 on the OSF page, 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. Because rearranged trials constitute false alarms in an associative recognition task, I compared whether this false alarm rate differed as a function of Cue Type and Cue Duration. The false alarm rate models followed identical structures to the hit rate models described in Experiment 1a.

2.3.2 Results

False Alarm Rate Model.

There was no significant main effect of Cue Type (R: M = .13, SE = .02; F: M = .12, SE = .02), χ^2 (1) = 3.29, p = .070, no main effect of Cue Duration (1 s: M = .13, SE = .02; 5 s: M = .13, SE = .02; 10 s: M = .12, SE = .02), χ^2 (2) = 2.40, p = .302, and no Cue Type x Cue Duration interaction, χ^2 (2) = .19, p = .909, on false alarm rates. Because false alarm rates from rearranged trials did not significantly differ across conditions, I removed rearranged trials from the rest of the analyses of recognition performance.

Hit Rate Model.

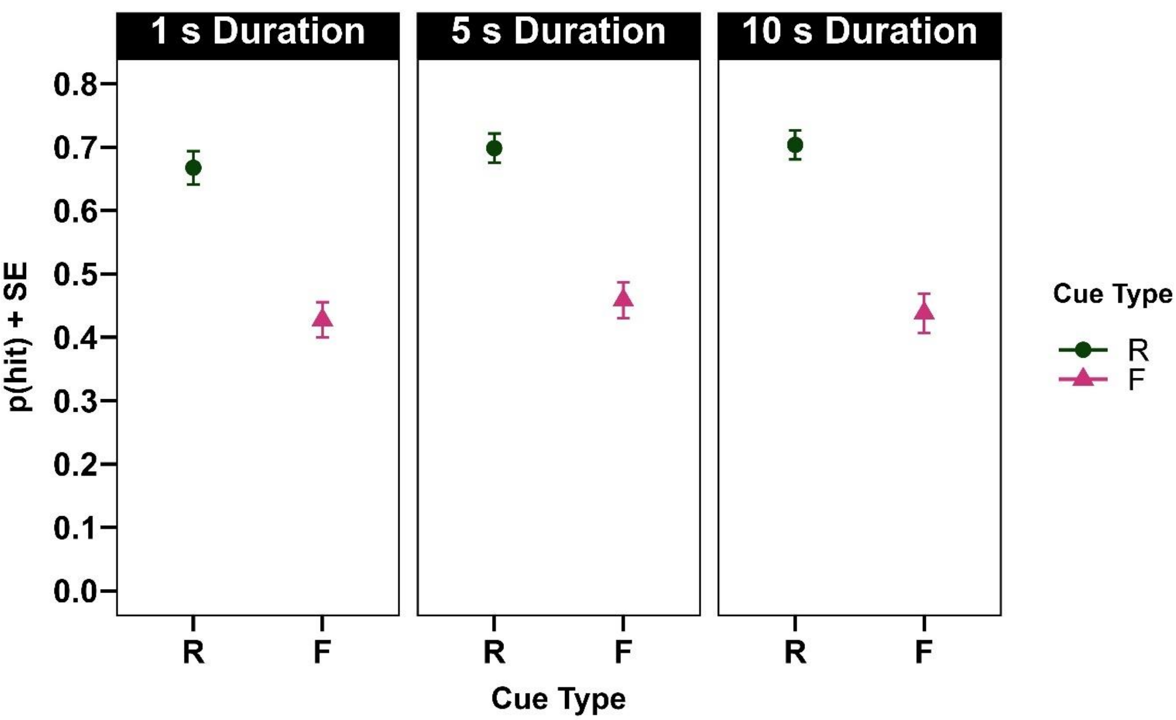
Figure 3 presents the proportion hit scores as a function of Cue Type and Cue Duration. There was a significant effect of Cue Type, χ^2 (1) = 218.29, p < .001, where participants overall recognized more R pairs (M = .72, SE = .02) than F pairs (M = .46, SE = .03), the usual directed forgetting effect. 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 x 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. However, collapsed across Cue Type, memory performance remained not significantly different between levels of the Cue Duration condition, *ORs* > .88, *CIs* = .65 – 1.60, *ps* > .305.

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

Figure 3. Experiment 2: Mean recognition performance represented as proportion hit on the associative recognition test.



Note. 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 error.

2.3.3 Discussion

When participants were explicitly instructed to encode associative information, there was a robust directed forgetting effect across all levels of cue duration. 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, Experiment 4). In Experiment 2, participants were instructed 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, there was no effect of increasing cue duration on associative recognition. This finding did not replicate the finding of Bancroft et al. (2013; Experiment 4). Moreover, again, there was 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 items showing a sharper rise in memorability with increasing cue duration.

2.4 Experiment 3

In considering the selective rehearsal account further, it became apparent that randomizing the cue durations as was done in Experiments 1a and 1b could have worked against any kind of rehearsal because participants did not know for any given item how long they might have available to rehearse. As a consequence, they may not have rehearsed much at all given the unpredictably of the duration that would be available. Previous investigations of cue duration and item-method directed forgetting had either blocked the study list by cue duration (Lee et al., 2007; Bancroft et al., 2013) or manipulated cue duration between-subjects (Allen & Vokey, 1998; Wetzel & Hunt, 1977). Experiments 1a, 1b, and 2 in this dissertation were the first investigations 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, it becomes clear that *how* cue duration is manipulated in the item-method directed forgetting paradigm is potentially important to target the time course of these processes. To provide a more straightforward comparison of the current intermixed investigations to the existing literature, in Experiment 3, the study phase design was switched to a blocked procedure: The study phase was divided into three blocks, each with a unique cue duration—again 1 s, 5s, or 10 s. Blocking cue duration should provide participants with a clearer picture regarding the time available to rehearse which may permit taking greater advantage of the cue duration.

2.4.1 Method

Participants.

The sample size was increased relative to Experiments 1a and 1b because manipulation of cue duration in those experiments had yielded a smaller effect size than that reported by Bancroft et al. (2013). One hundred and six participants were recruited from

Prolific, none of whom had participated in Experiments 1a, 1b, or 2. Their mean age was 28.65 (*SD* = 3.47), 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 condition. 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.

Materials.

The master word list was identical to the previous experiments.

***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 now, at the beginning of each block, they were told the cue duration for that block (i.e., 1 s, 5 s, or 10 s).

2.4.2 Results

Figure 4 presents the proportion hit scores 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* =

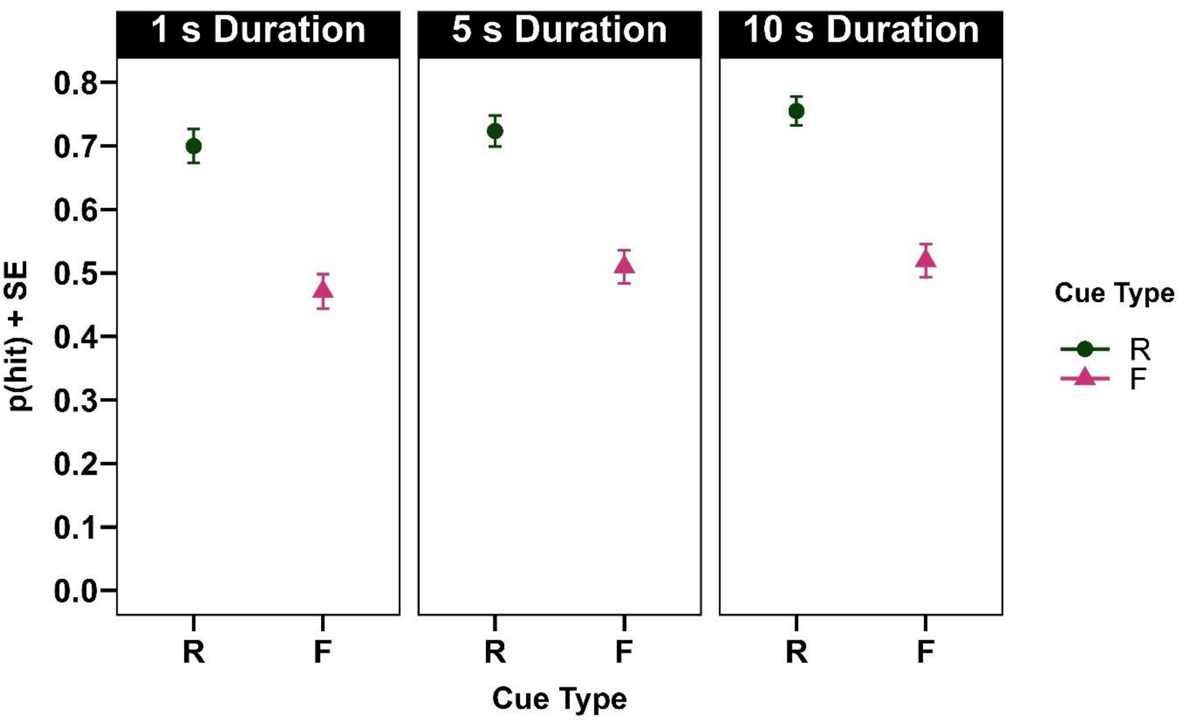
.03). Now there was also a significant main effect of Cue Duration, χ^2 (2) = 15.51, *p* < .001, where recognition differed between 1 s duration (*M* = .62, *SE* = .03), 5 s duration (*M* = .65, *SE* = .03), and 10 s duration (*M* = .68, *SE* = .03). There was, however, no Cue Type x Cue Duration interaction, 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.

Averaged over the levels of Cue Type, participants recognized significantly more items in the 10 s cue duration condition than in the 1 s cue duration condition, *OR* = .76, *CI* = .65 – .89, *p* < .001. Recognition was marginally significantly better in the 5 s cue duration condition than in the 1 s cue duration condition, *OR* = .86, *CI* = .73 – 1.00, *p* = .058. Recognition did not significantly differ between the 5 s and 10 s duration conditions, *OR* = .89, *CI* = .76 – 1.05, *p* = .202. Separated by Cue Type, for F items, recognition was significantly better with the 10 s cue duration than with the 1 s cue duration, *OR* = .80, *CI* = .64 – .99, *p* = .040. However, recognition did not significantly differ between the 5 s and 10 s duration conditions, *OR* = .95, *CI* = .76 – 1.18, *p* = .827, nor between the 1 s and 5 s duration conditions, *OR* = .84, *CI* = .68 – 1.05, *p* = .151. A similar pattern was also found for the R items, where participants recognized significantly more R items with the 10 s than with the 1 s cue duration, *OR* = .72, *CI* = .57 – .92, *p* = .004. Recognition did not significantly differ between the 5 s and 10 s duration conditions, *OR* = .83, *CI* = .65 – 1.06, *p* = .182, nor between the 1 s and 5 s duration conditions, *OR* = .87, *CI* = .68 – 1.10, *p* = .342.

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

Figure 4. Experiment 3: Mean recognition performance represented as proportion hit on the yes/no recognition test.



Note. 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 error.

2.4.3 Discussion

I again observed robust directed forgetting effects across all cue duration levels. Blocking by cue duration highlighted the differences in cue duration for the participants, and recognition now improved 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 replicated Lee et al. (2007) and Bancroft et al. (2013) and extended 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. This equivalence runs contrary to the selective rehearsal account that participants rehearse only (or at least primarily) R items. I reason that blocking by cue duration posed a challenge to the forgetting of F items: Removing these items became more difficult with increasing cue duration. As a result, slippage occurred such that F items also benefitted from some rehearsal. That the functions for R items and F items are essentially parallel across cue duration does, however, suggest that there is no differential cumulative rehearsal favoring R items.

2.5 Experiment 4

In Experiment 4, the goal was to determine whether the associative recognition findings of Experiment 2 would replicate when the study list was switched to a blocked procedure like that in Experiment 3 or whether there now would be an effect of increasing cue duration. Randomizing cue durations in Experiment 2 may have discouraged participants from elaboratively forming associations between the word pairs or from rehearsing those associations because they did 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 influences the respective processes of intentional remembering and forgetting for associative memory.

2.5.1 Method

Participants.

A sample size calculation was conducted using the effect size from Experiment 2, given that it was larger than Bancroft et al. (2013), η^2_p = .039. This 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.

Sixty participants were recruited from Prolific. Their mean age was 30.32 (*SD* = 2.34), 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 remained identical to the previous experiments.

Materials.

The master word list was 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 8 participants assigned to each. Participants were given the same instructions regarding the cue presentation and were informed of the cue duration before each block (i.e., 1 s, 5 s, or 10 s).

2.5.2 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 R pairs ($M = .29$, $SE = .04$) more than to F pairs ($M = .23$, $SE = .03$). There was no main effect of Cue Duration, χ^2 (2) = 1.29, $p = .525$, and no Cue Type x Cue Duration interaction, χ^2 (2) = 1.88, $p = .391$, on false alarm rates.

Hit Rate Model.

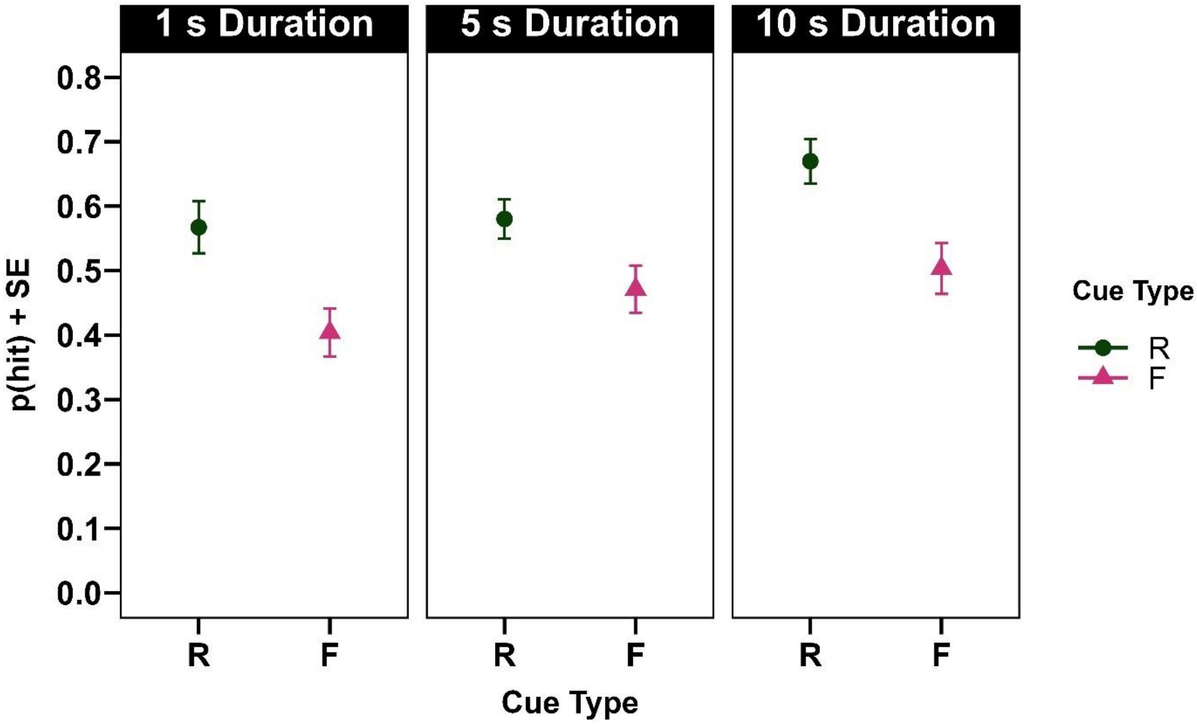
Figure 5 presents the proportion hit scores as a function of Cue Type and Cue Duration. As usual, there was a significant effect of Cue Type, χ^2 (1) = 42.71, $p < .001$, where participants overall recognized more R pairs ($M = .62$, $SE = .03$) than F pairs ($M = .46$, $SE = .03$). This time, there 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 x 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: $M = .46$, $SE = .03$), $OR = .47$, $CI = .33 – .66$, $p < .001$.

Averaged over the levels of Cue Type, participants recognized significantly more pairs in the 10 s than in the 1 s cue duration condition, $OR = .63$, $CI = .47 – .84$, $p < .001$. Recognition was marginally significantly better in the 10 s cue duration condition than in the 5 s cue duration condition, $OR = .75$, $CI = .56 – 1.00$, $p = .052$. Recognition did not significantly differ between the 1 s and 5 s duration conditions, $OR = .84$, $CI = .63 – 1.11$, $p = .299$. Separated by Cue Type, for F pairs, recognition was significantly better with the 10 than with the 1 s cue duration, $OR = .64$, $CI = .43 – .95$, $p = .023$. However, recognition did not significantly differ between the 5 s and 10 s duration conditions, $OR = .87$, $CI = .58 – 1.29$, $p = .676$, nor between the 1 s and 5 s duration conditions, $OR = .74$, $CI = .49 – 1.10$, $p = .171$. A similar pattern was found for the R pairs, where participants recognized significantly more R pairs with the 10 s than with the 1 s cue duration, $OR = .62$, $CI = .41 – .94$, $p = .018$, or with the 5 s cue duration, $OR = .65$, $CI = .43 – .99$, $p = .041$. Recognition of R pairs did not significantly differ between the 1 s and 5 s duration conditions, $OR = .95$, $CI = .64 – 1.42$, $p = .951$.

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, $b = .01$, $p = .928$.

Figure 5. Experiment 4: Mean recognition performance represented as proportion hit on the associative recognition test.



Note. 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 error.

2.5.3 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 trends. These findings replicated Bancroft et al. (2013; Experiment 4) and extended their findings to cue durations beyond 3 s for associative memory. 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. Critically, though, they apparently do this for both R pairs and F pairs, inconsistent with the selective rehearsal account. Instead, these findings fit with the explanation that I 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.

2.6 Interim Discussion of Experiments 1-4

In this chapter, I examined how varying the time allocated to an intent to remember or to forget information influences directed forgetting. Specifically, I carried out an in-depth investigation of the time course of 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, Lee, & Tsai, 2007). According to the selective rehearsal account, providing participants with more time to elaboratively rehearse R items should improve memory for those items in particular because participants engage in elaborative rehearsal—and potentially in cumulative elaborative rehearsal of the rehearsal set—of R items but not of F items throughout the study phase. However, with a larger range of cue durations to examine 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, I used a consistent range of cue durations: 1 s, 5 s, and 10 s. In Experiments 1a, 1b, and 3, I examined how increasing cue duration influenced item memory. There was a memory benefit with increasing cue duration only when participants were aware of how much time they could allocate to each item—when the cue durations were blocked—and this benefit was equivalent for R items and for F items. In Experiments 2 and 4, I examined whether associative memory is influenced by increasing cue duration and observed the same pattern of results as seen for item information in Experiments 1a, 1b, and 3: There was an effect of cue duration when it was blocked during the study phase and, even there, that benefit was apparent to the same extent for R items and for F items.

Why were there different effects of cue duration depending on the design of the study phase—intermixed over cue durations versus blocked by cue duration? In an intermixed design, participants cannot predict how long they will have to remember or forget the just-presented item. Thus, participants must deploy processes to successfully remember or forget as quickly as possible, given that the length of time to do so is unpredictable.

Dames and Oberauer (2022) provide support for this idea. They suggest that upon receipt of an R cue, participants attend to an item in their focus of attention in working memory, thereby boosting its memory strength. This boosting could be accomplished by short-term consolidation, a process assumed to firmly establish an initially weak and fragile trace in working memory (Jolicœur & Dell'Acqua, 1998; Ricker et al., 2018). Following an instruction to forget, the F item is removed from the focus of attention without being fully encoded into working memory. This removal process must be rapid for two reasons. First, the present finding that memory for F items did not differ across the wide range of cue durations indicates that participants are fast to remove F items without letting them remain in the focus of attention. If removal was not a rapid process, and possibly took several seconds, then I should have observed that memory for F items increased with increasing cue duration, which was not the case. Second, Oberauer (2018) varied the time interval between the F cue for the preceding word and the presentation of the following word and found that participants still demonstrated robust directed forgetting effects even at the shortest interval (i.e., 33 ms).

Now consider a blocked design. Here, cue duration is the same for every trial in a block so participants are fully aware of how long each cue will be presented. Consequently, participants know the length of time that they can dedicate to rehearsal—if they choose to rehearse at all. In the blocked situation, I observed that memory increased both for R items and for F items with increasing cue duration. The deployment of elaborative rehearsal mechanisms appears to have a time-dependent quality, where additional rehearsal is more successfully engaged 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 a kind of "slippage"—that the likelihood of non-strategic rehearsal becomes greater with the passage of time and independent of the item's cue. At any rate, there is no evidence in my data of differential rehearsal favoring R items: The functions for R and F items were parallel over cue duration in all four of my 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 I observed. Instead, these findings are in line with a removal process in working memory. Removal is proposed to operate rapidly (i.e., sometimes under 1 s) so that items do not remain in the focus of attention for too long (Oberauer, 2018). 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, as I found. In addition, some researchers have suggested that this removal process requires an active (potentially inhibitory) and resource-demanding mechanism that expunges irrelevant F items from memory (Fawcett & Taylor 2008; Fawcett & Taylor, 2012; Festini & Reuter-Lorenz, 2013, 2014). Contrarily, in my previous work (Tanberg & Hockley, under review; Tanberg, Fernandes, & MacLeod, 2022; Tan et al., 2020), I have questioned the idea that an intention to forget imposes a heavier cognitive demand than does an intention to remember. These findings from my previous work challenge the resource-demanding assumption of the removal mechanism but do not, of course, negate it.

The current experiments extend the findings of Bancroft et al. (2013) to broader and longer cue durations. Using blocked cue durations, Bancroft et al. found that increasing the 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 I observed for even longer cue durations in my 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 length of cue duration in item-method directed forgetting: It is possible that these strategies shift over cue duration.

It is also important to note that the longest cue duration of 10 s may have unintentionally induced free time in working memory. Mizrak and Oberauer (2021) demonstrated that increasing free 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, the memory for subsequent items is better compared to when the resource has less time to recover. This encoding-resource account (see also Popov & Reder, 2020) could explain overall better memory as cue duration increases. In the blocked design, as subsequent items 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, I have documented the role of increasing instructional cue duration on item recognition and on associative recognition beyond the common shorter durations from past studies. Although a selective rehearsal account predicts that increasing cue duration should selectively increase memory for R items but not for F items, I did not observe 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 consistent 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, I suggest that the directed forgetting effect derives from a quick retrieval check selectively done for R items but not for F items. A similarly quick removal process may operate on F items, as others have argued. In contrast, when cue durations are blocked, making the time available for rehearsal predictable, there is evidence of limited rehearsal improving memory, but that rehearsal appears to be applied equivalently to R and to F items. Finally, I saw no evidence of cumulative rehearsal selectively devoted to a rehearsal set containing only R items.

The goal of Chapter 3 was to investigate and elaborate on the quick retrieval check mechanism that is yet to be proposed in the item-method directed forgetting literature. Importantly, in Chapter 3, I propose the *selective retrieval* account.

Chapter 3:
Characterizing the Retrieval in Intentional Forgetting at Encoding

In the previous chapter, I examined how varying the time allocated to an intent to remember or to forget information influences directed forgetting in item memory and associative memory. For item memory, I observed a memory benefit with increasing cue duration only when participants were aware of how much time they could allocate to each item—when the cue durations were blocked—and this benefit was equivalent for R items and for F items. For associative memory, I observed a very similar pattern of results to that seen for item information. These findings contradicted the key assumptions of one of the dominant accounts of item-method directed forgetting—the selective rehearsal account. First, my findings challenged the time-based passive decay assumption of the selective rehearsal account where, upon presentation of an F cue, participants terminate maintenance rehearsal and allow the item to decay from working memory (Basden, Basden, & Gargano, 1993; Tan et al., 2020). If intentional forgetting is explained by passive decay, then memory for F items should decrease progressively the longer the cue duration. These findings were also in line with the considerable evidence against time-based decay in verbal working memory (Barrouillet et al., 2004; Dames & Oberauer, 2022; Oberauer & Lewandowsky, 2013).

Second, my findings were not in line with the cumulative rehearsal assumption—that participants elaboratively rehearse the R items throughout the study phase. While I provided evidence for cumulative rehearsal in the blocked design, my findings for the intermixed design did not support cumulative rehearsal. Instead, my findings were consistent with the role of a mechanism that operates to boost R item representation in long-term memory, thereby contributing to the directed forgetting effect.

I suggest that for R items a rapid boost—due to a quick *retrieval check*—plays the key role in driving the directed forgetting effect in the intermixed design. Upon receipt of an R cue, participants routinely allocate attention to confirming whether they have retained the to-be-remembered item adequately for the future test, a step that is not carried out 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—or shifting attention to the item to additionally attend to it—boosts its representation. When the length of the cue duration is unpredictable, this retrieval check for R items and the removal of F items may be all that participants do. But so far this new explanation is speculative: Does such a rapid retrieval check of only the R items actually contribute to the directed forgetting effect? In Chapter 3, I examine whether the act of retrieval drives the memory boost to R items and, thus, whether it underlies the directed forgetting effect.

3.1 The Role of Retrieval in Item-Method Directed Forgetting

Traditionally, retrieval is not perceived as an encoding event because most memory experiments treat encoding and retrieval as distinct stages or processes, yoked respectively to the study and test phases. Introducing the idea that episodic retrieval occurs during encoding can be seen as complicating experiments and theorizing by blurring the boundaries between encoding and retrieval (Benjamin & Ross, 2010; Hintzman, 2011; Tullis et al., 2014). Yet every act of encoding necessarily involves retrieval, as Kolers and Roediger (1984) articulated in their proceduralist account of memory (indeed, the converse is also true).

Numerous studies have shown that retrieval of information increases its long-term retention, possibly even more than further elaborative encoding of the information (Karpicke & Blunt, 2011; Karpicke, Butler, & Roediger, 2009; Pastotter & Bauml, 2014). Referred to as the *testing effect* (e.g., Bjork, 1975; see Roediger & Butler, 2011, for review), the act of retrieving information, even during study, enhances memory, possibly by producing additional retrieval paths to the information in memory so that it is more easily recovered subsequently. In fact, research on the testing effect in educational settings supports that testing oneself on the to-be-remembered material improves memory when studying for an upcoming test (Greving & Richter, 2018; Pollack & Miller, 2022).

In a typical testing effect study, participants first study a series of items during the initial study phase. Afterward, during the re-exposure phase, studied items are either repeated for additional re-exposure or they are subjected to a memory test or they do not recur in either form. In a later memory test, participants show better memory for studied items that were subjected to a memory test than for items that were simply repeated, or not repeated at all. The ample literature on the testing effect (Rowland, 2014) has provided a rich understanding of the role of retrieval at the time of learning and shown that this results in strengthening of memory representations.

Theoretical explanations of the testing effect have focused on the extent to which a successful retrieval contributes to the memorial benefit. The *desirable difficulties model* proposes that the difficulty induced by retrieval drives the testing effect: The more difficult it is to retrieve an item, the larger the magnitude of the testing effect should be (Bjork, 1994; Jacoby, 1978; Karpicke & Roediger, 2007). Another account, the *elaborative retrieval* *hypothesis*, suggests that testing results in the elaboration of a memory trace by engaging a successful retrieval (Carpenter, 2009). This elaboration results in better memory for tested items than for repeated or non-tested items because the act of retrieval elaborates the tested item's representation by activating semantic associates of that tested item (Carpenter, 2009). The theoretical characterization of the testing effect continues to be debated in the literature, but all of the accounts emphasize one common implication: The act of successful retrieval during learning, or encoding, strengthens memory of the retrieved information.

Previous research on the *reminding effect* also supports the idea that repeating an item as a reminder during the study phase enhances its encoding (Benjamin & Tullis, 2010). According to this framework, study phase retrievals (i.e., reminders) reduce interference and augment encoding of information through the promotion of elaboration (Benjamin & Ross, 2010; Hintzman, 2011; Jacoby, 1991). Indeed, accounts of the reminding effect maintain that the effortful retrieval of the first presentation at the time of the second presentation enhances memory for the first presentation (MacLeod, Pottruff, Forrin, & Masson, 2012; Tullis et al., 2014). A generally accepted explanation for the reminding effect suggests that the act of retrieving an item at its second presentation incorporates earlier studied item presentations into their memory traces (Benjamin & Tullis, 2010; Hintzman, 2010).

Taken together, the testing effect and the reminding effect literatures challenge the view that encoding and retrieval are separable processes. Instead, there is abundant evidence supporting the occurrence of retrieval during encoding (Carpenter, 2009; Karpicke & Grimaldi, 2012; Roediger & Karpicke, 2006a,b; Rowland, 2014; Siler & Benjamin, 2020).

The main goal of Chapter 3 is to examine whether, and if so how, retrieval contributes to the directed forgetting effect in the item-method paradigm.

Although retrieval has not been investigated as a potential mechanism for intentional forgetting at encoding, the term of *quick boost* has recently been introduced to the intentional forgetting at encoding literature in investigations of directed forgetting and working memory (Dames & Oberauer, 2022; Oberauer, 2018; Oberauer & Greve, 2022; Page & Norris, 1998). However, the exact definition of this quick boost in tests of long-term memory remains unclear. Critically, whether a quick boost is sufficient to provide a memorial advantage for R items in long-term memory is yet to be determined. Oberauer and Greve (2022) supported the role of a quick boost that enhances the representation of to-be-remembered information but found no directed forgetting effects in their long-term memory conditions. However, it is possible that their findings cannot be interpreted in the context of theoretical accounts of item-method directed forgetting because, during their study phase, they used a semantic orienting task that may have unintentionally forced participants to encode items more strongly than usual, prior to the R/F cue presentation. In addition, they incorporated surprise long-term memory tests after participants' memory for R and F items had already been routinely tested for working memory. Thus, although they proposed a quick boost to R items, they were only able to examine this in working memory.

In their work, Dames and Oberauer (2022) examined directed forgetting in working memory and proposed an attentional boost and automatic updating of working memory mechanism. Dames and Oberauer argued that each item is encoded into working memory upon presentation. Following an instruction to remember, participants rapidly boost the representation of the R item by attending to that item. They argued that this quick boost—which they referred to as an *attentional boost*—results in the directed forgetting effect. However, Dames and Oberauer also only tested working memory immediately following an intent to remember and an intent to forget. Their findings provide a fundamental building block to investigate whether a quick boost during encoding contributes to the directed forgetting effect in long-term memory using the item-method paradigm.

As one final motivation, two previous studies have shown that the directed forgetting effect is eliminated when elaborative encoding is undertaken prior to the instructional cue. MacLeod and Daniels (2000) showed in a mixed-list design that the directed forgetting effect was present for words that were simply read aloud but not for words that were generated from a short definitional sentence; Hourihan and MacLeod (2008) showed, also in a mixed-list design, that the directed forgetting effect was present for words read silently but not for words read aloud. It may be that, by necessitating a pre-cue retrieval, generation or production is sufficient to defeat the instruction to forget. In contrast, simply reading the word prior to the cue leads only to maintenance rehearsal (without retrieval), after which retrieval occurs only for R items and not for F items, this selective retrieval resulting in the directed forgetting effect.

With this background, I present a retrieval-based explanation of intentional forgetting at encoding. Here, I directly measured how selective retrieval during the study phase influences memory for R items versus for F items. In this paradigm, participants engage in retrieval of just-studied R items but not of just-studied F items. This “selective retrieval” promotes quick “one shot” refreshing of the R items, improving memory for them. In contrast, F items do not benefit from such retrieval. I present four experiments testing this retrieval explanation of item-method directed forgetting. My principal prediction was that, if the act of retrieval ordinarily provides R items with the memory boost, then including a condition where some F items also are actively retrieved should reduce, or perhaps eliminate, the directed forgetting effect.

3.2 Experiment 5

Following the item-method directed forgetting paradigm, I presented participants with a list of words, each word followed by an R cue or an F cue. To encourage the retrieval of items during initial encoding, some of the words—both some of the R items and some of the F items—were repeated during the study phase. Participants were made aware that there would be some Unrepeated items (i.e., items that were presented only once) and some Repeated items (i.e., items that were presented twice) during study. The rationale was that when a word was presented a second time, this would invite a retrieval (of its first presentation). I expected repetition to improve memory for all items, given its well documented memory benefit (see, e.g., Hintzman, 1976; Hintzman & Block, 1971), but I was particularly interested in how the benefit would compare for R and F items. Based on the retrieval explanation just outlined, I predicted that—relative to the Unrepeated items—the directed forgetting effect should be reduced for the Repeated items for two reasons. First, because R items would ordinarily be retrieved anyway after the instructional cue, repetition should benefit them less. Second, a retrieval boost should be selectively applied to repeated F items, which ordinarily would not be retrieved subsequent to the instructional cue.

3.2.1 Method

***Participants*.**

Seventy-one participants (ages 18-24; *M* = 22.4) from the University of Waterloo Department of Psychology participant pool took part for course credit. All reported normal or corrected-to-normal vision. To assess their English fluency, they completed Set A of the Mill Hill Vocabulary Scale, all scoring higher than 30% (*M* = 32.45, *SD* = 2.32). In addition, all participants were fluent English speakers and had learned English before the age of 9.

There were three exclusion criteria based on post-experiment questions: (1) responding yes to “Were you doing anything else while completing this task (e.g., watching videos, music, other)?”, (2) responding no to “Did you follow directions to the best of your ability?”, and (3) responding yes to “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.

The sample size was based on an a priori power analysis conducted using G*Power software (Faul et al., 2007) using $\eta^2 = .021$, which yielded an effect size f = .15 to power for a repeated-measures design with a desired power of ~.80 ($\alpha = .05$, two-tailed). This yielded a required sample size of *n* = 65. Data were collected from 71 participants. Three participants who switched tabs on their screens more than twice during the study, indicative of divided attention, were removed. Two participants were removed due to admitting having written down the words. After replacing these participants, the final sample size was 66 participants[3]. All experiments in Chapter 3 received ethics approval from the University of Waterloo Office of Research Ethics (Protocol #43083).

Materials.

The master word list and study and test list randomization procedures were identical to the previous experiments. The study list consisted of 68 words. Two additional words at the beginning and two at the end of the list, all four given R cues, served as primacy and recency buffers; these were not included in analyses. The critical study list therefore included 64 words: 16 Unrepeated R words, 16 Unrepeated F words, 16 Repeated R words, and 16 Repeated F words. Because 32 words (16 R and 16 F) were repeated, participants studied a total of 96 words in the study phase. Importantly, the same cue followed both presentations of repeated words.

The test phase consisted of a yes/no recognition test containing 128 words, the 64 old words (32 Unrepeated words and 32 Repeated words) from the study phase and 64 novel, unstudied words.

***Procedure*.**

All participants took part online via a jsPsych program hosted on the University of Waterloo Faculty of Arts server. Words—both during study and during test—were displayed in lowercase 20-pt Open Sans font. R/F cues during study were displayed in uppercase 32-pt

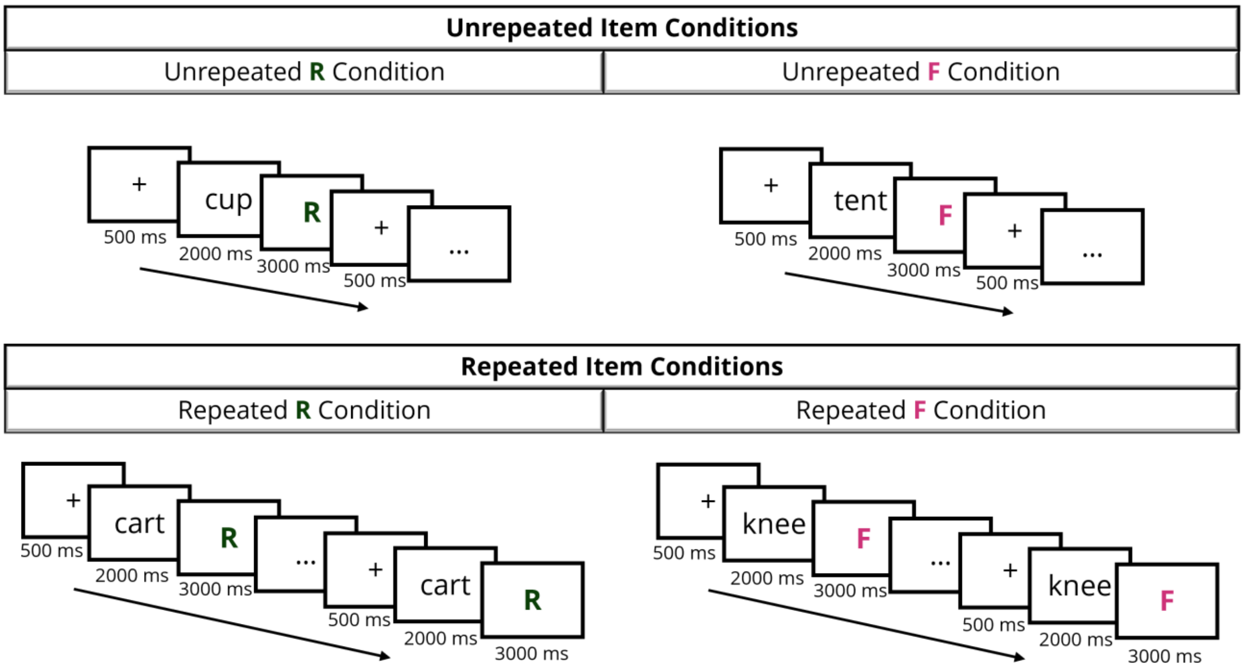
---

[3] The final sample size is one participant greater than the calculated sample size. During data collection, the number of people to be replaced was miscalculated and an extra person participated in the study. Results did not change with the inclusion of this participant, so I decided that I had no reason to remove them.

Open Sans font. The cues were RR and FF. All displays were in black on a white background and were centered on the screen.

In the study phase, each trial began with a 1-s fixation (+) followed by a single word shown in the center of the screen for 2 s. The word was followed immediately by a 3-s cue, either R or F, after which the next trial began immediately. Participants were told that their memory would be tested only for R items on a subsequent memory test and that they should forget all items followed by an F cue. They were also explicitly asked not to write down any words or to say them aloud. Finally, the critical instructions regarding repetition were presented: “Sometimes you will see a word twice. When this happens, please disregard the previous instruction for that word and instead apply the current instruction.” This instruction was included to avoid any confusion regarding what to do when a participant recognized that a word was presented twice. Figure 1 illustrates trials for the Unrepeated and Repeated conditions.

Figure 6. Experiment 5: Illustration of the Encoding Type conditions during the encoding phase. The three dots represent multiple additional trials following the example trial.



Immediately following the study phase, participants completed an old/new item recognition test in which they were instructed that, contrary to what they were told prior to the study phase, they would in fact be tested on both R and F words. Each test word remained visible until the participant responded to indicate whether that word was old or new. An old word was defined as a word seen during the study phase, regardless of whether it was followed by an R or F cue; a new word was defined as a word not presented at all during the study phase. Participants made an old response by pressing the *z* key or a new response by pressing the *m* key on their keyboard. They were informed that the test phase was self-paced and that they should be as accurate as possible. No feedback was provided, and the next test word immediately followed the response to the previous test word.

Following the test phase, participants were given a post-experiment questionnaire. This included self-reports of whether participants had followed the instructions (i.e., “Did you write down the words?”) and subjective reports of R-cue and F-cue strategies (i.e., “For items followed by an F, please briefly describe your strategy.”). The post-experiment questionnaire and the experimental instructions are available on the OSF page https://osf.io/bh3vp/.

Data Analysis.

For all experiments in Chapter 3, reported analyses were pre-registered. Similar to Chapter 2, 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). 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, I included a model with by-participant and by-item random intercepts. I additionally included by-participant and by-item random slopes for the effects of Cue Type and Encoding Type, only if including these random slopes improved model fit. If the model was singular, I continued to reduce the model complexity to report the best-fitting maximal model and I state the random effects structure used for that model. After fitting the models, I 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., 2019) with the Tukey 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) 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 were presented on the response scale, back-transformed from the log odds ratio scale for 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 Encoding Type (Unrepeated vs. Repeated) and were coded in all models using sum-to-zero contrasts. Data and analysis code for all experiments in Chapter 3 are available at OSF (https://osf.io/35v7f/). The complete project OSF page is available at https://osf.io/7xg8c/.

3.2.2 Results

Figure 7 presents the proportion hit scores as a function of Cue Type and Encoding Type. The mean collective False Alarm Rate for all conditions was 0.20 (*SE* = .02).

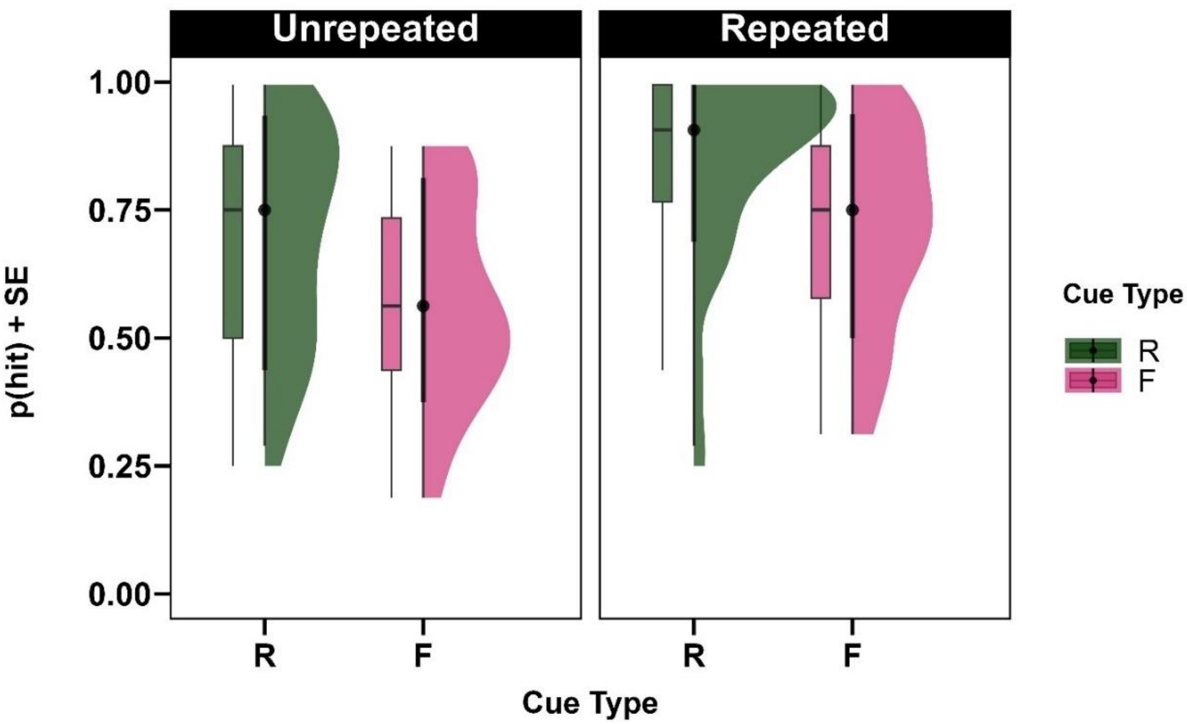
There was a significant effect of Cue Type, χ^2 (1) = 102.19, $p < .001$, where participants overall recognized more R items (*M* = .81, *SE* = .02) than F items (*M* = .67, *SE* = .03). There was also a significant main effect of Encoding Type, χ^2 (1) = 130.35, $p < .001$, where overall recognition was better for the Repeated items (*M* = .82, *SE* = .02) than for the Unrepeated items (*M* = .66, *SE* = .03). There was no Cue Type x Encoding Type interaction, χ^2 (2) = 2.46, $p = .117$.

To follow up the significant main effect of Cue Type, I conducted separate comparisons for each Encoding Type condition. There were significant directed forgetting effects both for the Unrepeated condition (R: *M* = .73, *SE* = .03; F: *M* = .58, *SE* = .03), *OR* =

.51, *CI* = .42 – .62, *p* < .001, and for the Repeated condition (R: *M* = .88, *SE* = .02; F: *M* = .75, *SE* = .03), *OR* = .40, *CI* = .32 – .51 *p* < .001.

Separated by Cue Type, for F items, recognition was significantly better in the Repeated condition than in the Unrepeated condition, *OR* = 2.16, *CI* = 1.77 – 2.63, *p* < .001. Similarly, for R items, participants recognized significantly more R items in the Repeated condition than in the Unrepeated condition, *OR* = 2.75, *CI* = 2.19 – 3.45, *p* < .001.

Figure 7. Experiment 5: Raincloud plot for recognition performance represented as proportion hit on the yes/no recognition test.



Note. The plot displays participant memory performance distribution of descriptive proportion hit for each Cue Type condition, separated by Encoding Type condition, and a boxplot. Error bars denote descriptive standard error.

3.2.3 Discussion

Participants recognized more R items than F items, demonstrating the usual robust directed forgetting effect. There also was the expected robust repetition effect: Participants recognized more Repeated items than Unrepeated items regardless of the item's cue. These findings were not consistent with the prediction that repetition should reduce the directed forgetting effect because F items would be selectively retrieved during study. Instead, I observed a directed forgetting effect that was equivalent for the Unrepeated items and the Repeated items. These findings indicate three possible mechanisms of action. First, it is possible that retrieval alone is not sufficient to drive the directed forgetting effect. Perhaps, retrieval needs to be coupled with another mechanism (i.e., elaborative rehearsal) to promote a benefit to the R items. Second, it is also possible that simply repeating an item throughout the study phase was not sufficient to induce a retrieval check, believed to occur for the R items. In fact, if this is true, this pattern of results may be explained by common theories of repetition memory that suggest that repetition creates a second memory trace rather than directly strengthening the first memory trace. Indeed, this would explain why both memory for R items and F items increased with the repetition. Finally, it is also possible that participants did not notice the repetition of an item without a direct incentive (i.e., button press). In Experiment 6, I address these issues by varying the Experiment 5 paradigm.

3.3 Experiment 6

In Experiment 6, I repeated the basic methods of Experiment 5, but with one key addition. In Experiment 5, participants were not required to report recognizing whether a word was being presented for a second time. Thus, I was not able to confirm whether participants actually retrieved the item upon the second presentation—or indeed that they even recognized repetitions. To determine how the act of retrieval influences item-method directed forgetting, memory performance for Repeated items that participants noticed versus those that they missed (i.e., did not notice) must be differentiated. The idea is that noticing a repetition indicates that a retrieval has occurred.

Consequently, the critical addition in Experiment 6 was to incorporate a button press to be made each time that a participant recognized an item's second presentation. This procedure was inspired by previous research on memory consequences of noticing change (Jacoby, Wahlheim, & Kelley, 2015; Robbins & Bray, 1974; Wahlheim, Maddox, & Jacoby, 2014;). In earlier work, Bruce and Weaver (1973) had presented word-pairs (e.g., A-B) and informed half of the participants that some pairs would change during the study list (e.g., A-D). The other half of the participants did not receive an instruction regarding change. Bruce and Weaver found an implicit repetition effect when change was noticed. Noticing a change, or repetition, requires a reminding of the original presentation of that item. This reminding entails a retrieval of that item's memory representation. Similarly, Jacoby et al. (2015) suggested the *memory-for-change framework*: For change detection to occur, new information must cue retrieval of existing memory traces (Jacoby & Wahlheim, 2013; Jacoby, Wahlheim, & Yonelinas, 2013; Putnam, Wahlheim, & Jacoby, 2014; Wahlheim & Zacks, 2019; Wahlheim, Smith, & Delaney, 2019). According to this framework, simply repeating an item without active change detection is not sufficient to induce a reminding, or retrieval in the context of this investigation, that provides a memorial benefit to the repeated items.

In Experiment 6, change detection was measured by requiring participants to press a button when they recognized an item upon the reminding (i.e., second presentation). Participants were told that some words would be repeated during the study phase, and they were instructed to press the space bar whenever they spotted this repetition. This allowed for the classification of repeated items into two separate categories: 1) *Missed Repetition items* that included items where participants did not notice their repetition, and 2) *Noticed Repetition items* that included items where participants did notice their repetition. Based on earlier findings showing that detecting change, or remindings, improves memory performance, I predicted that the directed forgetting effect should be reduced, or eliminated, for the noticed items. The act of successfully retrieving the item should augment subsequent recognition performance. However, I predicted that the directed forgetting effect might not be reduced for the missed items. If participants failed to recognize an item's repetition, then presumably they did not retrieve the repeated item. That is, the absence of reminding should lead to no change detection for the missed items, ultimately resulting in no memorial advantage. Thus, the recognition of missed items should be similar to the unrepeated items because of failure to retrieve the fact that the item was repeated.

I further predicted that the effect of noticing a repetition should be larger for the F items than for the R items. I derived this hypothesis from the fact that if receiving an R instruction results in a rapid boost due to a retrieval check, this act of retrieval should routinely occur for R items. Thus, the additional benefit of retrieval should already be occurring for these items. Of critical interest is how memory for F items will differ depending on whether the second presentation is noticed or missed, and how this will compare to Unrepeated F items. If noticing an item repetition constitutes a memory benefit associated with reminding, where detecting a reminder strengthens the first presentation as it is retrieved (Jacoby, Wahlheim, & Kelley, 2015), then noticing an F item repetition should improve memory for those items compared to when a repetition is not noticed or when the item is not repeated.

3.3.1 Method

***Participants*.**

Ninety-eight participants (ages 18-35; M = 22.4) from Prolific (https://prolific.com/) were recruited. Their mean age was 27 (*SD* = 3.45) and 64% identified as female and 12% identified as nonbinary. Fourteen participants were replaced because they did not follow the instructions and made no button presses during the study phase. To corroborate this, these participants also reported during the post-experimental questionnaire that they were not aware of the button press instruction before the study phase, demonstrating lack of attention to the study instructions. Three participants were excluded because they reported that they used external aids to remember the items, ultimately scoring 100% on the recognition test. Two participants were excluded because they switched tabs more than twice throughout the study. The final sample included 79 participants. All participants passed the Mill Hill Vocabulary Scale (*M* = 37.85, *SD* = 2.62).

The sample size was based on an a priori power analysis conducted using G*Power software. I calculated sample size using $\eta^2 = .025$, which yielded an effect size f = .16 using the G*Power calculation, with a desired power of ~.80 ($\alpha = .05$, two-tailed). This resulted in a required sample size of *n* = 79.

***Materials*.**

All materials were identical to those of Experiment 5.

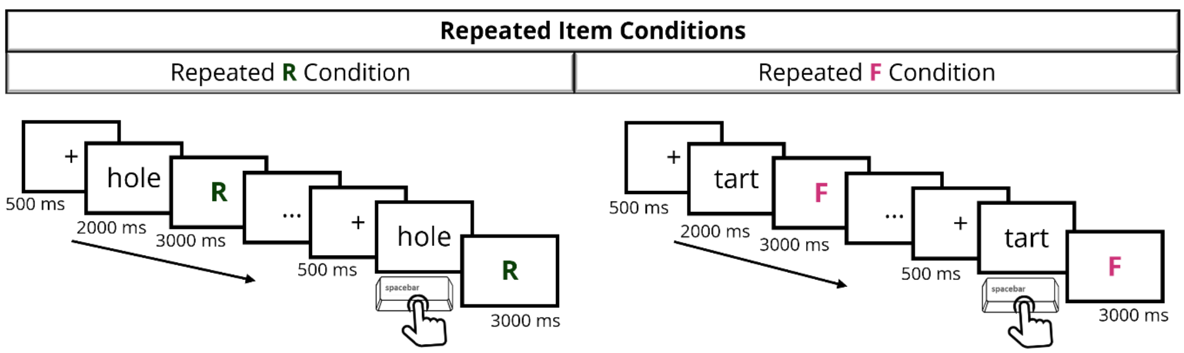
Procedure.

The procedure was similar to Experiment 5, with two key changes. First, prior to starting the experiment, participants were given practice with the button press instruction. Practice trials were identical to study phase trials (described below), except that the study word was replaced with a letter. Participants were instructed that they should press the spacebar when they noticed the repetition of a letter. Each letter was followed by R or F. At least 20 trials were presented, and the trials were repeated until the participant made a button press to at least one R item repetition and one F item repetition.

Following the practice phase, participants began the study phase after reading the study instructions again. Participants were informed that some words would be repeated, and that they should press the spacebar if they noticed a repeated word (even if they had been told to forget it). Participants were instructed to press the spacebar as quickly as possible upon noticing the repetition. The study list was identical to that used in Experiment 5.

The test phase was identical to Experiment 5. However, during the post-experiment questionnaire, participants were asked whether they were aware of the button press response during the study phase when they noticed a repetition. In addition, participants were asked whether they were instructed to make this button press upon noticing a Repeated R item, a Repeated F item, or regardless of the item cue presentation.

Figure 8. An illustration of the spacebar button presses during the second presentation of a Repeated R/F item. In contrast with Figure 6, a spacebar button press was implemented during the second presentation. Consequently, according to the proposed retrieval account, there should be no directed forgetting effect on the trials where participants pressed the button during the second presentation of an R/F item to indicate recognizing the repetition of the item (i.e., noticed repetition trials). Unrepeated item conditions were identical to Experiment 5, as depicted in Figure 6.



3.3.2 Results

The results are separated into two sections. First, I report analyses for the 2 (Cue Type: R vs. F) x 2 (Encoding Type: Unrepeated vs. Repeated) within-subjects design. These analyses also serve as a replication of the Experiment 5 design. Second, I perform more fine-grained conditional analyses of the Repeated item condition, where I defined Missed Repetition items as instances when participants did not recognize the second presentation of a Repeated R/F item because no button press was made, and Noticed Repetition items as instances when participants recognized the second presentation of an R/F repeated item by making a button press. This yielded a 2 (Cue Type: R vs. F) x 3 (Encoding Type: Unrepeated vs. Missed Repetition vs. Noticed Repetition) within-subject design.

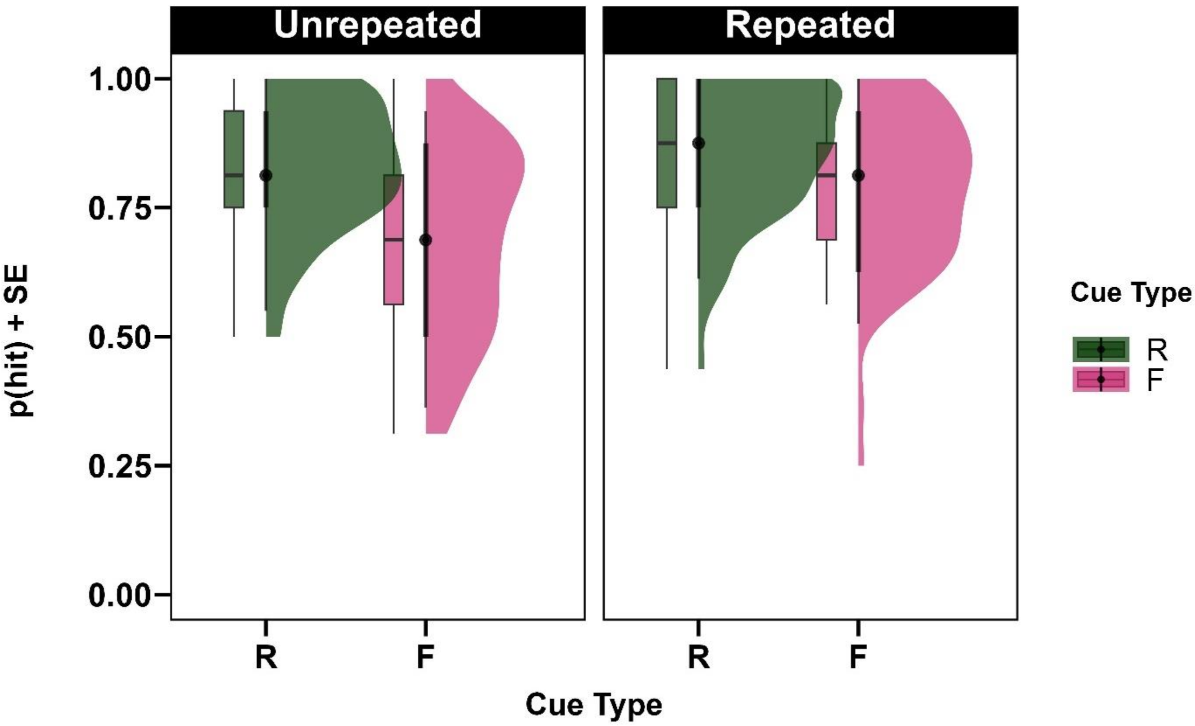
Replication of Experiment 5.

Figure 9 presents the proportion hit scores as a function of Cue Type and Encoding Type. The mean collective False Alarm rate for all conditions was 0.16 (*SE* = .03). The model included fixed effects of Cue Type and Encoding Type. The model was identical to Experiment 5, where the design was 2 (Cue Type: R vs. F) x 2 (Encoding Type: Unrepeated vs. Repeated).

There was a significant effect of Cue Type, χ^2 (1) = 61.68, *p* < .001, where participants overall recognized more R items (*M* = .87, *SE* = .01) than F items (*M* = .76, *SE* = .02). There was also a significant main effect of Encoding Type, χ^2 (1) = 15.94, *p* < .001, where overall recognition was better for Repeated items (*M* = .84, *SE* = .01) than for Unrepeated items (*M* = .76, *SE* = .02). There was no Cue Type x Encoding Type interaction, χ^2 (2) = 3.02, *p* = .082.

Separated by Encoding Type condition, there were significant directed forgetting effects for both the Unrepeated condition (R: *M* = .85, *SE* = .02; F: *M* = .71, *SE* = .02), *OR* = .43, *CI* = .34 – .55, *p* < .001, and the Repeated condition (R: *M* = .88, *SE* = .01; F: *M* = .80, *SE* = .02), *OR* = .59, *CI* = .45 – .76 *p* < .001. Separated by Cue Type, for F items, recognition was significantly better in the Repeated condition than in the Unrepeated condition, *OR* = 1.64, *CI* = 1.30 – 2.08, *p* < .001. However, recognition of R items did not significantly differ between the Repeated and Unrepeated conditions, *OR* = 1.64, *CI* = .93 – 1.57, *p* = .163.

Figure 9. Experiment 6: Raincloud plot for the recognition performance represented as proportion hit on the yes/no recognition test.



Note. The plot displays participant memory performance distribution of descriptive proportion hit for each Cue Type condition, separated by Encoding Type condition, and a boxplot. Error bars denote descriptive standard error.

Conditionalized Encoding Type Analysis.

For evidence of retrieval during the study phase, I separated the Repeated condition into the Noticed Repetition condition and the Missed Repetition condition. Reliance on conditionalized analyses raises the possibility that results may be influenced by item and participant selection effects (i.e., random effects). To address these concerns, the principal statistical analyses are mixed-effects logistic regression analyses that allow examination of the effects of Cue Type and Encoding Type on the likelihood of recognition while including participants and items as random effects (Brown, 2021). Thus, in my investigation, conditionalizing the repeated item condition does not pose a problem because my principal analyses control for the common issues reported in the literature on interpreting conditionalized results (Negley, Kelley, & Jacoby, 2018; Postman & Stark, 1969; Wahlheim & Jacoby, 2013; Wahlheim, Smith, & Delaney, 2019).

The final model included fixed effects of Cue Type (R vs. F) and Encoding Type (Unrepeated vs. Missed Repetition vs. Noticed Repetition), controlling for by-participant and by-item intercepts. Figure 10 presents the proportion hit scores as a function of Cue Type and Encoding Type.

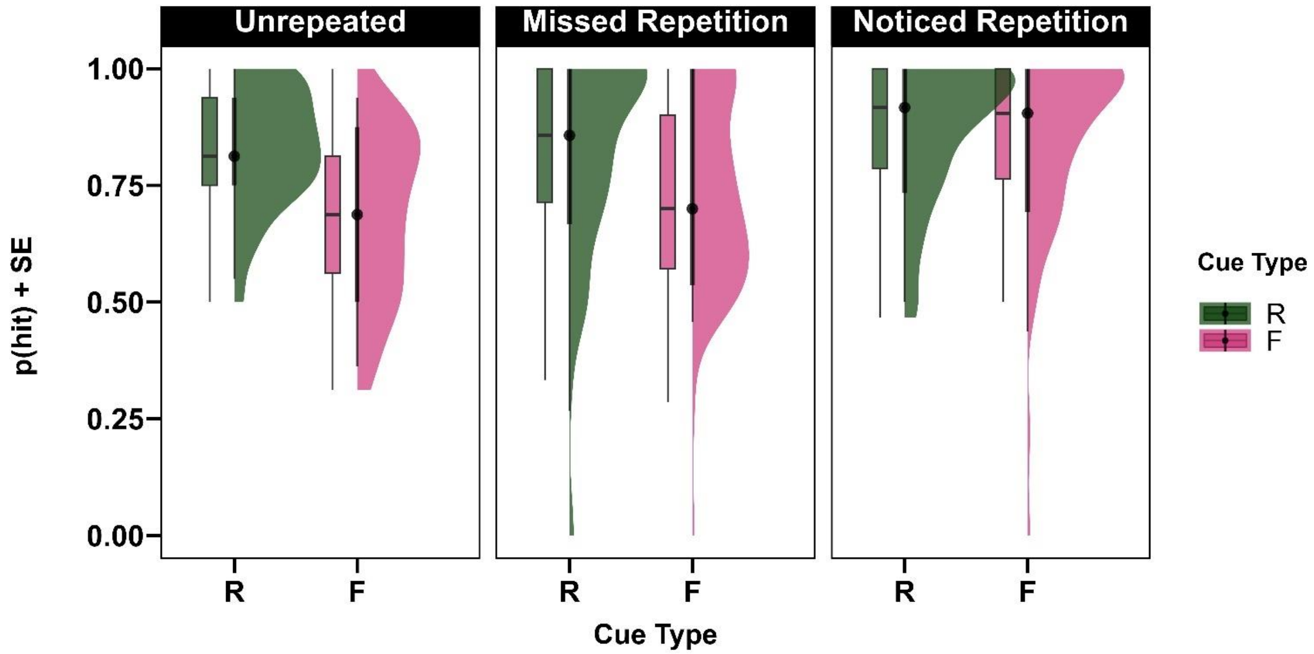
There was a significant effect of Cue Type, χ^2 (1) = 48.15, $p < .001$, indicating an overall robust directed forgetting effect (R: $M = .86$, $SE = .01$; F: $M = .78$, $SE = .02$). There was also a significant effect of Encoding Type, χ^2 (2) = 49.55, $p < .001$, where, averaged over the levels of Cue Type, recognition significantly differed between the Unrepeated ($M = .79$, $SE = .02$), Missed Repetition ($M = .78$, $SE = .02$), and Noticed Repetition ($M = .88$, $SE = .01$) conditions. There was a significant Cue Type x Encoding Type interaction, χ^2 (2) = 8.06, $p = .018$.

Separated by Encoding Type condition, there was a significant directed forgetting effect only for two conditions: the Unrepeated condition (R: $M = .85$, $SE = .02$; F: $M = .71$, $SE = .02$), $OR = .43$, $CI = .34 – .55$, $p < .001$, and the Missed Repetition condition (R: $M = .82$, $SE = .03$; F: $M = .73$, $SE = .03$), $OR = .59$, $CI = .41 – .86$, $p = .007$. Notably, there was no directed forgetting effect for the Noticed Repetition condition (R: $M = .89$, $SE = .01$; F: $M = .87$, $SE = .02$), $OR = .78$, $CI = .56 – 1.09$, $p = .143$.

Averaged over the levels of Cue Type, not surprisingly participants recognized significantly more items in the Noticed Repetition condition than in either the Unrepeated condition, *OR* = 1.95, *CI* = 1.52 – 2.51, $p < .001$, or the Missed Repetition condition, *OR* = .47, *CI* = .36 – .63, $p < .001$. However, recognition did not significantly differ between the latter two conditions, *OR* = .93, *CI* = .71 – 1.21, $p = .777$.

Follow-up comparisons aggregated across Cue Type revealed that, for F items, recognition was better for the Noticed Repetition condition than for either the Unrepeated condition, *OR* = 2.63, *CI* = 1.84 – 3.75, $p < .001$, or the Missed Repetition condition, *OR* = .41, *CI* = .29 – .60, $p < .001$. However, recognition did not differ between the Unrepeated and Missed Repetition conditions, *OR* = 1.09, *CI* = .79 – 1.51, $p = .815$. Similarly, for R items, recognition was better for the Noticed Repetition condition than for either the Unrepeated condition, *OR* = 1.45, *CI* = 1.02 – 2.05, $p < .001$, or the Missed Repetition condition, *OR* = .54, *CI* = .35 – .84, $p < .001$. Again, recognition did not differ between the Unrepeated and Missed Repetition conditions, *OR* = .77, *CI* = .51 – 1.21, $p = .385$.

Figure 10. Experiment 6: Raincloud plot for the recognition performance represented as proportion hit on the yes/no recognition test.



Note. The plot displays participant memory performance distribution of descriptive proportion hit for each Cue Type condition, separated by the conditionalized Encoding Type condition, where the Repeated condition was classified into Missed Repetition and Noticed Repetition conditions, and a boxplot for each condition. Error bars denote descriptive standard error.

3.3.3 Discussion

First, I replicated the findings of Experiment 5, where there was an overall repetition benefit to the Repeated items and robust directed forgetting effects for both Unrepeated and Repeated items. In the present experiment, my goal was to introduce a measure of retrieval upon repetition by incorporating a button press at the second presentation of Repeated items. This button press procedure allowed me to divide the Repeated condition into Missed Repetition and Noticed Repetition conditions, where I predicted that noticing the repetition of a Repeated item should induce retrieval of the first presentation of that item, ultimately reducing the directed forgetting effect for the Noticed Repetition items.

My main prediction was confirmed: The act of successfully retrieving an item eliminated the directed forgetting effect. The finding that the directed forgetting effect was only eliminated for the Noticed Repetition items and not for the Missed Repetition items supports the validity of the button press procedure as a confirmation of retrieval engaged during repetition of a Repeated item. Noticing the repetition clearly mattered: Participants demonstrated robust directed forgetting effects for the Missed Repetition condition, illustrating that when the second presentation of an item goes unnoticed, no retrieval has occurred.

Memory for R items did not differ between Unrepeated and Missed Repetition items. That is, when participants did not notice the repetition, memory for R items was similar to that when the item was not repeated. Simply repeating an item was not sufficient to induce a retrieval nor the repetition benefit (Hintzman, 2003; 2011). Indeed, when participants noticed repetition, memory for R items significantly improved during the recognition task. It is possible that R items receive their routine retrieval check immediately upon the instruction to remember. Then, upon the second presentation of that R item, assuming that repetition is noticed, detecting the repetition strengthens the R item in memory by reminding the participant of the original presentation.

In addition, for F items, memory for Unrepeated and Missed Repetition items did not differ, indicating that when participants missed the repetition of an item, their memory did not differ compared to when the item was simply not repeated. Contrarily, participants recognized more Noticed Repetition items than both Unrepeated and Missed Repetition items. This finding confirms that when repetition is recognized, it acts as an effective reminder of an F item that ultimately induces a retrieval. This act of retrieval provides a memory benefit associated with noticing a reminder, as supported by the previous literature on memory for change (Benjamin & Tullis, 2010; Wahlheim & Zacks, 2019; Wahlheim, Smith, & Delaney, 2019). A primary assumption of the memory for change framework is that reminding of existing memories (i.e., first presentation) is necessary to encode those memories together with new information. When participants notice the second presentation of an item, this leads to the retrieval of that item's first presentation, serving as an effective reminder. Noticing and reporting a repetition brings about better memory for both the original presentation and the repetition responsible for the reminding of the original presentation (Jacoby & Wahlheim, 2013; Hintzman, 2011; Putnam et al., 2014).

Experiment 6 provides evidence clearly consistent with my selective retrieval account. There is, however, an alternative explanation to consider. It is also possible that separating items into Noticed and Missed Repetition conditions essentially constitutes sorting items with respect to how well participants initially encoded them during their first presentation. If so, then the findings of Experiment 6 could also be explained by a strength account wherein Noticed Repetition items were initially encoded more strongly than Missed Repetition items. In Experiment 7, I put forward another method to test the proposed retrieval boost mechanism. This method provides a more direct investigation of the selective retrieval account.

3.4 Experiment 7

In Experiments 5 and 6, I forced the retrieval of F items by repetitions that occurred a few trials after the first presentation of the item. These experiments confirmed that repeating an F item improves memory, but only when the repetition is noticed. The probability of successfully retrieving an F item depends on that item serving as an effective reminder of its first presentation. Retrieving an F item at any point during the study phase improves later recognition of that item. Repetition is capable of overriding an instruction to forget, depending on noticing the second presentation of that item. While Experiments 5 and 6 provide an investigation of how inducing the retrieval of F items influences the directed forgetting effect, they do not entirely capture the true nature of my proposed retrieval check mechanism. These experiments support the motivation that retrieval contributes to item-method directed forgetting.

The type of retrieval that I propose occurs instantaneously after the R cue presentation: Participants perform a rapid retrieval check to remind themselves of the item and now label it as to-be-remembered. This retrieval check is not performed for F items, which instead are quickly removed from working memory (Dames & Oberauer, 2022; Oberauer, 2018). Essentially, retrieval of an R item immediately after cue presentation is a kind of recognition test of the just-presented item that is now to-be-remembered. Based on this interpretation, implementing a recognition test following the F cue should force a similar kind of retrieval check to occur for the F items, which should consequently reduce the directed forgetting effect. It is this manipulation that is the focus of Experiment 7.

In Experiment 7, I altered the way that I induced a retrieval during the study phase, the goal being to generalize the way that the retrieval check was carried out. Instead of repeating the words, on half of the trials, I incorporated a single-item recognition test immediately after the cue. The goal of doing this was to force the retrieval of the preceding item, which is predicted to occur routinely for R items. I separated the trials immediately followed by a recognition test into two categories: *Mismatched* and *Matched*. On Mismatched trials, the test item was an entirely new item that did not match the target item, so the participant should respond “no.” On Matched trials, the test item was identical to the target item that preceded the R/F cue, so the participant should respond “yes.” I refer to the trials where there was no recognition task immediately after the cue as the No Test trials.

Critically, I predicted that the directed forgetting effect should be reduced for the Matched condition, where the test item is the target item as opposed to a new word in the Mismatched condition. This effectively forces a retrieval for F items, a retrieval that would ordinarily only take place for R items. Essentially, repetition of the target item in the recognition test should induce the retrieval of that item, a retrieval that would not take place for Mismatched or for No Test trials. For Mismatched test items, retrieval would not have occurred when a new word was presented during the encoding phase recognition task because the test item does not match the target item. It is also possible, however, that retrieval sometimes does occur in Mismatched trials, but its efficacy would be less than in Matched trials. In Mismatched trials, a broad comparison of the new word to the preceding target word may happen but this comparison would be shallow enough to immediately detect that the new word does not match the preceding target word. This would not bring about a retrieval benefit, unlike what occurs with successful matching (Rowland, 2014; Pollack & Miller, 2022).

3.4.1 Method

Participants.

Eighty-four participants were recruited from Prolific. Their mean age was 28.50 (*SD* = 2.67) with 55% identifying as female and 11% identifying as nonbinary. The exclusion criteria were identical to the previous experiments.

A sample size calculation was conducted for a repeated measures within-factor design, using $\eta^2 = .019$ (based on Experiment 5), which yielded an effect size of f = .14 and the calculation was conducted with a desired power of power of ~.80 ($\alpha = .05$, two-tailed). This yielded a required sample size of $n = 72$.

Seven participants were excluded because they reported on the post-experiment questionnaire that they did not follow the instructions for the recognition task during study. Three further participants were excluded for switching tabs more than twice during the study. One participant was excluded due to a programming error while saving data online. The final sample therefore included 73 participants, all of whom passed the Mill Hill Vocabulary Scale (*M* = 37.85, *SD* = 2.62)[4].

Materials.

---

[4] The final sample size is one participant greater than the calculated sample size. During data collection, I miscalculated how many people were to be replaced and ran one extra person. Results did not change with the inclusion of this participant, so I decided that there was no reason to remove them.

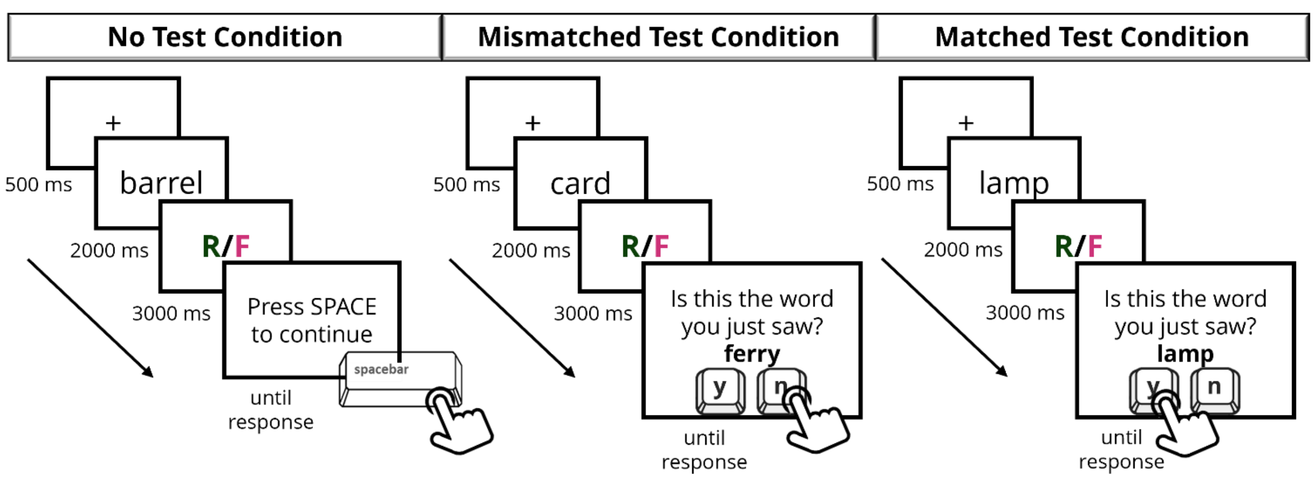
The master word list was identical to previous experiments. The study list consisted of 68 words. Two additional words at the beginning and two at the end of the list, all four given R cues, served as primacy and recency buffers; these were not included in analyses. The critical study list therefore included 64 words, which were split into 32 items in the Encoding Test condition and 32 items in the No Test condition. The words in the Encoding Test condition were followed by a yes/no recognition test immediately after the R/F cue presentation. Of the 32 items in the Encoding Test condition, 16 were cued to-be-remembered and 16 were cued to-be-forgotten. For 8 of the R items in the Encoding Test condition, the just-presented target word appeared on the encoding recognition task (i.e., Matched Test condition). For the other 8 R items in the Encoding Test condition, a new unstudied word was presented on the encoding recognition task (i.e., Mismatched Test condition). The same setup was followed for the F items in the Encoding Test condition. Lastly, 16 items in the No Test condition were F-cued and 16 were R-cued.

The randomizations were identical to previous experiments. The test phase consisted of a yes/no recognition test containing 128 words, the 64 old words (32 Encoding Test words and 32 No Test words) from the study phase and 64 novel, unstudied words. The novel words presented in the final test phase were different from the novel words presented during the yes/no recognition test during the study phase.

Procedure.

During the study phase, each word was presented at the center of the screen, one a time, and immediately followed by an R/F cue. The item and cue durations were identical to those of Experiments 5 and 6. In the Encoding Test condition, a recognition test immediately followed the cue presentation. Participants were asked "Is this the word you just saw?" and were provided with a bolded, single word in the center of the screen. The exact instructions were: "After some of the words, you will be asked to determine whether you just saw that word or a different word. When this happens, if you believe the displayed word is the word that you just saw immediately before, press the Y key for YES. If you believe that the displayed word is not the word that you just saw immediately before, press the N key for NO." On No Test trials, participants were given an on-screen prompt—an asterisk—and were to press the space bar as quickly and accurately as possible to move on. Participants were instructed to place their left index finger on the Y key, their right index finger on the N key, and their thumbs on the space bar. There was no time limit for responding, but participants were not made aware of this. The order of the Encoding Test and No Test trials was randomized. The procedure for the test phase was identical to that of Experiments 5 and 6. Figure 11 displays the study phase procedure for Experiment 7.

Figure 11. Experiment 7: Illustration of the conditions during the study phase.



Data Analysis.

The model comparisons and criteria for mixed-effects logistic regression were identical to previous experiments. For each analysis, the final model that converged included the random intercepts for participants and words, excluding all correlations among random effects. If this was not the case for an analysis, I report the converged random effects structure in the Results section.

I also examined the response time (RT) differences across the Encoding Type conditions as a function of the Cue Type. On a trial level, I excluded trials with RTs above or below +/-3 standard deviation from the individual cell mean[5]. I log-transformed RTs and analyzed data using linear mixed-effects models assuming a Gaussian distribution. The final model that converged included the random intercepts for participants and for words, excluding all correlations among random effects. The linear mixed-effects models were fitted using maximum likelihood estimation for model selection and restricted maximum likelihood estimation for slope estimates. *p*-values for effects were obtained using the Kenward-Roger (Kenward & Roger, 1997) approximation for denominator degrees of freedom. Within text means and standard errors for RTs were presented on the model-based log odds ratio scale. I visualized model-based log-transformed RTs for the plots but report the untransformed RTs for each condition as a table in the Results section.

3.4.2 Results

The results are separated into two sections. First, I report the analyses for the performance on the yes/no recognition task during the test phase, highlighted under the *Test*

---

[5] The RT results for all experiments in Chapter 3 did not change when I included the excluded +/-3 standard deviation trials.

Phase section. Next, I report the analyses for the performance on the yes/no recognition task embedded in the study phase. Finally, I report the analyses of RT to the asterisk probe following the R/F cue during the study phase. These results are highlighted under the *Study Phase* sections for the study phase analyses.

3.4.2.1 Test Phase

Replication of Experiment 5 with Collapsed Encoding Type Condition.

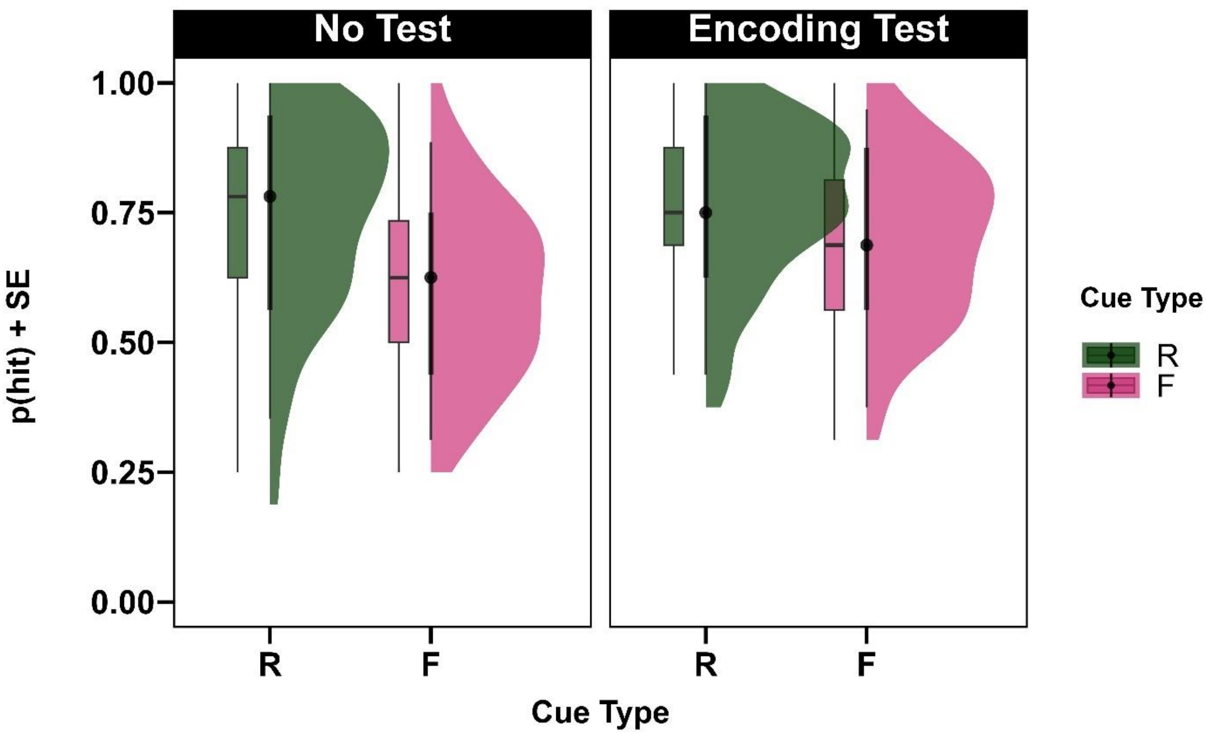
Figure 12 presents the proportion hit scores as a function of Cue Type and Encoding Type. The mean collective False Alarm rate for all conditions was 0.27 (*SE* = .03). The model was identical to that in Experiment 5, where I included fixed effects of Cue Type (R vs. F) and Encoding Type (No Test vs. Encoding Test). Here, the Mismatched Test and Matched Test conditions were averaged to create the Encoding Test condition that was analogous to the Repeated condition from Experiment 5.

There was a significant effect of Cue Type, χ^2 (1) = 71.68, $p < .001$, where participants overall recognized more R items (*M* = .78, *SE* = .02) than F items (*M* = .67, *SE* = .02). There was also a significant effect of Encoding Type, χ^2 (1) = 24.98, $p < .001$, where recognition was better for the Encoding Test condition (*M* = .76, *SE* = .02) than for the No Test condition (*M* = .69, *SE* = .02). There was a significant Cue Type x Encoding Type interaction, χ^2 (1) = 5.65, $p = .018$.

Separated by Encoding Test, there was a significant directed forgetting effect in the No Test condition (R: *M* = .76, *SE* = .02; F: *M* = .61, *SE* = .02), *OR* = .48, *CI* = .40 – .58, $p < .001$, and in the Encoding Test condition (R: *M* = .79, *SE* = .02; F: *M* = .72, *SE* = .02), *OR* =

.67, *CI* = .55 – .81, *p* < .001. Separated by Cue Type, for F items, recognition was significantly better in the Encoding Test condition (*M* = .72, *SE* = .02) than in the No Test condition (*M* = .61, *SE* = .02), *OR* = 1.62, *CI* = 1.36 – 1.94, *p* < .001. However, recognition of R items in the Encoding Test condition (*M* = .79, *SE* = .02) did not significantly differ from that of R items in the No Test condition (*M* = .76, *SE* = .02), *OR* = 1.18, *CI* = .97 – 1.43, *p* = .107.

Figure 12. Experiment 7: Raincloud plot for the recognition performance represented as proportion hit on the yes/no recognition test.



Note. The plot displays each participant's distribution of proportion hit for each Cue Type condition, separated by Encoding Type condition, and a boxplot. Error bars denote descriptive standard error.

Performance on the Yes/No Recognition Test separated by Encoding Test Condition.

To investigate whether recognition differed depending on the type of probe received immediately after the R/F cue presentation, I included fixed effects of Cue Type (R vs. F) and Encoding Type (No Test vs. Mismatched Test vs. Matched Test). Figure 13 presents proportion hit scores as a function of Cue Type and Encoding Type.

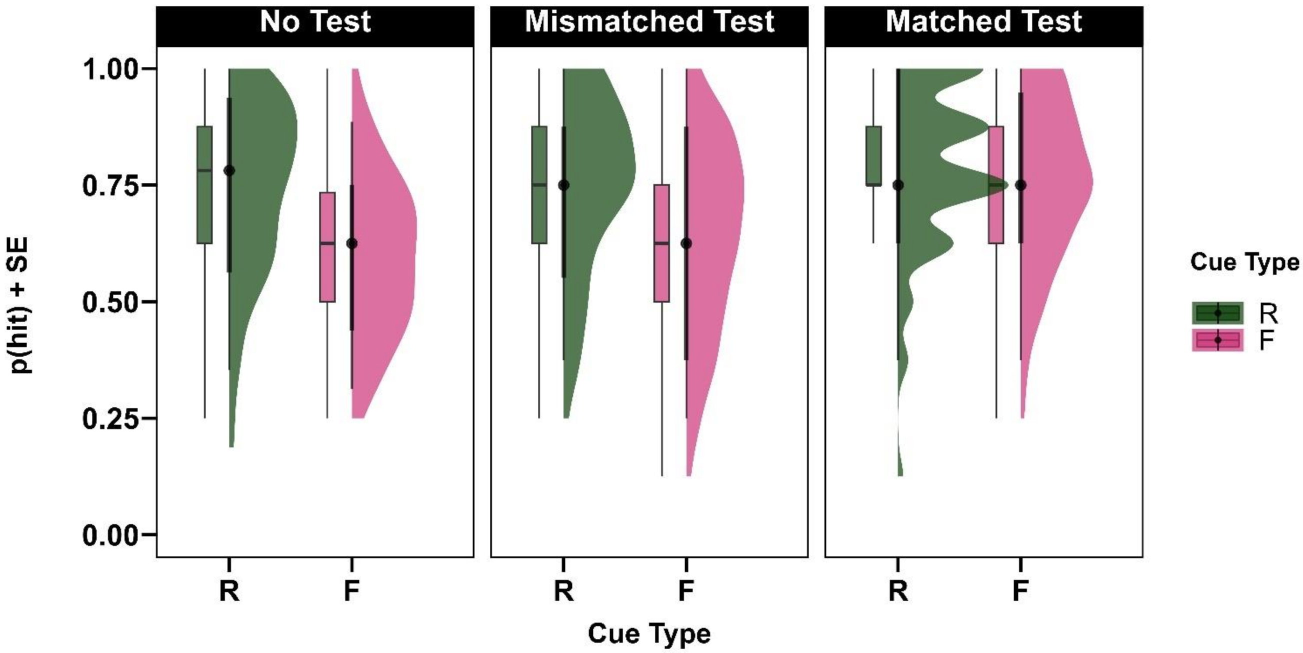
There was a significant effect of Cue Type, χ^2 (1) = 71.90, $p < .001$, where participants overall recognized more R items (M = .78, SE = .02) than F items (M = .69, SE = .02). There was also a significant effect of Encoding Type, χ^2 (2) = 43.27, $p < .001$, where recognition significantly differed between the No Test condition (M = .69, SE = .02), the Mismatched Test condition (M = .72, SE = .02) and the Matched Test condition (M = .80, SE = .02). There also was a significant Cue Type x Encoding Type interaction, χ^2 (2) = 9.45, p = .009.

Separated by Encoding Test, there was a significant directed forgetting effect in the No Test condition (R: M = .77, SE = .02; F: M = .61, SE = .02), OR = .48, CI = .40 – .58, $p <$ .001, and in the Mismatched Test condition (R: M = .77, SE = .02; F: M = .65, SE = .03), OR = .55, CI = .43 – .72, $p < .001$. However, there was no significant directed forgetting effect in the Matched Test condition (R: M = .81, SE = .02; F: M = .78, SE = .02), OR = .82, CI = .61 – 1.09, p = .167.

Averaged over the levels of Cue Type, participants recognized significantly more items in the Matched Test condition than in the No Test condition, OR = 1.73, CI = 1.41 – 2.12, $p < .001$, and than in the Mismatched Test condition, OR = .65, CI = .52 – .82, $p <$ .001. However, recognition did not significantly differ between the No Test condition and the

Mismatched Test condition, *OR* = 1.12, *CI* = .93 – 1.36, *p* = .332. Follow-up comparisons aggregated across Cue Type revealed that for F items, participants recognized significantly more items in the Matched Test condition than in the Mismatched Test condition, *OR* = .53, *CI* = .39 – .73, $p < .001$, or the No Test condition, *OR* = 2.27, *CI* = 1.71 – 2.98, $p < .001$. However, recognition did not significantly differ between the Mismatched Test condition and the No Test condition, *OR* = 1.21, *CI* = .93 – 1.56, *p* = .207. In contrast, there were no differences in R item recognition performance across any levels of the Encoding Test condition, between the Matched Test and the Mismatched Test conditions, *OR* = .79, *CI* = .56 – 1.11, *p* = .207, between the Matched Test and the No Test conditions, *OR* = 1.33, *CI* = .99 – 1.78, *p* = .062, or between the Mismatched Test and the No Test conditions, *OR* = 1.05, *CI* = .79 – 1.39, *p* = .924.

Figure 13. Experiment 7: Raincloud plot for the recognition performance showing proportion hit on the test phase yes/no recognition test.



Note. The plot displays each participant's distribution of proportion hit for each Cue condition, separated by Encoding Test condition, and a boxplot. Mean proportion hit denote the descriptive response proportion to yes/no recognition test. Error bars denote descriptive standard error.

3.4.2.2 Study Phase

Performance on the Post-Cue Yes/No Recognition Test Probe. To examine recognition performance on the study phase yes/no single-item recognition test that occurred immediately after the cue presentation, I ran a model with fixed effects of Cue Type (R vs. F) and Encoding Test Type (Mismatched Test vs. Matched Test). The No Test condition did not have a recognition test to analyze. While I report the results for this analysis, it is important to take caution when interpreting these results, as performance is at ceiling for both

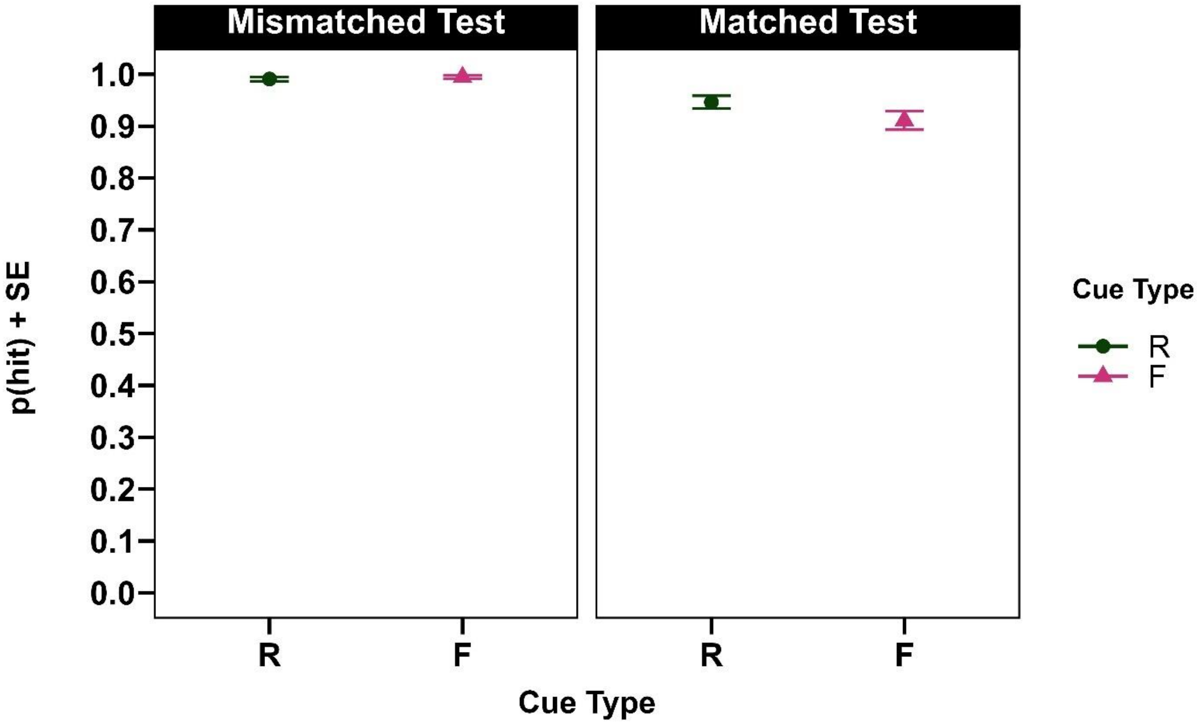
Mismatched Test and Matched Test condition. Figure 14 presents proportion hit scores as a function of Cue Type and Encoding Type.

There was a significant effect of Cue Type, χ^2 (1) = 4.47, *p* = .034, where participants correctly identified target words that preceded an R cue (*M* = .97, *SE* = .00) more often than those that preceded an F cue (*M* = .92, *SE* = .01). There was also a significant effect of Encoding Test Type, χ^2 (1) = 38.96, *p* < .001, where participants correctly identified whether the target word matched the word preceding the cue in the Mismatched Test condition (*M* = .99, *SE* = .00) more often than in the Matched Test condition (*M* = .95, *SE* = .01). There was no Cue Type x Encoding Test Type interaction, χ^2 (1) = 2.18, *p* = .140.

Separated by Cue Type, there was a significant directed forgetting effect in the Matched Test condition (R: *M* = .97, *SE* = .00; F: *M* = .94, *SE* = .01), *OR* = .54, *CI* = .33 – .88, *p* = .013, but no significant directed forgetting effect in the Mismatched Test condition (R: *M* = .99, *SE* = .02; F: *M* = .99, *SE* = .02), *OR* = 1.70, *CI* = .40 – 7.23, *p* = .472.

Separated by Encoding Type, for F items, participants correctly identified more often whether the target word preceding the cue was the same as the test word in the Mismatched Test condition than in the Matched Test condition, *OR* = 21.97, *CI* = 6.70 – 72.10, *p* < .001. An analogous pattern was found for the R items, *OR* = 6.95, *CI* = 2.64 – 18.30, *p* < .001.

Figure 14. Experiment 7: Proportion hit for the recognition performance on the study phase yes/no recognition test probe, separated by Cue Type and Encoding Type.



Note. Mean proportion hit denotes the descriptive response proportion to the study phase yes/no recognition test probe. Error bars denote descriptive standard error.

Response Times to Post-Cue Probes. To investigate RTs to the probe following R/F cue presentation, I ran a model that included fixed effects Cue Type (R vs. F) and Encoding Type (No Test vs. Mismatched Test vs. Matched Test). In the No Test condition, the post-cue probe was an asterisk. In the Mismatched Test and Matched Test conditions, the post-cue probe was a word requiring a yes/no recognition decision. The final model that converged included only the random by-participant intercept, excluding all correlations among random effects. Table 1 presents untransformed RTs as a function of Cue Type and Encoding Type. Figure 15 presents model-based log-RTs as a function of Cue Type and Encoding Type.

Table 1

Untransformed RTs to the Encoding Probe in Experiment 7

	Encoding Type		
	No Test	**Mismatched Test**	**Matched Test**
R	1280 (.02)	1676 (.01)	1537 (.02)
F	1177 (.02)	1597 (.02)	1694 (.01)

Note. Standard errors represented in parentheses.

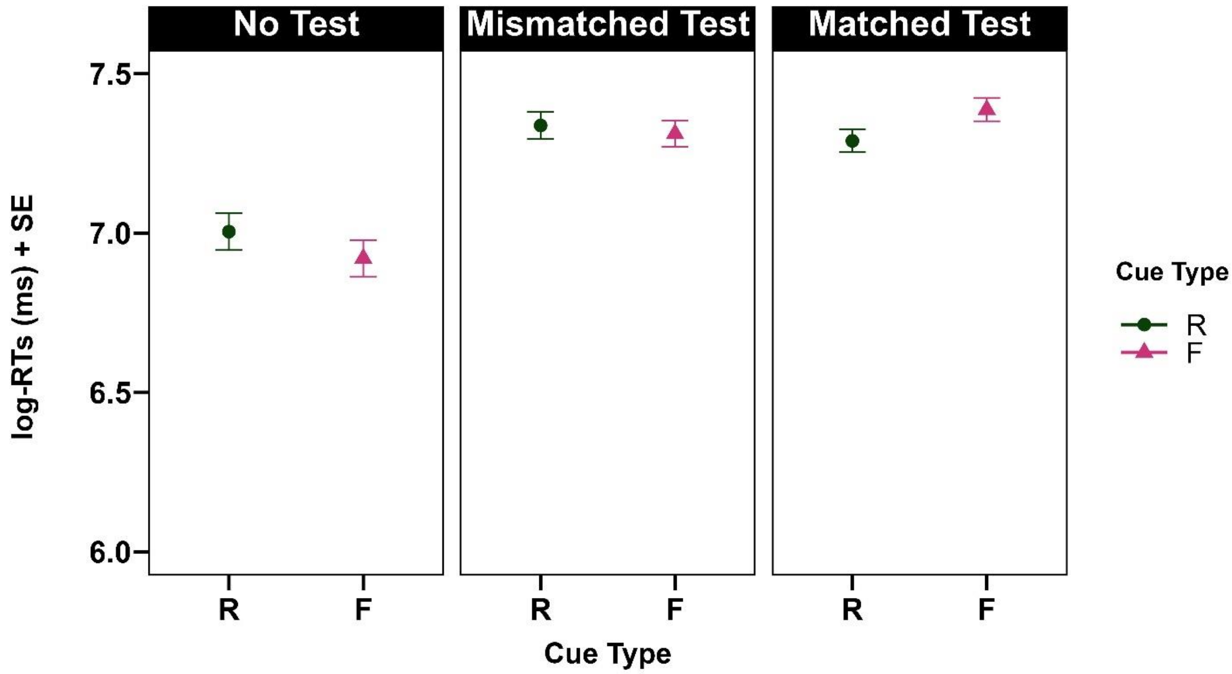
There was no overall significant effect of Cue Type, χ^2 (1) = 3.00, *p* = .083. There was a significant effect of Encoding Type, χ^2 (2) = 1128.86, *p* < .001, where overall participants were slower to respond to a probe in the Matched Test condition (*M* = 7.27, *SE* = .03) than in the Mismatched Test condition (*M* = 7.24, *SE* = .03) and than in the No Test condition (*M* = 6.79, *SE* = .03). There was a significant Cue Type x Encoding Type interaction, χ^2 (2) = 25.13, *p* < .001.

Separated by Encoding Type, in the No Test condition, participants were significantly slower to respond to the asterisk probe following an R cue (*M* = 6.83, *SE* = .03) than following an F cue (*M* = 6.75, *SE* = .03), *b* = .09, *SE* = .02, *p* < .001. In the Mismatched Test condition, RTs to the recognition task did not significantly differ following an R cue (*M* = 7.25, *SE* = .03) and an F cue (*M* = 7.24, *SE* = .03), *b* = .01, *SE* = .03, *p* = .750. In the Matched Test condition, participants were significantly slower to respond to the recognition task following an F cue (*M* = 7.31, *SE* = .03) than following an R cue (*M* = 7.23, *SE* = .03), *b* = .08, *SE* = .03, *p* = .003.

Averaged over the levels of Cue Type, RTs were significantly faster in the No Test condition than in the Mismatched Test condition, *b* = .45, *SE* = .02, *p* < .001, or in the

Matched Test condition, *b* = .48, *SE* = .02, *p* < .001. However, RTs did not significantly differ between the Mismatched Test condition and the Matched Test condition, *b* = .02, *SE* = .02, *p* = .443. Follow-up comparisons aggregated across Cue Type condition revealed that for F items, participants were significantly faster to respond to the asterisk probe in the No Test condition (*M* = 6.75, *SE* = .03) than to the recognition test in the Mismatched Test condition (*M* = 7.24, *SE* = .03), *b* = .49, *SE* = .02, *p* < .001, or in the Matched Test condition (*M* = 7.31, *SE* = .03), *b* = .56, *SE* = .02, *p* < .001. RTs were also significantly slower in the Matched Test condition (*M* = 7.27, *SE* = .03) than in the Mismatched Test condition (*M* = 7.24, *SE* = .03), *b* = .07, *SE* = .03, *p* = .031. For R items, the results were similar. Participants were significantly faster to respond to the asterisk probe in the No Test condition (*M* = 6.83, *SE* = .03) than to the recognition test in the Mismatched Test condition (*M* = 7.25, *SE* = .03), *b* = .41, *SE* = .02, *p* < .001, or in the Matched Test condition (*M* = 7.23, *SE* = .03), *b* = .39, *SE* = .02, *p* < .001. RTs did not significantly differ between the Mismatched Test condition (*M* = 7.25, *SE* = .03) and the Matched Test condition (*M* = 7.23, *SE* = .03), *b* = .02, *SE* = .03, *p* = .700.

Figure 15. Experiment 7: Mean log RTs to an encoding probe across Cue Type for each Encoding Test Type condition.



Note. Mean model-based log-RTs are plotted as a function of Cue Type and Encoding Type. Error bars denote the model-based standard deviations.

3.4.3 Discussion

In Experiment 7, my primary goal was to induce the retrieval of an F item immediately after the cue presentation. To do this, on some trials, I incorporated a recognition test following the cue presentation, where participants were asked to determine whether the word was the one that they had just studied. The goal of this manipulation was to mimic the natural retrieval check process that I propose for the R items and to instigate this process for the F items. In this way, I was able to investigate the role of an immediate retrieval mechanism while inducing retrieval via repetition using an additional method compared to Experiments 5 and 6.

Foremost, I conceptually replicated the findings of Experiments 5 and 6, where I observed a repetition benefit to the Encoding Test items. Next, I split the Encoding Test items into the Mismatched Test and the Matched Test conditions. The critical difference between these two conditions was that in the Mismatched Test condition, participants were presented with a new item on the encoding recognition test whereas in the Matched Test condition, the target word was identical to the studied word that preceded the R/F cue. Thus, the Matched Test condition induced a repetition of the studied item by forcing the retrieval of that preceding item—a type of retrieval check to determine whether they have the item readily available in working memory.

My primary hypothesis was confirmed: For F items, successful retrieval (i.e., repetition of the target item on the encoding recognition test) of a Matched Test F item boosted memory for that item beyond the initial forgetting that occurs when an item is not repeated (i.e., when a new item was presented on the encoding recognition test). The directed forgetting effect was only eliminated for the Matched Test items and not for the Mismatched Test items, illustrating the importance of *successful* retrieval of an item in boosting memory.

In line with my other predictions, recognition of Mismatched Test F items did not differ from that of No Test F items. That is, presenting a new word on the encoding recognition test did not induce the retrieval of the target item. In contrast, participants recognized significantly more Matched Test F items compared to Mismatched Test F items and No Test F items, demonstrating a retrieval benefit. This finding confirms that repeating the target item on the encoding recognition test induced a retrieval check similar to the retrieval that I have argued routinely occurs for R items. Furthermore, recognition for R items did not differ across the different types of Encoding Test conditions. This finding confirms my proposed retrieval check mechanism: Inducing a retrieval check for R items via the encoding recognition test did not influence memory because this type of retrieval check would already be happening for R items upon the cue presentation.

Taken together with the finding that the Matched Test F item memory boost arose due to the forced retrieval check, these results suggest that inducing a rapid retrieval check on F items eliminated the directed forgetting effect, critically highlighting the important role of retrieval in driving the directed forgetting effect. These findings are also critical to my primary thesis: A quick retrieval boost routinely occurs for the R items but not for the F items. When such retrieval process is forced upon the F items, however, the directed forgetting effect is eliminated.

Secondary to my main hypotheses for my proposed selective retrieval account, incorporating a probe following cue presentation provided useful information regarding another account of item-method directed forgetting. One of the key components of previous investigations into the mechanisms of item-method directed forgetting has been examination of RTs to a probe, or to another task, following R/F cue presentation (Fawcett & Taylor, 2008, 2012). Specifically, as described in Chapter 1, Fawcett and Taylor (2008, 2012) asked participants to respond to a probe (e.g., asterisk or irrelevant word) during the study phase of item-method directed forgetting. They reported that participants were slower to respond to a probe following an F cue than following an R cue and suggested on this basis that the cognitive demands of intentional forgetting exceeded those of remembering. This finding has been the hallmark evidence in support of an effortful attentional withdrawal mechanism operating to remove irrelevant F information from memory. The selective rehearsal account makes the opposite prediction: RTs should be slower to a probe following an R cue than an F cue because the directed forgetting effect results from elaborative rehearsal of only the R items, which interferes with the representation of the F items. Therefore, the selective rehearsal account predicts that the cognitive demands for remembering should exceed those of forgetting.

In Experiment 7, participants were led to make one of two different responses to a probe during the study phase: (1) in the No Test condition, identical to Fawcett and Taylor (2008), participants simply had to press the space bar to an asterisk probe to proceed to the next trial, and (2) in the Matched/Mismatched Test condition, participants had to perform a recognition test. In the No Test condition, participants were slower to press the button following an R cue than following an F cue. In sharp contrast, in the Matched Test condition, this RT pattern reversed: Participants were slower to respond on the recognition test following an F cue than following an R cue. This difference in RT performance provides evidence for the effect of retrieval on forgetting in the item-method paradigm suggesting that my novel, retrieval-inducing manipulation was successful. That is, a Matched encoding recognition task successfully induced the retrieval check of F items, which would not ordinarily have occurred. When participants were forced to retrieve the preceding F item, the RT pattern reversed because it was harder to retrieve the already-removed F item that is not routinely subjected to an act of retrieval.

Second, the switch in the pattern of RT performance between the two conditions sheds light on the mechanisms of item-method directed forgetting. This pattern switch could be explained by process changes in the Matched Test condition compared to the No Test condition. The No Test condition provides a purer test of the cognitive demands of intentional remembering and intentional forgetting, where I did not induce a retrieval process. Thus, the RT data pattern for the No Test condition here, which was set up to be similar to the probe conditions in prior studies, conflicts with the attentional inhibition account and instead would support a non-inhibitory account of item-method directed forgetting. Before I elaborate on the critical theoretical debate regarding the underlying cognitive mechanisms, I present a final experiment to test a key assumption of the attentional inhibition account.

3.5 Experiment 8

The results from Experiments 6 and 7 suggest that retrieval plays an important role in item-method directed forgetting. In Experiment 6, retrieval was induced by repeating the items later in the study list and memory benefits depended on participants noticing such repetition. In Experiment 7, retrieval was induced immediately after the R/F cue presentation to mimic the routinely occurring retrieval check for R items. These two experiments provide a unified picture of how retrieval contributes to item-method directed forgetting in providing the benefit to the R items only.

In Experiment 7, by incorporating an asterisk probe (i.e., No Test condition) or a recognition task (i.e., Mismatched/Matched Test conditions) to force a retrieval check, I was able to investigate the cognitive demands of intentional remembering and intentional forgetting following an R/F instruction. According to the attentional inhibition account, intentional forgetting requires more resources than intentional remembering. Greater demands on limited working memory resources mean that there are fewer resources available to attend to a secondary task immediately after an F instruction than after an R instruction. Contrarily, the non-inhibitory, selective rehearsal account proposes that intentional remembering requires more resources than intentional forgetting, thus fewer resources should be available to attend to a secondary task following an R instruction compared to following an F instruction.

To date, there have been conflicting findings on the resource demands of intentional remembering and intentional forgetting. Fawcett and Taylor (2008, 2012) argued that intentional forgetting requires more resources because participants were slower to respond to an irrelevant probe following an F cue than following an R cue. They interpreted this finding as an indicator of greater processing in the immediate aftermath of an F instruction compared with an R instruction. In contrast, Popov et al. (2019) found that when an item was preceded by an R cue, memory for that item was worse than when an item was preceded by an F cue. That is, memory for items preceded by an F instruction was better than memory for items preceded by an R instruction. This effect was cumulative: The more preceding items that were F, the better the memory was for the current item. Lastly, in my previous work, Tan et al. (2020) argued that, because participants showed better memory for R items paired with F items compared to R items paired with R items, an intent to forget absorbed fewer resources than did an intent to remember. If forgetting was more resource-demanding than remembering, then pairing a F item with a R item should absorb more resources from the processing of that R item than should pairing a R item with that R item. Also inconsistent with F items using more resources was the Tan et al. finding that memory for F items did not differ from memory for baseline items (i.e., in a neutral/no-cue condition).

Overall, the findings of Popov et al. (2019) and Tan et al. (2020) do not support the idea that intentional forgetting imposes a heavier cognitive demand than does intentional remembering. In contrast, the active forgetting assumption of the attentional inhibition account (Fawcett & Taylor, 2008, 2012) proceeds to dominate item-method directed forgetting literature. Indeed, the notion that an instruction to forget engages inhibitory mechanisms involved in expunging the item is generally accepted and little debated (Anderson & Hulbert, 2021; Chiu et al., 2021; Fellner, Waldhauser, & Axmacher, 2020; Wylie et al., 2007).

The key evidence for the attentional inhibition account is the RT differences to a secondary task following an R/F instruction. Critically, Fawcett and Taylor (2008, 2012) introduced three inter-stimulus intervals (ISIs) between the cue presentation and the onset of the secondary task. Their ISI timing observations suggest that intentional forgetting takes less than 2,600 ms, because RT differences disappeared at the 2,600 ms ISI condition. Across their investigations of active forgetting, they incorporated these ISI manipulations and consistently found slower RTs on a secondary task while executing the F instruction at 1,400 ms and 1,800 ms ISIs, but not in the 2,600 ms ISI condition (Fawcett & Taylor, 2008, 2012, 2012; Fawcett, Lawrence, & Taylor, 2016). The RT differences together with the ISI timing observations have been the strongest argument in favour of the active, cognitively-demanding attentional inhibition account of item-method directed forgetting.

In Experiment 8, I incorporated into my Experiment 7 design ISI timing manipulations identical to those used by Fawcett and Taylor (2008). The purpose of this was two-fold: (1) My main goal was to replicate the findings of my retrieval-check-inducing paradigm to strengthen support for my selective retrieval account, and (2) an additional goal was to investigate how an active forgetting mechanism may fit within my selective retrieval account. So far, my primary interest has been to demonstrate evidence for the retrieval check mechanism that is ordinarily occurring for the R items. It is also possible that item-method directed forgetting can be explained by a combination of two processes: a facilitation process that rapidly boosts the memory strength of to-be-remembered information via a retrieval check and another removal process that reduces the representation of to-be-forgotten information via active inhibition. In this case, my selective retrieval account may provide a unified explanation by integrating an active forgetting component to explain how F items are removed from working memory.

Participants studied one item at a time, each followed by an R/F cue. Critically, the delay between the cue presentation and the onset of the asterisk probe (i.e., No Test condition) and the recognition test (i.e., Mismatched/Matched Test conditions) varied between 1,400 ms, 1,800 ms, and 2,600 ms. The structure of all conditions was otherwise identical to Experiment 7 (with the exception of introducing the three varying ISIs). For recognition memory performance on the test phase, I expected to replicate the findings of Experiment 7.

For RT performance during the study phase, I predicted two possible outcomes. First, it is possible that I would replicate the RT pattern from Experiment 7. According to a non-inhibitory account, in the No Test condition, participants should be slower to respond to the asterisk probe following an R instruction than following an F instruction. This is because the retrieval check that routinely occurs for the R items occupies more resources compared to no additional process occurring following an F cue. However, in the Matched Test condition, participants should be slower on the recognition test following an F instruction than following an R instruction. This is because the retrieval-inducing recognition test introduces an additional process that does not ordinarily occur for the F items. Having to retrieve an already-removed F item, participants should be slower than for R items, where the retrieval check is already occurring. In this first scenario, introducing ISIs should not have an effect on the RT pattern because the retrieval check occurs rapidly. It is possible, however, that I will replicate the findings of Fawcett and Taylor (2008). If so, at the ISIs of 1,400 ms and 1,800 ms, participants should be slower to respond to the asterisk probe following an F cue than an R cue, but this RT difference should disappear for the ISI of 2,600 ms. Such a finding would support the role of an active forgetting mechanism that is more cognitively-demanding than is intentional remembering.

3.5.1 Method

***Participants*.**

One hundred and fifty-one participants were recruited from Prolific. Their mean age was 25 (*SD* = 3.25) with 64% identifying as female and 4% identifying as nonbinary. The exclusion criteria were identical to those of the previous experiments.

The sample size calculation was identical to that of Experiment 7, but because I introduced a new ISI variable into the design, I doubled the required sample size to detect up to a 50% reduction in detecting a desired power of ~.80 (α = .05, two-tailed). This yielded a required sample size of n = 144.

Four participants were excluded for switching tabs more than twice during the study. Three participants were excluded because they reported writing down all of the words, ultimately scoring 100% for both R and F items. The final sample therefore included 144 participants, all of whom passed the Mill Hill Vocabulary Scale (M = 46.72, SD = 3.83).

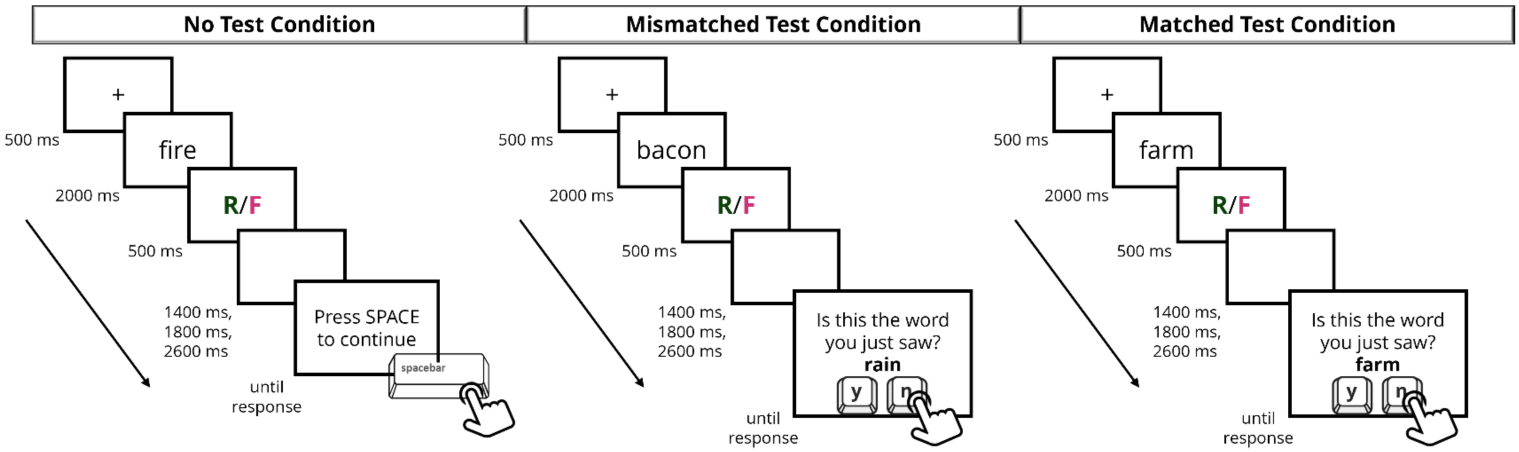
Materials.

The materials, encoding list, and test list were identical to those of Experiment 7.

Procedure.

The study phase was identical to that of Experiment 7, with a few key changes. First, the presentation timings were changed to be similar to those used by Fawcett and Taylor (2008). Each item was presented for 2 s, followed by an R/F cue for 500 ms. Immediately following the cue, a blank screen appeared for ISIs of either 1,400, 1,800, or 2,600 ms. Afterward, a probe was presented. The type of the probe was identical to those used in Experiment 7. Figure 16 displays the study phase procedure for Experiment 8.

Figure 16. Experiment 8: Systematic representation of the conditions during the study phase, including the ISI conditions.



Data Analysis.

Model comparisons and criteria for mixed-effects logistic regression were identical to those used in Experiment 7, with the exception that the model included a fixed effect of ISI (1,400 ms, 1,800 ms vs. 2,600 ms). Linear mixed-effects models for RT analyses also were identical to those in Experiment 7.

3.5.2 Results

The structure of the results section is identical to that of Experiment 7, where analyses reported under *Test Phase* represent the memory performance on the yes/no recognition task during the test phase. Next, I report the analyses for the performance on the single-item yes/no recognition task during the study phase. Finally, I report the analyses for RT to the probe following the R/F cue during the study phase. These results are highlighted under the *Study Phase* sections for the study phase analyses.

3.5.2.1 Test Phase

Replication of Experiment 5 with Collapsed Encoding Type Condition.

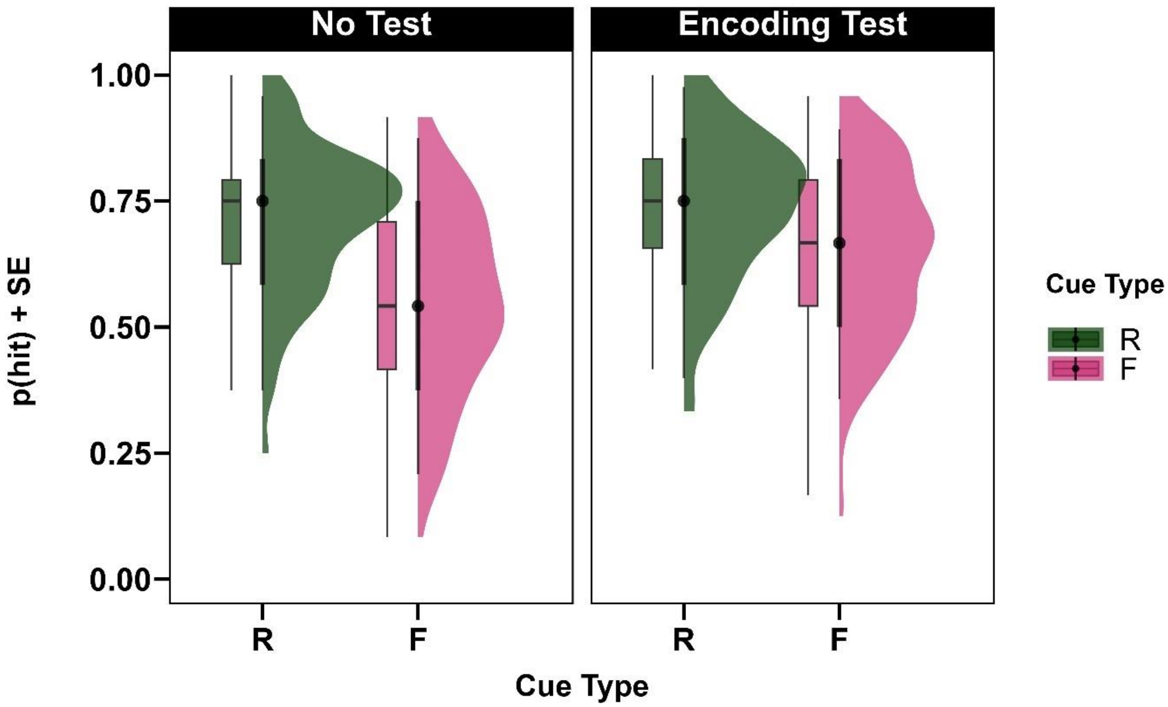
Figure 17 presents the proportion hit scores as a function of Cue Type and Encoding Type. The mean collective False Alarm rate for all conditions was 0.26 (*SE* = .02). The model was identical to that in Experiment 5, where I included fixed effects of Cue Type (R vs. F) and Encoding Type (No Test vs. Encoding Test). Here, the Mismatched Test and Matched Test conditions were averaged to create the Encoding Test condition that was analogous to the Repeated condition from Experiment 5.

There was a significant effect of Cue Type, χ^2 (1) = 267.71, $p < .001$, where participants overall recognized more R items (*M* = .74, *SE* = .01) than F items (*M* = .61, *SE* = .01). There was also a significant effect of Encoding Type, χ^2 (1) = 61.03, $p < .001$, where recognition was better for the Encoding Test condition (*M* = .70, *SE* = .01) than for the No Test condition (*M* = .65, *SE* = .01). There was also a significant Cue Type x Encoding Type interaction, χ^2 (1) = 17.81, $p < .001$.

Separated by Encoding Test, there was a significant directed forgetting effect both in the No Test condition (R: *M* = .73, *SE* = .01; F: *M* = .55, *SE* = .01), *OR* = .46, *CI* = .41 – .51, $p < .001$, and in the Encoding Test condition (R: *M* = .75, *SE* = .01; F: *M* = .66, *SE* = .01), *OR* = .63, *CI* = .57 – .70, $p < .001$.

Separated by Cue Type, for F items, recognition was significantly better in the Encoding Test condition (*M* = .72, *SE* = .02) than in the No Test condition (*M* = .61, *SE* = .02), *OR* = 1.56, *CI* = 1.41 – 1.72, $p < .001$. Recognition for R items was also significantly better in the Encoding Test condition (*M* = .72, *SE* = .01) than in the No Test condition (*M* = .75, *SE* = .01), *OR* = 1.13, *CI* = 1.01 – 1.26, $p = .028$.

Figure 17. Experiment 8: Raincloud plot for the recognition performance represented as proportion hit on the yes/no recognition test.



Note. The plot displays each participant's distribution of proportion hit for each Cue condition, separated by Encoding Test condition, and a boxplot.

Replication of Experiment 7 with Separated Encoding Type Condition.

To examine the influence of the encoding yes/no recognition test, I included fixed effects of Cue Type (R vs. F) and Encoding Type (No Test vs. Mismatched Test vs. Matched Test). Figure 18 presents the proportion hit scores as a function of Cue Type and Encoding Type, where Encoding Type was separated into the three levels.

There was a significant effect of Cue Type, χ^2 (1) = 266.75, p < .001, where participants overall recognized more R items (M = .75, SE = .01) than F items (M = .63, SE =

.01). There was also a significant effect of Encoding Type, χ^2 (2) = 117.57, $p < .001$, where recognition significantly differed across the levels of Encoding Type. There was also a significant Cue Type x Encoding Type interaction, χ^2 (2) = 58.07, $p < .001$.

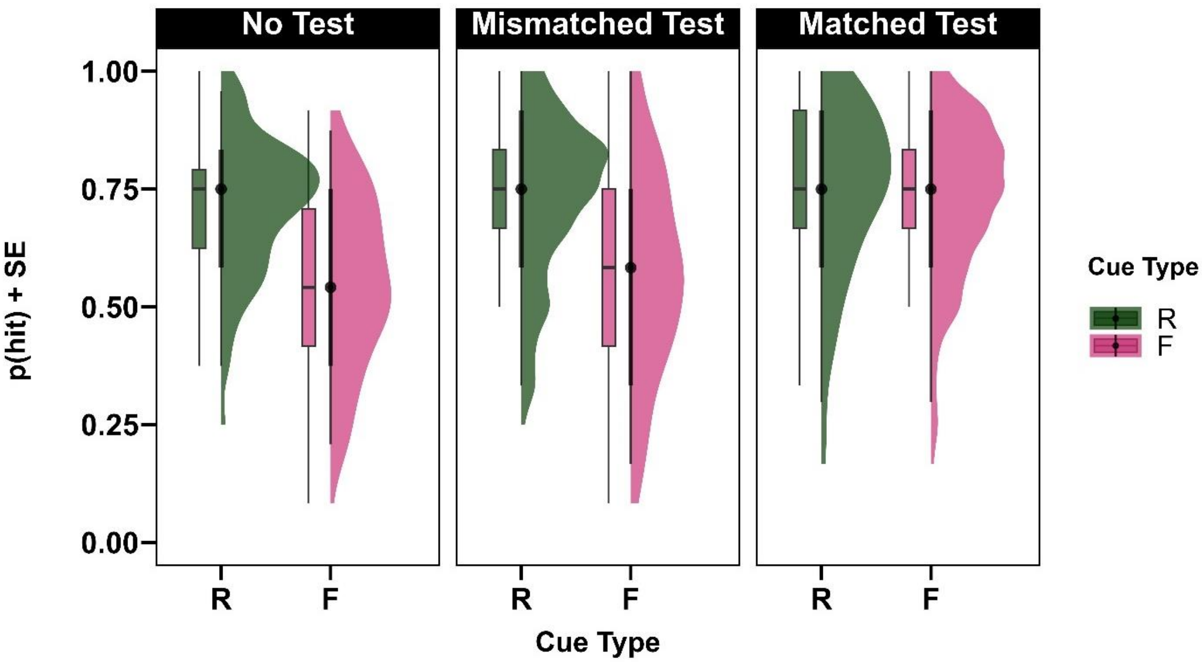
Separated by Encoding Test, there was a significant directed forgetting effect both in the No Test condition (R: *M* = .73, *SE* = .01; F: *M* = .55, *SE* = .02), *OR* = .46, *CI* = .41 – .51, $p < .001$, and in the Mismatched Test condition (R: *M* = .75, *SE* = .01; F: *M* = .57, *SE* = .02), *OR* = .45, *CI* = .39 – .53, $p < .001$. However, there was no significant directed forgetting effect in the Matched Test condition (R: *M* = .76, *SE* = .01; F: *M* = .74, *SE* = .01), *OR* = .91, *CI* = .78 – 1.07, $p = .250$.

Averaged over the levels of Cue Type, overall recognition was better for the Matched Test condition (*M* = .75, *SE* = .01) than for either the Mismatched Test condition (*M* = .67, *SE* = .01), *OR* = 1.65, *CI* = 1.48 – 1.85, $p < .001$, or the No Test condition (*M* = .65, *SE* = .01), *OR* = 1.09, *CI* = .98 – 1.21, $p < .001$. However, recognition did not significantly differ between the Mismatched Test condition and the No Test condition, *OR* = .66, *CI* = .58 – .75, $p < .001$.

Follow-up comparisons aggregated across Cue Type revealed that for F items, participants recognized significantly more items in the Matched Test condition than in either the Mismatched Test condition, *OR* = .46, *CI* = .39 – .55, $p < .001$, or the No Test condition, *OR* = 2.33, *CI* = 1.99 – 2.72, $p < .001$. However, recognition did not significantly differ between the Mismatched Test condition and the No Test condition, *OR* = 1.08, *CI* = .93 – 1.25, $p = .425$. In contrast, there were no differences in R item recognition performance across any levels of the Encoding Test condition—between the Matched Test and

Mismatched Test conditions, *OR* = .07, *CI* = .77 – 1.13, *p* = .661, between the Matched Test and No Test conditions, *OR* = 1.17, *CI* = .99 – 1.38, *p* = .057, or between the Mismatched Test and No Test conditions, *OR* = 1.09, *CI* = .93 – 1.28, *p* = .399.

Figure 18. Experiment 8: Raincloud plot for the recognition performance showing proportion hit on the test phase yes/no recognition test.



Note. The plot displays each participant's distribution of proportion hit for each Cue condition, separated by Encoding Test condition, and a boxplot. Mean proportion hit denote the descriptive response proportion to yes/no recognition test. Error bars denote descriptive standard error.

Performance on the Yes/No Recognition Test separated by Encoding Test Condition and ISI Condition.

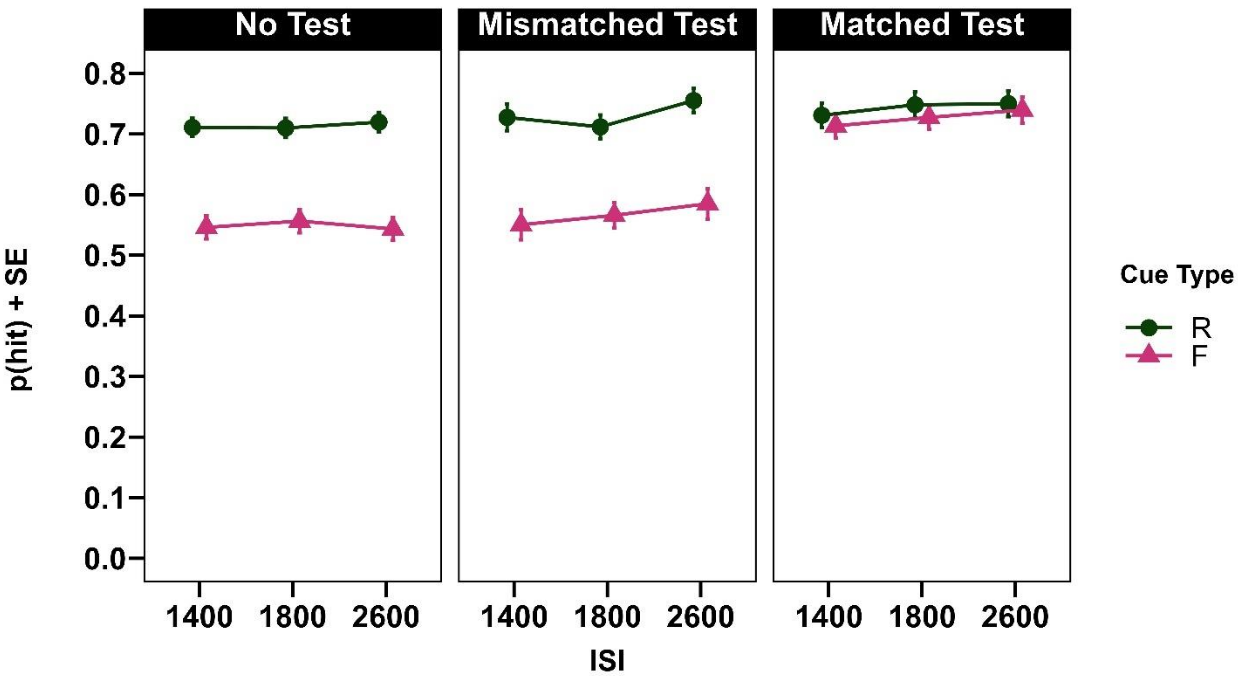
To examine how different durations of ISI between cue presentation and the onset of the probe affected performance, I included fixed effects of Cue Type (R vs. F), Encoding Type (No Test vs. Mismatched Test vs. Matched Test), and ISI (1,400 ms vs. 1,800 ms vs. 2,600 ms). Figure 19 presents the relevant proportion hit scores.

Similar to the other models, there was the same significant effect of Cue Type, χ^2 (1) = 266.78, $p < .001$, where participants overall recognized more R items ($M = .75$, $SE = .01$) than F items ($M = .63$, $SE = .01$). There was also a significant effect of Encoding Type, χ^2 (2) = 117.40, $p < .001$, where recognition was better for the Matched Test condition ($M = .75$, $SE = .01$) than for the Mismatched Test condition ($M = .67$, $SE = .01$) or the No Test condition ($M = .65$, $SE = .01$). However, there was no main effect of ISI, χ^2 (2) = 2.62, $p = .270$, indicating that recognition performance did not significantly differ across the ISI conditions (1,400ms: $M = .68$, $SE = .01$; 1,800ms: $M = .69$, $SE = .01$; 2,600ms: $M = .70$, $SE = .01$). Only one interaction was significant: Cue Type x Encoding Type, χ^2 (2) = 58.11, $p < .001$. There was no Cue Type x ISI interaction, χ^2 (2) = .71, $p = .703$, no Encoding Type x ISI interaction, χ^2 (4) = 3.34, $p = .502$, and no Cue Type x Encoding Type x ISI interaction, χ^2 (4) = .71, $p = .950$.

Because my main interest was to investigate recognition performance at each level of ISI condition, the following comparisons were pre-registered and analyzed to interpret my hypotheses. For each level of ISI, there were significant directed forgetting effects in the No Test condition (1,400 ms: $M_R = .73$, $SE_R = .02$; $M_F = .55$, $SE_F = .02$; 1,800 ms: $M_R = .73$, $SE_R = .02$; $M_F = .56$, $SE_F = .02$; 2,600 ms: $M_R = .74$, $SE_R = .02$; $M_F = .55$, $SE_F = .02$), *ORs* > .43, *CIs* > .33 – .57, *ps* < .001, and in the Mismatched Test condition, (1,400 ms: $M_R = .74$,

SE_R = .02; M_F = .55, SE_F = .02; 1,800 ms: M_R = .73, SE_R = .02; M_F = .57, SE_F = .02; 2,600 ms: M_R = .77, SE_R = .02; M_F = .59, SE_F = .02). There was no directed forgetting effect in the Matched Test condition (1,400 ms: M_R = .75, SE_R = .02; M_F = .73, SE_F = .02; 1,800 ms: M_R = .77, SE_R = .02; M_F = .74, SE_F = .02; 2,600 ms: M_R = .77, SE_R = .02; M_F = .76, SE_F = .02), *ORs* > .89, *CIs* > .68 – 1.17, *ps* > .398. Recognition did not significantly differ between the different levels of ISI, *ORs* > .85, *CIs* > .62 – 1.49, *ps* > .391.

Figure 19. Experiment 8: Proportion hit for the recognition performance on the test phase yes/no recognition test, separated by Cue Type, Encoding Type, and ISI.



Note. Mean proportion hit denote the descriptive response proportion to the test phase yes/no recognition test. Error bars denote descriptive standard error.

3.5.2.2 Study Phase

Replication of Experiment 7 with Performance on the Post-Cue Yes/No Recognition Test Probe.

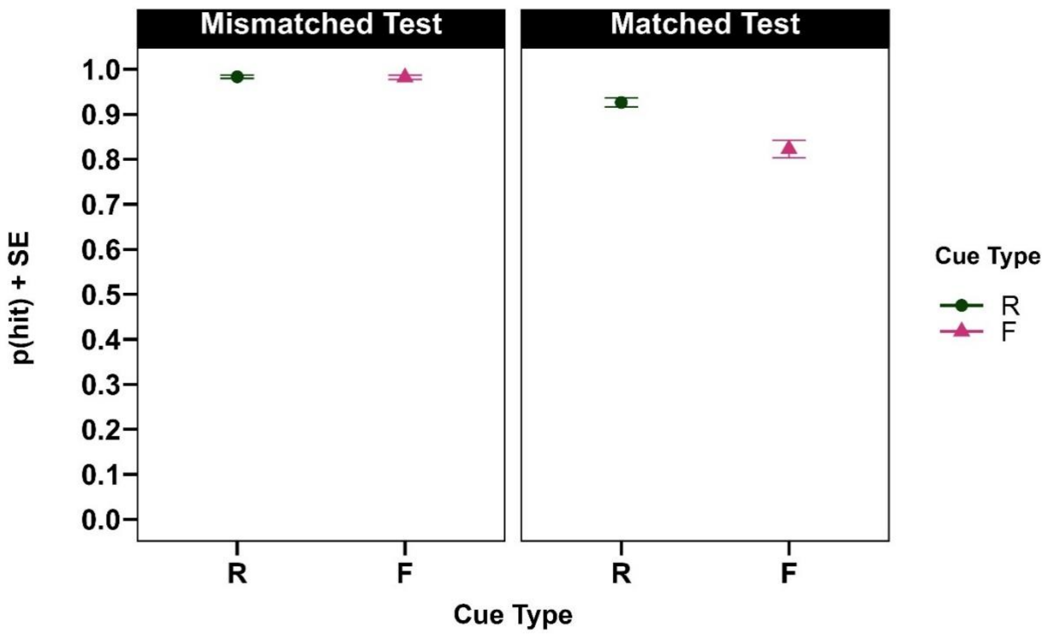
To examine recognition performance on the study phase single-item yes/no recognition test that occurred immediately after the cue presentation, I ran a model with fixed effects of Cue Type (R vs. F) and Encoding Test Type (Matched Test vs. Mismatched Test). I removed the No Test condition because these trials did not have a recognition test to analyze. Figure 20 presents proportion hit scores as a function of Cue Type and Encoding Type.

There was a significant effect of Cue Type, χ^2 (1) = 84.91, $p < .001$, where participants correctly identified target words that preceded an R cue (*M* = .98, *SE* = .00) more often than those that preceded an F cue (*M* = .95, *SE* = .00). There was also a significant effect of Encoding Type, χ^2 (1) = 248.60, $p < .001$, where participants correctly identified whether the target word matched the word preceding the cue better in the Mismatched Test condition (*M* = .99, *SE* = .01) than in the Matched Test condition (*M* = .93, *SE* = .00). There was a significant Cue Type x Encoding Type interaction, χ^2 (1) = 15.44, $p < .001$.

Separated by Encoding Type, there was a significant directed forgetting effect in the Matched Test condition (R: *M* = .96, *SE* = .01; F: *M* = .88, *SE* = .01), *OR* = .31, *CI* = .24 – .39, $p < .001$, but no significant directed forgetting effect in the Mismatched Test condition (R: *M* = .99, *SE* = .00; F: *M* = .99, *SE* = .00), *OR* = .93, *CI* = .56 – 1.55, $p = .785$.

Separated by Cue Type, for F items, participants correctly identified whether the target word matched the word preceding the cue in the Mismatched Test condition more often than in the Matched Test condition, *OR* = .06, *CI* = .04 – .09, *p* < .001. Similarly, for R items, participants correctly identified whether the target word matched the word preceding the cue in more often in the Mismatched Test condition than in the Matched Test condition, *OR* = .19, *CI* = .13 – .28, *p* < .001.

Figure 20. Experiment 8: Proportion hit for the recognition performance on the study phase yes/no recognition test, separated by Cue Type and Encoding Type.



Note. Mean proportion hit denote the descriptive response proportion to the test phase yes/no recognition test. Error bars denote descriptive standard error.

Performance on the Yes/No Recognition Test after the Cue Presentation with ISI.

To examine how recognition performance on the study phase yes/no recognition test was influenced by ISI and Encoding Type condition, I ran a model with fixed effects of Cue Type (R vs. F), Encoding Type (Matched Test vs. Mismatched Test), and ISI (1,400 ms vs.

1,800 ms vs. 2,600 ms). Figure 21 presents proportion hit scores as a function of Cue Type and Encoding Type.

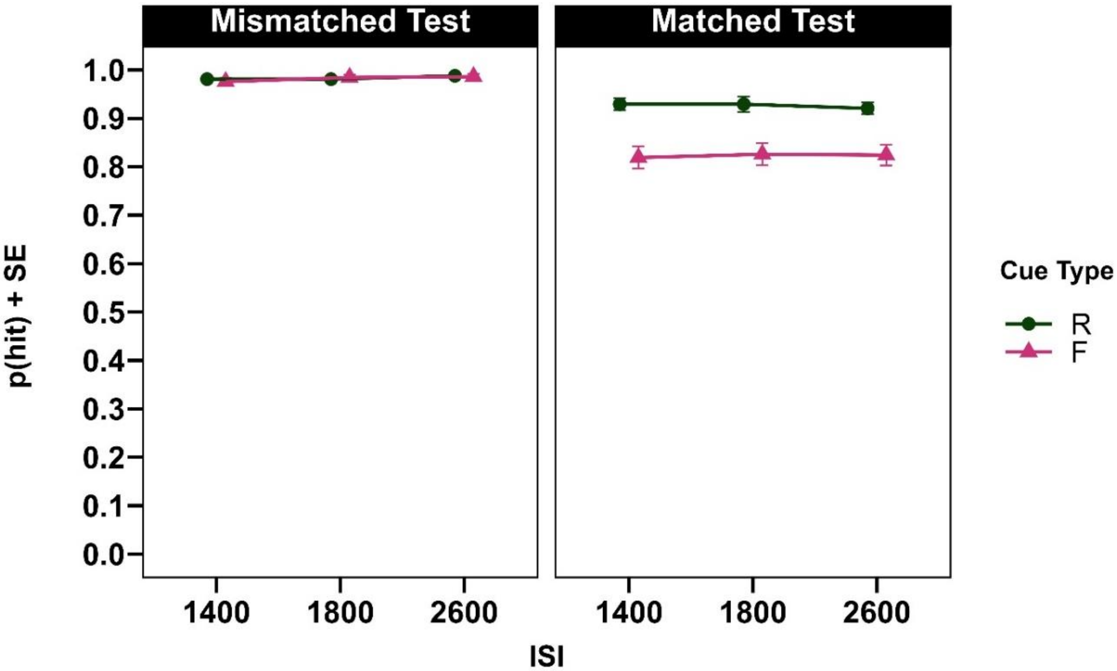
There was a significant effect of Cue Type, χ^2 (1) = 84.91, *p* < .001, where participants correctly identified target words that preceded an R cue (*M* = .98, *SE* = .00) more often than those that preceded an F cue (*M* = .95, *SE* = .00). There was also a significant effect of Encoding Type, χ^2 (1) = 242.11, *p* < .001, where participants correctly identified whether the target word matched the word preceding the cue more often in the Mismatched Test condition (*M* = .99, *SE* = .01) than in the Matched Test condition (*M* = .93, *SE* = .00). There was a significant Cue Type x Encoding Type interaction, χ^2 (1) = 15.61, *p* < .001. However, the main effect of ISI was not significant, χ^2 (2) = .43, *p* = .808, indicating that recognition performance did not significantly differ across the ISI conditions (1,400 ms: *M* = .97, *SE* = .00; 1,800 ms: *M* = .98, *SE* = .00; 2,600 ms: *M* = .98, *SE* = .00). None of the other interactions were significant: Cue Type x ISI, χ^2 (2) = .54, *p* = .763, Encoding Type x ISI, χ^2 (2) = 2.72, *p* = .257, or Cue Type x Encoding Type x ISI, χ^2 (2) = .53, *p* = .766.

To investigate the significant Cue Type x Encoding Type interaction, first I conducted follow-up comparisons without including the ISI condition. Separated by Encoding Type, there was a significant directed forgetting effect in the Matched Test condition (R: *M* = .96, *SE* = .01; F: *M* = .88, *SE* = .01), *OR* = .31, *CI* = .24 – .39, *p* < .001, but not in the Mismatched Test condition (R: *M* = .99, *SE* = .00; F: *M* = .99, *SE* = .00), *OR* = .93, *CI* = .56 – 1.55, *p* = .785.

Separated by Cue Type, for F items, participants correctly identified whether the target word matched the word preceding the cue more often in the Mismatched Test condition than in the Matched Test condition, *OR* = .06, *CI* = .04 – .09, $p < .001$. Similarly, for R items, participants correctly identified whether the target word matched the word preceding the cue more often in the Mismatched Test condition than in the Matched Test condition, *OR* = .19, *CI* = .13 – .28, $p < .001$.

The following analyses were pre-registered to interpret my main hypothesis of whether recognition performance differed at each level of ISI. For each level of ISI, there were significant directed forgetting effects in the Matched Test condition (1,400 ms: M_R = .96, SE_R = .01; M_F = .88, SE_F = .02; 1,800 ms: M_R = .97, SE_R = .01; M_F = .89, SE_F = .02; 2,600 ms: M_R = .96, SE_R = .01; M_F = .89, SE_F = .02), *ORs* > .28, *CIs* > .19 – .50, *ps* < .001, but not in the Mismatched Test condition (1,400 ms: M_R = .99, SE_R = .00; M_F = .9, SE_F = .00; 1,800 ms: M_R = .99, SE_R = .00; M_F = .99, SE_F = .00; 2,600 ms: M_R = .99, SE_R = .00; M_F = .99, SE_F = .00), *ORs* > .77, *CIs* > .32 – 2.98, *ps* > .524. However, at each level of ISI, recognition did not significantly differ between the different levels of ISI, *ORs* > .62, *CIs* > .20 – 2.75, *ps* > .373.

Figure 21. Experiment 8: Proportion hit for the recognition performance on the study phase yes/no recognition test, separated by Cue Type, Encoding Type, and ISI.



Note. Mean proportion hit denote the descriptive response proportion to the study phase yes/no recognition test. Error bars denote descriptive standard error.

Response Times to Post-Cue Probes.

To examine participant RTs to a probe following R/F cue presentation, I ran a model that included fixed effects Cue Type (R vs. F), Encoding Type (No Test vs. Mismatched Test vs. Matched Test), and ISI (1,400 ms vs. 1,800 ms vs. 2,600 ms). In the No Test condition, the post-cue probe was an asterisk. In the Mismatched Test and Matched Test conditions, the post-cue probe was a yes/no recognition test. The final model that converged included only the random by-participant intercept, excluding all correlations among random effects. Figure 22 presents the model-based log-RTs as a function of Cue Type and Encoding Type, providing a direct replication of Experiment 7. Table 2 presents the descriptive untransformed RTS as a function of Cue Type, Encoding Type, and ISI. Figure 23 presents the model-based log-RTs, separated by Cue Type, Encoding Type, and ISI.

Table 2

Untransformed RTs to the Encoding Probe in Experiment 8

	Encoding Type								
	No Test			Mismatched Test			Matched Test		
	1,400	1,800	2,600	1,400	1,800	2,600	1,400	1,800	2,600
R	1008 (.02)	1020 (.02)	955 (.03)	1304 (.02)	1283 (.02)	1290 (.02)	1238 (.02)	1290 (.03)	1330 (.01)
F	1003 (.01)	953 (.02)	876 (.02)	1319 (.02)	1327 (.01)	1392 (.01)	1522 (.01)	1444 (.02)	1498 (.01)

Note. Standard errors represented in parentheses.

There was a significant effect of Cue Type, χ^2 (1) = 6.45, *p* = .011, where, averaged over the levels of Encoding Type and ISI, participants were slower to respond to a probe following an R cue (*M* = 7.01, *SE* = .02) than following an F cue (*M* = 6.92, *SE* = .02). There was also a significant effect of Encoding Type, χ^2 (1) = 2473.29, *p* < .001, where, overall, participants were slower to respond to a probe in the Matched Test condition (*M* = 7.06, *SE* = .02) than in the Mismatched Test condition (*M* = 7.02, *SE* = .02) or in the No Test condition (*M* = 6.67, *SE* = .02). However, it is important to note that, averaged over the levels of Cue Type and ISI, RTs did not significantly differ between the Matched Test and the Mismatched Test conditions, *b* = .02, *SE* = .01, *p* = .131. RTs were significantly faster in the No Test condition than in the Matched Test condition, *b* = .38, *SE* = .01, *p* < .001, or in the Mismatched Test condition, *b* = .37, *SE* = .01, *p* < .001. There was a significant Cue Type x Encoding Type interaction, χ^2 (1) = 220.29, *p* < .001. However, there was no main effect of

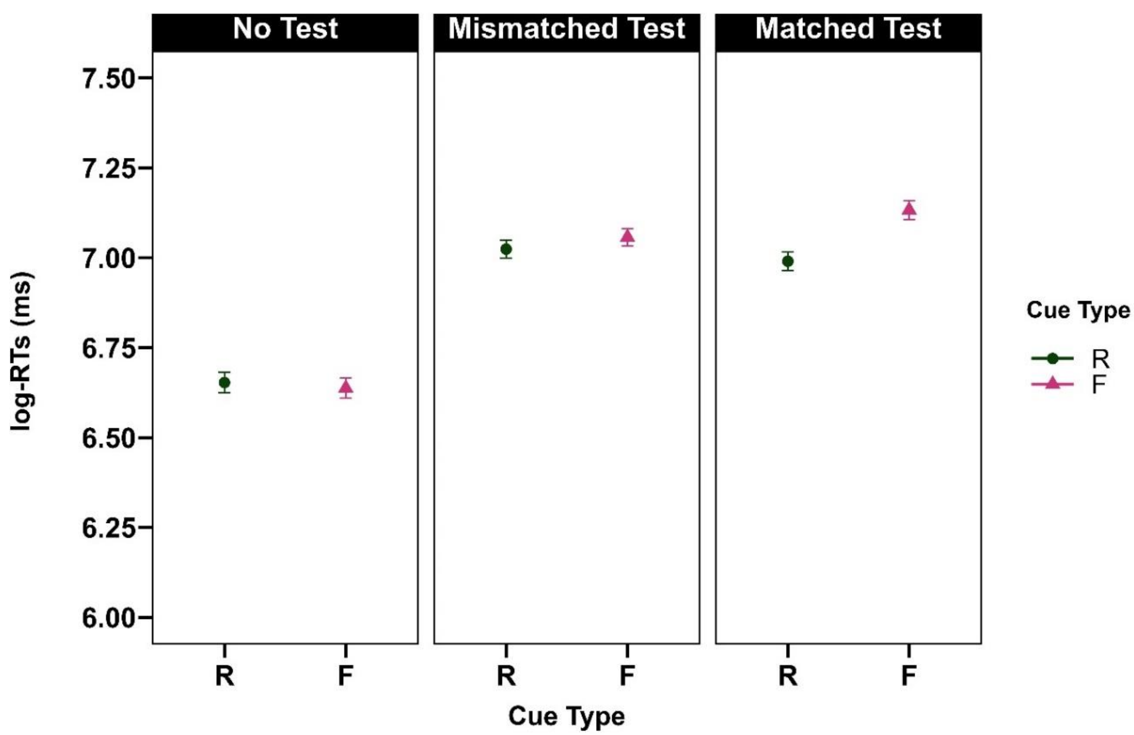
ISI, χ^2 (2) = 3.02, *p* = .271, indicating that RTs did not significantly differ across the ISI conditions (1,400 ms: *M* = 6.93, *SE* = .02; 1,800 ms: *M* = 6.91, *SE* = .02; 2,600 ms: *M* = 6.93, *SE* = .02). No other interactions were significant: Cue Type x ISI, χ^2 (2) = 1.24, *p* = .537, Encoding Type x ISI, χ^2 (4) = 7.24, *p* = .124, or Cue Type x Encoding Type x ISI, χ^2 (4) = 2.69, *p* = .611.

To investigate the significant Cue Type x Encoding Type interaction, first I conducted follow-up comparisons averaged over the levels of ISI. Separated by Encoding Type, in the No Test condition, participants were significantly slower to respond to the asterisk probe following an R cue (*M* = 6.73, *SE* = .02) than following an F cue (*M* = 6.61, *SE* = .02), *b* = .13, *SE* = .01, *p* < .001. In the Mismatched Test condition, however, the direction was the opposite: Participants were significantly slower to respond to the recognition task following an F cue (*M* = 7.06, *SE* = .02) than following an R cue (*M* = 7.02, *SE* = .02), *b* = .03, *SE* = .02, *p* = .032. Similarly, in the Matched Test condition, participants were significantly slower to respond to the recognition task following an F cue (*M* = 7.13, *SE* = .02) than following an R cue (*M* = 6.99, *SE* = .02), *b* = .14, *SE* = .02, *p* < .001. That is, participants were slower to correctly identify whether the target word in the encoding recognition task matched the word preceding the F instruction than preceding the R instruction.

Separated by Cue Type, for F items, participants were significantly faster to respond to the asterisk probe in the No Test condition (*M* = 6.61, *SE* = .02) than when there was a recognition test—in the Mismatched Test condition (*M* = 7.06, *SE* = .02), *b* = .45, *SE* = .01, *p* < .001, and in the Matched Test condition (*M* = 7.13, *SE* = .02), *b* = .53, *SE* = .01, *p* < .001.

Participants were slower to respond to the recognition test probe in the Matched Test condition than in the Mismatched Test condition, $b = .08$, $SE = .02$, $p < .001$. For R items, participants were significantly faster to respond to the asterisk probe in the No Test condition ($M = 6.73$, $SE = .02$) than in the recognition test in the Mismatched Test condition ($M = 7.02$, $SE = .02$), $b = .29$, $SE = .01$, $p < .001$, or in the Matched Test condition ($M = 6.99$, $SE = .02$), $b = .26$, $SE = .01$, $p < .001$. However, RTs did not significantly differ between the Mismatched Test and Matched Test conditions, $b = .03$, $SE = .02$, $p = .071$.

Figure 22. Experiment 8: Mean estimated (model-based) log RTs across Cue Type for each Encoding Type condition, where Encoding Test condition was averaged.



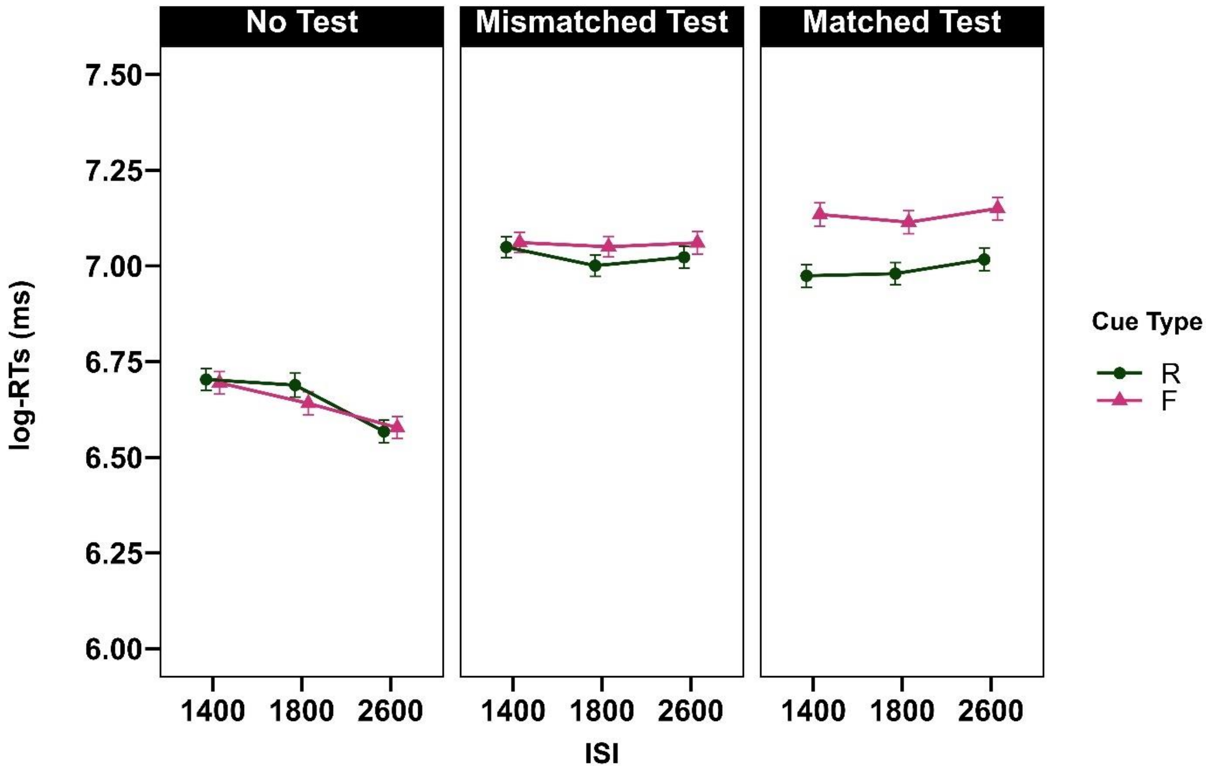
Note. RTs were collected for the encoding recognition task for the Mismatched Test and Matched Test conditions and via button press to an asterisk probe for the No Test condition. Error bars denote the model-based standard error.

As described earlier, examining RT patterns after an R/F cue has provided what has been interpreted as the hallmark evidence favouring the attentional inhibition account. Critically, Fawcett and Taylor (2008) observed that participants were slower to respond to a probe following an F instruction than following an R instruction, at least before 2600 ms. As a result, they argued for an active, cognitively-demanding, and possibly inhibitory forgetting mechanism that successfully expunged irrelevant F items from working memory before 2,600 ms elapsed. Of main interest was how the three levels of ISI incorporated in the Fawcett and Taylor investigation might influence my asterisk probe in the No Test condition and the recognition test probe in the Mismatched Test and Matched Test conditions. Thus, the follow-up comparisons were pre-registered to provide a direct investigation of the key evidence in support of the attentional inhibition account.

For each level of ISI, in the No Test condition, participants were slower to respond to the asterisk probe following an R cue (1,400 ms: M_R = 6.74, SE_R = .03; 1,800 ms: M_R = 6.73, SE_R = .03; 2,600 ms: M_R = 6.73, SE_R = .03) than following an F cue (1,400 ms: M_F = 6.63, SE_F = .03; 1,800 ms: M_F = 6.61, SE_F = .03; 2,600 ms: M_F = 6.58, SE_F = .03), *bs* > .11, *SE*s = .02, *ps* < .001. In the Mismatched Test condition, RTs to the recognition test probe did not significantly differ following an R cue (1,400 ms: M_R = 7.05, SE_R = .03; 1,800 ms: M_R = 7.00, SE_R = .03; 2,600 ms: M_R = 7.02, SE_R = .03) or following an F cue (1,400 ms: M_F = 7.04, SE_F = .03; 1,800 ms: M_F = 7.05, SE_F = .03; 2,600 ms: M_F = 7.06, SE_F = .03), *bs* > .01, *SE*s = .03, *ps* > .063. However, in the Matched Test condition, participants actually were slower to respond to the recognition test probe following an F cue (1,400 ms: M_F = 7.13, SE_F = .03; 1,800 ms: M_F = 7.11, SE_F = .03; 2,600 ms: M_F = 7.15, SE_F = .03) than following an R cue (1,400 ms: M_R = 6.97, SE_R = .03; 1,800 ms: M_R = 6.98, SE_R = .03; 2,600 ms: M_R = 7.02, SE_R = .03), *bs* > .13, *SE*s = .03, *ps* < .001.

Comparisons at each level of ISI revealed that for F items in the No Test condition, participants were marginally significantly slower to respond to the asterisk probe following the 1,400 ms ISI (*M* = 6.63, *SE* = .03) than following the 2,600 ms ISI (*M* = 6.58, *SE* = .03), *b* = .04, *SE* = .02, *p* = .046. For the rest of the levels of ISI, RTs did not significantly differ between the different levels of ISI, *bs* > .00, *SEs* = .03, *ps* > .164.

Figure 23. Experiment 8: Mean estimated (model-based) log RTs across Cue Type for each Encoding Type condition, where Encoding Test condition was averaged, separated by ISI.



Note. RTs were collected for the encoding recognition task for the Mismatched Test and Matched Test conditions and via button press to an asterisk probe for the No Test condition. Error bars denote the model-based standard error.

3.5.3 Discussion

Experiment 8 directly replicated the general pattern of findings from Experiment 7, confirming the elimination of the directed forgetting effect for Matched Test items. I attribute this result to the enhancement observed for F items when retrieval was forced. Once again, the directed forgetting effect in the Mismatched Test condition mirrored that in the No Test condition. Importantly, there were no discernible differences in memory for R items across the Encoding Test conditions, providing additional evidence for the inherent operation of the retrieval check mechanism for R items. Forcing such a check for R items exerted no effect because it was already routinely done.

The combined results of Experiments 7 and 8 robustly support my principal hypothesis that successful retrieval of F items enhances memory for those items beyond the initial forgetting associated with non-retrieval. By deliberately inducing the retrieval of F items, the directed forgetting effect was eliminated and, through replication, it was established that this retrieval check mechanism is the driving force behind the R item benefit in item-method directed forgetting.

In Experiment 8, an additional goal was to explore the cognitive demands of intentional remembering and intentional forgetting. Considering debates on the cognitive effort involved in intentional forgetting, I proposed that the retrieval check mechanism may simultaneously operate with a process actively removing F items from working memory. Contrary to Fawcett and Taylor's (2008, 2012) key findings supporting an active inhibitory removal mechanism, my RT results in Experiment 7 and, collapsing across the ISI timings, in Experiment 8 do not align with their observations. Specifically, in the No Test condition, participants were slower to respond to a secondary task following an R cue than an F cue, the reverse of Fawcett and Taylor's results. Furthermore, in the Mismatched Test and Matched Test conditions, the RT pattern reversed, with participants being slower to respond to a secondary probe following an F instruction than following an R instruction. This RT pattern reversal actually supports the effectiveness of the retrieval-inducing manipulation. The incorporation of an encoding recognition task successfully prompted the retrieval check of F items, a process that would not typically occur. The reversal in the pattern between the

Mismatched and Matched Test conditions confirms that the encoding task served as a rapid retrieval check, essentially querying, “What is the current item?”

Looking at the ISI timing manipulation, the RT results from Experiment 8 showed no support that intentional forgetting takes less than 2,600 ms, in fact, I found no RT differences across the three ISI timing observations. Importantly, this experiment did not replicate Fawcett and Taylor’s (2008, 2012) findings that RTs are slower following an F cue at 1,400 ms and 1,800 ms, but no RT differences at 2,600 ms. Instead, I found consistent RT differences across the three ISI conditions—and in the direction opposite to those reported by Fawcett and Taylor.

Taken together, the findings of Experiment 8 provide no support for an active inhibitory removal mechanism, but instead, provide further support for the crucial role of retrieval in item-method directed forgetting. Thus, there is no incentive for an active forgetting mechanism that is operating jointly with the proposed selective retrieval check mechanism.

Chapter 4:
General Discussion

In this dissertation, my primary goal was to carry out an in-depth exploration of the fundamental assumptions underpinning the most longstanding account of the directed forgetting effect using the item-method: the selective rehearsal account. Specifically, I examined how increasing the temporal duration devoted to intentional remembering and intentional forgetting, while the cue was available, influenced the directed forgetting effect. I have provided compelling evidence that the selective rehearsal account inadequately captures the complete dynamics of the item-method directed forgetting effect. My findings consistently refuted the two pivotal assumptions integral to the selective rehearsal account: (1) the continuous cumulative rehearsal of R items, and (2) the time-based decay of F items. My findings not only challenge the validity of the selective rehearsal account but also highlight its limitations in providing a comprehensive explanation for intentional forgetting at encoding.

In Experiments 1a, 1b, and 2, my findings revealed that cue duration, when randomly intermixed, did not influence either item memory or associative recognition memory. Contrary to the cumulative rehearsal assumption, increasing R cue duration did not improve memory for R items. Additionally, the time-based decay and interference predictions were not supported, given no influence at all of increasing cue duration, challenging the passive view of intentional forgetting at encoding. According to the time-based decay assumption, extended cue durations should reduce memory for F items due to the longer duration between the F instruction and the memory test, which I did not observe. According to the interference assumption, the cumulative rehearsal of R items should interfere with memory for F items, as motivated participants may reinstate the R item rehearsal set during the F cue presentation. While my investigate did not directly negate the interference prediction, there was no evidence of cumulative rehearsal. Thus, interference alone is not sufficient to explain the directed forgetting effect within the selective rehearsal account.

In Experiments 3 and 4, with blocked cue durations, I replicated Bancroft et al. (2013) and Lee et al. (2007) in that both R and F item and associative recognition increased reliably, albeit modestly, as cue duration increased. This finding challenged the selective rehearsal assumption that longer cue durations ought to exclusively benefit R item rehearsal. These results question the selective nature of cue duration effects indicating that, under certain conditions, R items may benefit from continuous cumulative rehearsal. Notably, this cumulative rehearsal is not exclusively confined to R items. The observed trend indicated that, with longer cue durations, F items might inadvertently be more likely to be incidentally rehearsed, thereby enhancing their memorability. This unintentional rehearsal introduced a complication: It interfered with the efficacy of the removal mechanism designed to expunge F items from working memory. The main insights from this finding are that intentional forgetting at encoding is a rapid process, and that even F items may receive accidental rehearsal if not promptly removed from working memory. This result also sheds light on a limitation in the removal mechanism, indicating that when F items linger too long in the focus of attention, they may not as readily undergo removal.

Furthermore, the selective rehearsal account struggled to explain the conditions under which item-method directed forgetting arises. When cue durations were unpredictable, rapid strengthening of R items and quick removal of F items from working memory contributed to the directed forgetting effect (Experiments 1a, 1b, and 2). However, when cue durations were blocked and predictable, cumulative rehearsal occurred equivalently for R and F items, highlighting the limitations of the time-dependent rehearsal being applied selectively to R items (Experiments 3 and 4).

Next, in this dissertation, I proposed and tested a novel account that emphasized the critical role of retrieval in the item-method directed forgetting paradigm. Specifically, I explored a rapid retrieval check mechanism that selectively strengthens the representation of R items. My findings confirmed that intentional remembering arises when a retrieval benefit occurs for R items, while intentional forgetting of F items occurs as a by-product of not receiving this retrieval benefit.

In Chapter 3, I found that forcing the retrieval check mechanism for F items through a button press (Experiment 6) or recognition retrieval check (Experiments 7 and 8) eliminated the directed forgetting effect. Thus, a retrieval check is sufficient to boost F items back up to the level of R items. Additionally, I examined the cognitive demands post R/F cues, evaluating RTs to post-cue probes to assess whether the rapid removal of F items required more processing resources than did the retrieval boost to R items, as has been argued in the content of the attentional inhibition account. Notably, I found no evidence supporting an active effortful removal mechanism that demands more cognitive resources than intentional remembering. I discuss the implications of this finding for the item-method directed forgetting literature shortly in the Post-Cue Probe RTs and Mechanisms of Intentional Forgetting subsection.

4.1 Retrieval in Item-Method Directed Forgetting

In Chapter 3, I explored a vital memory process—retrieval—that is typically overlooked as a contributor to processing during encoding. My goal was to test my proposed retrieval-based explanation: In the item-method directed forgetting paradigm, participants retrieved each R item immediately following an instruction to remember. This "retrieval check" serves as a verification to label the just-presented item as to-be-remembered. Critically, this act of retrieval promotes the elaborative encoding and strengthening of R items in long-term memory. This retrieval check is exclusive to the R items and absent for F items. An instruction to Forget labels these items as task-irrelevant (the task being the intention to remember all R items for the later memory test). According to this *selective retrieval* account, participants do not retrieve the just-presented item upon an F instruction because they do not need to bring F items back in the focus of attention to strengthen them. Thus, the directed forgetting effect is driven by the retrieval boost to the R items, at the expense of F item retrieval.

In Chapter 3, my primary goal was to test this retrieval check mechanism favouring R items. I predicted that inducing retrieval checks for F items would reduce or eliminate the directed forgetting effect. In Experiment 5, inducing retrieval by repeating items did not produce results supportive of a retrieval benefit, possibly because mere repetition failed to induce retrieval of the first presentation. Participants may have regarded the second presentation of a repeated item as a new presented word, creating a new memory. Indeed, previous research on repetition and memory suggested that one of the possible mechanisms to explain a repetition benefit is the *multiple trace hypothesis*: Each presentation of a repeated item is encoded as a new memory, suggesting that repetition does not strengthen the first presentation of the repeated item. Instead, repetition generates a new memory that coexists with memories of other presentations of the same repeated item (Hintzman, 1971, 1986, 2010). This explanation aligns with my Experiment 5 findings, supporting the claim that simply repeating an F item did not induce its retrieval.

In Experiment 6, requiring a button press upon noticing the second presentation of an item eliminated the directed forgetting effect. Conversely, when no button press occurred, signalling a missed repetition, the directed forgetting effect was comparable to the effect for unrepeated items. Moreover, inducing retrieval through immediate recognition tasks after the R/F cue presentation, as in Experiments 7 and 8, again eliminated the directed forgetting effect when the recognition target matched the just-presented item. In sharp contrast, when the recognition target was a new word, there was a typical directed forgetting effect. In fact, this effect was equivalent to the effect for items that did not receive an immediate recognition task after the R/F cue. This finding provides additional support for my retrieval-based explanation by showing that the elimination of the directed forgetting effect when the recognition target matched the just-presented item was due to the rapid retrieval boost being applied to the F items. Indeed, the absence of an F item performance difference between when the recognition target was a new word and when there was no recognition task after the R/F cue suggests that an act of retrieval has taken place.

In Experiments 7 and 8, that R item performance did not differ under my encoding recognition test manipulation also confirms a critical assumption of my retrieval-based explanation: R items are already routinely retrieved throughout the study phase, which underlies the benefit to the R items in the directed forgetting effect, so forcing such retrieval is effectively redundant. This assumption is further supported by the performance on the encoding recognition test, where memory was better for R items than for F items in the Matched condition, indicating that since R items were already readily retrieved and available, memory was better on the encoding recognition test for these items. These novel results demonstrate the critical role of retrieval in the item-method directed forgetting effect.

4.2 Post-Cue Probe RTs and Mechanisms of Intentional Forgetting at Encoding

In their articles, Fawcett and Taylor (2008, 2012) found that participants were slower to respond to a probe (e.g., an asterisk or an irrelevant word) following an F cue than following an R cue. This led them to suggest that the cognitive demands of intentional forgetting exceeded those of remembering. This finding has been the paramount evidence in support of an effortful attentional withdrawal, or inhibition, mechanism operating to remove irrelevant F information from memory. A non-inhibitory account, such as selective rehearsal, makes the opposite prediction: RTs should be longer to a probe following an R cue than an F cue because the directed forgetting effect results from elaborative rehearsal of only the R items. No additional process occurs for F items so cognitive demands for intentional remembering should exceed those of intentional forgetting.

In Experiment 7, participants demonstrated longer response times when responding to a task-irrelevant probe (i.e., asterisk) immediately after the R/F cue, following an instruction to Remember than following an instruction to Forget. In this case, where no additional process is induced after the R/F instruction, this finding provides a clear assessment of resource allocation following intentional remembering and intentional forgetting. It may be best, then, to think of this condition as a closer replication of the Fawcett and Taylor procedure. Thus, this result contradicts the attentional inhibition account and supports a non-inhibitory explanation to how F items are removed from working memory (e.g., Dames & Oberauer, 2022; Oberauer, 2018).

Further insight into the resource demands of intentional forgetting arise from Experiment 8. While replicating the RT pattern observed in Experiment 7, I found that varying the ISI between the onset of the R/F cue and the task-irrelevant probe—using the very ISIs used by Fawcett and Taylor (2008, 2012)—did not influence the RT pattern. This finding contradicts the key RT observation by Fawcett and Taylor (2008, 2012). According to their account, intentional forgetting consumes cognitive resources for an extended period, with RTs slower following an F cue at 1,400 ms and 1,800 ms but not at 2,600 ms. At this point, resources should be redirected to processing R items. However, my findings reveal no RT differences across ISI conditions, challenging both the attentional inhibition account and the proposed timings for an active inhibitory process to remove F items.

4.3 The Selective Retrieval Account

Although decades of research demonstrate robust directed forgetting effects in the item-method paradigm, the cognitive mechanisms governing intentional forgetting at encoding are still in question. In fact, much of the theoretical interest concerning item-method directed forgetting revolves around the debate concerning the mechanisms involved in intentional remembering and intentional forgetting. This dissertation challenges the assumptions of the selective rehearsal account and introduces a new perspective, the selective retrieval account, aimed at overcoming the limitations inherent in the selective rehearsal account.

As described earlier, Chapter 2 provided no support for the cumulative rehearsal and time-based decay assumptions of the selective rehearsal account. In their article, Oberauer and Lewandowsky (2013) offered an insightful perspective on theory development in literature dominated by decay theories: “Decay theories nevertheless do not face away – rather, they adapt” (p. 16). Indeed, their perspective guided my approach to theory development in intentional forgetting at encoding. In Chapter 3, I have provided consistent evidence supporting my retrieval-based explanation to highlight the critical role of retrieval in item-method directed forgetting. This solidified that the benefit to the R items arises from a rapid retrieval boost—a retrieval check—that strengthens R items in long-term memory. Further, I examined the nature of the removal mechanism that operates to intentionally forget F items. Experiments 7 and 8 provided no support for an active inhibitory mechanism, thus I propose to retain the selective retrieval account as a non-inhibitory explanation to item-method directed forgetting, that hinges on the important role of retrieval.

In the selective retrieval view, participants maintain each item in memory via maintenance rehearsal in anticipation of its instructional cue. Following an R cue, participants check that they know the just-presented item, a retrieval that boosts its memory strength. Typically, only R items benefit from this retrieval checking operation. Given that they do not need to be remembered, F items ordinarily are not retrieved and do not receive this additional boost of memory strength. Instead, F items are rapidly removed from working memory to render their accessibility for further encoding. The selective retrieval account incorporates a combination of two processes to explain the directed forgetting effect: (1) a rapid boost to R items that operates in the form of a retrieval check mechanism, and (2) a quick removal mechanism applied to F items that arises as a by-product of not receiving retrieval and removes them from the focus of attention in working memory. These processes working jointly explains why sometimes, it is harder for individuals to forget information on purpose. For example, if you are going to a friend's house, and they told you that you should turn right onto Front Street, you would intentionally remember this street name for future travel. However, if your friend later tells you that "I was wrong, you need to turn right onto Wellington Street", it might be harder for you to intentionally forget Front Street, because this directional information was retrieved and lingered for too long to effectively removed. Likewise, if your friend told you their new unit number for their apartment, and at the time of receiving this information, you did not have the intent to remember it for the future, it might be harder for you to remember it because at the time of encoding, you did not perform a retrieval check to make sure you have saved the unit number in your long-term memory.

4.4 Limitations and Future Directions

Although Chapter 2 provided a detailed investigation of how varying cue durations influence the item-method directed forgetting effect, these cue durations were limited to 1 s, 5 s, and 10 s. While previous investigations have explored other intervals of cue durations (Bancroft et al., 2013; Lee et al., 2007), it is possible that specific cue duration points other than the intervals that I used could produce different results. For example, if the retrieval check mechanism occurs rapidly upon receiving the R cue, as I have argued, then varying the time of the cue presentation using much shorter durations (i.e., 100 ms) could provide additional information about the operation of this mechanism. Perhaps introducing very short durations would prevent participants from engaging even in the rapid retrieval check, reducing memory for R items and reducing the overall directed forgetting effect. Future work should introduce other cue durations to further compare these mechanistic differences in intentional remembering and intentional forgetting.

Next, this dissertation examined how intentional remembering and intentional forgetting influences long-term memory. However, given that item-method directed forgetting is a paradigm to measure intentional forgetting at encoding, it would be worthwhile for future investigations to assess these investigations in working memory. For example, if F items are rapidly removed from the focus of attention in working memory, would their removal free up more working memory capacity resources for the processing of relevant information? In this case, incorporating a working memory test (e.g., a test of binding memory as a function of location of the word on the screen; Dames & Oberauer, 2022; Oberauer, 2018) may pinpoint specific mechanisms that are operating immediately after an instruction to Remember or Forget.

Another limitation pertains to the tests of my proposed retrieval check mechanism. Experiment 6, and Experiments 7 and 8, differ notably in the timing and nature of retrieval processes. In Experiment 6, retrieval of F items occurred 9 to 11 trials after the initial presentation, representing delayed retrieval. This delayed retrieval check was conducted after the item had been in memory for an extended period, diverging from the immediacy of the proposed retrieval check mechanism. Conversely, in Experiments 7 and 8, retrieval took place immediately after the cue presentation, aligning more closely with the retrieval check mechanism that I proposed for the selective retrieval account. This immediate retrieval mimics the prompt retrieval that routinely occurs for R items. Although the findings from Experiment 6 and from Experiments 7 and 8 both support the important role of retrieval in item-method directed forgetting, the immediate retrieval mechanism in Experiments 7 and 8 provides a better look at how this mechanism operates.

The delayed retrieval process in Experiment 6 does, however, raise a potentially interesting avenue for future research. One possible direction includes assessing whether the retrieval check mechanism occurs throughout the study list. Inspired by the rehearsal set idea from selective rehearsal, a possible assumption of the selective retrieval account could be that, in addition to the rapid retrieval check that boosts memory for R items, under some circumstances, R items may also receive routine retrieval throughout the study list, maybe even during the presentation of an F cue. Future research could assess the validity of this delayed retrieval check mechanism by incorporating encoding recognition tasks throughout the study list.

4.5 Conclusion

Based on the findings presented in this dissertation, I propose the substitution of a retrieval mechanism for a rehearsal mechanism as the key element in both intentional remembering and intentional forgetting during encoding. Specifically, I have introduced a novel account of intentional forgetting at encoding to elucidate the directed forgetting effects observed in the item-method directed forgetting paradigm.

In the *selective retrieval account*, a directed forgetting effect emerges because of a swift retrieval check that enhances memory for to-be-remembered items but that is not applied to the to-be-forgotten items. Put simply, based on the current results, I conclude that the intention to remember engages a cognitive process that rapidly retrieves to-be-remembered information, ensuring its accessibility for future use. This retrieval check serves as a recognition test and is selective, being exclusively activated when there is an intention to remember. Consequently, the absence of this retrieval check for to-be-forgotten information negates further strengthening of this information into long-term memory. As well, to-be-forgotten information, upon an intent to forget, is rapidly removed from the focus of attention in working memory. Therefore, intentional forgetting arises from the omission of item retrieval and a removal mechanism, not from the mere passage of time resulting in decay.

This work emphasizes how we intentionally update the contents of our memory with relevant information through the effective use of rapid retrieval. That is, with an intention to remember, we boost the representation of relevant information by quickly checking what we can remember about a just-presented event. So, with the goal of intentionally remembering the relevant brown sugar measurement, retrieving the new relevant measurement that your partner just provided ensures proper access to it such that memory for that recipe material should benefit. This quick “double-checking” retrieval is adequate to improve memory for the recipe, thus that person does not need to continue rehearsing the new measurement and can proceed to bake delicious cookies. Likewise, who people who tend to worry and sometimes cannot remember if they have turned off their stove or straightener, establishing a routine of quickly doing a retrieval check of unplugging the device can enhance their memory for an unplugged and turned off device, providing assurance that you they will not burn down their home. On the other hand, to intentionally forget irrelevant brown sugar measurement (e.g., if your partner indicates that previously told brown sugar measurement was wrong and irrelevant), not retrieving this measurement should result in more successful forgetting. While memory for the irrelevant brown sugar measurement may linger (benefitting from maintenance until told that it is irrelevant), the absence of retrieval results in poorer memory for the to-be-forgotten brown sugar amount compared to the to-be-remembered brown sugar amount. This dissertation provides insight into the fundamental cognitive mechanisms that operate to selectively strengthen memory for goal-relevant information while intentionally removing goal-irrelevant information to reduce its accessibility. Importantly, this dissertation provides a novel account introducing a unified explanation of item-method directed forgetting and incorporating retrieval as the core process in item-method directed forgetting.

References

Ahmad, F. N., Tan, P., & Hockley, W. E. (2019). Directed forgetting for categorised pictures: Recognition memory for perceptual details versus gist. *Memory, 27*(7), 894–903. https://doi.org/10.1080/09658211.2019.1591456

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

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.

Baddeley, A. D. (1986). Working memory. Oxford, UK: Oxford University Press.

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

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Bartsch, L. M., & Oberauer, K. (2023). The contribution of episodic long-term memory to working memory for bindings. *Cognition*, *231*, 105330. https://doi.org/10.1016/j.cognition.2022.105330

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition*, *46*, 769-808. https://doi.org/10.3758/s13421-018-0805-9

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: further comparisons of the item and list methods. *Memory*, *4*(6), 633–653. https://doi.org/10.1080/741941000

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

Benjamin, A. S. (Ed.). (2011). The causes and consequences of reminding. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 71–87). Psychology Press.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228. https://doi.org/10.1016/j.cogpsych.2010.05.004

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. (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, Inc.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123-144). New York: Wiley.

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

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), Article 2515245920960351. doi:10.1177/2515245920960351

Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology, 100*(1), 9–17. https://doi.org/10.1037/h0035488

Burgess, N., Hockley, W. E., & Hourihan, K. L. (2017). The effects of context in item-based directed forgetting: Evidence for "one-shot" context storage. *Memory & Cognition*, *45*(5), 745–754. https://doi.org/10.3758/s13421-017-0692-5

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.

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

Chiu, Y.-C., Wang, T. H., Beck, D. M., Lewis-Peacock, J. A., & Sahakyan, L. (2021). Separation of item and context in item-method directed forgetting. *Neuroimage*, *235*, 117983. https://doi.org/10.1016/j.neuroimage.2021.117983

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*, *151*12), 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. doi: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: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., Taylor, T. L., & Nadel, L. (2013). Intentional forgetting diminishes memory for continuous events. *Memory, 21*(6), 675–694. https://doi.org /10.1080/09658211.2012.748078

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. (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. (2019). *An R companion to applied regression*, Third edition. Sage, Thousand Oaks, CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*(1), 58–72. https://doi.org/10.1037/0096-3445.112.1.58

Golding, J. M., Long, D. L., & MacLeod, C. M. (1994). You can't always forget what you want: Directed forgetting of related words. *Journal of Memory and Language*, *33*(4), 493–510. https://doi-org.proxy.lib.uwaterloo.ca/10.1006/jmla.1994.1023

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

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin, 102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, *9*, 411657. https://doi.org/10.3389/fpsyg.2018.02412

Hauswald, A., & Kissler, J. (2008). Directed forgetting of complex pictures in an item method paradigm. *Memory, 16*(8), 797–809. https://doi.org/10.1080/09658210802169087

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth." *Perspectives on Psychological Science*, *6*(3), 253–271. https://doi.org/10.1177/1745691611406924

Hintzman, D. L. (1976). Repetition and memory. *Psychology of Learning and Motivation, 10*, 47-91. https://doi.org/10.1016/S0079-7421(08)60464-8

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*(3), 297–306. https://doi.org/10.1037/h0030907

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.

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

Hourihan, K. L., Ozubko, J. D., & MacLeod, C. M. (2009). Directed forgetting of visual symbols: Evidence for nonverbal selective rehearsal. *Memory & Cognition*, *37*(8), 1059–1068. https://doi.org /10.3758/MC.37.8.1059

Hourihan K. L. (2021). The influence of cue probability on item and source judgments in item method directed forgetting. *Memory*, *29*(9), 1136–1155. https://doi.org/10.1080/09658211.2021.1967400

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1354–1365. https://doi.org/10.1037/0096-1523.32.6.1354

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology*, *62(4)*, 242-246. https://doi.org/10.1037/1196-1961.62.4.242

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior, 17*(6), 649–667. https://doi.org/10.1016/S0022-5371(78)90393-6

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41*(5), 625–637. https://doi.org/10.3758/s13421-013-0298-5

Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory & Cognition, 41*(5), 638–649. https://doi.org/10.3758/s13421-013-0313-x

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(5), 12821297. https://doi.org/10.1037/xlm0000123

Jolicœur, P., & Dell' Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138–202. https://doi.org/10.1006/cogp.1998.0684

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborate studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., Butler, A. C., & Roediger, H. L., 3rd (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own?. *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24*(3), 401–418. https://doi.org /10.1007/s10648-012-9202-2

Karpicke, J. D., & Roediger, H. L. III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. https://doi.org/10.2307/2533558

Kolers, P. A., & Roediger, H. L. III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23*(4), 425–449. https://doi.org/10.1016/S0022-5371(84)90282-2

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

Lee, Y., Lee, H. & Fawcett, J. M. (2013). Intentional forgetting reduces colour-naming interference: Evidence from item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*, 220-236. https://doi.org /10.1037/a0028905

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, (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.

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129. https://doi.org/10.3758/BF03210819

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. 43 (pp. 163–214). Elsevier Science.

MacLeod, C. M., & Daniels, K. A. (2000). Direct vs. indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review*, *7(2)*, 354-359. https://doi.org/10.3758/BF03212993

MacLeod, C. M., Pottruff, M. M., Forrin, N. D., & Masson, M. E. J. (2012). The next generation: The value of reminding. *Memory & Cognition*, *40*(5), 693-702. https://doi.org/10.3758/s13421-012-0182-8

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.

Montagliani, A., & Hockley, W. E. (2019). Item-based directed forgetting for categorized lists: Forgetting of words that were not presented. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 73*(3), 135–143. https://doi-org /10.1037/cep0000177

Muter, P. (1980). Very rapid forgetting. *Memory & Cognition, 8*(2), 174-179. https://psycnet.apa.org/doi/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

Negley, J. H., Kelley, C. M., & Jacoby, L. L. (2018). The importance of time to think back: The role of reminding in retroactive effects of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(9), 1352–1364. https://doi.org/10.1037/xlm0000512

Oberauer, K. (2009a). Design for a working memory. *Psychology of Learning and Motivation*, *51,* 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*, 346–357. https://doi.org/10.3758/MC.37.3.346

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*(3), 368–387. https://doi.org/10.1037/0096-3445.134.3.368

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. (2019). Working memory and attention – A conceptual analysis and review. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.58

Oberauer, K., & Greve, W. (2021). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology: General.* Advance online publication. 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

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., & Vockenberg, K. (2009). Updating of working memory: Lingering bindings. *The Quarterly Journal of Experimental Psychology*, *62*(5), 967–987. https://doi.org/10.1080/17470210802372912

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

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

Pastötter, B., & Bäuml, K. H. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*, *5*, 286. https://doi.org/10.3389/fpsyg.2014.00286

Polack, C. W., & Miller, R. R. (2022). Testing improves performance as well as assesses learning: A review of the testing effect with implications for models of learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 48*(3), 222–241. https://doi.org/10.1037/xan0000323

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

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology*, *79*(1, Pt.1), 168–177. https://doi.org/10.1037/h0026932

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition*, *42*(7), 1198–1210. https://doi.org/10.3758/s13421-014-0419-9

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *64*(1), 41–46. https://doi.org/10.1037/a0016569

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

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory: An overview of working memory consolidation. *Annals of the New York Academy of Sciences*, *1424*(1), 8–18. https://doi.org/10.1111/nyas.13633

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology*, *69*(10), 1969–1995. https://doi.org/10.1080/17470218.2014.914546

Robbins, D., & Bray, J. F. (1974). The spacing effect and the A-B, A-C paradigm: Evidence for retroactive facilitation. *Journal of Experimental Psychology, 103*(3), 420–425. https://doi.org/10.1037/h0037203

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

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

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

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., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1346–1354. https://doi.org/10.1037/a0020121

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. *Psychology of Learning and Motivation*, *59*, 131-189. http://doi.org/10.1016/B978-0-12-407187-2.00001-0

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.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072. https://doi.org/10.1037//0278-7393.28.6.1064.

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). Tokyo: Springer.

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

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229. https://www.ncbi.nlm.nih.gov/pubmed/20842279

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

Tanberg, P., Fernandes, M. A., & MacLeod, C. M. (2022). Aging and directed forgetting: Evidence for an associative deficit but no evidence for an inhibition deficit. *Canadian journal of experimental*, *76*(3), 210–217. https://doi.org/10.1037/cep0000292

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

Tullis, M., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General, 143*(4), 1526–1540.

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *40*(1), 94–105. https://doi.org/10.1037/a0034055

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology. General*, *148*(1), 30–50. https://doi.org/10.1037/xge0000458

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: a memory-for-change perspective. *Memory*, *27*(6), 849–867. https://doi.org/10.1080/09658211.2019.1582677

Whitlock, J., Chiu, J. Y.-C., & Sahakyan, L. (2022). Directed forgetting in associative memory: Dissociating item and associative impairment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 29–42. https://doi.org/10.1037/xlm0001027

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

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

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method--directed forgetting. *Cerebral Cortex*, *18*(3), 670–682. https://doi.org/10.1093/cercor/bhm101

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241–264). Academic Press.

Zacks, R.T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(1), 143-156. https://doi.org/10.1037/0278-7393.22.1.143